大模型数据合成

——从通用到行业的对齐实践

刘学博

哈尔滨工业大学（深圳）

2025年8月22日

01 大模型与数据合成背景

02 基础：通用与垂域数据合成

03 核心：高效数据学习与利用

04 进阶：“数据-模型”能力对齐

05 领域瓶颈与未来展望

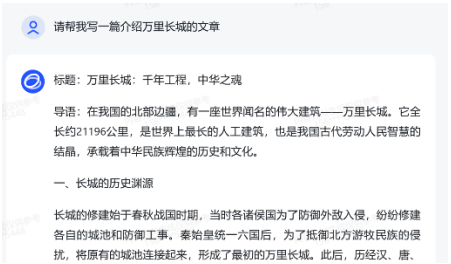
大模型成为当前人工智能的巅峰

- 大模型是指包含超大规模参数（大于10^9）的神经网络模型，在内容生成等方面能够完成过去只有人才能完成的任务，被认为是人工智能从弱向强的标志。

【文生文】

ChatGPT

“请帮我写一篇介绍万里长城的文章”



【文生图】

DALL·E

“请帮我给文章中的长城配个好看的图”

【文生视频】

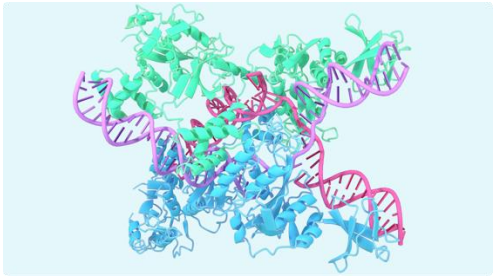
SORA

“请再帮我生成一段关于万里长城的多角度航拍视频”

【科学计算】

AlphaFold

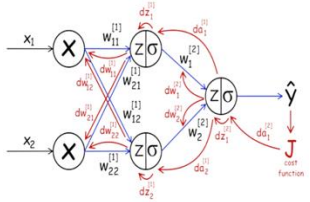
将过去人工需要数月或数年的蛋白质精确结构预测任务缩短至只需几秒钟



大模型因何强大?

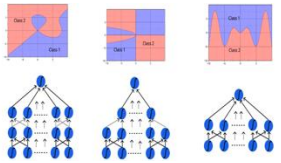
- **大模型本质是运用强大算法消耗大量算力在海量数据训练出的复杂概率分布函数。**

解决了神经网络(包括大模型)的训练问题。



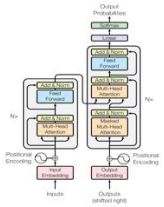
反向传播算法

理论上证明了神经网络的强大拟合能力。



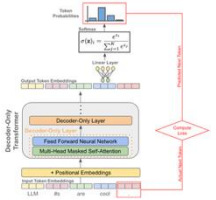
万能逼近定理(1989)

能捕获长距离依赖关系。



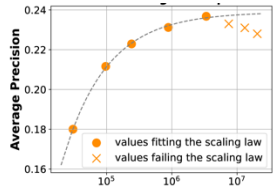
Transformer (2017)

解决了用无标注数据训练的问题。



自监督预训练机制(2018)

揭示出参数量、数据量、算力与性能的正相关性。



神经标度律 (2020)

强算法（模型）

模型参数越多，需要更多数据来拟合!

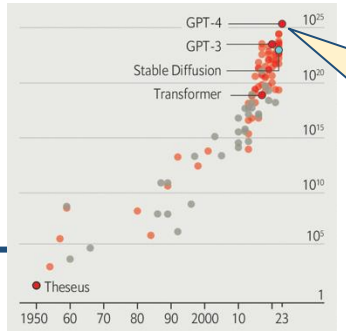
Chinchilla使用1.4万亿Token训练，全球可用文本量估计为3.2万亿Token。

大模型

大数据

大算力

GPU算力消耗每年增加一个量级!



GPT-4所需算力估计高达 $2.15*10^{25}$ FLOPs。

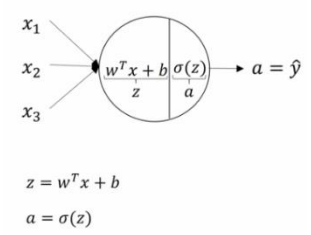
大模型因何强大?

- **万能逼近定理**：只要给予网络足够数量的神经元，便可以拟合任何复杂的连续函数。

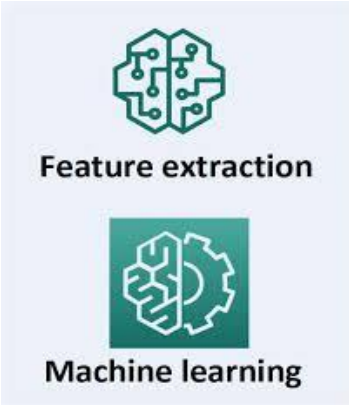
  (Kornik etal., 1989; Cybenko, 1989)

【从神经网络到大模型的技术演化】

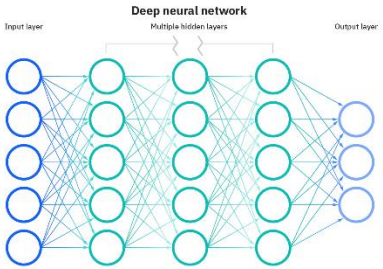
从零训练 ← → 利用基础模型适配



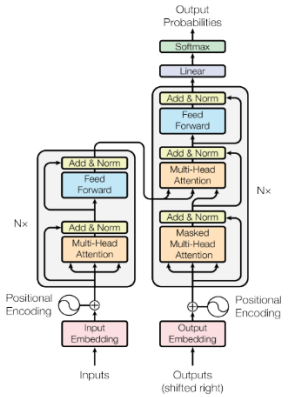
浅层感知器



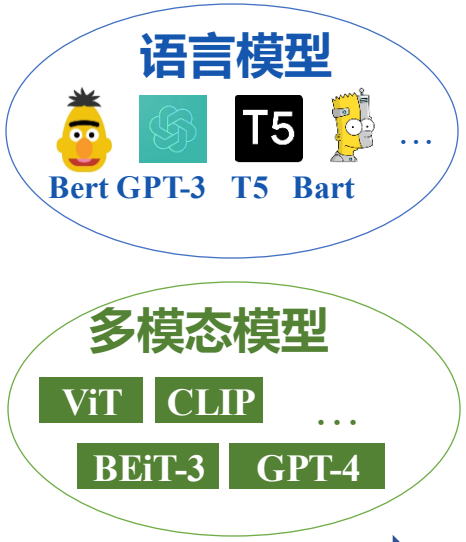
传统机器学习



传统深度学习



预训练+微调



网络层数越来越深、参数规模越来越多、训练数据越来越大

大模型因何强大?

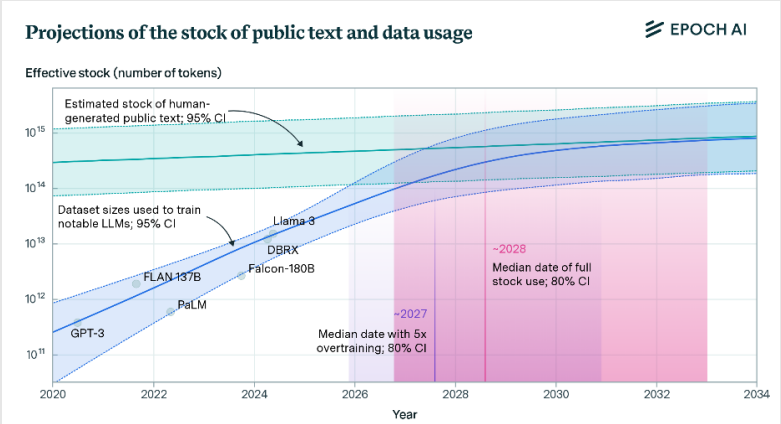
$$\hat{\theta} = \arg\min_{\theta} \left\{ \sum_{(x,y)\in\mathcal{D}} L(f(x;\theta), y) \right\}$$

算力　数据　算法（模型）

模型参数越多，需要更多数据来拟合!

模型名称	参数量	训练数据量
LaMDA	1370亿	1680亿
GPT-3	1750亿	3000亿
Jurassic	1780亿	3000亿
Gopher	2800亿	3000亿
MT-NLG 530B	5300亿	2700亿
Chinchilla	700亿	14000亿
Falcon	400亿	10000亿
LLAMA	630亿	14000亿
LLAMA-2	700亿	20000亿

2028年或耗尽高质量文本资源



网络层数越来越深、参数规模越来越多、训练数据越来越大

数据红利即将耗尽

- **数据合成是提升与对齐大模型能力最直接、高效的手段**

一个最直接的手段：引入专家合成数据

- 据The Information报道，Open AI 至少找了300位生物学博士以每小时100美元的价格让他们回答复杂科学问题，生产推理数据

成本高

- 数据公司 Labelbox 以时薪200美元雇佣会计师，让他们根据股票表现等数据，修正大模型分析特定公司前景的报告

Twitter首席执行官
埃隆·马斯克

“在AI训练中，我们现在基本上耗尽了人类知识的累积总和。”

OpenAI首席执行官
萨姆·奥尔特曼

“AI模型最终应该能够生成足够高质量地合成数据，以有效地完成自我训练。”

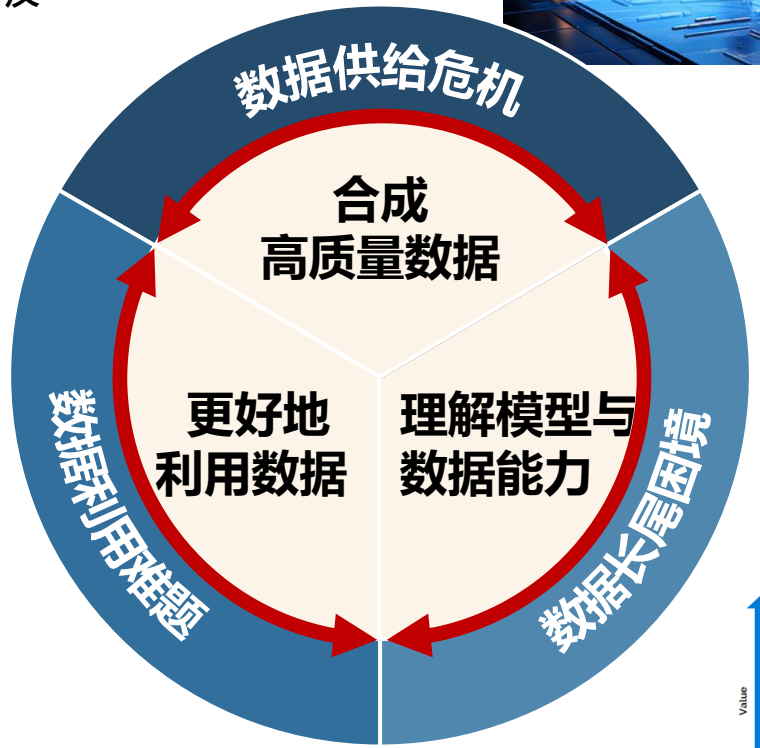
“我们已经达到了数据的峰值，必须利用现有的数据，因为互联网只有一个。”

大模型顶尖研究者
伊尔亚·苏茨克维

AI训练数据挑战及应对措施

➢ 数据供给危机

- 全球3.2万亿词元高质量文本消耗殆尽
- 高质量数据资源可能在2028年枯竭
- 必须依赖合成数据技术的突破

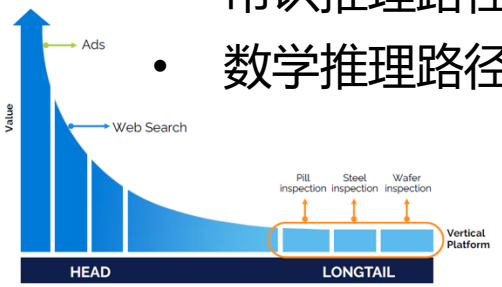


➢ 数据利用难题

- 低质量合成数据会引入噪声与偏差
- 干扰模型泛化能力、输出可靠性
- 数据资源竞争显著提升训练成本

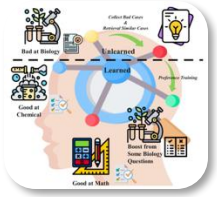
➢ 数据长尾困境

- 特定领域数据稀缺
- 专业场景标注成本高昂
- 常识推理路径重要但稀缺
- 数学推理路径合成质量有限

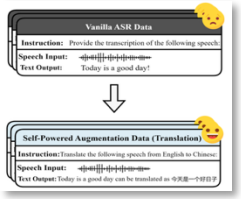


报告总览

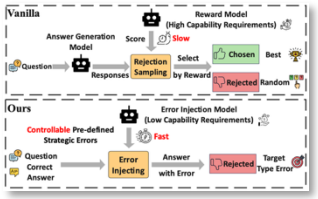
合成高质量数据



APT：弱点数据生成与迭代式能力对齐



Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

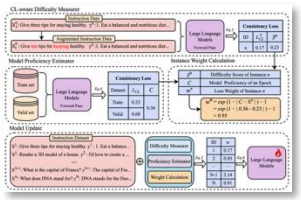


SeaPO：策略性错误放大的偏好数据合成

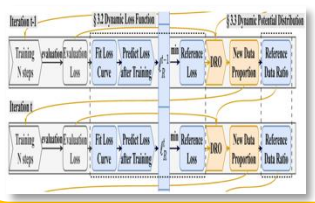
LongMT: CoT偏好数据广域搜索与细粒度策略合成

AQuilt：逻辑与反思增强的指令对齐数据合成

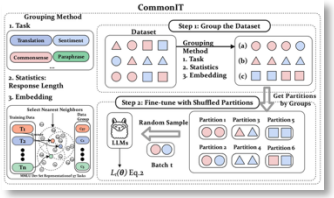
更好地利用数据



CCL：数据驱动的课程一致性学习

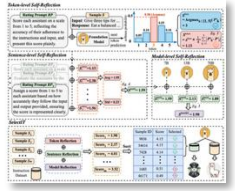


DRPruning：基于数据分布鲁棒的模型剪枝

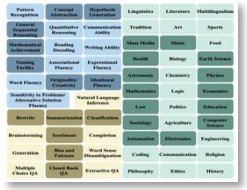


CommonIT：基于数据划分的共性感知指令微调方法

理解数据与模型能力之间的关系



SelectIT：不确定性感知的选择性指令数据微调

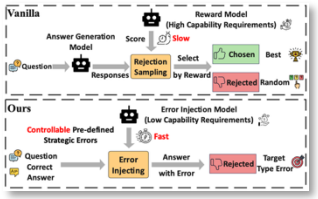


CDT：多维度的数据驱动大语言模型能力框架

本次报告介绍其中4个代表性的工作

- 4个工作以文本模态为主，期待加入多模态火花🔥！

合成高质量数据



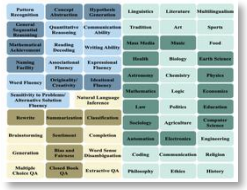
SeaPO：策略性错误放大的偏好数据合成

LongMT: CoT偏好数据广域搜索与细粒度策略合成

AQuilt：逻辑与反思增强的指令对齐数据合成

更好地利用数据

理解数据与模型能力之间的关系



CDT：多维度的数据驱动大语言模型能力框架

01 大模型与数据合成背景

02 基础：通用与垂域数据合成

03 核心：高效数据学习与利用

04 进阶：“数据-模型”能力对齐

05 领域瓶颈与未来展望

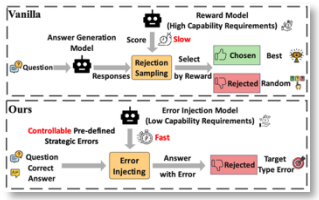
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

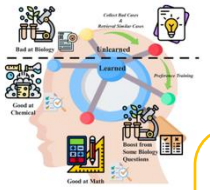
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

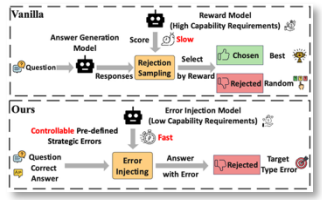
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

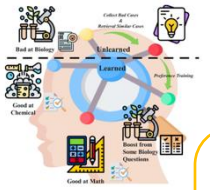
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

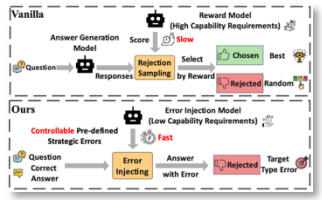
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

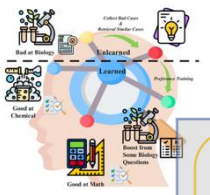
偏好数据

指令微调数据

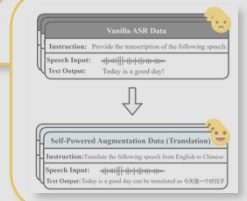
通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐



Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

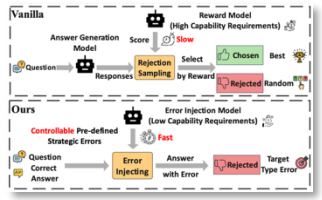
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

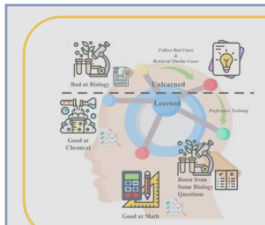
偏好数据

指令微调数据

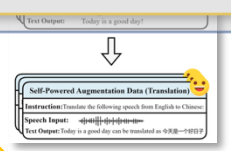
通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐



Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

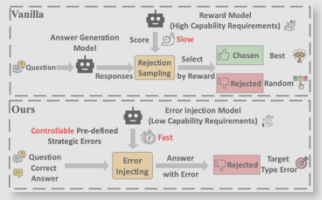
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

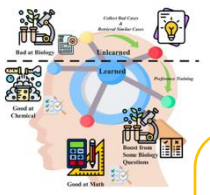
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

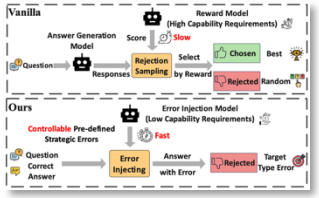
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

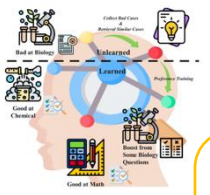
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



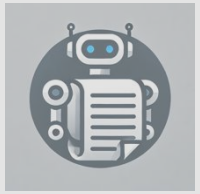
APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务



LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

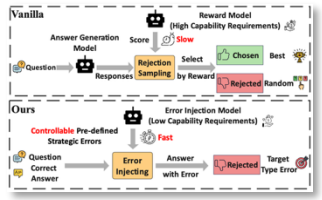
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

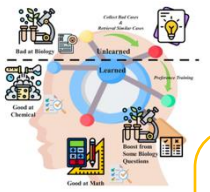
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

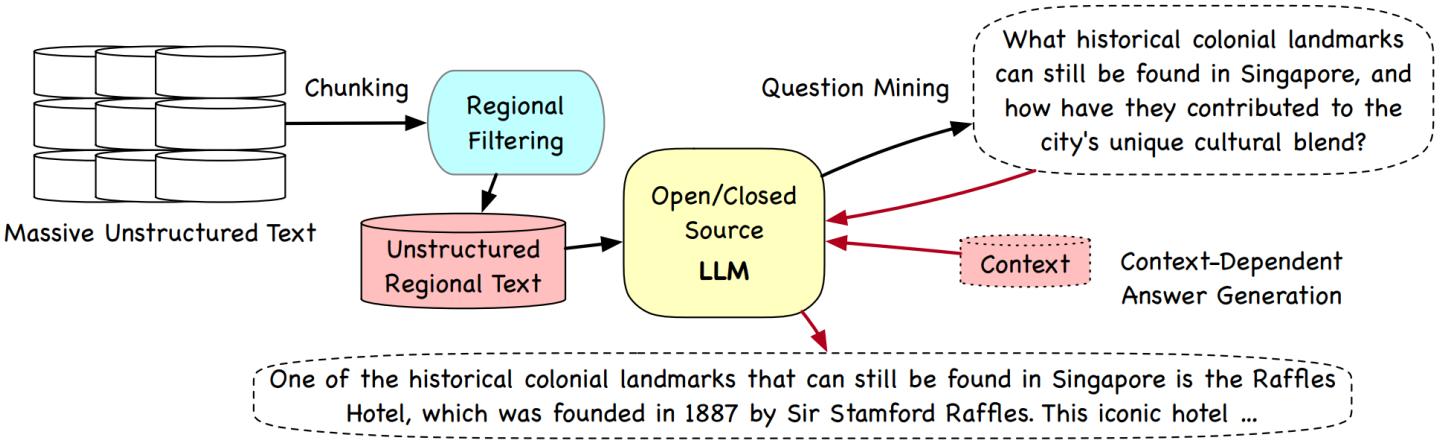
AQuilt: Weaving Logic and Self-Inspection into Low-Cost, High-Relevance Data Synthesis for Specialist LLMs

Xiaopeng Ke[1], Hexuan Deng[1], Xuebo Liu[1], Jun Rao[1], Zhenxi Song[1], Jun Yu[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen

EMNLP 2025

研究背景：基于无标签文本的数据合成技术

- **通用指令数据合成**
  - **不同场景和任务的无标签文本较为充足**
- **优点** 😊
  - 对齐现实数据风格
  - 数据源丰富，合成数据多样性强
- **缺点** 😭
  - 合成数据文本依赖性强
  - 需要根据实际需求进行过滤
  - 高性能大模型（如72B或以上）合成成本高、效率低

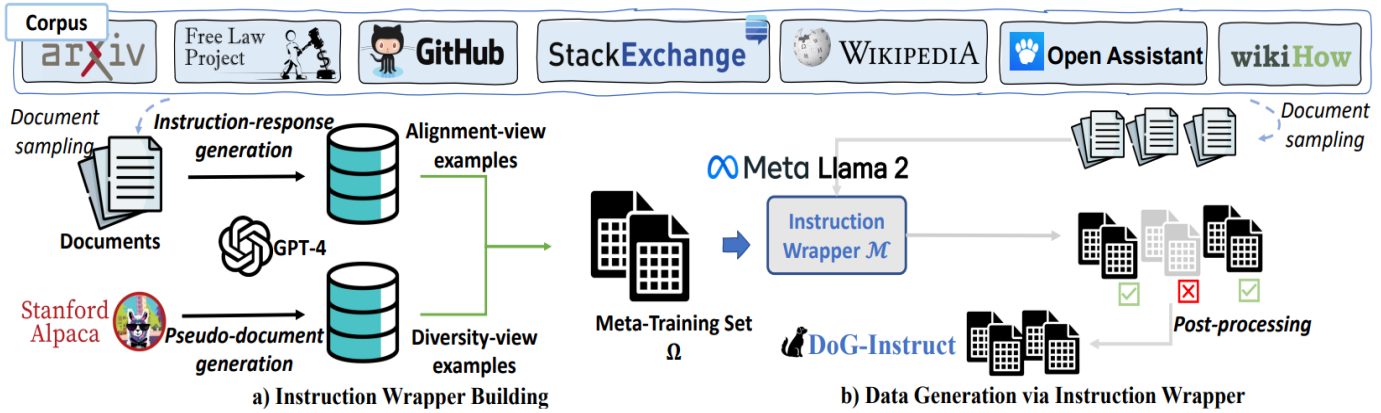


Wang et al., 2024[1]

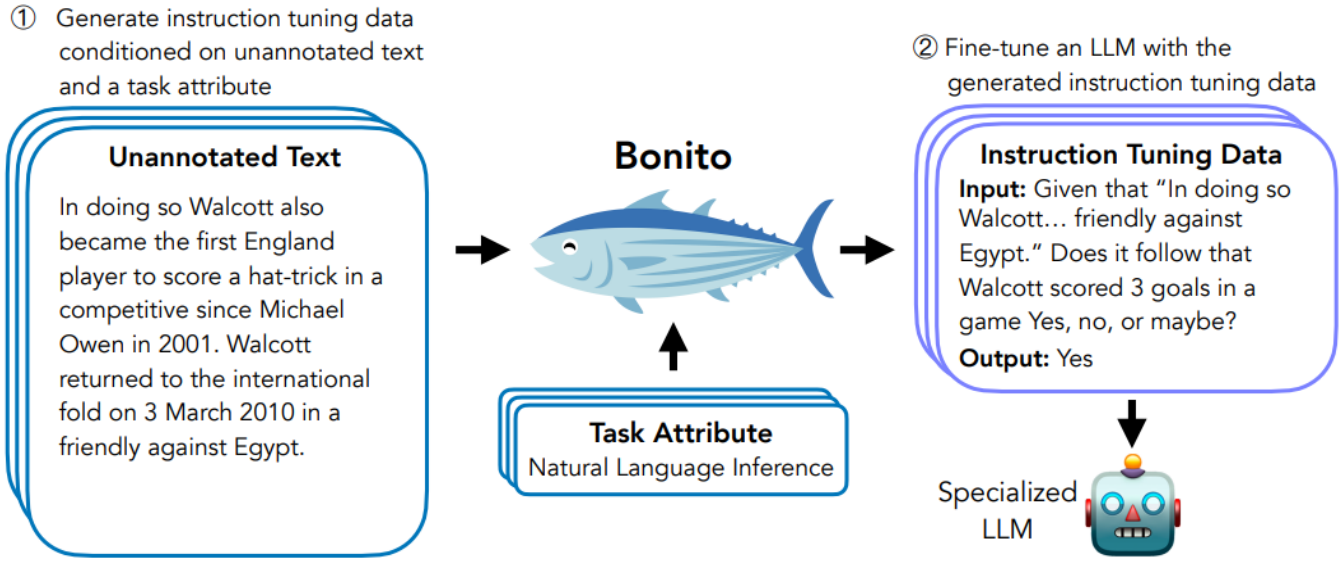
- Wang et al.通过few-shot的方式提示大模型基于无标签文本合成相关问答对
  - 实现简单，合成数据可控性较差

研究背景：数据合成能力蒸馏

- **轻量级数据合成模型蒸馏与训练**
  - 适用于大批量数据合成场景
  - 一般为7B左右大小
- **优点** 😊
  - 数据合成成本低
  - 合成效率高
- **缺点** 😭
  - 通用性和鲁棒性有待增强
    - Dog-Instruct无法指定任务类型
    - Bonito仅能合成阅读理解相关数据



Chen et al., NAACL2024[2]



Nayak et al., ACL2024 [3]

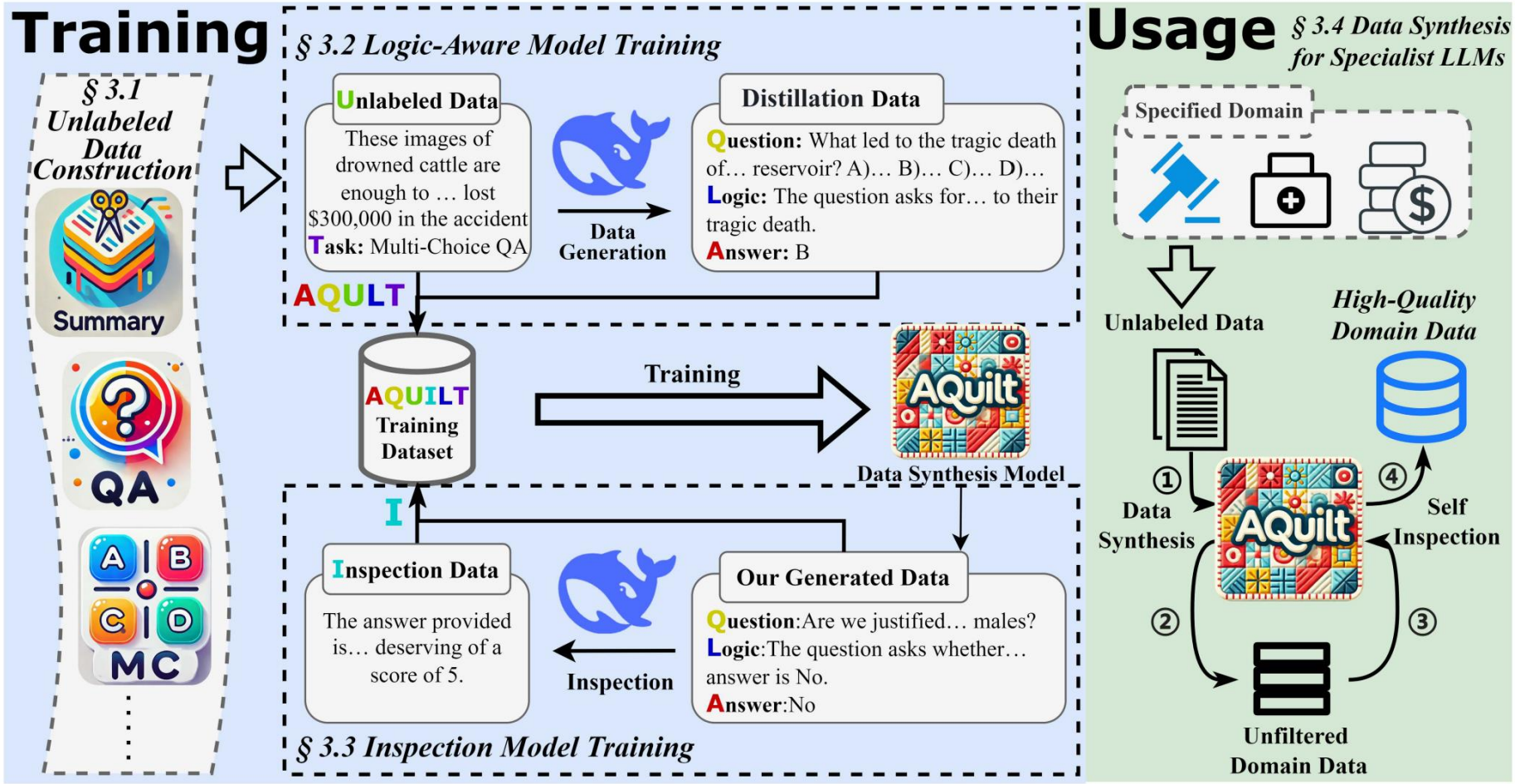
逻辑与反思增强的指令对齐数据合成（AQuilt）

- AQuilt框架图

- 无标签文本收集
  - 涵盖多领域文本

- 逻辑合成能力训练
  - 强化合成数据质量

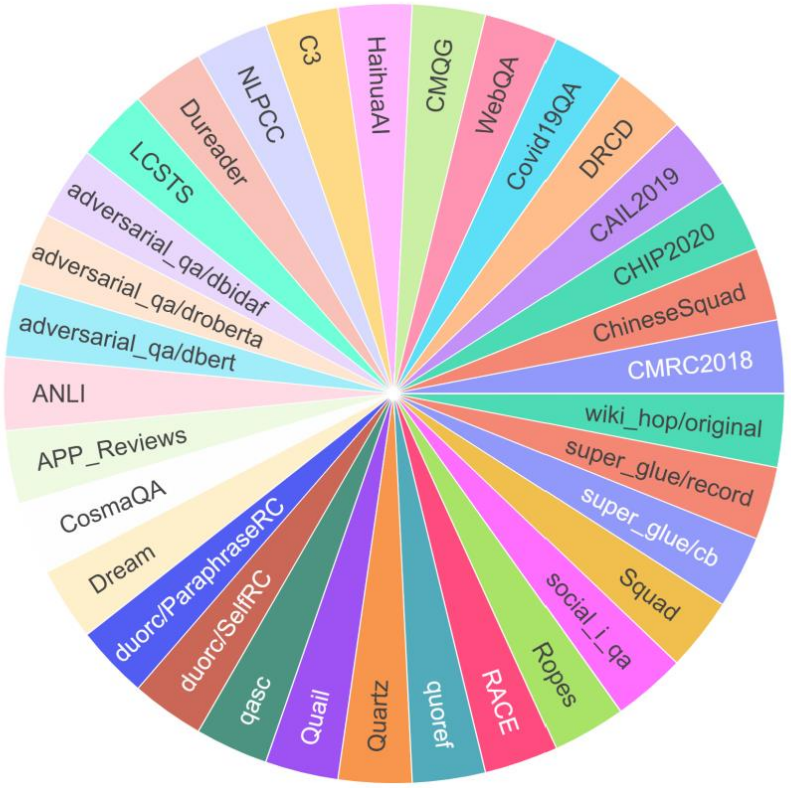
- 垂域任务微调
  - 多垂域任务验证有效性



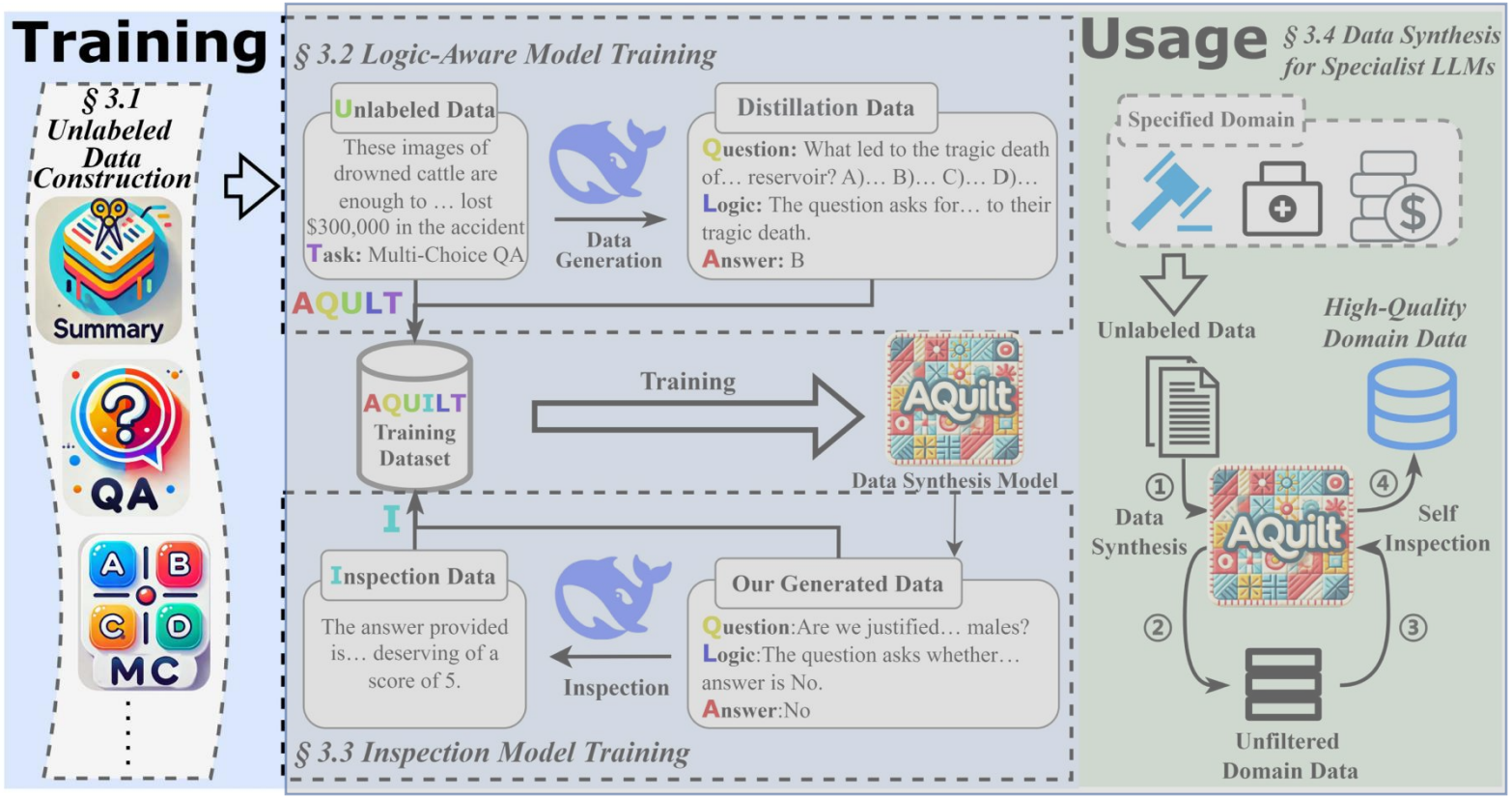
- 数据质量评估能力训练
  - 去除低分数据

逻辑与反思增强的指令对齐数据合成（AQuilt）

- **无标签文本收集**
  - 目的：收集无标签文本，为合成问答数据做准备
  - 需要覆盖多领域文本，提升模型面对不同形式文本的鲁棒性

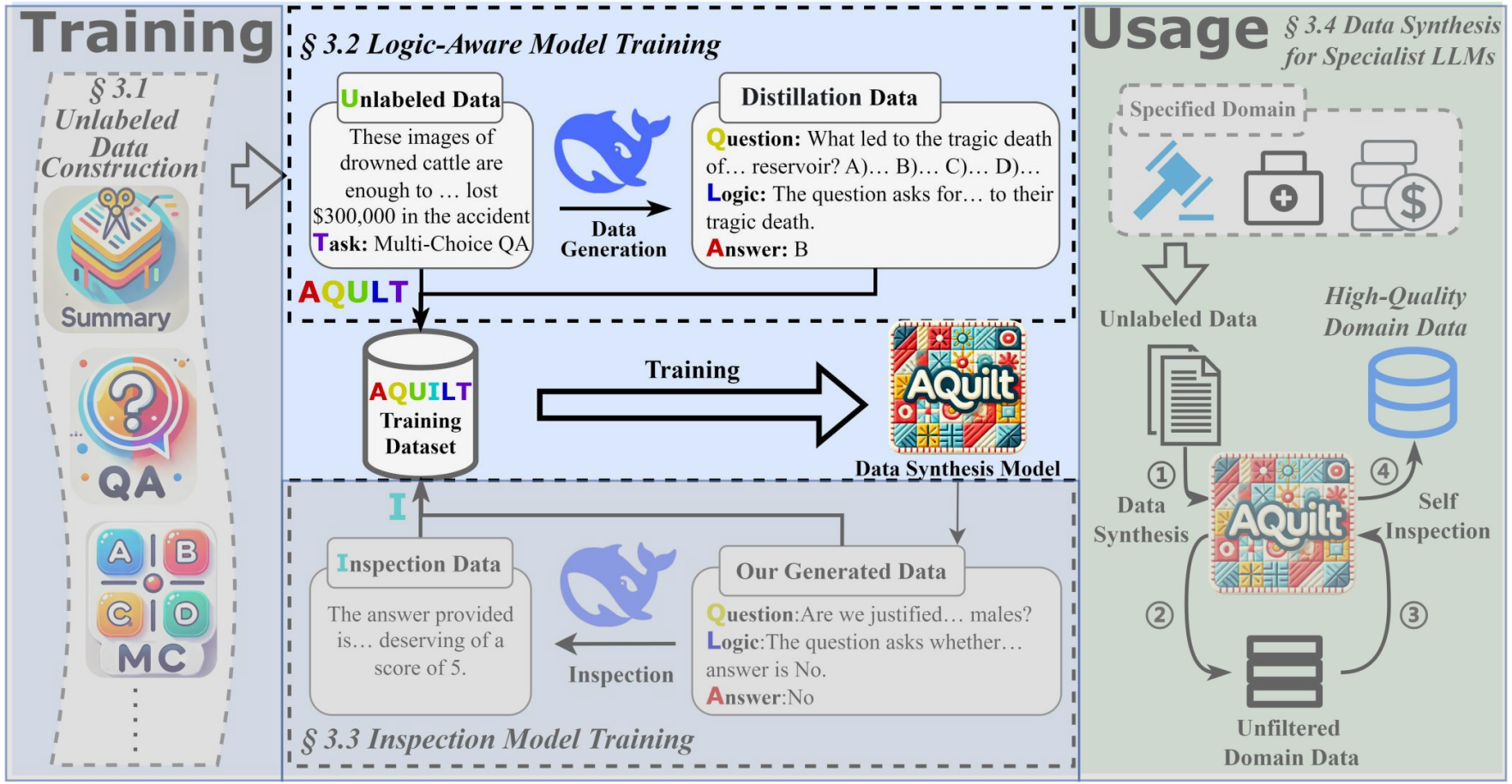


无标签文本来源



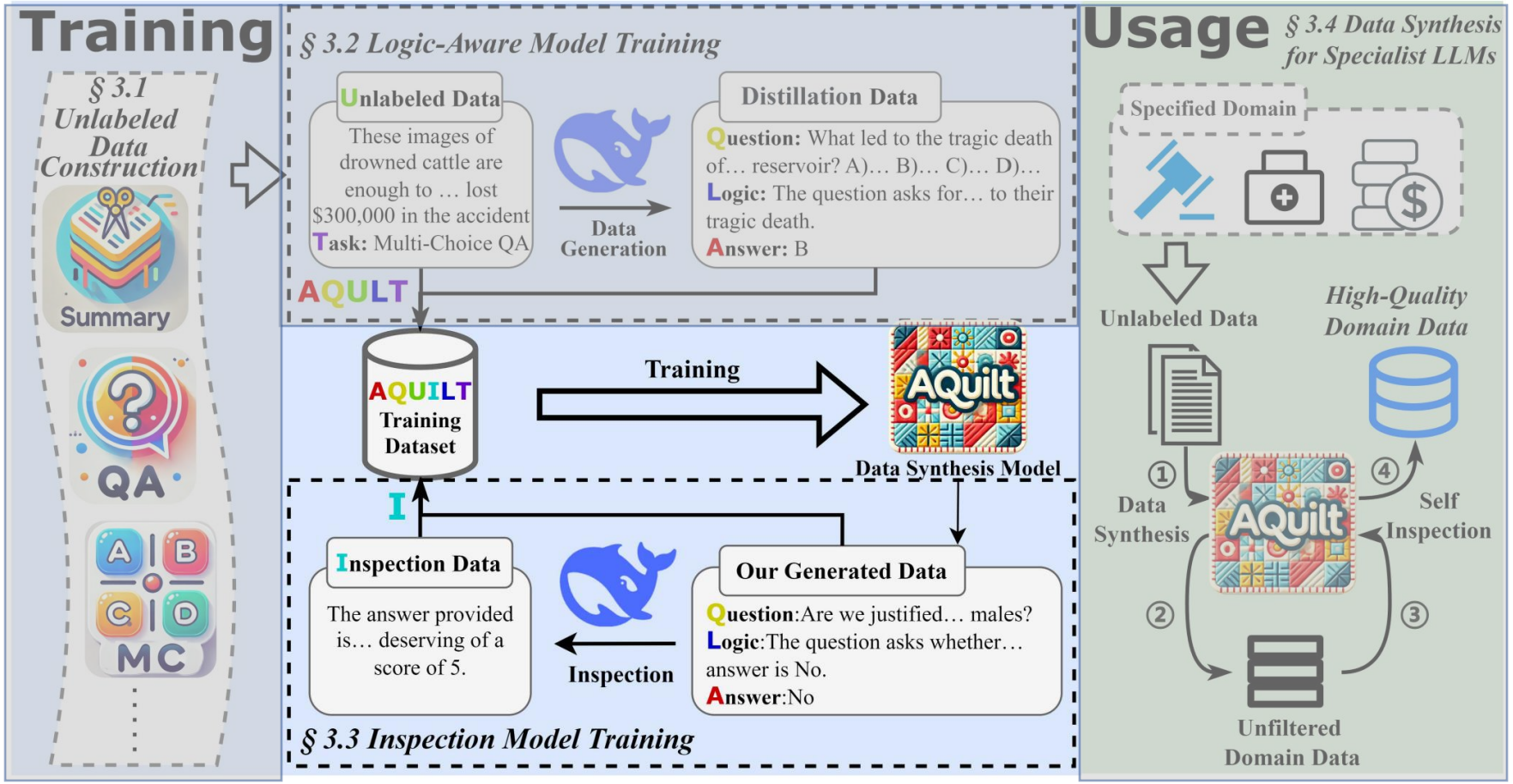
逻辑与反思增强的指令对齐数据合成（AQuilt）

- **逻辑合成能力训练**
  - $(a, q, l) = LLM_{Strong}^{GenData}(u, t)$ 借助强模型基于无标签文本(u)和任务类型(t)合成带有逻辑链(l)的指令对齐(q, a)数据
  - 同时进行多样性和文本依赖性的双层过滤以取得最终的元训练集



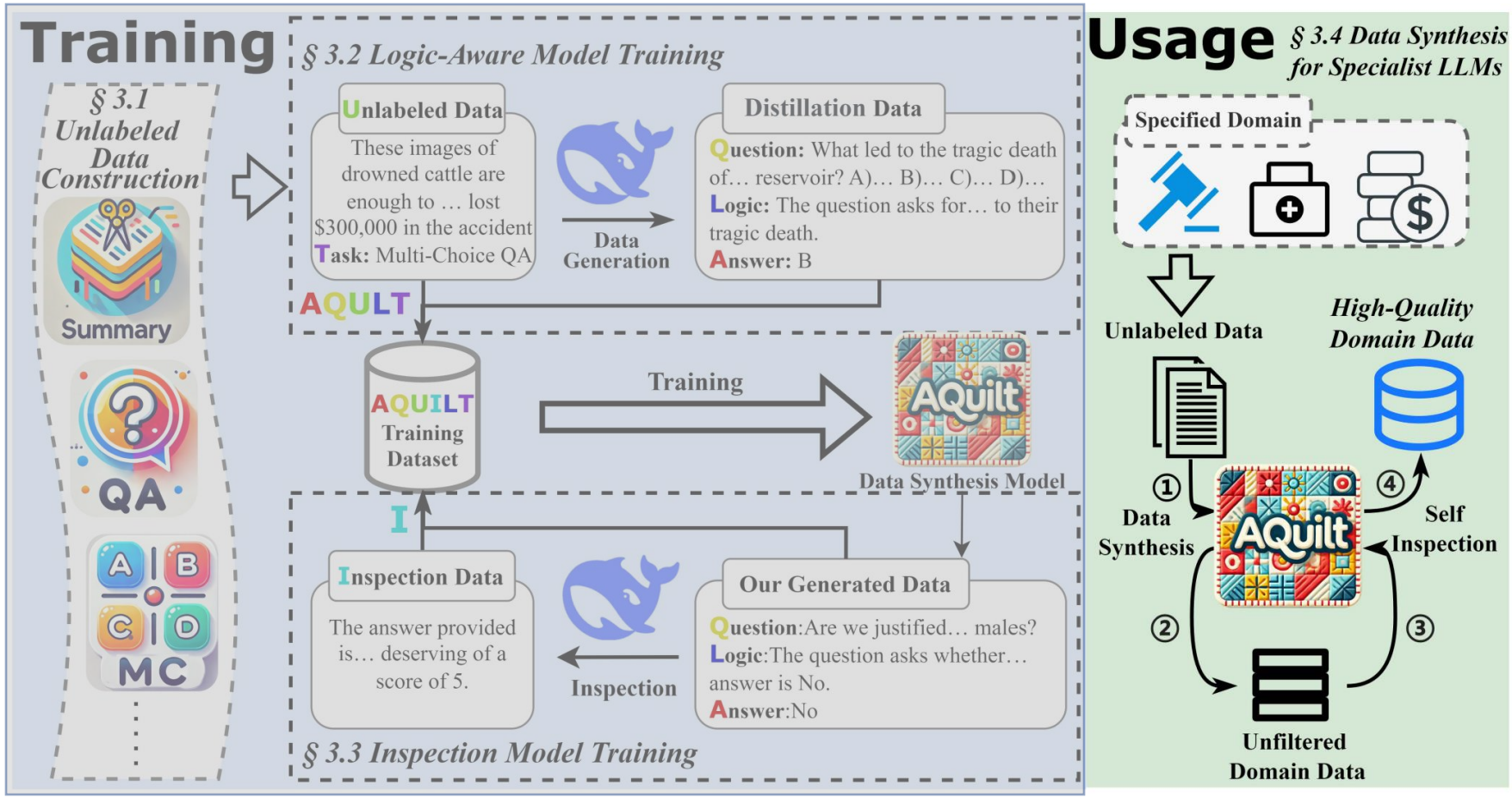
逻辑与反思增强的指令对齐数据合成（AQuilt）

- **数据质量评估能力训练**
  - 目的：赋予数据合成模型数据评估能力，不依赖外部模型自动去除低分数据
  - $i = LLM_{Strong}^{GenInsp}(a, q, u, l, t)$ 借助强模型基于AQuilt合成的数据构建质量评分数据
  - i表示评分数据，共分为5个质量等级，对每个质量等级进行下采样以均衡评分训练数据的分布



逻辑与反思增强的指令对齐数据合成（AQuilt）

- **垂域任务微调**
  - 目的：验证AQuilt数据合成模型在多垂域任务的有效性
  - 基于垂域文本和任务使用AQuilt合成数据，再通过AQuilt的自检模块过滤低分数据



AQuilt垂域任务微调实验效果

- **在多个领域任务数据集上进行测试：**
- 使用Qwen2.5-7B-Instruct和Llama3-8B-Instruct作为基模型

Model	Source	SquadQA		PubMedQA		CEVAL		Translation		EssayQA		Avg.	
		Score	Cost	Score	Cost	Score	Cost	Score	Cost	Score	Cost	Score	Cost
Qwen2.5-7B	None	3.12	0	56.60	0	87.46	0	32.23	0	19.21	0	39.72	0
	TAPT	3.16	0.90	56.40	1.13	87.72	2.32	33.00	1.88	19.11	1.61	39.88	1.57
	Bonito	22.78	1.19	71.40	1.42	NA	NA	NA	NA	NA	NA	NA	NA
	DeepSeek-V3 w/ Self-Instruct	NA	NA	74.00	10.40	87.81	16.73	36.95	20.31	**24.07**	26.02	NA	NA
	DeepSeek-V3 w/ Unlabeled Data	16.09	3.91	75.80	4.65	**88.55**	7.21	36.89	9.14	20.73	12.96	47.61	7.57
	DeepSeek-V3 w/ SI+UD	30.20	6.33	**76.80**	7.55	88.34	12.88	36.81	18.32	23.22	24.47	**51.47**	13.91
	AQuilt	**34.69**	1.48	74.00	1.75	88.44	2.90	**38.00**	2.44	22.11	2.25	51.45	2.16
Llama3-8B	None	3.68	0	73.60	0	58.32	0	27.79	0	15.37	0	35.75	0
	TAPT	3.67	1.22	73.60	1.56	58.50	3.73	28.13	3.07	15.20	2.38	35.82	2.39
	Bonito	23.05	1.51	72.20	1.85	NA	NA	NA	NA	NA	NA	NA	NA
	DeepSeek-V3 w/ Self-Instruct	NA	NA	74.80	10.75	61.96	17.95	34.07	21.50	21.26	26.57	NA	NA
	DeepSeek-V3 w/ Unlabeled Data	16.69	4.23	75.80	4.91	59.25	8.43	34.67	10.33	19.27	13.52	41.14	8.28
	DeepSeek-V3 w/ SI+UD	32.01	6.65	**76.40**	7.81	**64.16**	14.10	**35.57**	19.51	**21.93**	25.03	46.01	14.62
	AQuilt	**40.89**	1.79	75.20	2.10	63.16	3.91	34.50	3.35	19.65	2.77	**46.68**	2.78

使用蒸馏后的数据合成模型以显著更低的成本达到了和DeepSeek-V3可比的效果

消融实验设置

- **分析思维链和质量评估模块的影响：**
- 使用Llama3-8B-Instruct作为基模型

Model	SquadQA	CEVAL	Translation	Avg.
AQuilt	**40.89**	**63.16**	**34.50**	**46.18**
w/o Logic	40.68	59.64	33.61	44.64
w/o Self-Inspection	40.00	60.95	34.22	45.06
w/ Low-Quality	39.81	59.22	33.15	44.06

质量评估和思维链均会影响最终微调效果，是否包含思维链影响更大

合成数据特性分析

- **分析合成数据特性（合成数据的文本依赖性分析）：**
  - 基于无标签文本合成的问题容易出现文本依赖问题（依赖无监督文本中的特定知识）
- CEVAL、Translation、SquadQA三个不依赖文本的测试任务上进行分析

GPT-4o Prompts for Independence Analysis in Synthetic Data

```
You are a professional question analysis assistant, responsible for determining
whether a question relies on a text for its answer based on the provided question.
Criteria for Judgment:
The question contains some obvious keywords that indicate reliance on a text,
such as “the above content,” “according to the text,” “the above text,” “in the
text,” “in the passage,” etc.
If the question is about understanding or inquiring about the content of a
certain text, then it is also considered a question that relies on a text for
its answer.
Formatting Requirements:
Please carefully review the above criteria.
Determine whether the question provided by the user has text dependency.
If it does, please answer directly with ’Yes.’
If it does not, please answer directly with ’No.
```

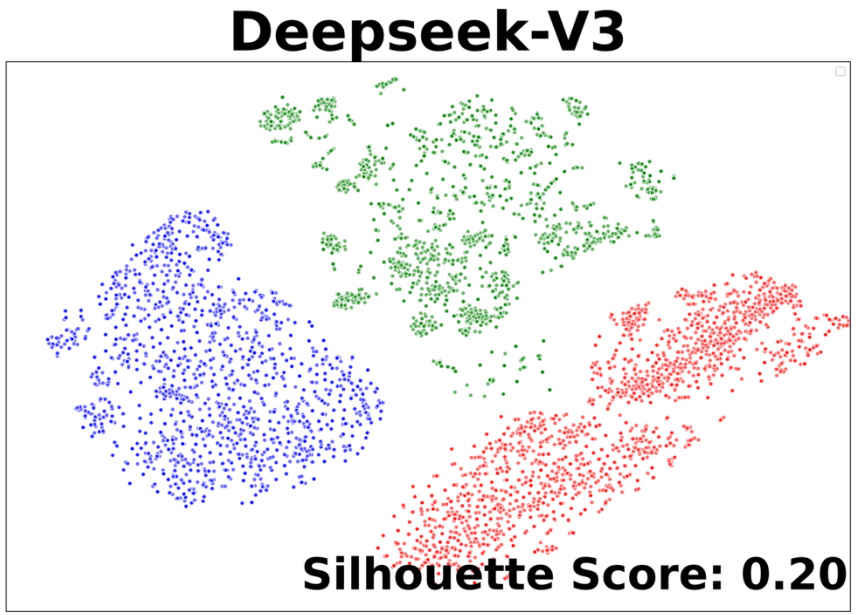
文本依赖性的评估Prompt (GPT-4o)

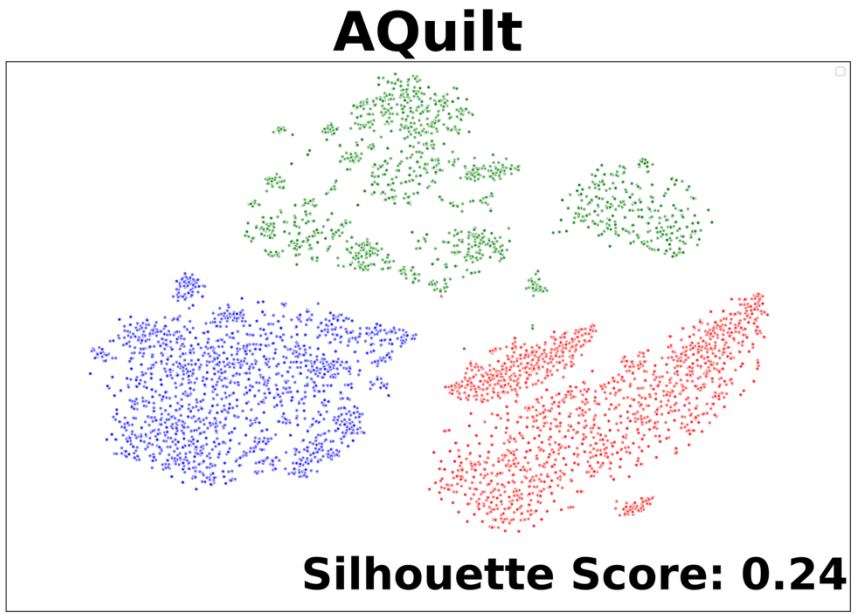
Model	SquadQA	CEVAL	Translation	Avg.
DeepSeek-V3	6.90%	8.18%	0.00%	5.23%
AQuilt	**0.40%**	**5.15%**	0.00%	**1.85%**

AQuilt合成的数据文本依赖性更弱
更适合不需要文本依赖的任务

合成数据特性分析

- **分析合成数据特性（使用t-SNE对数据向量进行降维）：**
- 红、绿、蓝分别代表CEVAL、Translation、SquadQA三个任务的合成数据





AQuilt合成的数据具备更强的领域相关性

与基座模型的对比分析

- **AQuilt的数据合成能力与基座模型的关系**
  - 对比分析更强的Qwen2.5-72B-Instruct

- **两个维度进行对比**
  - 性能和成本

- **Cost计算方式**
  - 相同实验条件下的时间花费

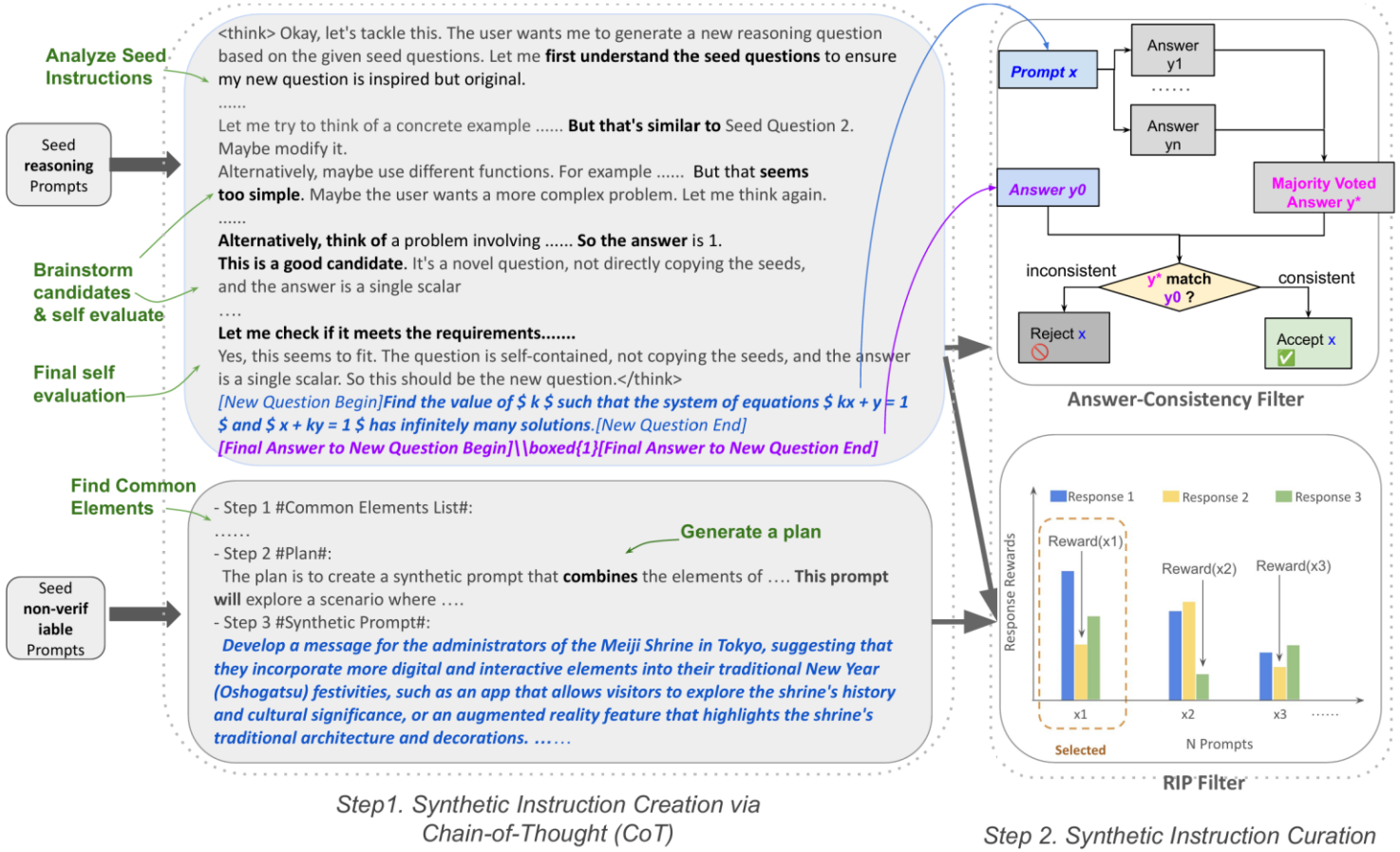
Source	SquadQA	CEVAL	Translation	Avg.	
				Score	Cost
Qwen2.5-72B	21.19	59.06	**34.82**	38.36	19.92×
AQuilt	**40.89**	**63.16**	34.50	**46.18**	**1**×

AQuilt合成的数据具备更优的性能和显著的成本优势

Meta论文简介：Cot-Self-Instruct

COT-SELF-INSTRUCT: BUILDING HIGH-QUALITY SYNTHETIC PROMPTS FOR REASONING AND NON-REASONING TASKS（2025年7月）

- **思维链引导的合成数据生成**
  - 先规划再合成，数据质量更加可控
- **任务感知的质量过滤策略**
  - 根据不同任务分别采用多数投票和外部奖励模型进行质量过滤



Meta AI, 2025.7 [4]

AQuilt 总结

- AQuilt是一种多领域数据合成模型与框架
  - 小规模数据合成模型构建 → 低成本数据合成，便捷高效
  - 思维链融入问答数据合成 → 提升问答逻辑性与推理能力
  - 自我检查能力训练 → 不依赖外部模型确保合成数据质量
- 多模态火花🔥
  - 低成本生成：轻量模型降低跨模态标注成本
  - 推理可解释性：结构化逻辑链增强多模态合成数据质量
  - 质量可控性：自检机制保障跨模态一致性

参考文献

[1] Ingo Ziegler, Abdullatif Köksal, Desmond Elliott, Hinrich Schütze. CRAFT Your Dataset: Task-Specific Synthetic Dataset Generation Through Corpus Retrieval and Augmentation. arXiv 2024.

[2] Yongrui Chen, Haiyun Jiang, Xinting Huang, Shuming Shi, Guilin Qi. DoG-Instruct: Towards Premium Instruction-Tuning Data via Text-Grounded Instruction Wrapping. NAACL 2024.

[3] Nihal Nayak, Yiyang Nan, Avi Trost, Stephen Bach. Learning to Generate Instruction Tuning Datasets for Zero-Shot Task Adaptation. ACL 2024.

[4] Ping Yu, Jack Lanchantin, Tianlu Wang, Weizhe Yuan, Olga Golovneva, Ilia Kulikov, Sainbayar Sukhbaatar, Jason Weston, Jing Xu. Cot-Self-Instruct: Building High-Quality Synthetic Prompts for Reasoning and Non-Reasoning Tasks. arXiv 2025.

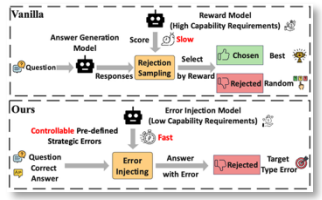
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

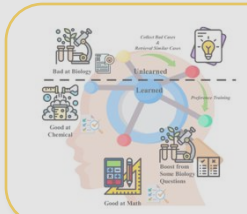
偏好数据

指令微调数据

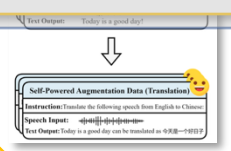
通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐



Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

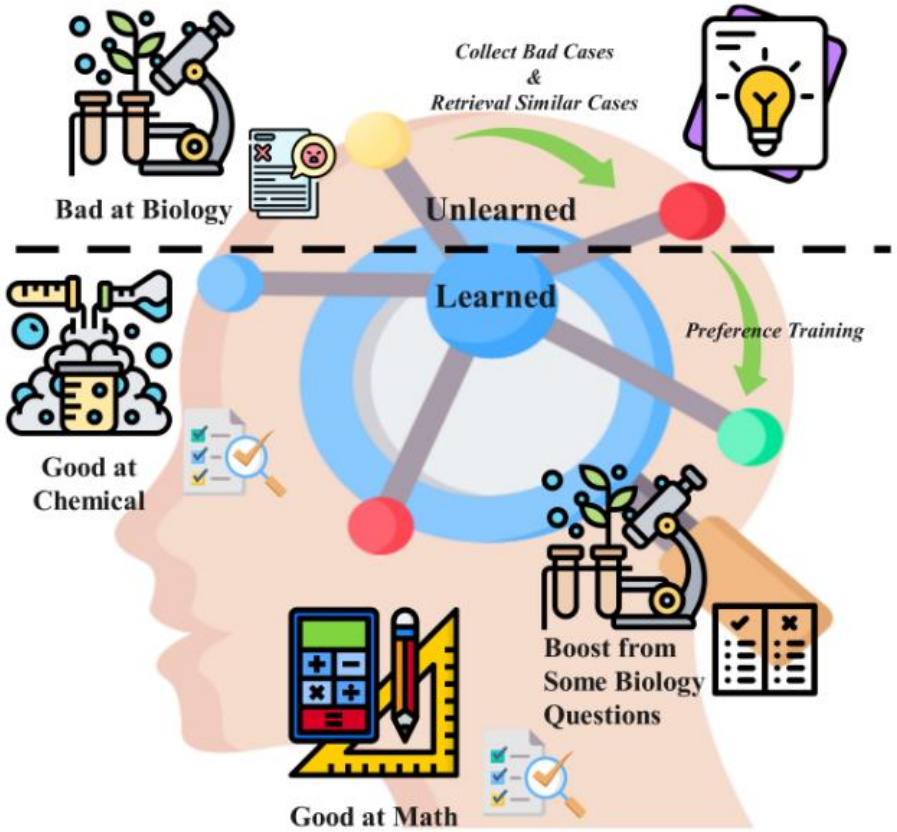
APT: Improving Specialist LLM Performance with Weakness Case Acquisition and Iterative Preference Training

Jun Rao[1], Zepeng Lin[1], Xuebo Liu[1], Xiaopeng Ke[1], Lian Lian[2], Dong Jin[2], Shengjun Cheng[2], Jun Yu[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]Huawei Cloud Computing Technologies Co., Ltd.

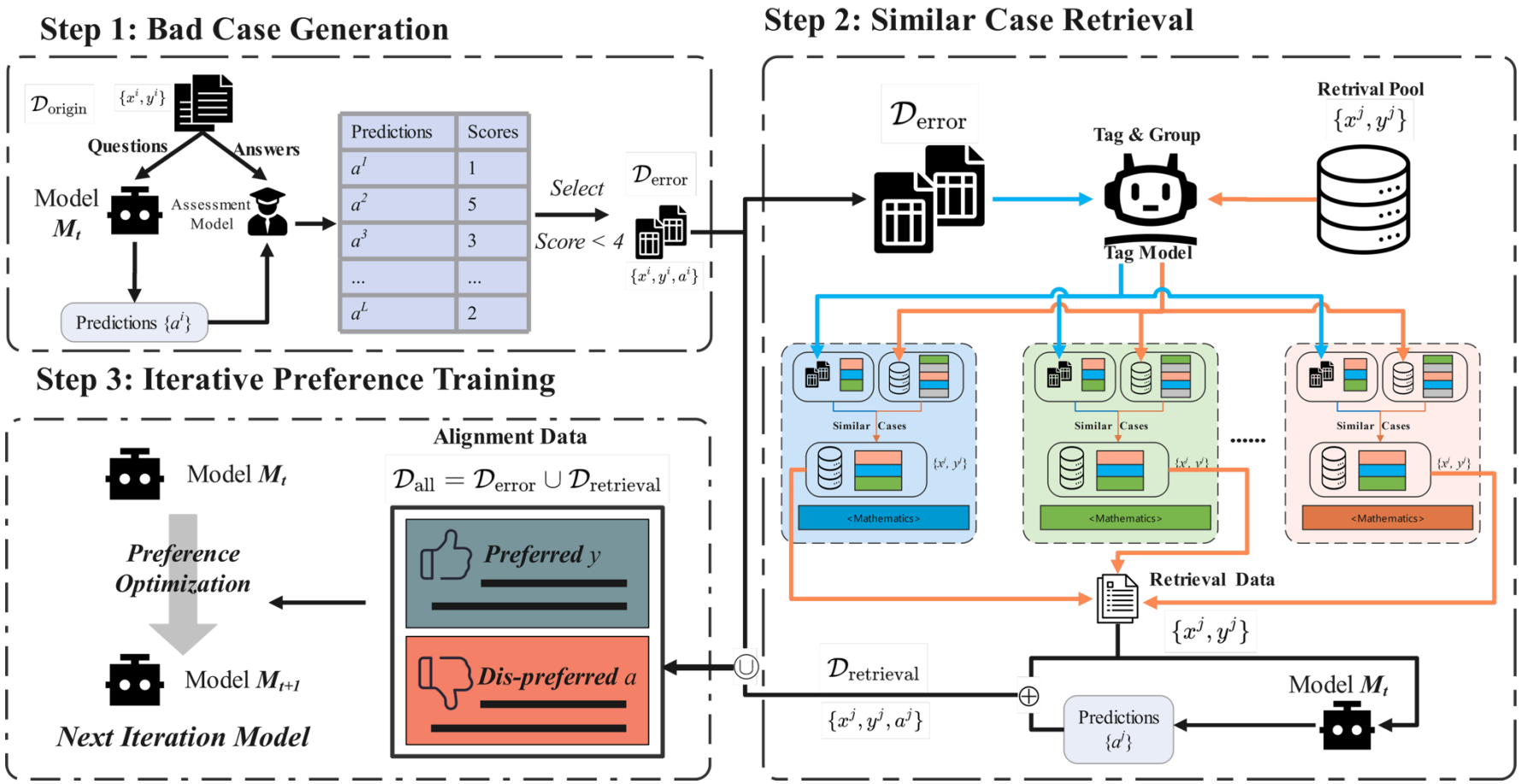
ACL 2025

背景：模型在垂直领域的精细化训练和能力提升

- **模型垂域精细化训练**
  - 通过精细化构建垂直领域数据，微调通用模型在垂域的进一步适配和优化。
  - **优点**
    - 😊 简单有效、收敛快
  - **缺点**
    - 😭 泛化能力减弱，模型过度拟合
- **APT**
  - **目标**
    - 😊 识别模型弱点数据，迭代能力提升

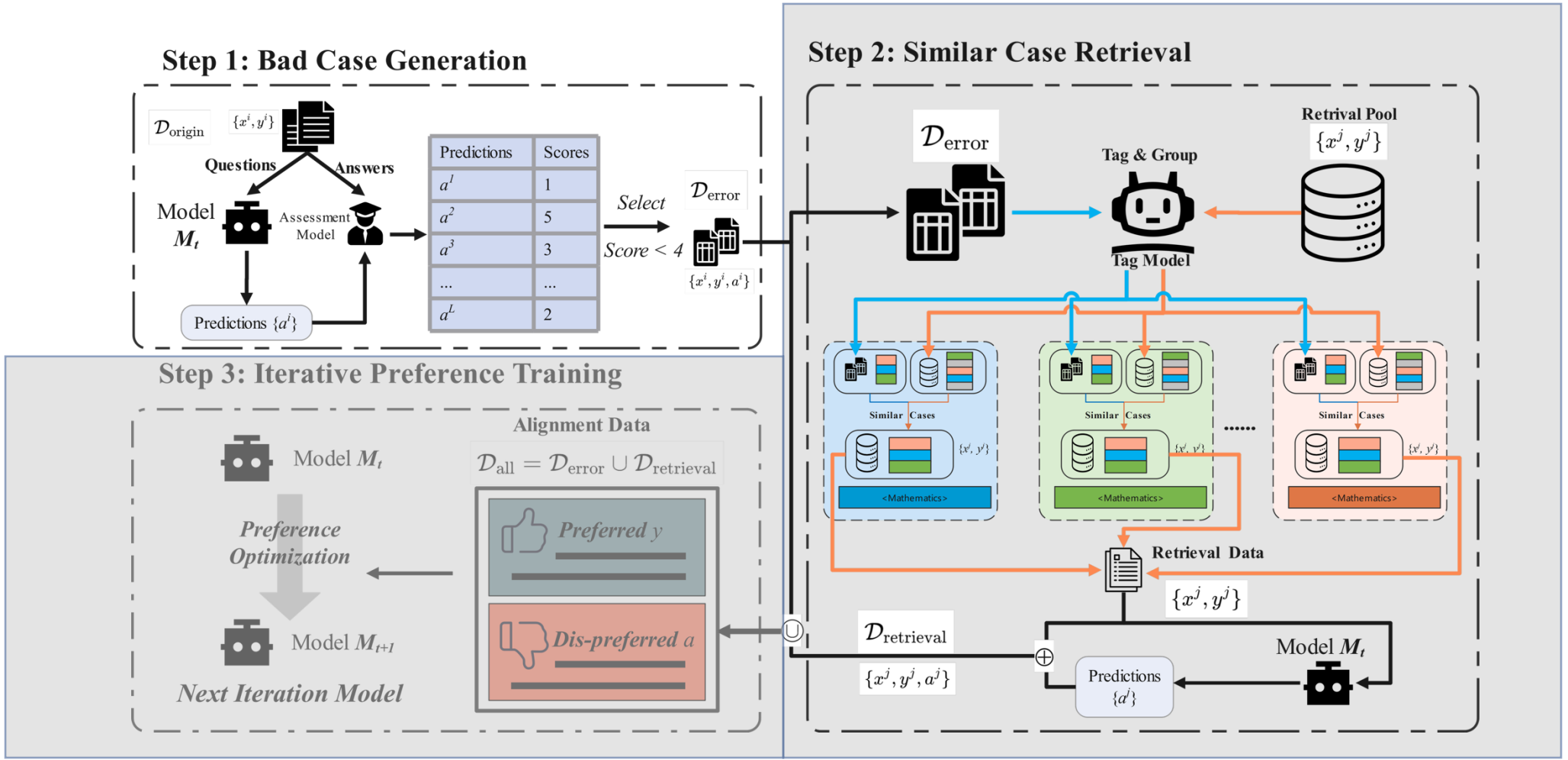


- **弱点数据生成**
  - 评估模型识别低分数据
- **相似弱点数据构建**
  - TAG级检索
- **偏好迭代优化**
  - SFT loss正则



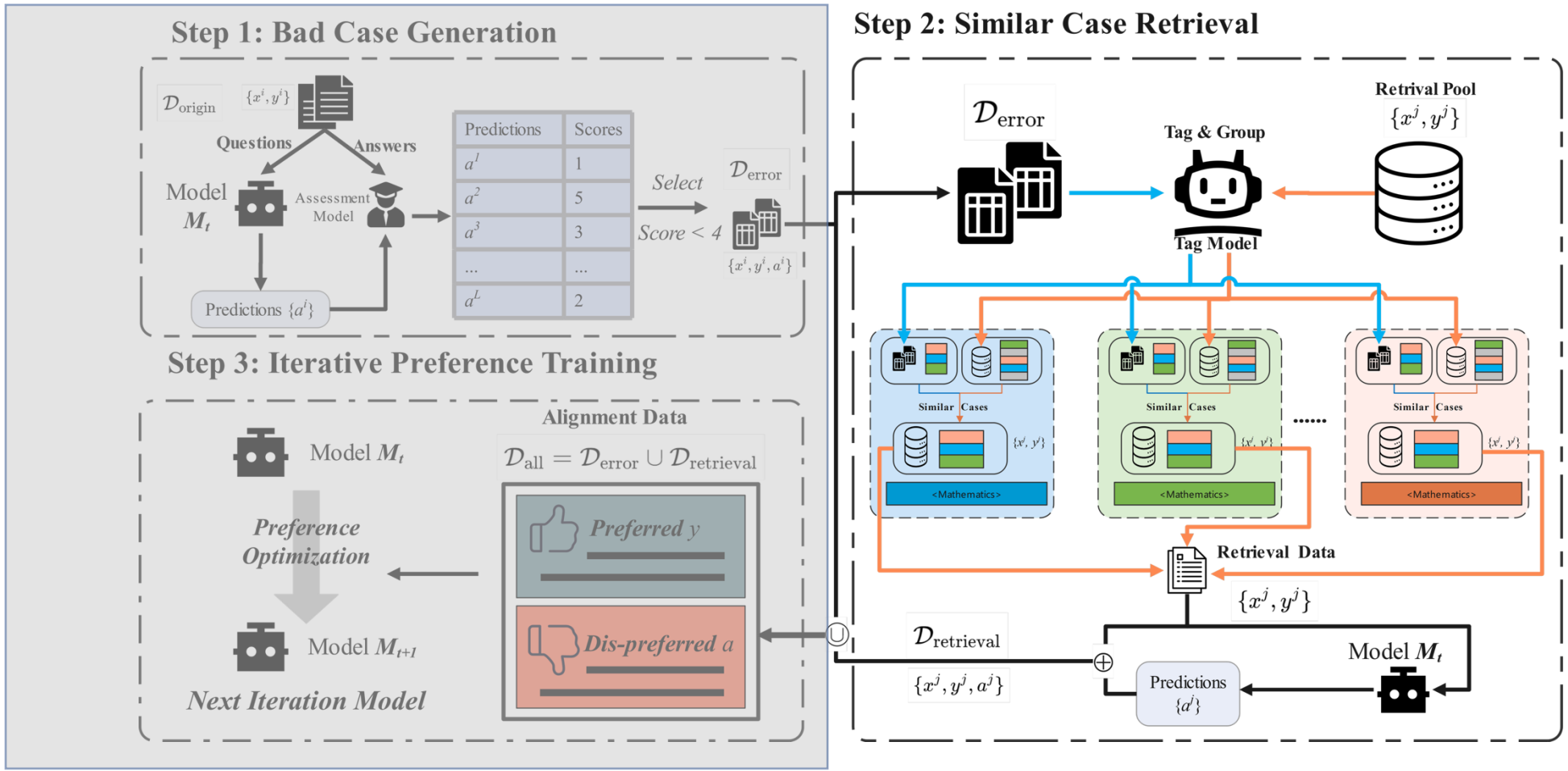
评估：弱点数据偏好迭代样本构造

- **弱点数据生成**构建过程
  - 目的：识别模型弱点数据，为迭代偏好训练做准备
  - 通过评估模型评价模型响应，选择不良响应构建偏好迭代数据



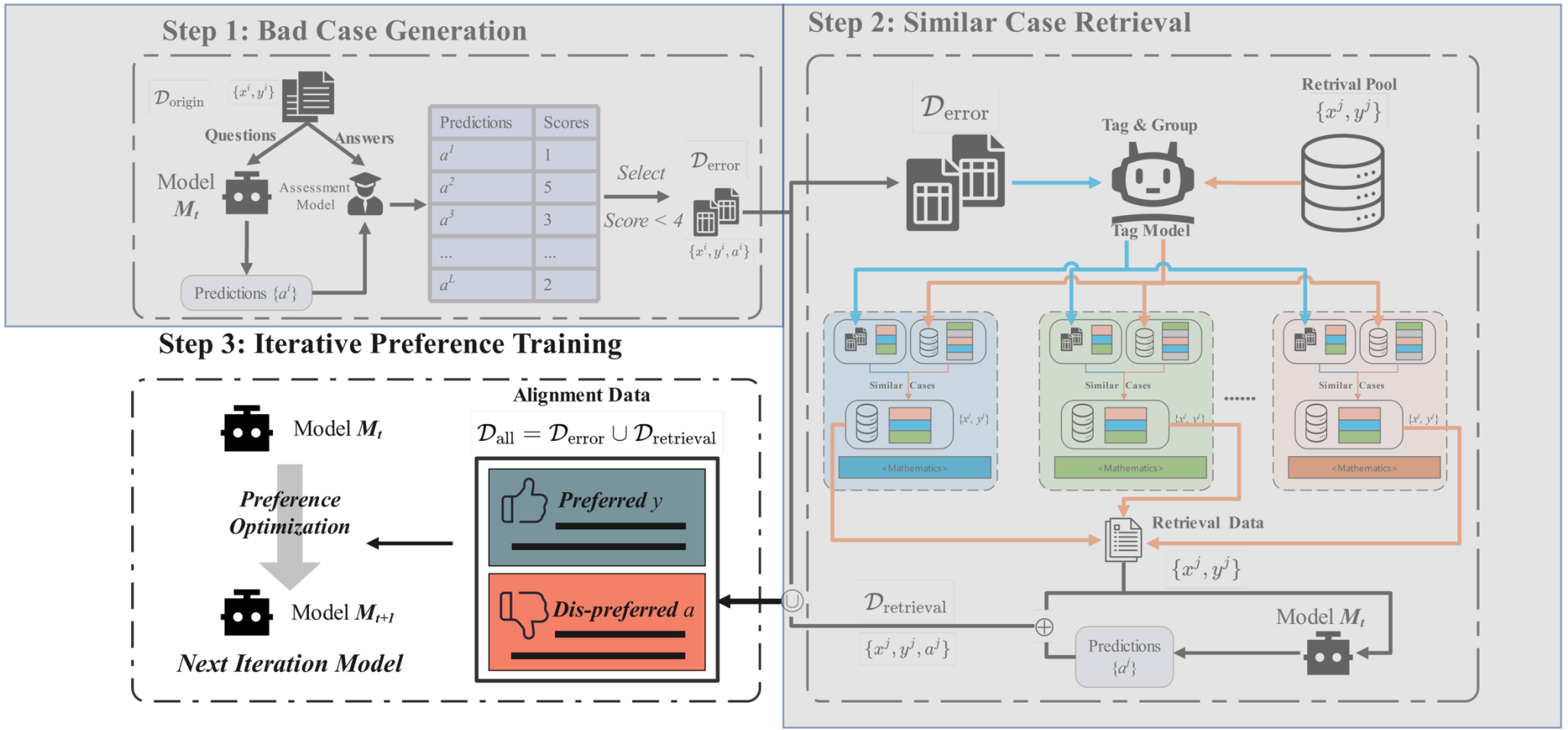
检索：相似数据扩展构造

- **相似弱点数据构建**
  - 目的：通过**扩展数据集**来提升模型垂直领域的能力和**泛化性能**
  - 通过TAG操作实现更细粒度的分组检索，提升检索准确性



迭代偏好训练：迭代提升模型垂域性能

- **迭代偏好训练**
  - 目的：通过迭代偏好训练提升模型垂域能力
  - 引入SFT损失以增强训练的鲁棒性：$L(\theta) = L_{\text{DPO}}(\theta, \theta_{ref}) + \alpha L_{\text{SFT}}(\theta)$



垂域性能提升

- **APT在数学、代码、对话垂域的效果**

APT在不同基座模型和垂域上显著超过基线模型与已有方法（DMT）

Model	Method	Domain Dataset	Math Reasoning	Coding	Instruction Following	General Capability	AVG
LLama2-7B	**Base**	-	15.5	19.2	7.5	56.9	24.8
	SFT	GSM	32.1	18.5	34.3	58.5	35.9
		CodeAlpaca	17.7	23.6	29.2	58.7	32.3
		Dolly	19.3	20.9	16.9	59.1	29.1
	Mixed Training	-	34.1	25.0	19.1	59.1	34.3
	+Continued SFT	GSM	32.6	27.4	19.5	59.0	34.6
		CodeAlpaca	31.9	28.1	21.1	58.9	35.0
		Dolly	33.2	28.1	18.0	59.1	34.6
	+DMT (Dong et al., 2024)	GSM	34.0	26.9	18.5	59.0	34.6
		CodeAlpaca	33.4	26.4	18.3	58.5	34.2
		Dolly	33.7	26.0	18.0	58.4	34.0
	+Ours	GSM	**39.2** (+5.1)	26.1	24.2	**59.4**	**37.2** (+2.9)
		CodeAlpaca	34.8	**28.4** (+3.4)	24.3	**59.3**	**36.7** (+2.4)
		Dolly	34.7	27.4	**25.0** (+5.9)	**59.3**	**36.6** (+2.3)
Mistral-7B-V0.3	**Base**	-	40.6	33.1	16.4	64.5	38.7
	SFT	GSM	56.9	41.1	33.8	**66.5**	49.6
		CodeAlpaca	46.4	43.8	30.3	66.1	46.7
		Dolly	43.8	41.0	26.0	**66.8**	44.4
	Mixed Training	-	58.2	43.7	30.0	66.4	49.6
	+Continued SFT	GSM	58.9	43.6	29.4	66.3	49.6
		CodeAlpaca	58.9	44.6	31.2	66.3	50.3
		Dolly	58.7	44.1	27.2	66.3	49.1
	+DMT (Dong et al., 2024)	GSM	59.3	44.5	30.1	66.1	50.0
		CodeAlpaca	59.3	42.9	29.0	65.9	49.3
		Dolly	59.1	43.6	29.1	66.0	49.5
	+Ours	GSM	**61.8** (+3.6)	45.9	30.6	**66.5**	**51.2** (+1.6)
		CodeAlpaca	60.2	**49.7** (+6.0)	31.2	**66.7**	**52.0** (+2.4)
		Dolly	59.6	47.6	**35.0** (+5.0)	**66.8**	**52.3** (+2.7)

泛化性能的保持

- **APT在多个通用数据集上的测试结果**

APT在广泛的通用能力项的衡量上基本保持通用能力不下降

Method	Domain Dataset	MMLU	BBH	ARC-e	ARC-c	Boolq	OpenBookQA	WinoGrande	Avg.
LLama2-7B	-	45.2	40.7	74.6	46.3	77.7	44.2	69.3	56.9
SFT	GSM	46.7	39.2	76.8	49.7	79.9	47.4	69.7	58.5
	CodeAlpaca	47.2	40.6	76.9	50.9	79.3	46.8	69.1	58.7
	Dolly	48.2	39.3	78.3	52.5	79.3	47.0	69.5	59.1
Mixed Training	-	45.6	39.2	79.0	53.1	80.2	47.0	69.6	59.1
+Continued SFT	GSM	45.8	39.4	78.5	51.9	80.2	47.2	69.9	59.0
	CodeAlpaca	45.6	39.8	78.2	51.3	80.4	46.8	70.3	58.9
	Dolly	45.5	40.0	78.8	52.4	80.5	46.8	69.6	59.1
+DMT (Dong et al., 2024)	GSM	45.4	37.8	78.7	53.0	81.4	47.2	69.8	59.0
	CodeAlpaca	45.5	37.8	76.9	51.4	80.7	47.4	69.5	58.5
	Dolly	45.3	38.1	77.0	50.9	80.6	47.2	69.9	58.4
+Ours	GSM	45.4	39.4	78.9	52.6	81.0	48.0	70.9	**59.4**
	CodeAlpaca	45.2	39.7	78.7	52.7	80.7	48.0	70.2	**59.3**
	Dolly	45.1	39.0	79.4	53.4	80.5	47.8	69.9	**59.3**
Mistral-7B-V0.3	-	62.5	58.0	78.4	52.6	82.2	44.0	73.6	64.5
SFT	GSM	61.9	58.2	81.8	56.4	84.5	46.8	75.5	**66.5**
	CodeAlpaca	62.1	58.2	80.4	56.1	84.5	46.4	75.3	66.1
	Dolly	60.7	58.8	82.1	58.2	85.1	47.0	75.5	**66.8**
Mixed Training	-	61.0	58.5	81.7	57.3	85.6	46.4	74.4	66.4
+Continued SFT	GSM	60.8	58.5	80.9	56.7	85.4	46.2	75.3	66.3
	CodeAlpaca	61.0	58.5	81.2	56.7	85.5	46.2	74.8	66.3
	Dolly	61.0	58.2	81.3	56.6	85.4	46.2	75.1	66.3
+DMT (Dong et al., 2024)	GSM	61.0	59.5	79.0	55.4	85.9	46.8	74.7	66.1
	CodeAlpaca	60.9	58.5	78.6	55.3	86.2	47.2	74.7	65.9
	Dolly	61.0	59.6	78.9	55.6	86.0	46.6	74.7	66.0
+Ours	GSM	60.8	58.2	80.6	57.9	86.5	46.6	75.0	**66.5**
	CodeAlpaca	60.9	59.1	81.6	57.5	86.0	46.8	74.7	**66.7**
	Dolly	60.7	58.8	81.9	58.5	86.2	46.4	74.8	**66.8**

- **评估模型的选择**
  - 自评估 vs. 先进的评估模型

先进的评估模型效果更优

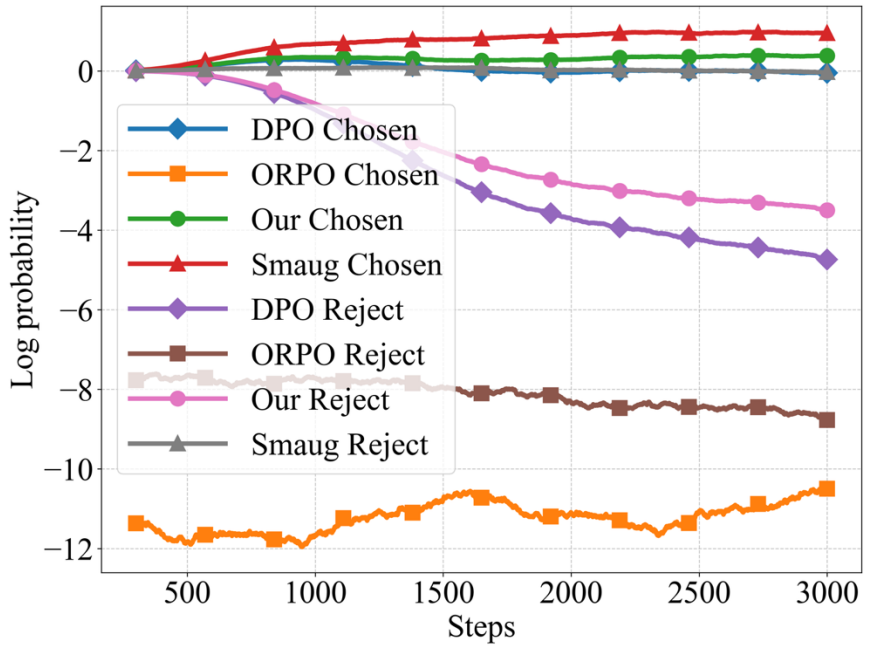
Domain	Self		Prometheus	
	Domain	AVG	Domain	AVG
GSM	34.8	34.6	**39.2**	**37.2**
CodeAlpaca	26.0	34.7	**28.4**	**36.7**
Dolly	18.7	34.9	**25.0**	**36.6**

- **检索方法的选择**
  - 相似性检索 vs. TAG检索

TAG检索效果更优

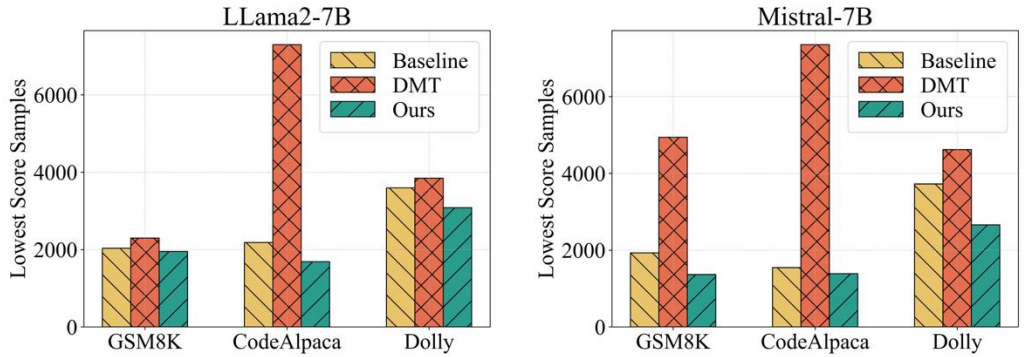
Domain Dataset	Mixed Training	Only Error	Mean Vector	Cluster Based	Tag Based
GSM	34.1	36.3	38.1	38.1	**38.1**
CodeAlpaca	25.0	26.6	27.7	26.9	**28.2**
Dolly	19.1	23.5	21.4	21.7	**23.7**

偏好训练优化目标



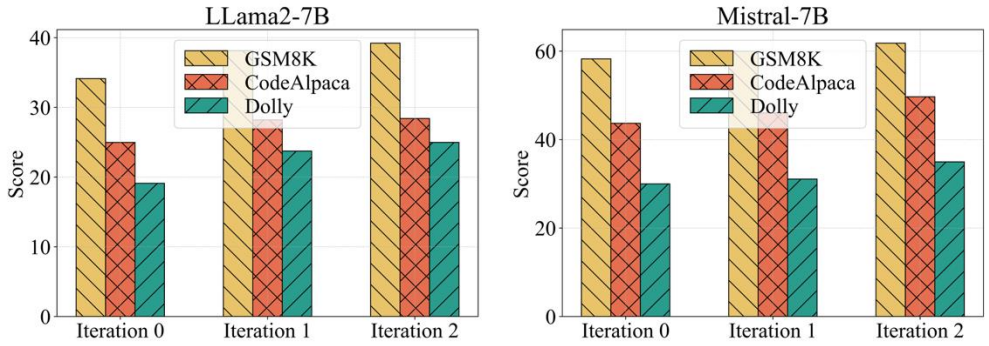
APT提升选定奖励，降低拒绝奖励

降低弱点数据比例



APT有效降低模型的弱点数据比例

迭代稳定提升



APT持续提升性能

APT 总结

- APT 是一种迭代偏好训练方法
  - 低成本识别弱点数据 → 避免多余无效数据训练
  - 相似数据检索构建 → 扩充检索数据提升模型泛化
  - 迭代偏好提升 → 多轮迭代减少缺陷数据
- APT落地：
  - 模型：7B-V3版本的盘古金融L1行业模型
  - 测评：相较于GPT4，行业平均分位值从96提升至99

参考文献

[1] Guanting Dong, Hongyi Yuan, Keming Lu, Chengpeng Li, Mingfeng Xue, Dayiheng Liu, Wei Wang, Zheng Yuan, Chang Zhou, Jingren Zhou. How abilities in large language models are affected by supervised fine-tuning data composition. ACL 2024.

[2] Seungone Kim, Juyoung Suk, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo. Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models. ACL 2024.

[3] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, John Schulman. Training verifiers to solve math word problems. arXiv 2021.

[4] Yifan Xu, Xiao Liu, Xinghan Liu, Zhenyu Hou, Yueyan Li, Xiaohan Zhang, Zihan Wang, Aohan Zeng, Zhengxiao Du, Zhao Wenyi, Jie Tang, and Yuxiao Dong. ChatGLM-Math: Improving Math Problem-Solving in Large Language Models with a Self-Critique Pipeline. ACL 2024.

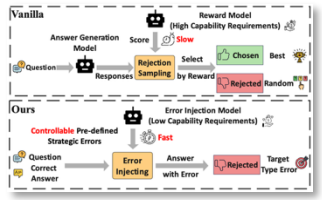
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

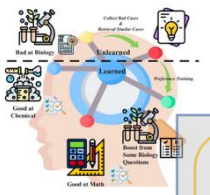
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

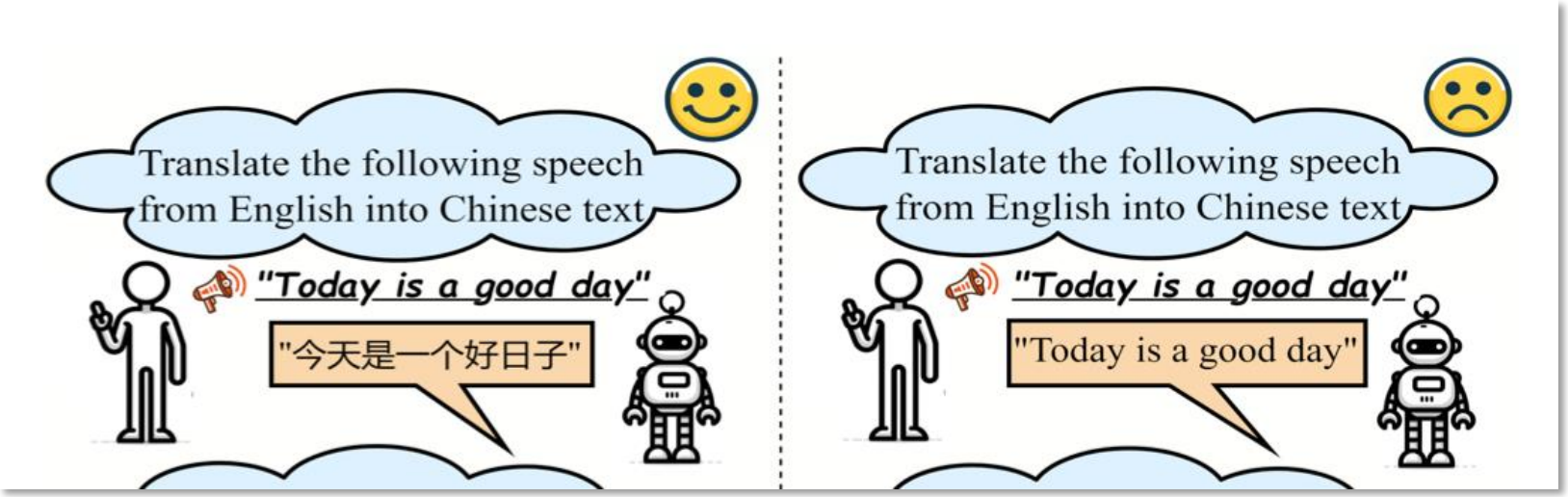
Self-Powered LLM Modality Expansion for Large Speech-Text Models

Tengfei Yu[1], Xuebo Liu[1*], Zhiyi Hou[2], Liang Ding[3], Dacheng Tao[4], Min Zhang[1]
[1]Institute of Computing and Intelligence, Harbin Institute of Technology, Shenzhen, China
[2]Faculty of Computing, Harbin Institute of Technology, Harbin, China
[3]The University of Sydney [4]Nanyang Technological University

EMNLP 2024

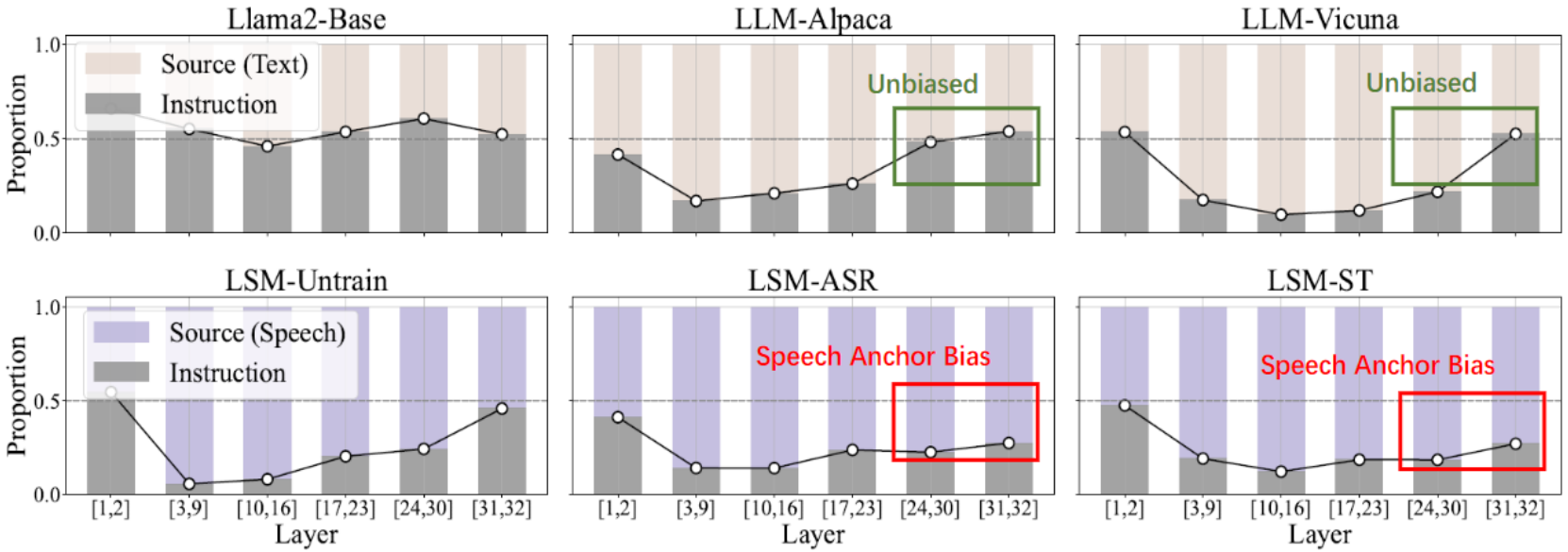
背景：语音-文本多模态模型（LSMs）的挑战

- **模态融合难题**
  - 语音和文本模态的联合预训练：需要海量计算资源，成本高昂
  - 微调 LLM 来得到 LSM ： 对数据质量要求严格，需要人工精心设计语音指令数据集（能否尝试合成数据？）
- **简单使用大量ASR数据进行指令微调会出现指令跟随缺陷**



背景：语音-文本多模态模型（LSMs）的挑战

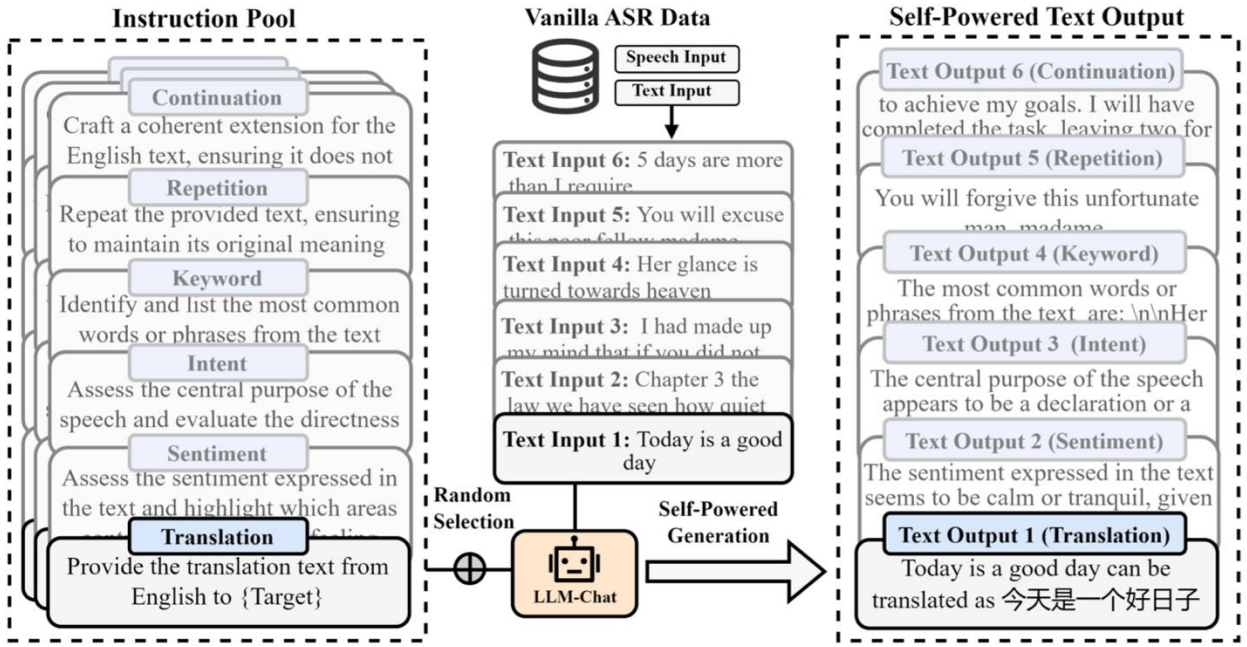
- **大语音模型的模态不平衡问题分析**
  - LLMs：早期层在指令和源输入之间均匀分配注意力；中间层优先考虑源信息；而更深的层则重新关注指令
  - LSMs： 在所有层次上模型始终偏好语音输入而非指令。

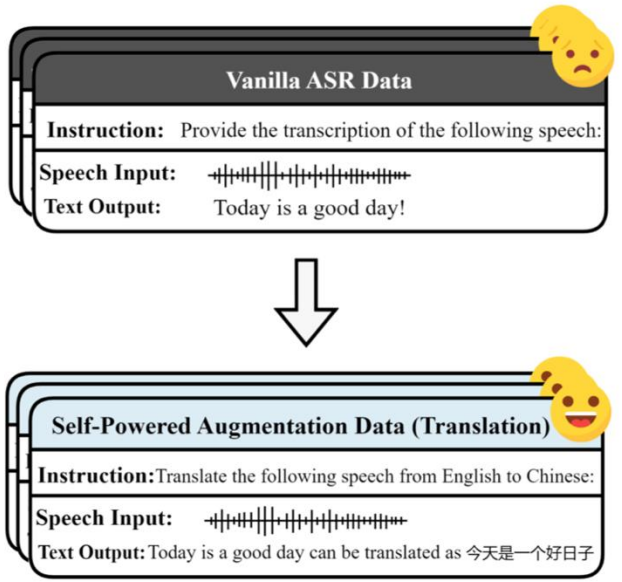


定位问题：LSMs过度关注语音输入，**导致模型在推理过程中将整个语音模态解读为训练阶段出现的指令**

Self-Powered LSM 概览

- **指令池构建**
  - 为不同类型任务准备若干指令实例
- **形成增强样本**
  - 三元组 $(i, s, \hat{t})$ 构成训练数据



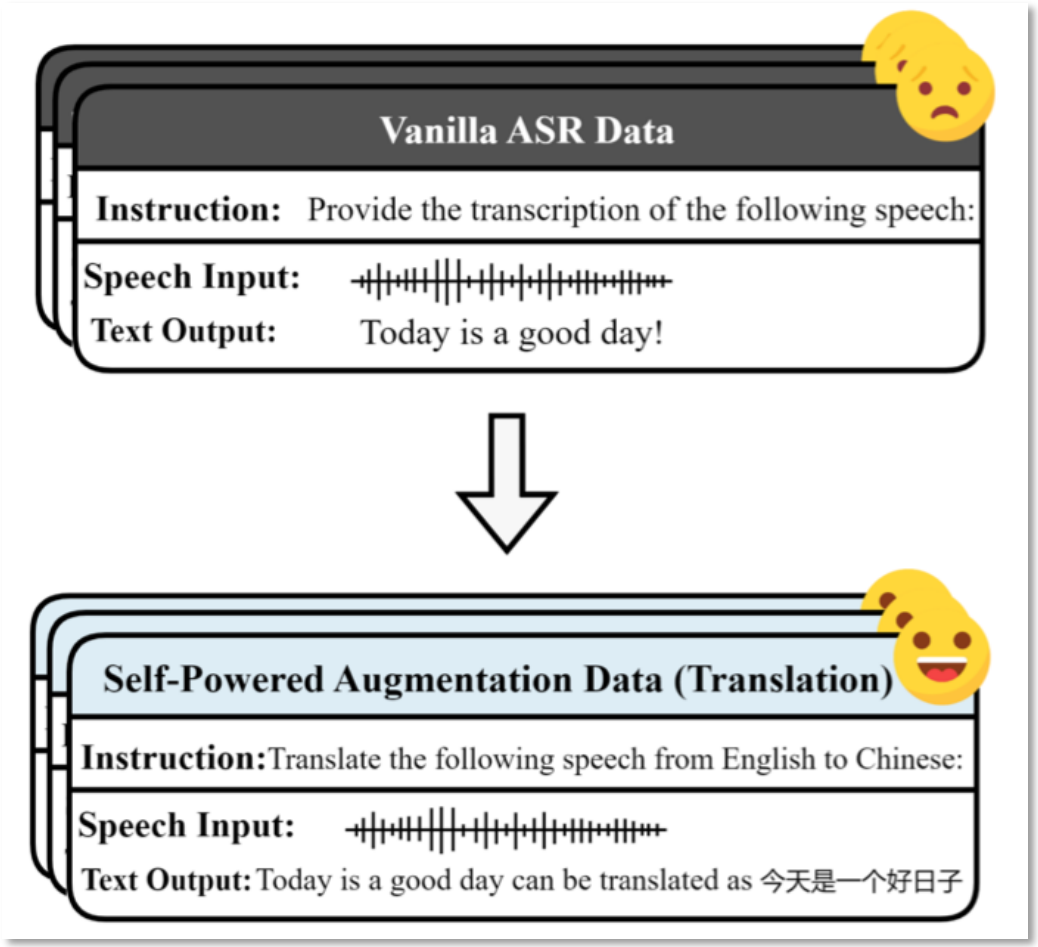


- **自驱动数据生成**
  - 随机选择指令实例 i 和 ASR 数据样本 (s, t) 的文本 t，得到模型输出 $\hat{t}$

自驱动数据生成

- **指令池构建**
  - 定义 K 类语音任务
  - 每类任务预设 m 种指令模板
- **数据生成**
  - 任取 ASR 语音-文本对 (s, t)
  - 从指令池中随机选取一类任务
  - 随机选取这类任务中的一条指令实例 i
  - 模型根据指令实例 i 和 t 生成增强文本 $\hat{t}$
  - 三元组 $(i, \mathrm{s}, \hat{t})$ 构成训练数据
- **核心思想：**

$$
\begin{aligned}
\mathcal{L}(\theta) &= -\log P(\mathbf{t}|\mathbf{s}, \mathbf{i}; \theta) \\
&\doteq -\log P(\mathbf{t}|\mathbf{s}; \theta) \qquad \textit{(Speech Anchor Bias)} \\
&\doteq -\log P(\hat{\mathbf{t}}|\mathbf{s}; \theta) \qquad \textit{(Self-Powered Generation)} \\
&\doteq -\log P(M(\mathbf{t}, \mathbf{i})|\mathbf{s}; \theta) \qquad \textit{(Modified Objective)} \\
&\text{where } M(\mathbf{t}, \mathbf{i}) = \arg\max_{\hat{\mathbf{t}}} P(\hat{\mathbf{t}}|\mathbf{t}, \mathbf{i}; \theta).
\end{aligned}
$$

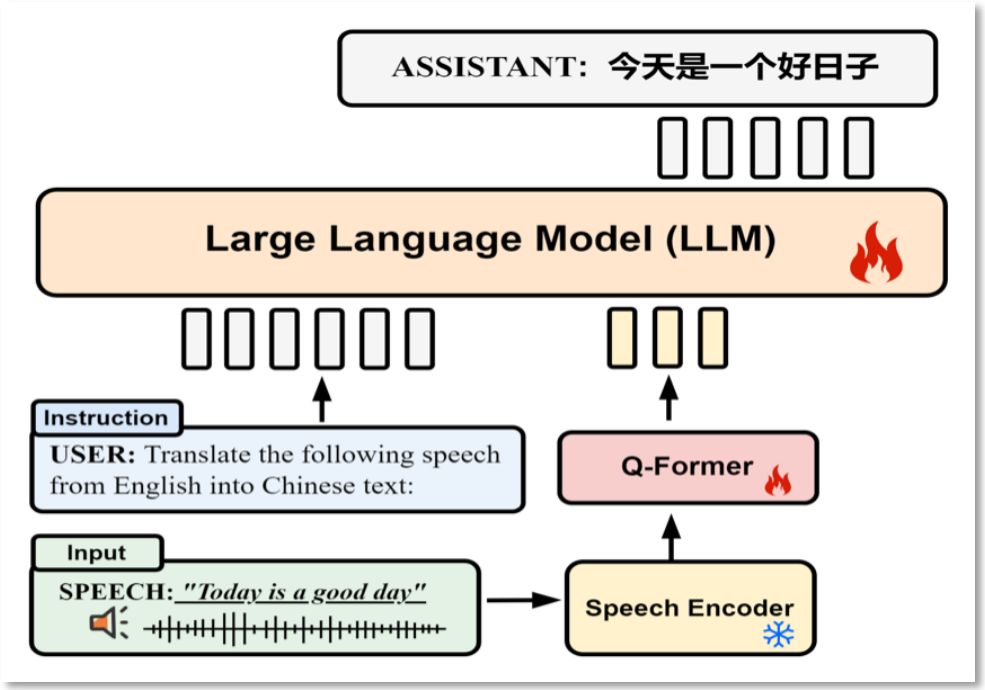


模型训练

- **相关模块**
  - **语音编码器 $\boldsymbol{\theta_s}$**: Whisper Encoder（冻结❄）
  - **LLM $\boldsymbol{\theta_l}$** : Vicuna-7B-1.5（训练🔥）
  - **连接模块 $\boldsymbol{\theta_q}$** ： Q-Former（训练🔥 ）
- **训练目标**
  - 优化目标（模态对齐与领域适应）：

$$\hat{\theta}_q, \hat{\theta}_l = \arg\min_{(\theta_q, \theta_l)} \left(-\log P(\hat{\mathbf{t}} | \mathbf{s}, \mathbf{i}_{div}; \theta_s^{❄}, \theta_q^{🔥}, \theta_l^{🔥}) \right)$$

  - s 表示语音输入
  - $\hat{t}$ 表示目标文本句子（由自驱动数据构造而来）
  - i_{div} 表示文本指令



实验效果

- 对比 Vanilla Instruction Tuning 和现有方法（Qwen-Audio、BLSP）
- Self-Powered LSM 在 ASR、ST、SLU 和 QA 全面领先

ID	Method	#Para	ASR↓		ST↑		ER↑	KE↑		IC↑	QA↑	
			Clean	Other	CoVoST	MuSTC	MELD	Light	Water	FSC	WebQ	BoolQ
	Existing Method											
1	Qwen-Audio* (Chu et al., 2023)	8.4B	**2.6**	**5.1**	**24.5**	22.0	49.1	0.4	1.1	38.4	71.4	12.2
2	BLSP* (Wang et al., 2023a)	6.9B	16.8	22.3	7.5	14.7	32.1	3.0	**57.7**	60.8	70.0	60.4
	Implemented Method											
3	Vanilla IT	6.9B	7.8	13.1	0.2	0.0	0.0	0.0	0.0	0.0	44.3	0.0
4	BLSP	6.9B	26.6	35.6	8.8	14.6	30.0	4.6	10.3	41.9	71.1	33.6
	Our Method with BackBone Model: Vicuna-7B-1.5											
5	Self-Powered LSM with Whisper-small	6.9B	3.8	8.0	17.5	21.4	51.4	30.7	34.9	47.2	72.4	59.8
6	Self-Powered LSM with Whisper-medium	7.1B	4.0	7.9	19.2	23.7	49.6	32.3	42.5	58.2	73.4	60.4
7	Self-Powered LSM with Whisper-large	7.4B	3.3	6.1	20.7	**23.8**	**51.5**	**36.2**	47.1	**61.0**	**73.6**	**61.4**

Self-Powered LSM 显著提升多项语音任务性能

自生成数据的效果分析

- **真实标注数据仅具备任务内能力**
  - “Emotion + Intent”在 ER/IC 任务表现好，但在 ST/QA 任务完全失效
  - Translation 仅能完成少量特定翻译任务

- **自生成数据具备广泛迁移能力**
  - 使用少量自生成数据即可覆盖多任务
  - 完全使用自生成数据训练后，模型在大部分任务上领先

Method	ASR↓		ST↑		ER↑	KE↑		IC↑	QA↑	
	Clean	Other	CoVoST	MuST-C	MELD	Light	Water	FSC	WebQ	BoolQ
Emotion + Intent	100.0	100.0	0.0	0.0	**71.7**	0.0	0.0	**99.6**	43.2	0.0
Translation	100.0	100.0	4.5	0.0	0.0	0.0	0.0	0.0	42.2	0.0
w/ Self-Powered Aug.	90.8	94.1	10.2	6.2	29.4	15.5	15.1	2.0	50.7	48.6
Self-Powered LSM	**3.8**	**8.0**	**17.5**	**21.4**	51.4	**30.7**	**34.9**	47.2	**72.4**	**59.8**

Self-Powered LSM 总结

- Self-Powered LSM 是一种通过自生成数据增强语音-文本能力的方法
  - 显著缓解语音锚定偏差 → 提升指令对齐能力
  - 通过 LLM 自生成多任务指令数据 → 增强泛化性，支持跨任务迁移
  - 完全基于现有多模态数据全自动合成 → 合成高效，低成本

- 核心
  - 充分利用大语言模型的内在语言能力（表征）解决模态对齐问题
  - 多模态数据合成当前阶段应充分挖掘大语言模型文本表征能力

[1] Yunfei Chu, Jin Xu, Xiaohuan Zhou, Qian Yang, Shiliang Zhang, Zhijie Yan, Chang Zhou, Jingren Zhou. Qwen-Audio: Advancing Universal Audio Understanding via Unified Large-Scale Audio-Language Models. arXiv 2023.

[2] Chen Wang, Minpeng Liao, Zhongqiang Huang, Jinliang Lu, Junhong Wu, Yuchen Liu, Chengqing Zong, Jiajun Zhang. BLSP: Bootstrapping Language-Speech Pre-training via Behavior Alignment of Continuation Writing. arXiv 2023.

[3] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever. Robust Speech Recognition via Large-Scale Weak Supervision. ICML 2023.

[4] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, Eric P. Xing. Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90% ChatGPT Quality. LMSYS Blog 2023.

[5] Junnan Li, Dongxu Li, Silvio Savarese, Steven Hoi. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. arXiv 2023.

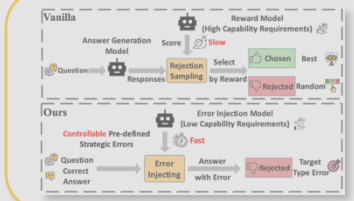
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

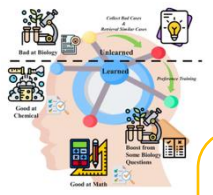
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务

LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

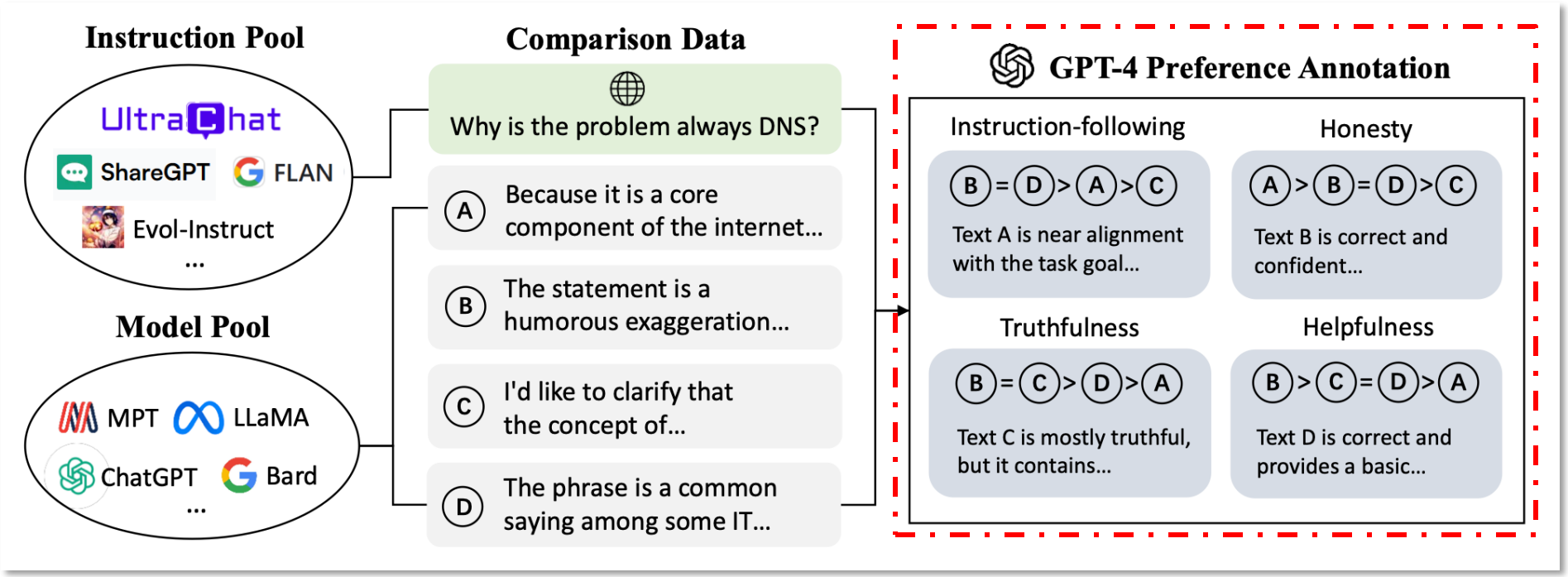
SeaPO: Strategic Error Amplification for Robust Preference Optimization of Large Language Models

Jun Rao[1], Yunjie Liao[1], Xuebo Liu[1], Zepeng Lin[1], Lian Lian[2], Dong Jin[2], Shengjun Cheng[2], Jun Yu[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]Huawei Cloud Computing Technologies Co., Ltd.

EMNLP 2025 Findings

背景：偏好数据合成

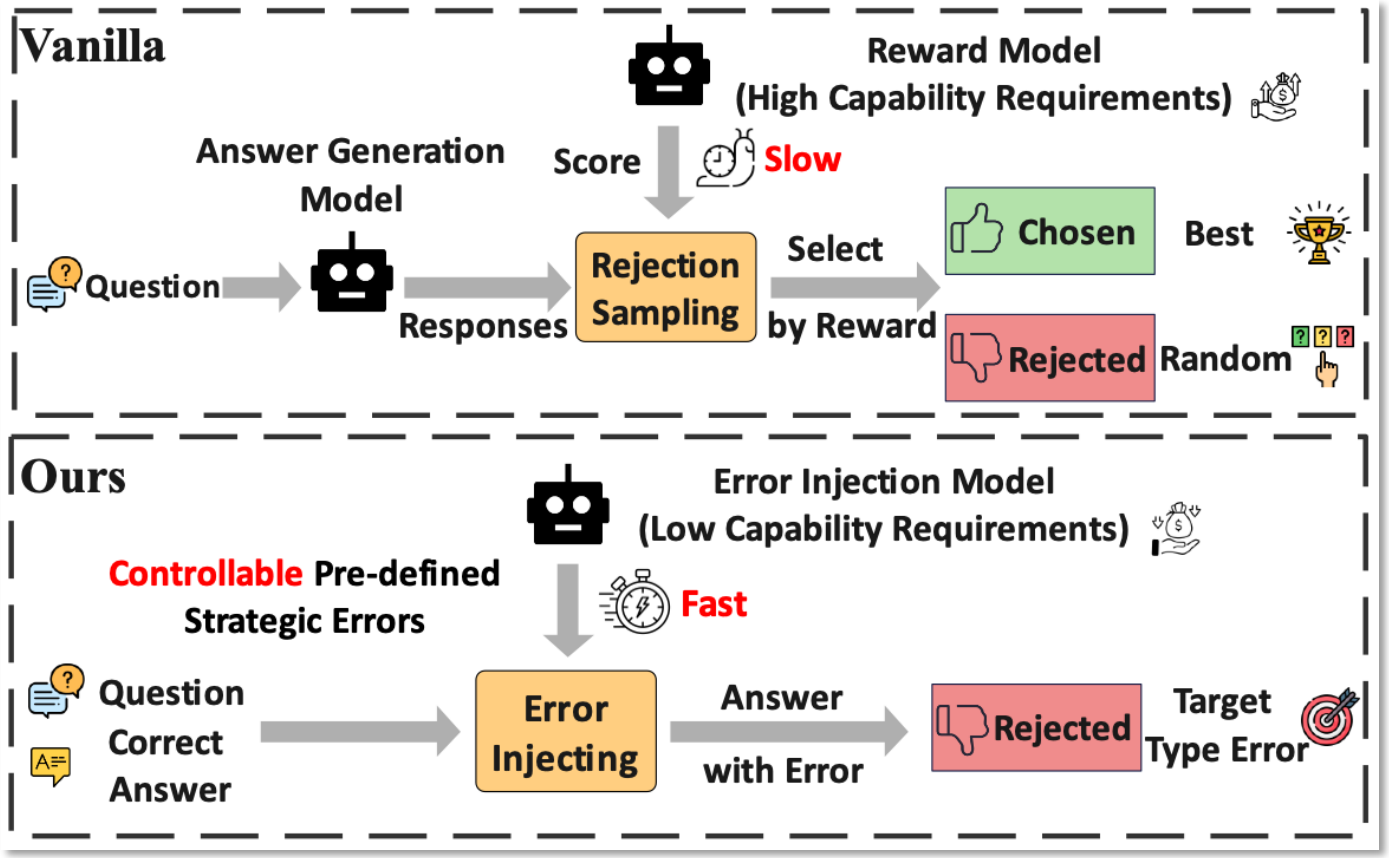
- **拒绝采样（Rejection Sampling）**
  - 对同一输入生成多个回复，使用 Reward Model 评估回复质量，根据分数选择最好的作为正样本，剩下的回复中随机选择作为负样本
  - ❓样本随机性强，不可控
  - 💰需要 Reward Model（如GPT-4），造成成本增加



SeaPO 优势

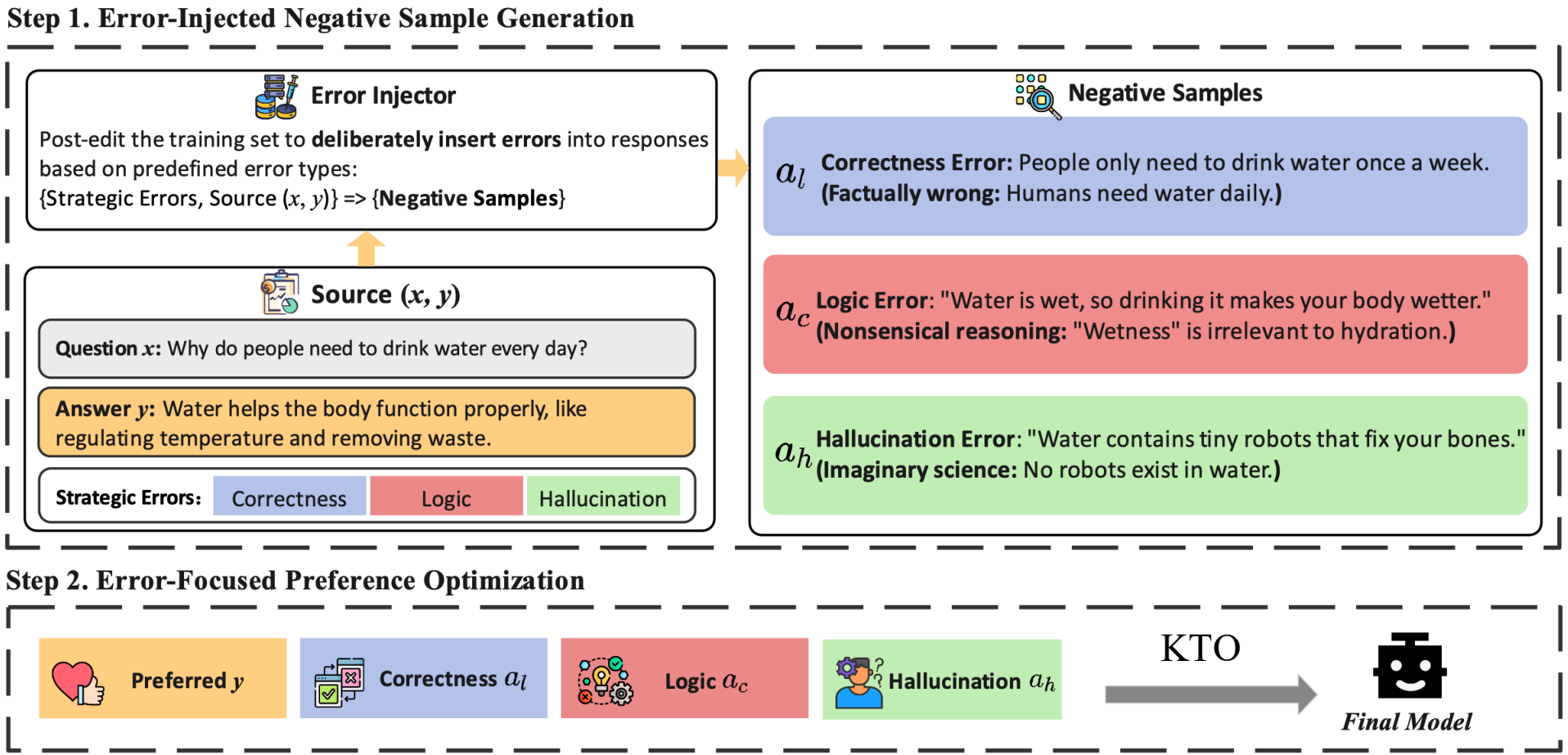
- **样本质量可控**
  - 预定义错误类型
  - 将指定错误类型引入原正确答案（SFT数据）中构造负样本
  - 错误答案与原正确答案构成偏好数据对

- **不依赖外部模型**
  - 利用目标模型自身进行负样本数据合成
  - 降低数据合成成本



SeaPO 框架概览

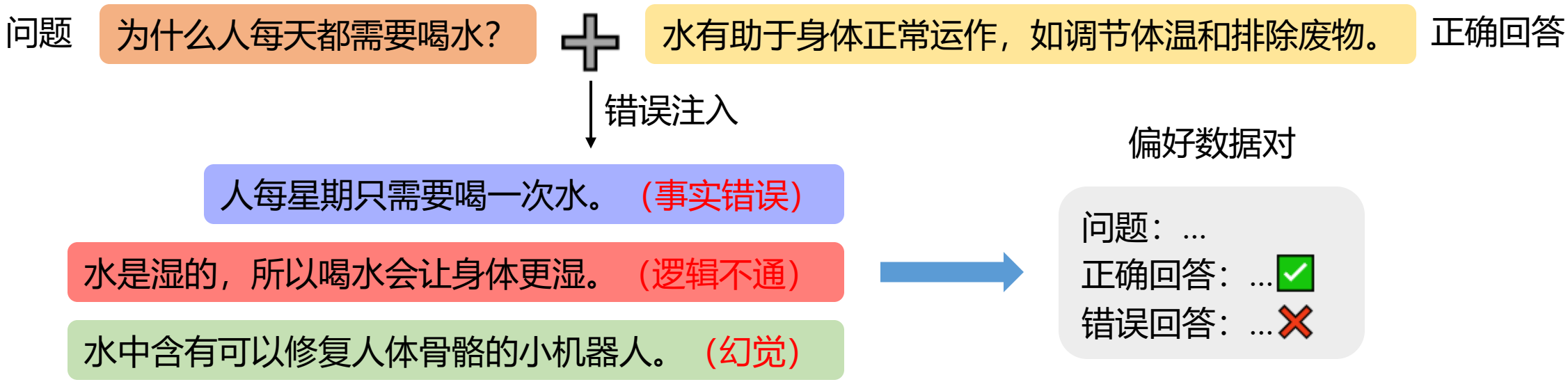
- 负样本构造
  - 对正确答案注入错误
- 预定义错误类型
  - 正确性错误、逻辑性错误、幻觉



- 偏好优化
  - KTO

错误类型定义

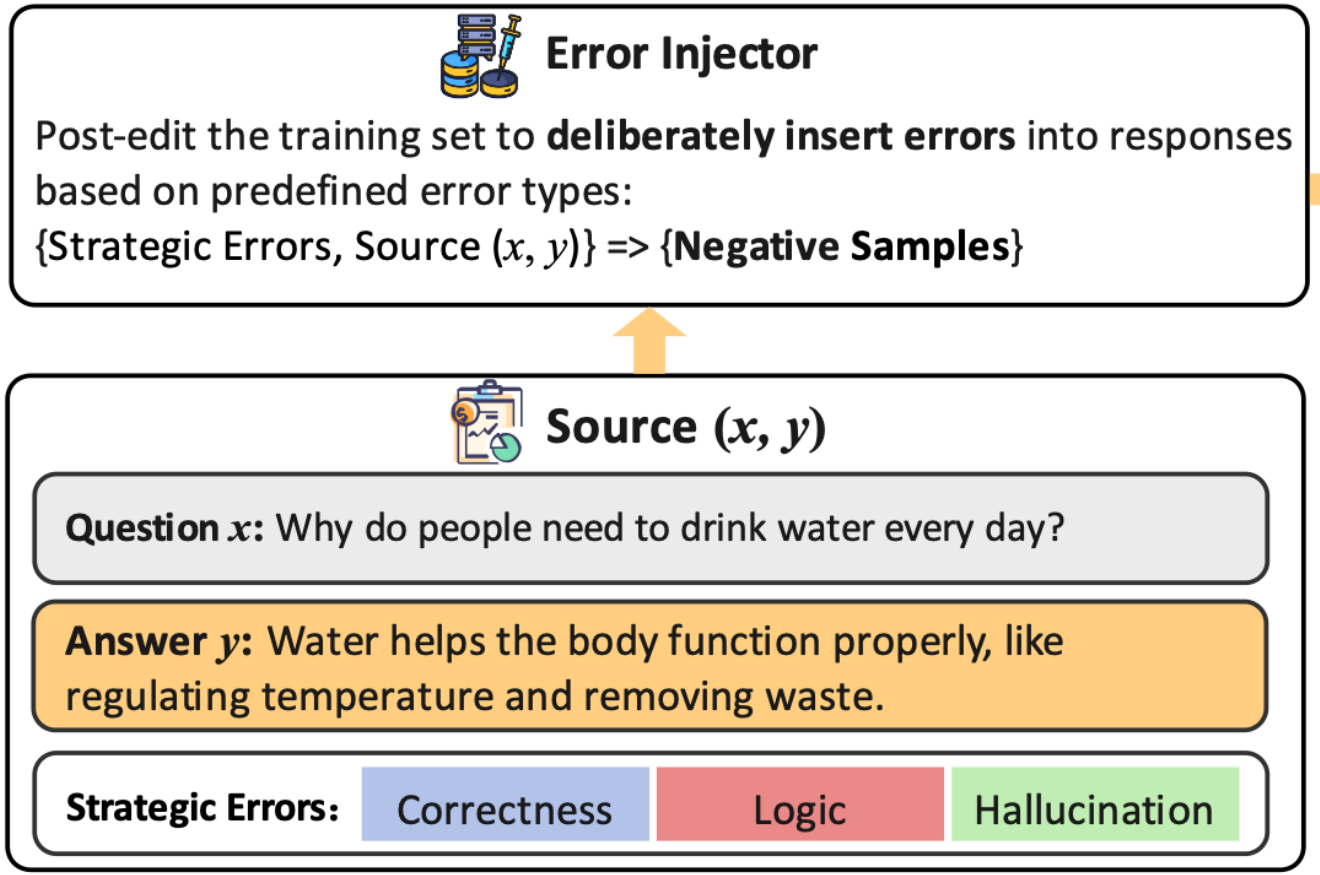
- 大模型生成响应中存在的准确性缺陷、逻辑漏洞或幻觉输出，根据响应缺陷，总结并定义三种错误类型：
  - 正确性错误：事实性偏差或计算失误
  - 逻辑错误：推理矛盾、结论缺乏支撑或逻辑链断裂
  - 幻觉：生成内容与输入信息或客观事实脱节



基于错误注入的负样本生成

- **公式定义**
  - 设 $E = \{e_c, e_l, e_h\}$ 表示预定义的三种错误类型
  - 给定的问题 x 和正确回答 y
  - 指定错误类型 $e \in E$
  - 由模型自身作为 Error Injector
  - 得到基于错误注入的负样本生成公式：

$$a_{error} = Injector(x, y, e)$$



偏好优化：Why KTO?

- **前景理论**
  - KTO 的损失函数通过调节正负样本权重，平衡损失和收益对模型的贡献，确保模型更关注负面反馈的修正。
- **公式表示**
  - 价值函数：模拟人类对收益/损失的非线性感知

$$v_{KTO}(x,y) = \begin{cases} \sigma\left(r_{KTO}(x,y) - z_{ref}\right) & \text{正样本} \\ \sigma\left(z_{ref} - r_{KTO}(x,y)\right) & \text{负样本} \end{cases}$$

  - 损失函数：

$$L_{KTO} = E_{x,y\sim D}[w(y)(1 - v_{KTO}(x,y))]$$

实验效果

- 在数学、推理、代码、知识、真实性五大领域的测试集进行测试
- 使用Qwen2.5-Instruct系列1.5B-14B，以及Llama3-8B-Instruct模型

Model	Method	Math		Reasoning	Coding	Knowledge	Truthful		Avg.
		MATH	GSM	BBH	HumanEval	MMLU	MC1	MC2	
Llama3-8B	Instruct	28.0	75.1	**68.5**	75.5	**65.7**	36.2	51.8	57.3
	+Vanilla	**28.1**	**75.3**	67.6	77.5	65.3	38.9	56.3	58.4
	+SeaPO	25.8	72.2	62.9	**79.5**	64.7	**55.0**	**66.1**	**60.9**
Qwen2.5-1.5B	Instruct	30.4	61.0	**45.6**	**70.5**	**59.5**	30.5	46.0	49.1
	+Vanilla	27.8	56.4	43.4	**70.5**	59.2	31.5	48.0	48.1
	+SeaPO	**30.9**	**64.1**	44.7	69.2	59.3	**39.3**	**55.0**	**51.8**
Qwen2.5-7B	Instruct	47.2	72.5	**67.5**	88.7	72.8	42.4	59.3	64.3
	+Vanilla	46.8	73.4	59.4	**90.9**	**72.9**	51.0	65.3	65.7
	+SeaPO	**56.5**	**75.4**	67.0	88.6	72.5	**60.2**	**70.9**	**70.2**
Qwen2.5-14B	Instruct	50.1	85.6	73.7	89.1	**79.7**	51.5	69.2	71.3
	+Vanilla	50.1	86.8	73.1	**89.2**	**79.7**	52.0	**69.6**	71.5
	+SeaPO	**53.9**	**90.7**	**78.9**	88.5	79.3	**52.5**	69.1	**73.3**

在多个模型上均优于原始模型和基于原 UltraFeedback 数据集训练的模型

错误类型数据混合

- 在不同任务中各类错误类型的性能结果分析表明：针对最高频错误类型进行训练可有效提升任务表现。
- 数学、代码和推理类任务易出现正确性（correctness）和逻辑性（logic）错误，而基于知识的TruthfulQA任务则更容易产生幻觉（hallucination）。
- 通过构建相关错误类型进行优化后，各任务性能均获得显著提升模型

ID	Error Type	Math		Reasoning	Coding	Knowledge	Truthful		Avg.
		MATH	GSM	BBH	HumanEval	MMLU	MC1	MC2	
1	None (Instruct)	47.2	72.5	67.5	88.7	72.8	42.4	59.3	64.3
2	Untargeted	56.6	70.8	68.6	87.6	72.9	55.6	67.4	68.5
3	Logic	54.5	71.8	65.2	89.0	72.7	59.5	70.7	69.1
4	Correctness	54.7	69.6	65.4	**89.3**	72.8	57.8	68.8	68.3
5	Hallucination	53.7	74.4	66.7	**89.3**	72.5	60.8	72.0	69.9
6	3+4	52.9	**79.8**	**69.4**	87.7	**73.2**	52.6	64.9	68.6
7	3+5	52.5	77.5	67.4	**89.3**	72.6	**62.8**	**72.9**	**70.7**
8	4+5	53.9	75.7	67.7	88.0	72.7	61.8	71.7	70.2
9	3+4+5	**56.5**	75.4	67.0	88.6	72.5	60.2	70.9	70.2

SeaPO 总结

- SeaPO 是一种偏好数据合成方法
  - 根据模型回复缺陷定义错误 → 适用性广，可拓展
  - 构造指定错误类型的负样本 → 负样本质量可控
  - 无需外部模型 → 降低数据合成成本
- 多模态火花🔥
  - 融合图像、文本、音频等多模态特征，构建跨模态的策略性错误类型库
  - 通过模拟多模态交互来生成策略性负样本
  - 结合多模态反馈信号优化偏好学习，以提升模型与用户意图的对齐度

[1] Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Bingxiang He, Wei Zhu, Yuan Ni, Guotong Xie, Ruobing Xie, Yankai Lin, Zhiyuan Liu, Maosong Sun. UltraFeedback: Boosting Language Models with Scaled AI Feedback. ICML 2024.

[2] Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, Douwe Kiela. KTO: Model Alignment as Prospect Theoretic Optimization. ICML 2024.

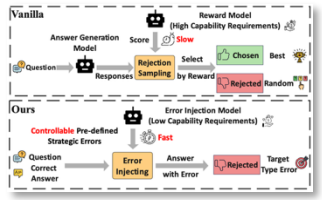
报告总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
  - 通用场景、特定任务
  - 特定场景、特定任务

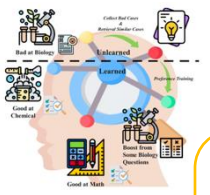
偏好数据

指令微调数据

通用场景



SeaPO：策略性错误放大的偏好数据合成



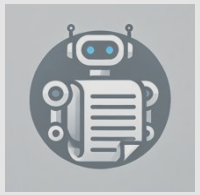
APT：弱点数据生成与迭代式能力对齐

Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

AQuilt：逻辑与反思增强的指令对齐数据合成

特定任务

通用任务



LongMT: CoT偏好数据广域搜索与细粒度策略合成

特定场景

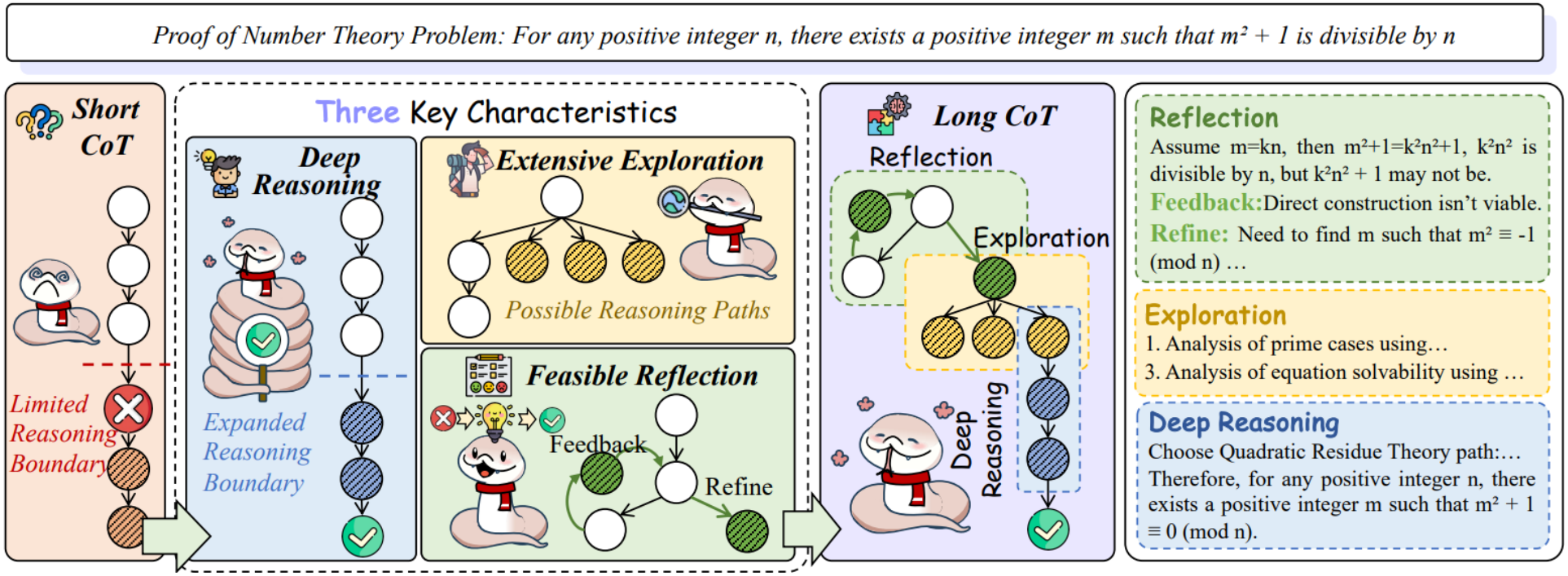
LongMT: Towards Humen-Like Document-Level Translation Agent

Yutong Wang, Zepeng Lin, Yunjie Liao, Xuebo Liu, Min Zhang
Harbin Institute of Technology, Shenzhen

Ongoing Work

CoT偏好数据广域搜索与细粒度策略合成

- **长思维链推理通过测试时间扩展，在给定问题空间内进行更详细、迭代的探索和反思**
  - ➢ 与传统短思维链的区别在于：**【深度推理、广泛探索、反馈传播】**
  - Agent面对复杂任务可以拆分成多步处理，其中每步根据具体情况采用不同行动策略，直到任务完成。整个过程可视为一条超长的CoT
  - **【深度推理】**能否合成一些CoT推理轨迹，对Agent进行**偏好优化**?
  - **【广泛探索】**如何兼顾合成CoT推理轨迹时的**广域性 + 高质量**?
  - **【反馈传播】**如何将整条CoT获得的最终任务结果分解为**每个单步策略的得分（细粒度）**？



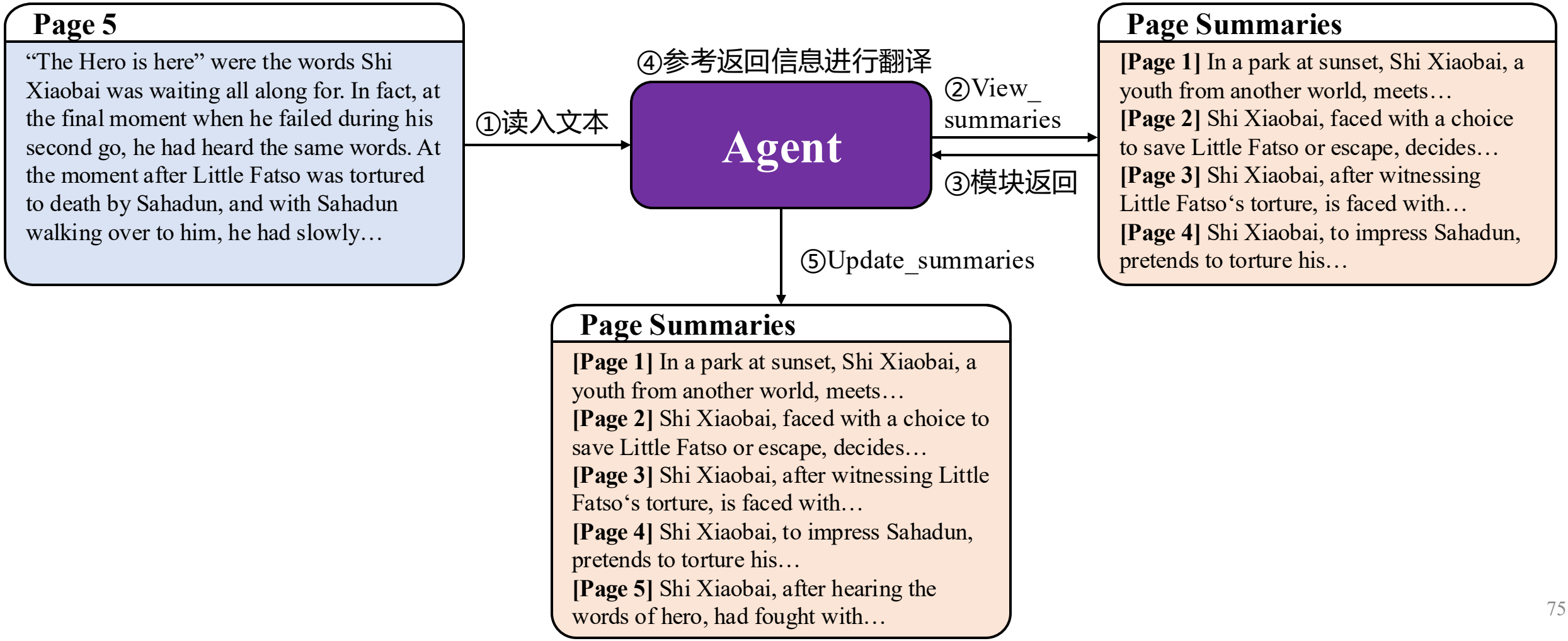
(Chen et al., 2025)

实际优化场景：文档翻译

- 翻译一篇长文档时，如何像人类翻译的一样好？ **【深度推理】**
  - 分段翻译：将整篇文档分成**多个片段**（Pages）
  - 翻译过程中做**阅读式的总结**，把握每段的大致内容，适当**回看前文**
  - 记录文中出现的**关键实体信息**（人物、地名、事件等）
- 模仿人类翻译者设计智能体模块：
  - **View_summaries**: 查看前文某几页的摘要
  - **View_pages**: 查看前文某几页的原文
  - **Lookup_entities**: 从实体记录中查找指定实体
  - **Update_summaries**: 更新摘要
  - **Update_records**: 更新实体记录

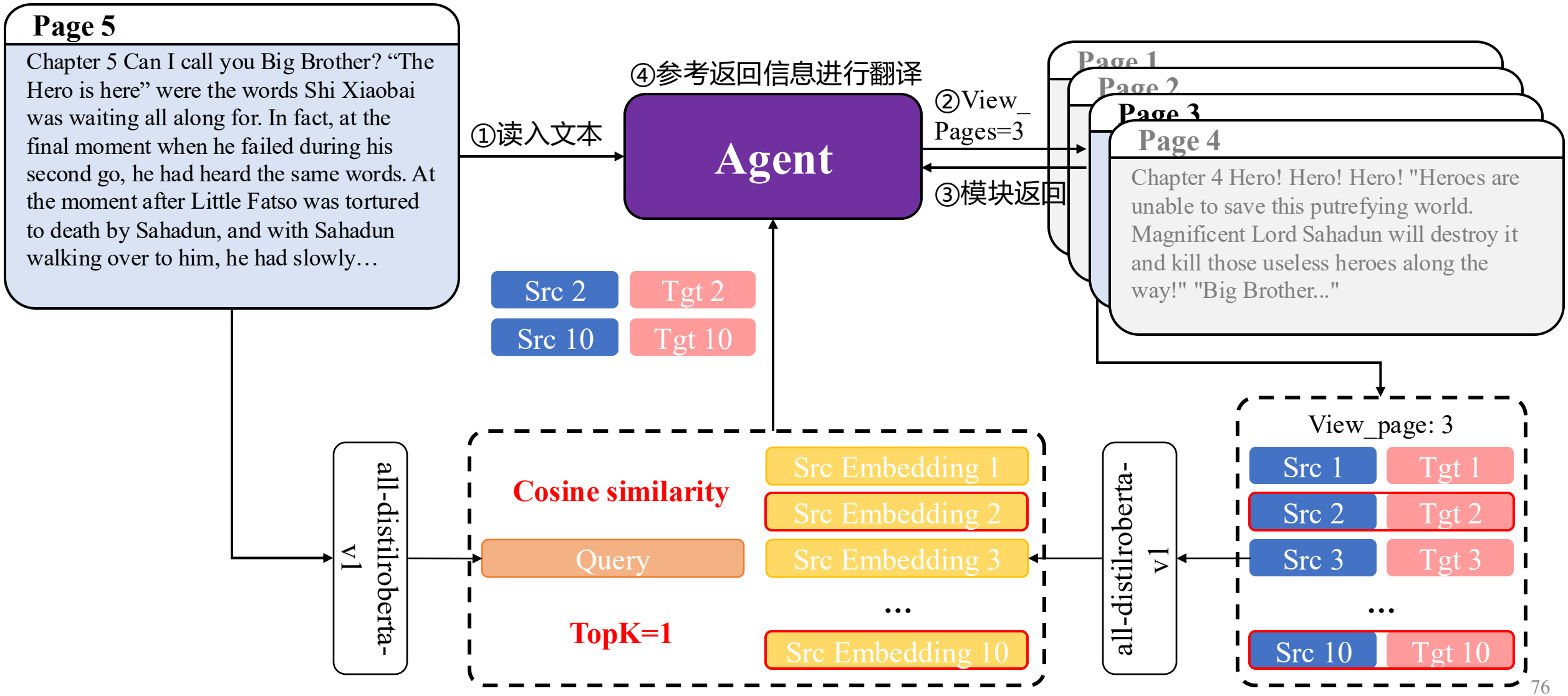
文档翻译智能体模块定义

- **View_summaries**: 查看前文某几页的摘要
- **Update_summaries**: 更新摘要



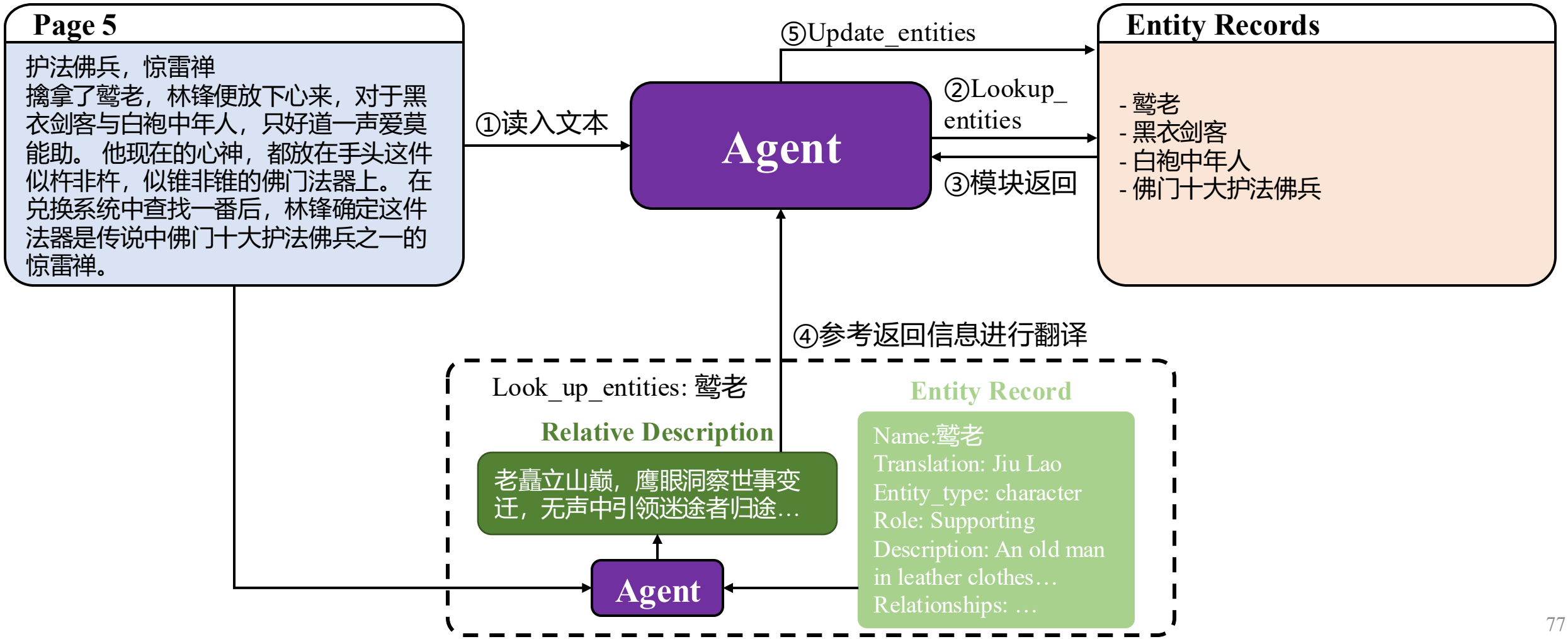
文档翻译智能体模块定义

- **View_pages**: 查看前文某几页的原文，从中提取最相关的句子



文档翻译智能体模块定义

- **Lookup_entities**: 从实体记录中查找指定实体
- **Update_entities**: 更新实体记录



长文档翻译问题分析

- 长文档翻译智能体整体思路
  - **分治**：将大篇幅文档分为若干页（人为粗分 + LLM调整）
  - **查询**：LLM根据当前页内容调用模块查询相关信息
  - **翻译**：LLM利用查询到的额外信息对当前页进行翻译
  - **更新**：LLM调用模块对摘要、实体记录等记忆体进行更新

  → **多步推理的CoT过程**

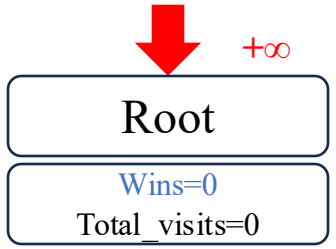
  → **每步候选的行动策略**
- 存在的挑战
  - 翻译时需要的额外信息因当前页的具体内容而异
  - 在翻译一些简单内容时，引入过多模块的信息会带来**性能干扰**和**token开销**
- 回到刚才的技术问题：
  - **【深度推理】**能否合成一些CoT推理轨迹，对文档翻译Agent进行**偏好优化**？
  - **【广泛探索】**如何兼顾合成CoT推理轨迹时的**广域性 + 高质量**？
  - **【反馈传播】**如何将整条CoT获得的最终翻译分数分解为**每个单步策略的得分**？

  → **MCTS**

数据采样方法：MCTS

① Select：选择当前待展开的结点【广泛探索】

UCB1兼顾探索少的点保证广域性，得分高的点保证高质量



1. View_summaries
2. View_pages
3. Look_up_entities
4. Translate

当前策略比基线强多少

$$\text{wins} = \begin{cases} 0, if\ \text{cur\_score} > \text{baseline\_score} \\ \dfrac{\text{cur\_score}}{\text{baseline\_score}}, otherwise \end{cases}$$

结点质量

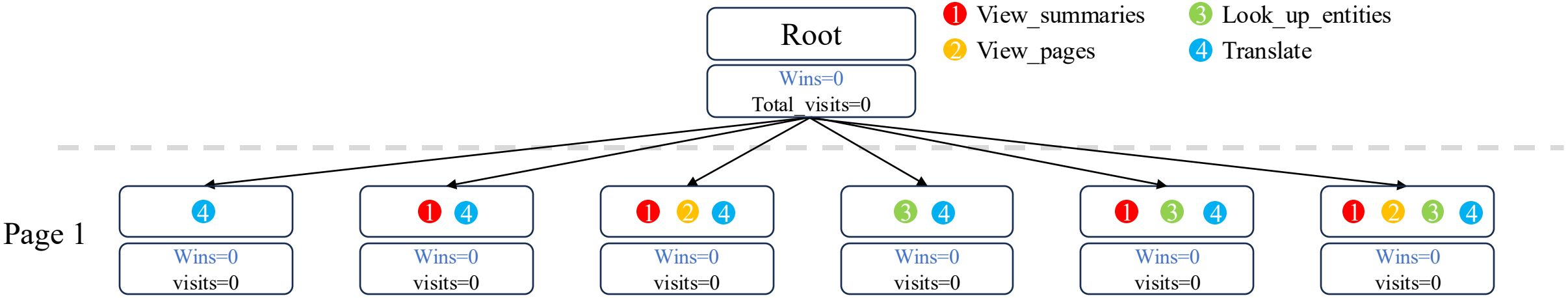
$$Q = \frac{\text{wins}}{\text{visits}}$$

平衡质量与探索度

$$\text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

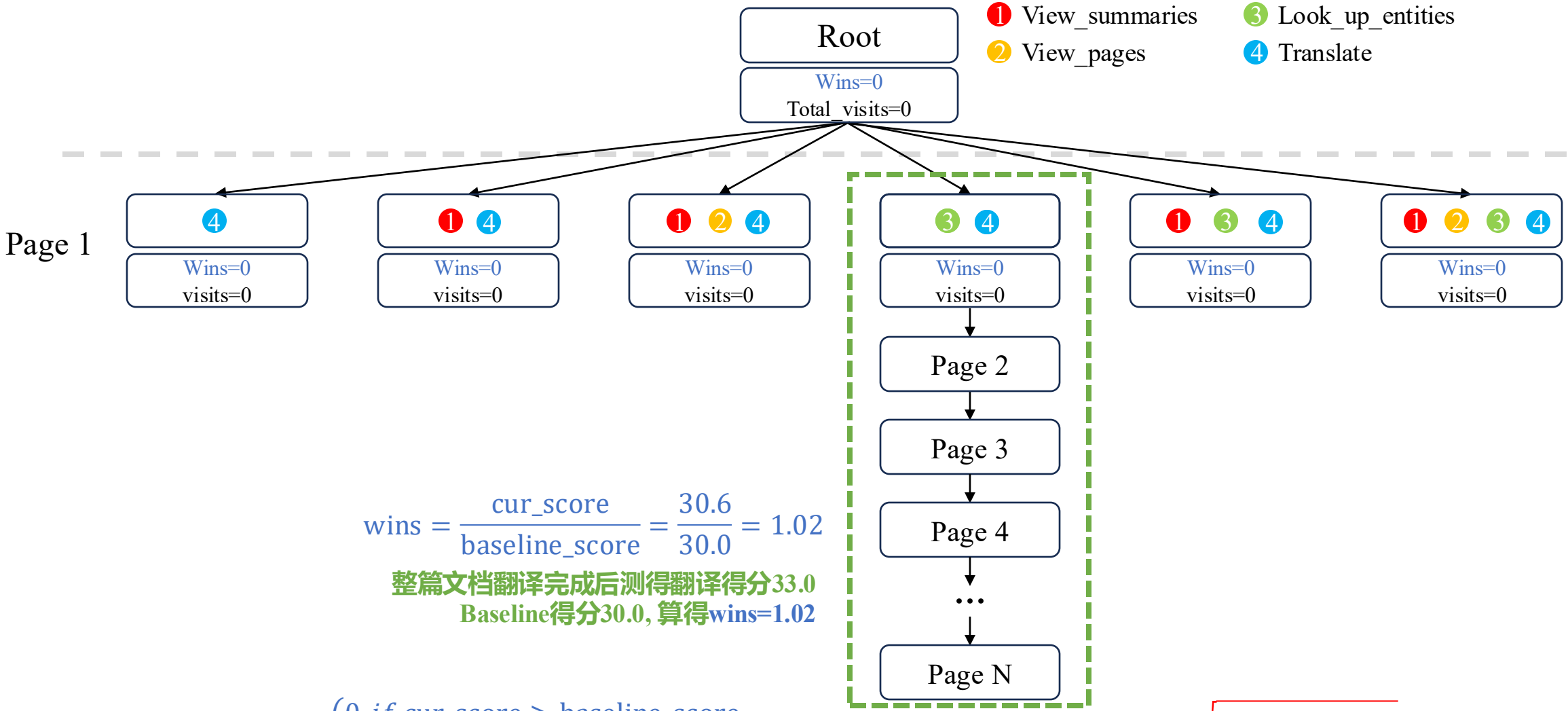
② Expand：展开当前结点，使用不同的模块调用策略



$$\text{wins} = \begin{cases} 0, if\ \text{cur\_score} > \text{baseline\_score} \\ \dfrac{\text{cur\_score}}{\text{baseline\_score}}, otherwise \end{cases} \qquad Q = \frac{\text{wins}}{\text{visits}} \qquad \text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

③ Simulate：任取一新结点，随机生成CoT直至所有页翻译完成



$$wins = \frac{cur\_score}{baseline\_score} = \frac{30.6}{30.0} = 1.02$$

整篇文档翻译完成后测得翻译得分33.0
Baseline得分30.0, 算得wins=1.02

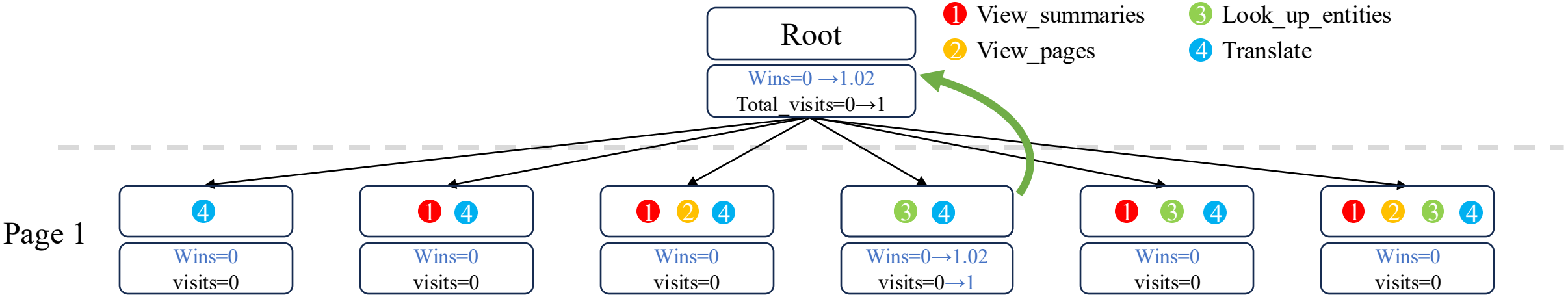
$$wins = \begin{cases} 0, if\ cur\_score > baseline\_score \\ \frac{cur\_score}{baseline\_score}, otherwise \end{cases}$$

$$Q = \frac{wins}{visits}$$

$$UCB1 = Q + C * \sqrt{\frac{\ln(total\_visits)}{visits}}$$

数据采样方法：MCTS

④ Backpropagate：将wins数值反向传播，更新所有父节点【反馈传播】



$$wins = \frac{cur\_score}{baseline\_score} = \frac{30.6}{30.0} = 1.02$$

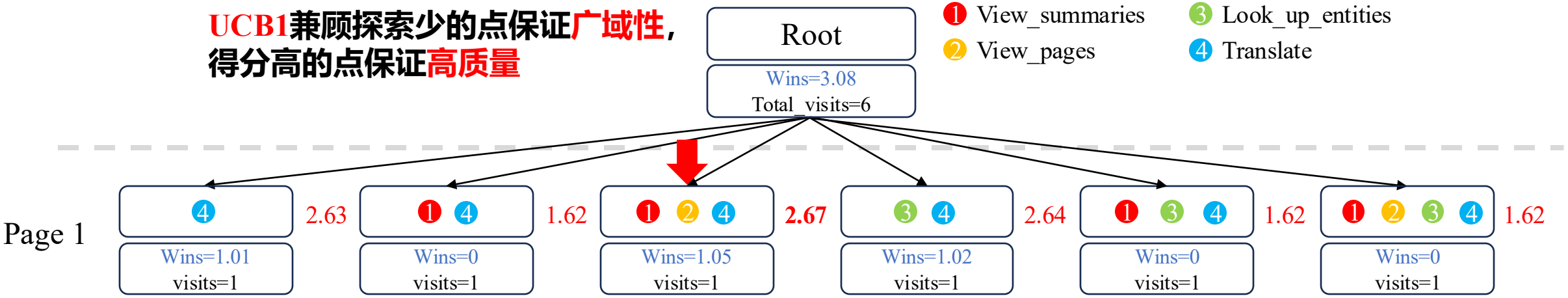
整篇文档翻译完成后测得翻译得分33.0
Baseline得分30.0, 算得wins=1.02

$$wins = \begin{cases} 0, if\ cur\_score > baseline\_score \\ \frac{cur\_score}{baseline\_score}, otherwise \end{cases} \qquad Q = \frac{wins}{visits} \qquad UCB1 = Q + C * \sqrt{\frac{\ln(total\_visits)}{visits}}$$

数据采样方法：MCTS

① Select：选择当前待展开的结点【广泛探索】

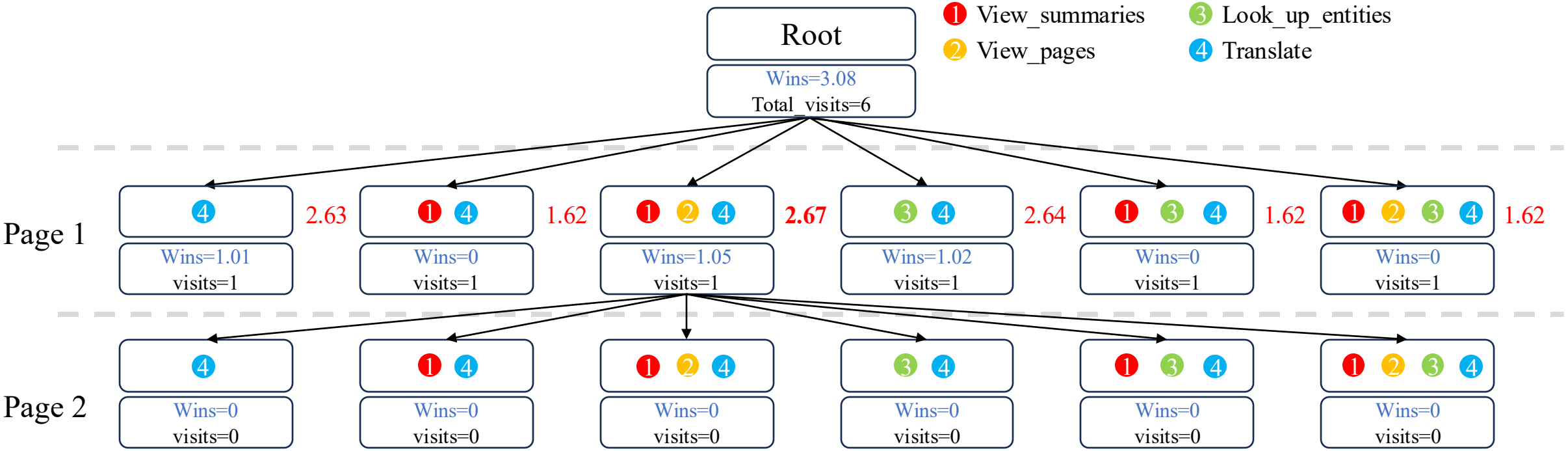
UCB1兼顾探索少的点保证广域性，得分高的点保证高质量



$$\text{wins} = \begin{cases} 0, if\ \text{cur\_score} > \text{baseline\_score} \\ \dfrac{\text{cur\_score}}{\text{baseline\_score}}, otherwise \end{cases} \qquad Q = \frac{\text{wins}}{\text{visits}} \qquad \text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

② Expand：展开当前结点，使用不同的模块调用策略

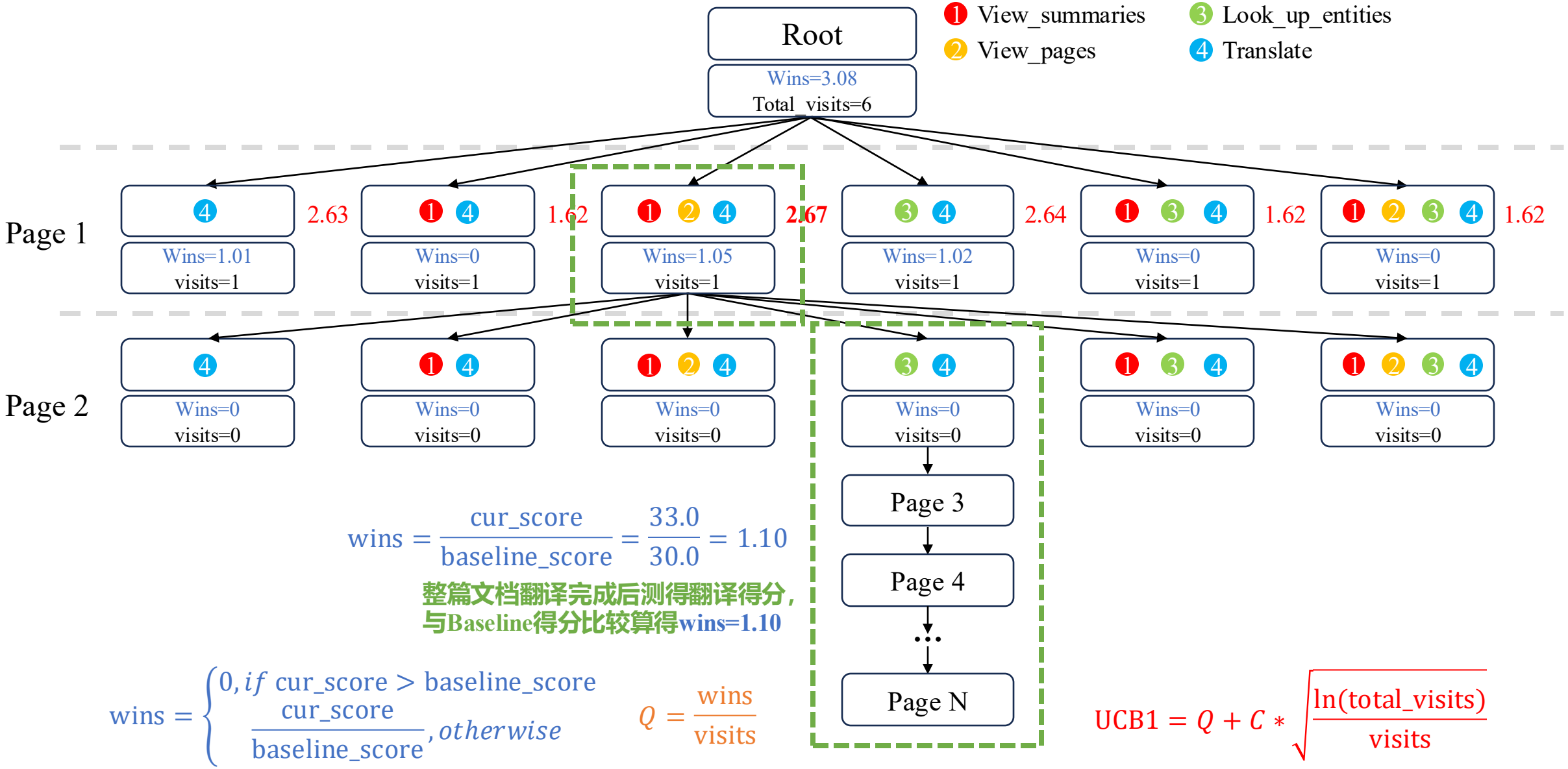


$$\text{wins} = \frac{\text{cur\_score}}{\text{baseline\_score}} \qquad Q = \frac{\text{wins}}{\text{visits}}$$

$$\text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

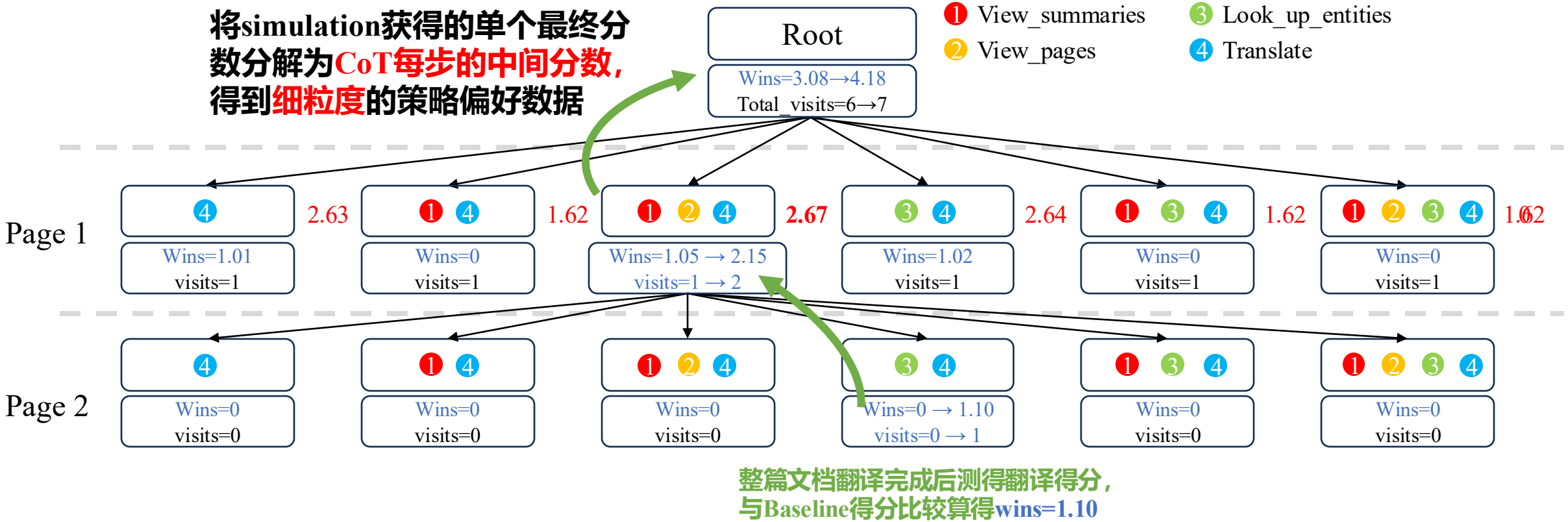
③ Simulate：任取一新结点，随机生成CoT直至所有页翻译完成



数据采样方法：MCTS

④ Backpropagate：将wins数值反向传播，更新所有父节点【反馈传播】

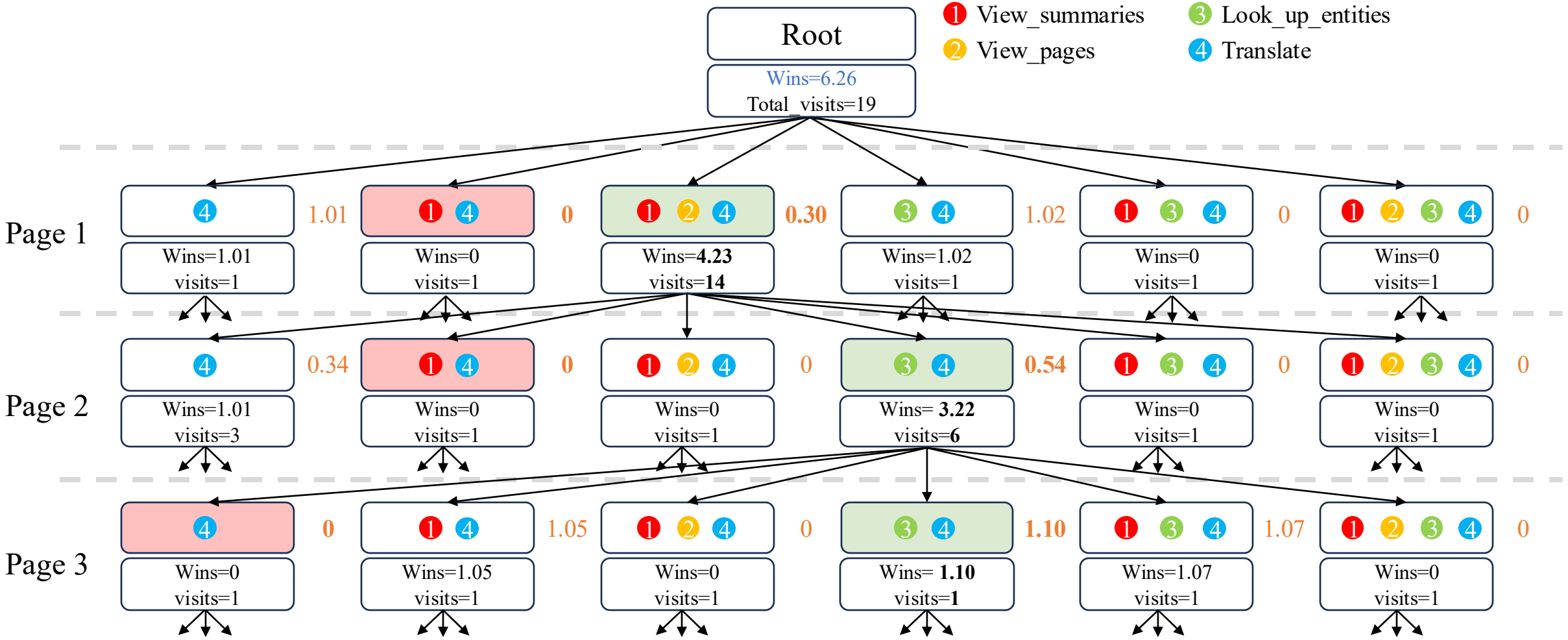
将simulation获得的单个最终分数分解为CoT每步的中间分数，得到细粒度的策略偏好数据



$$\text{wins} = \begin{cases} 0, if\ \text{cur\_score} > \text{baseline\_score} \\ \dfrac{\text{cur\_score}}{\text{baseline\_score}}, otherwise \end{cases} \qquad Q = \frac{\text{wins}}{\text{visits}} \qquad \text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

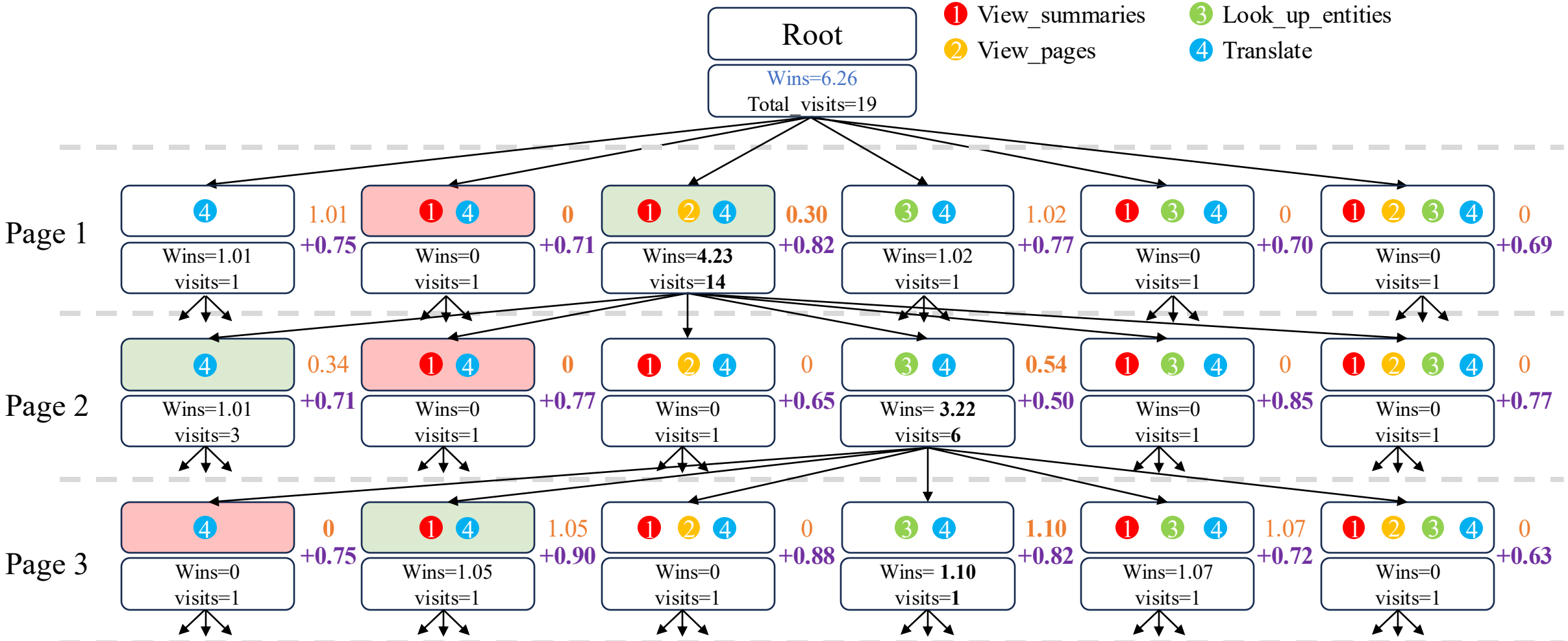
⑤数据生成：每层挑选Q值最大的结点作为Win样例，最小的作为Lose样例



$$\text{wins} = \begin{cases} 0, if\ \text{cur\_score} > \text{baseline\_score} \\ \dfrac{\text{cur\_score}}{\text{baseline\_score}}, otherwise \end{cases} \qquad Q = \frac{\text{wins}}{\text{visits}} \qquad \text{UCB1} = Q + C * \sqrt{\frac{\ln(\text{total\_visits})}{\text{visits}}}$$

数据采样方法：MCTS

改进方向：翻译结果每页都可测得**独自的得分**



数据采样方法：MCTS

- 采样得到的CoT训练样本

每页都有单独的偏好数据→细粒度【反馈传播】

Win——高Q值样本

System: You are translating a document…

User: Page 5

Assistant: view_summaries

Module:
[Page 1] In a park at sunset, Shi Xiaobai…
[Page 2] Shi Xiaobai, faced with a choic…
[Page 3] Shi Xiaobai, after witnessin…
[Page 4] Shi Xiaobai, to impress…

Assistant: view_pages

Module:
[Page 3] Chapter 3 Have you…
[Page 4] Chapter 4 Hero! Hero!...

Assistant: translate

Assistant: update_summaries

Assistant: update_entities

Lose——低Q值样本

System: You are translating a document…

User: Page 5

Assistant: view_pages

Module:
[Page 3] Chapter 3 Have you…
[Page 4] Chapter 4 Hero! Hero!...

Assistant: translate

Assistant: update_summaries

Assistant: update_entities

Agent训练方法：DPO

- 训练调用模块的Agent模型
  - DPO优化目标

$$L_{DPO}(\pi_\theta|\pi_{\text{ref}}) = -\min_{\pi_\theta} E_{(x,y_w,y_l)\sim D}[\log\sigma(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)})]$$

y_l：采样的模块调用**CoT负例（Loss）**

y_w：采样的模块调用**CoT正例（Win）**

x：待翻译的**当前页（Page）**

  - 优化结果
    - Agent倾向于选择对全局翻译有利的模块调用策略
    - 最大化模块调用的收益，避免额外信息干扰和token开销
    - 模仿人类翻译过程，获得更高的翻译质量分数

Agent训练方法：DPO

- 测试结果
  - 数据集：Guofeng Test 2 En→Zh
  - 基座模型：Qwen2.5-7B-Instruct

	Id	System	COMET
不加任何模块	**1**	W/o Modules	82.51
LLM自动调用模块	**2**	Auto Modules	81.13
强制只调用单个模块	**3**	W/ Entity_Records	82.15
	4	W/ View_Pages	82.30
	5	W/ View_Summaries	82.52
MCTS训练LLM调用模块	**6**	MCTS Modules	**82.75**

- CoT偏好数据广域搜索与细粒度策略合成
  - 复杂任务分解为多个步骤 → 长 CoT 推理轨迹
  - MCTS 采样 CoT 数据 → 兼顾广域性与高质量，为每步生成细粒度中间策略
  - DPO 偏好优化智能体 → 学会动态调整策略，防止冗余信息产生干扰和开销
- 多模态火花🔥
  - 将复杂多模态任务分解为可解释的子模块（如场景分析、实体检索、跨模态对齐），并以序列化方式组织执行。
  - 利用MCTS探索并采样最优的多模态模块调用路径，以平衡结果的广度与生成质量

参考文献

[1] Yuxi Xie, Anirudh Goyal, Wenyue Zheng, Min-Yen Kan, Timothy P. Lillicrap, Kenji Kawaguchi, Michael Shieh. Monte Carlo Tree Search Boosts Reasoning via Iterative Preference Learning. NeurIPS 2024.

[2] Weize Chen, Jiarui Yuan, Chen Qian, Cheng Yang, Zhiyuan Liu, Maosong Sun. Optima: Optimizing effectiveness and efficiency for llm-based multi-agent system. arXiv 2024.

[3] Jiayi Zhang, Jinyu Xiang, Zhaoyang Yu, Fengwei Teng, Xionghui Chen, Jiaqi Chen, Mingchen Zhuge, Xin Cheng, Sirui Hong, Jinlin Wang, Bingnan Zheng, Bang Liu, Yuyu Luo, Chenglin Wu. Aflow: Automating agentic workflow generation. arXiv 2024.

[4] Zhirui Deng, Zhicheng Dou, Yutao Zhu, Ji-Rong Wen, Ruibin Xiong, Mang Wang, Weipeng Chen. From Novice to Expert: LLM Agent Policy Optimization via Step-wise Reinforcement Learning. arXiv 2024.

[5] Yutong Wang, Jiali Zeng, Xuebo Liu, Derek F. Wong, Fandong Meng, Jie Zhou, Min Zhang. DelTA: An Online Document-Level Translation Agent Based on Multi-Level Memory. ICLR 2025.

通用与垂域数据合成总览

- 5种针对不同场景和任务的数据合成方法
  - 通用场景、通用任务
    - AQuilt：使用简便高效、具有泛化性、成本比DeepSeek、Qwen72B合成显著降低
  - 通用场景、特定任务
    - Self-Powered-LSM：自驱动合成语音-文本跨模态数据，增强指令遵循能力
    - APT：针对性合成与检索数据，解决行业特殊错误
    - SeaPO：针对性合成数据解决通用错误或特定错误
  - 特定场景、特定任务
    - LongMT：理清楚长思维链的从哪方面推理、每方面探索什么、反思什么，找到可靠任务反馈，合成长思维链数据
    - 当前阶段的高质量数据合成仍然需要领域专家参与！！！

01 大模型与数据合成背景

02 基础：通用与垂域数据合成

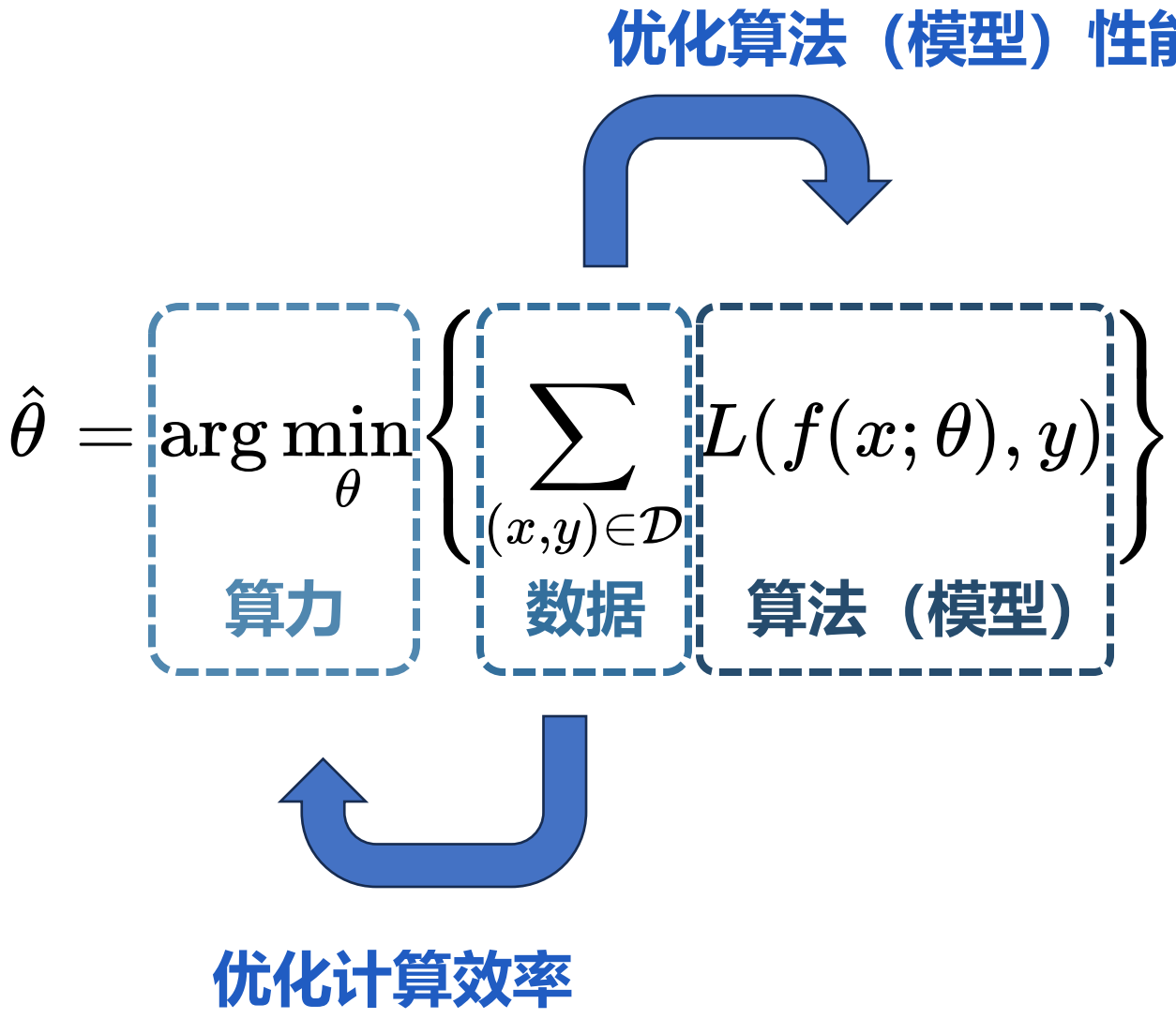
03 核心：高效数据学习与利用

04 进阶：“数据-模型”能力对齐

05 领域瓶颈与未来展望

背景：高效数据学习与利用

- **数据合成为大模型训练提供“食材”**
  - 更好的自动化特征提取
  - 更好地捕捉语法、语义和上下文信息
  - 更好地提高模型泛化能力和准确性
- **数据利用是大模型训练的“食谱”**
  - 更高效的资源调度与计算优化
    - 优化算力部分
  - 更稳定的迭代能力与模型进化
    - 优化算法（模型）部分



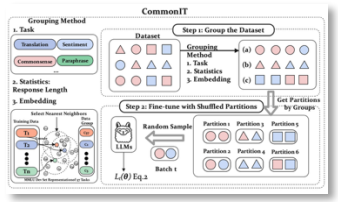
报告总览

- 3种方式帮助模型**更好地学习数据**

模型性能优化

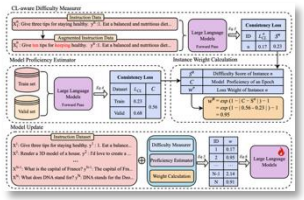
计算效率优化

捕获数据共性



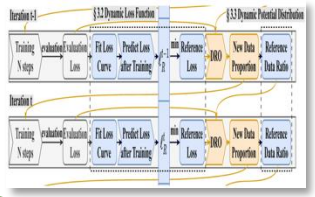
CommonIT：基于数据划分的共性感知指令微调方法

捕获模型反馈



CCL：数据驱动的课程一致性学习

捕获下游领域特性



DRPruning：基于数据分布鲁棒的模型剪枝

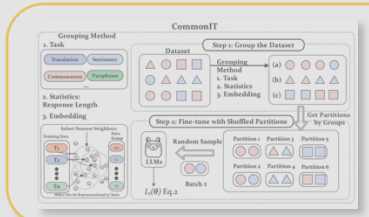
报告总览

- 3种方式帮助模型**更好地学习数据**

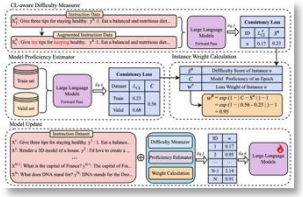
模型性能优化

计算效率优化

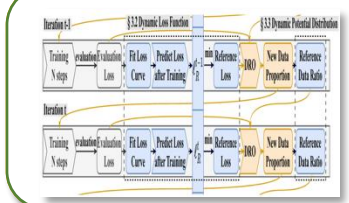
捕获数据共性



CommonIT：基于数据划分的共性感知指令微调方法



CCL：数据驱动的课程一致性学习



DRPruning：基于数据分布鲁棒的模型剪枝

捕获模型反馈

捕获下游领域特性

报告总览

- 3种方式帮助模型更好地学习数据

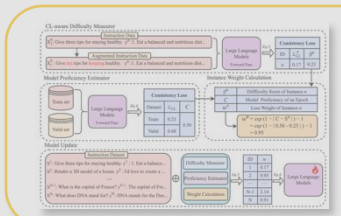
模型性能优化

计算效率优化

捕获数据共性

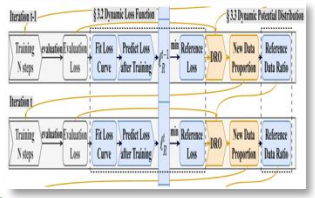
CommonIT：基于数据划分的共性感知指令微调方法

捕获模型反馈



CCL：数据驱动的课程一致性学习

捕获下游领域特性



DRPruning：基于数据分布鲁棒的模型剪枝

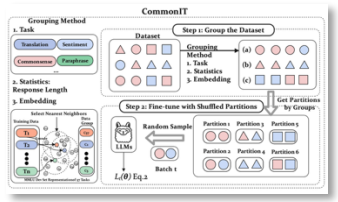
报告总览

- 3种方式帮助模型**更好地学习数据**

模型性能优化

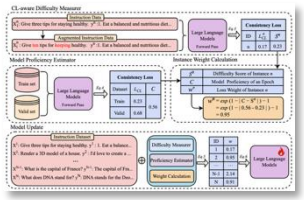
计算效率优化

捕获数据共性



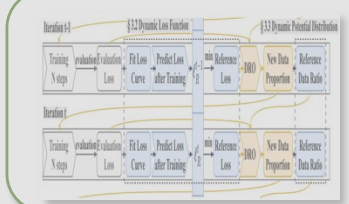
CommonIT：基于数据划分的共性感知指令微调方法

捕获模型反馈



CCL：数据驱动的课程一致性学习

捕获下游领域特性



DRPruning：基于数据分布鲁棒的模型剪枝

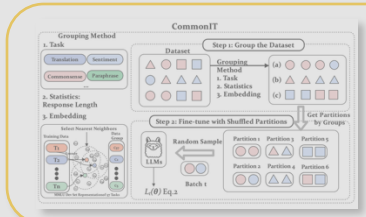
报告总览

- 3种方式帮助模型**更好地学习数据**

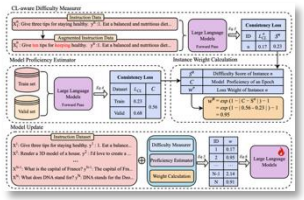
模型性能优化

计算效率优化

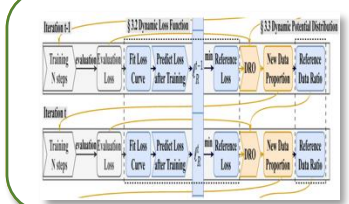
捕获数据共性



CommonIT：基于数据划分的共性感知指令微调方法



CCL：数据驱动的课程一致性学习



DRPruning：基于数据分布鲁棒的模型剪枝

捕获模型反馈

捕获下游领域特性

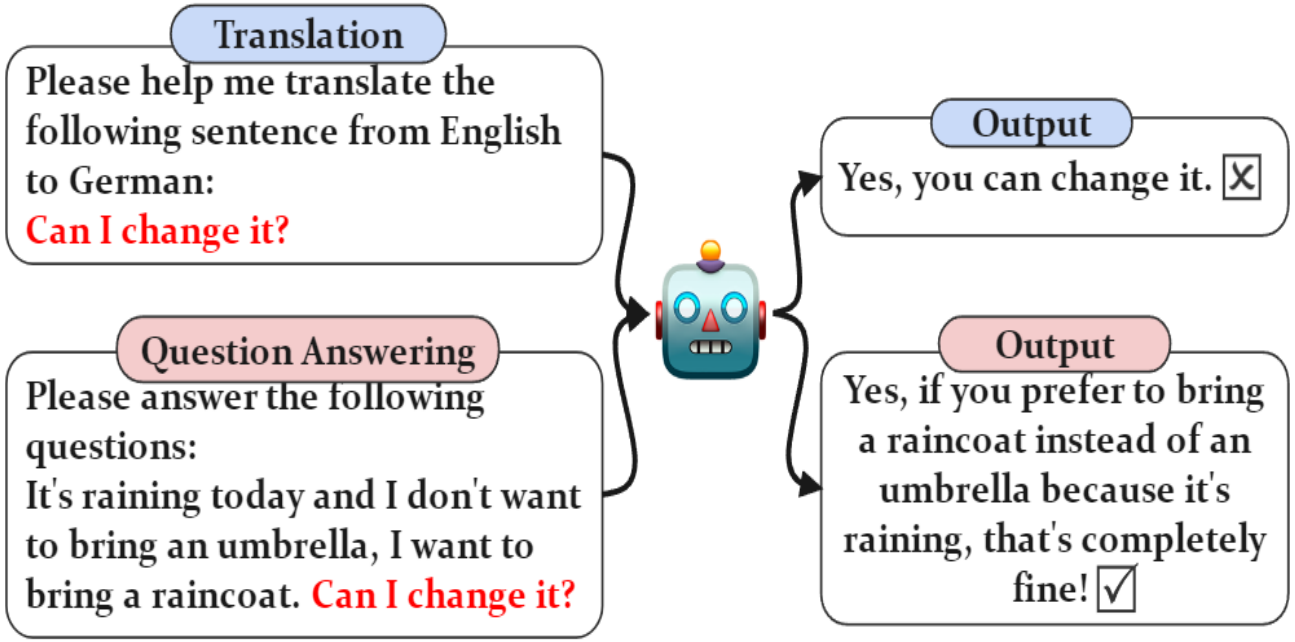
CommonIT: Commonality-Aware Instruction Tuning for Large Language Models via Data Partitions

Jun Rao[1], Xuebo Liu[1], Lian Lian[2], Shengjun Cheng[2], Yunjie Liao[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]Huawei Cloud Computing Technologies Co., Ltd.

EMNLP 2024

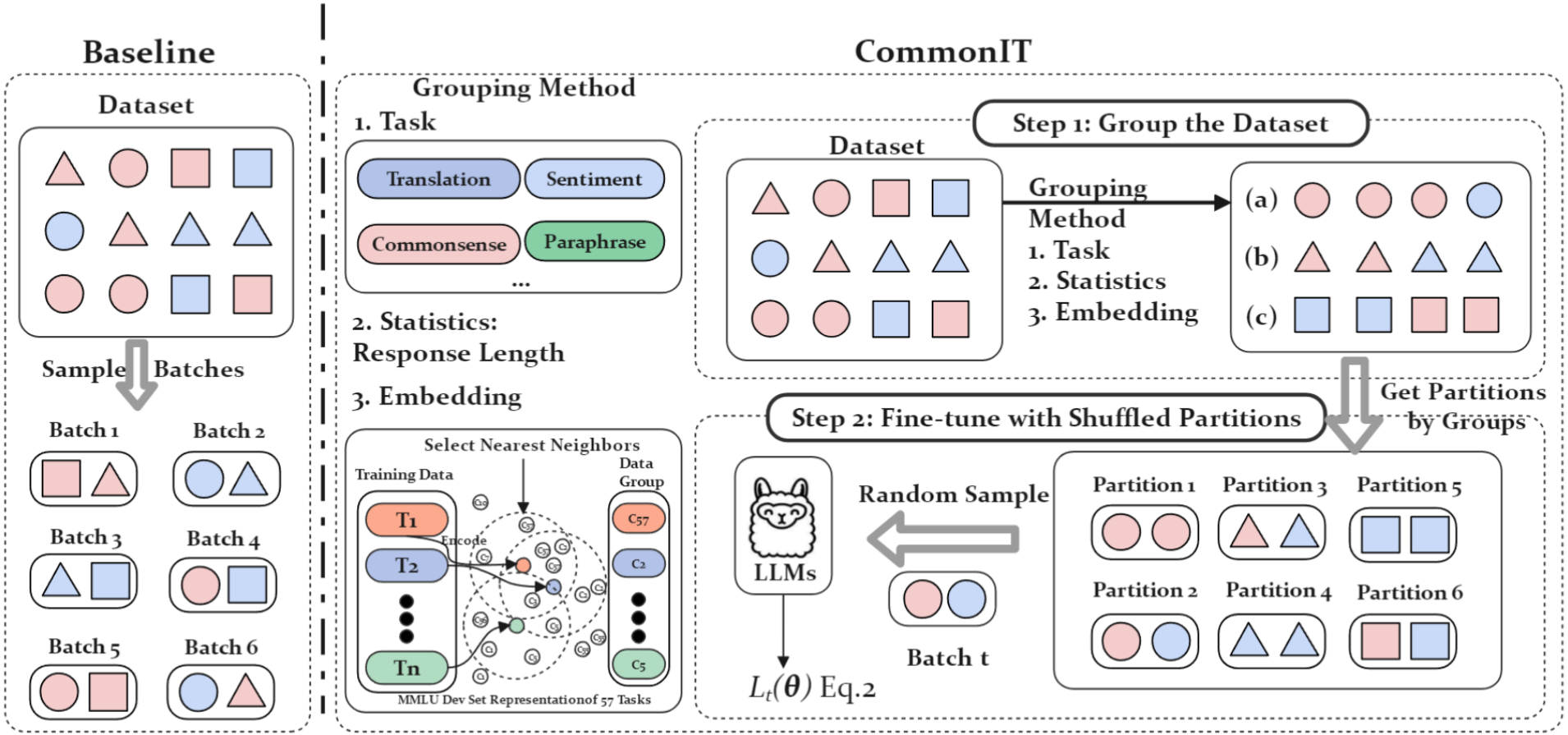
问题引入

- 指令微调数据特点：数据多样性和数据混合策略
  - 问题：混合了多种指令和任务，导致部分情况下指令遵循能力降低
  - ※ **捕获数据共性，提升学习效果；学习更加对齐预训练阶段**



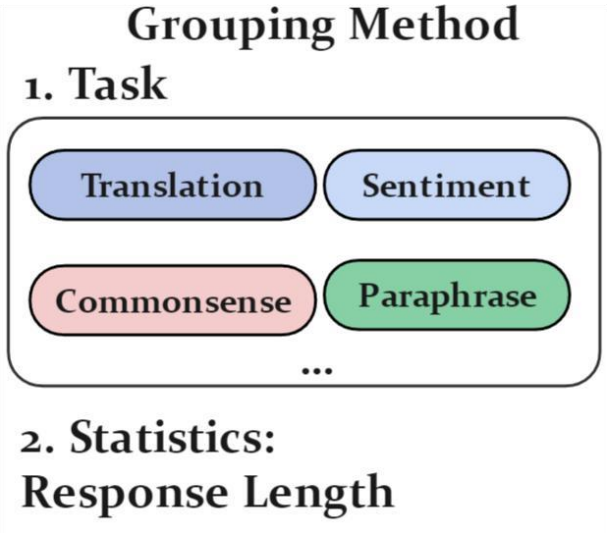
LLM 对任务指令理解不清容易导致性能下降

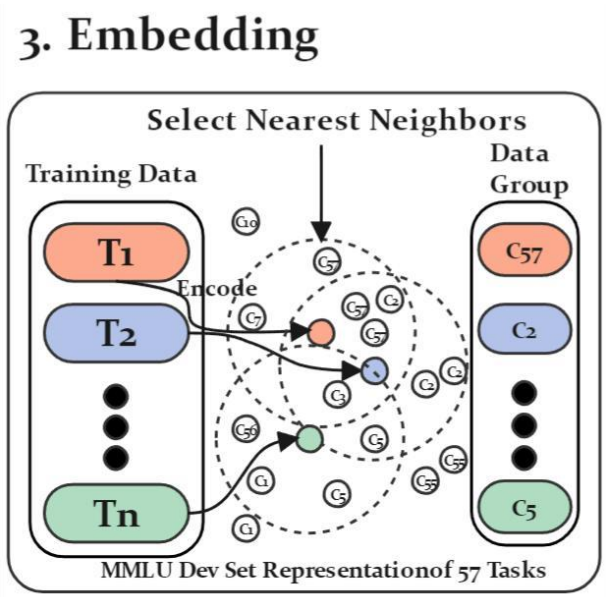
- **任务类别划分**
  - 把数据按特定规则归类
- **特定Batch采样策略**
  - 保证Batch内数据属于同一类别



类别定义

- 对指令进行聚类，以特定方式对数据进行微调，从而捕获数据共性：
  - 任务：数据属于的类别，如翻译、问答、写代码等
  - 统计指标：一些常用的统计信息，如回复的长度
  - 表征聚类：数据中问题的表征，利用聚类算法得到类别



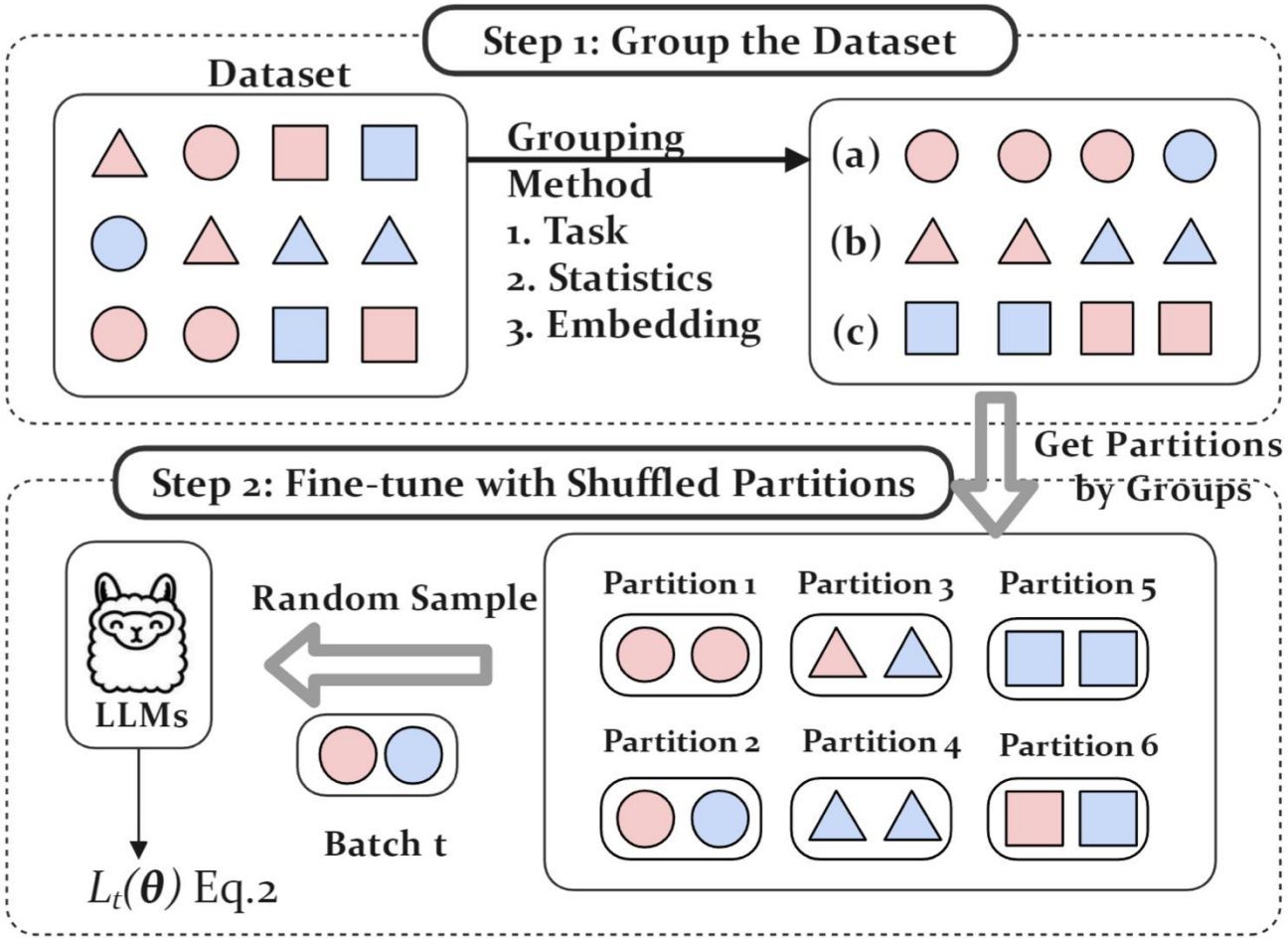


- **训练时约束**
  - 保证数据训练时每个batch内的数据都来自一个类别
  - 如右图所示：每批数据只有 a 或 b 或 c

$$\mathcal{B}_1^*, \cdots, \mathcal{B}_t^*, \cdots, \mathcal{B}_T^* = \text{sample}^*(\mathcal{D})$$

- 最终训练目标：原始SFT loss

$$L_t(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N} \log P(\boldsymbol{y}_t^{(i)} | \boldsymbol{x}_t^{(i)}; \boldsymbol{s}_t^{(i)}; \boldsymbol{\theta})$$



实验效果

- 使用3个指令微调数据集
- 在知识、推理、多语言、代码四大领域的测试集进行测试
- 使用多种类型模型，包括Llama 1及2系列7B-13B，以及Bloom、Qwen模型

Dataset	Model	MMLU		BBH		TydiQA	Codex-Eval	AVG.
		0-shot	5-shot	Direct	CoT	F1	P@10	
-	LLaMa 7B	32.1	35.2	34.0	33.3	37.0	18.3	31.7
FLAN CoT	IT*	41.3	42.5	33.7	31.3	44.4	17.3	35.1
	IT	37.1	38.3	32.9	34.1	47.5	19.3	34.9
	CommonIT							
	By Embedding	**40.2**	41.4	**36.1**	33.5	45.8	17.4	35.7
	By Length	38.7	**42.3**	33.4	**35.2**	**47.9**	**20.2**	**36.3**
Alpaca	IT*	42.6	38.3	28.5	32.3	23.6	25.0	31.7
	IT	34.8	36.4	32.6	33.0	37.4	22.6	32.8
	CommonIT							
	By Embedding	**41.1**	40.1	**33.6**	33.8	38.7	23.0	35.1
	By Length	40.4	**40.1**	33.5	**34.6**	**38.9**	**24.7**	**35.4**
FLAN	IT*	45.4	47.1	38.6	36.1	45.0	12.9	37.5
	IT	44.2	45.2	38.3	37.2	45.1	16.8	37.8
	CommonIT							
	By Task	46.6	47.4	38.9	37.2	45.7	19.6	39.2
	By Embedding	**47.2**	**48.5**	38.9	37.9	44.5	**21.8**	39.8
	By Length	46.7	47.9	**39.7**	**39.9**	**47.2**	19.3	**40.1**

在多个模型上均优于原始指令微调的模型，长度度量在通用领域中最有效

实验效果

- 使用3个指令微调数据集
- 在特定任务的测试集进行测试（通用知识、数学、工具调用、代码）
- 使用LLama系列上进行实验

Dataset	Model	MMLU (0-shot/5-shot)				
		Humanities	Social.	STEM	Other	AVG.
-	LLaMa 7B	31.5/31.5	31.2/37.3	29.7/32.3	36.1/41.3	32.1/35.2
FLAN CoT	IT	34.6/36.7	41.1/42.3	30.8/31.1	42.7/43.5	37.1/38.3
	CommonIT					
	By Length	34.9/**39.2**	**44.4/48.2**	**32.7/33.9**	44.5/**49.1**	38.7/**42.3**
	By Embedding	**38.3**/38.3	43.5/46.1	32.1/33.7	**47.3**/48.6	**40.2**/41.4
Alpaca	IT	34.4/35.4	35.3/35.4	28.2/30.5	41.0/41.3	34.8/36.4
	CommonIT					
	By Length	38.3/**38.5**	43.4/42.8	32.2/32.2	48.3/**47.1**	40.4/40.1
	By Embedding	**39.6**/37.8	**44.1/43.2**	**32.5/33.3**	**48.5**/46.9	**41.1/40.1**
FLAN	IT	42.7/42.1	49.9/50.7	34.1/37.0	50.5/52.1	44.2/45.2
	CommonIT					
	By Task	44.0/45.2	53.8/54.3	36.9/38.0	52.5/52.6	46.6/47.4
	By Length	44.5/44.5	53.6/54.7	35.9/39.5	53.3/54.2	46.7/47.9
	By Embedding	**44.8/45.3**	**53.9/55.6**	**37.3/39.8**	**53.7/54.7**	**47.2/48.5**

基于特定领域， Embedding 指标最有效

Model/Domain	GSM	OpenFunctions	Code
IT	39.0	30.4	23.6
Common IT			
By Length	36.0 (-3.0)	31.3 (+0.9)	28.2 (+4.6)
By Embedding	39.0 (+0.0)	34.8 (+4.4)	28.3 (+4.7)
By Task	**43.5 (+4.5)**	**35.7 (+5.3)**	**29.3 (+5.7)**

基于特定任务，任务指标最有效

CommonIT 总结

- CommonIT是一种基于数据共性的指令微调方法
  - 根据数据的共性定义类别 → 适用性广，可拓展
  - 多个模型任务在不同指标下均有效 → 泛化性强
  - 相同微调数据下效率更高 → 降低数据训练成本
- 如何模拟其他类型的数据共性?
  - 数据用途：事务数据、报告数据
  - 数据来源：内部数据、外部数据
  - 适合用于提升不同的下游任务（仅用于内部的大模型，充分捕获内部数据共性）

[1] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, Quoc V. Le. Finetune Language Models are Zero-Shot Learners. ICLR 2022.

[2] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, Tatsunori B. Hashimoto. Alpaca: A Strong, Replicable Instruction-Following Model. https://crfm.stanford.edu/2023/03/13/alpaca.html.

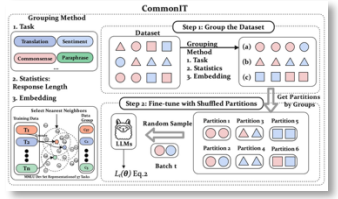
报告总览

- 3种方式帮助模型**更好地学习数据**

模型性能优化

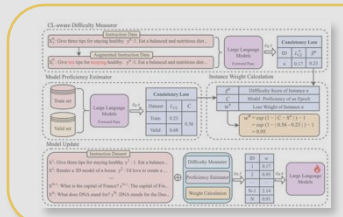
计算效率优化

捕获数据共性



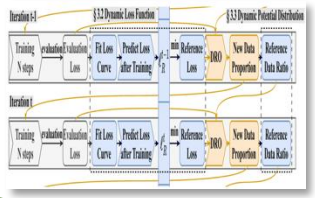
CommonIT：基于数据划分的共性感知指令微调方法

捕获模型反馈



CCL：数据驱动的课程一致性学习

捕获下游领域特性



DRPruning：基于数据分布鲁棒的模型剪枝

Curriculum Consistency Learning for Conditional Sentence Generation

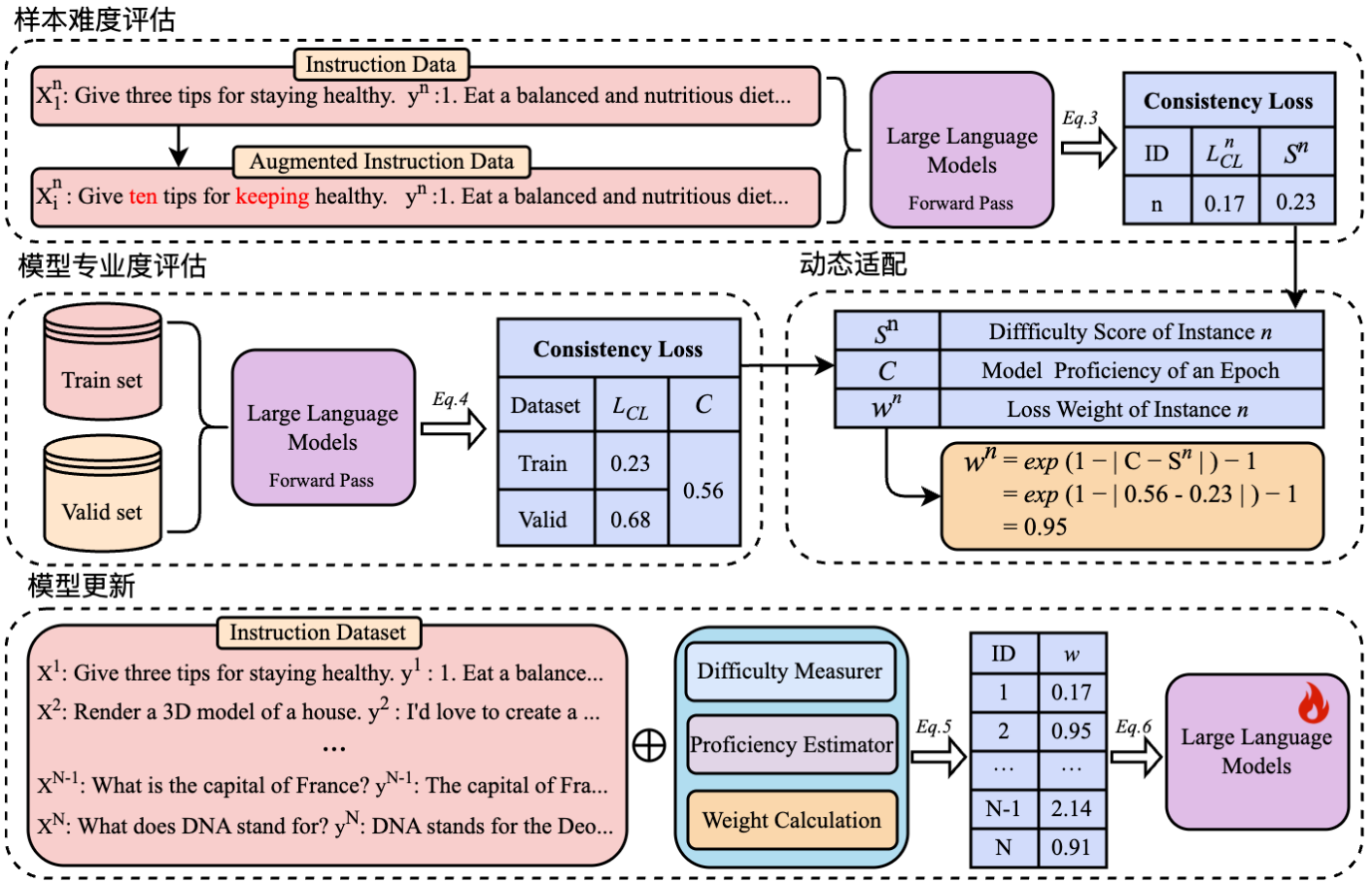
Liangxin Liu[1], Xuebo Liu[1], Lian Lian[2], Dong Jin[2], Shengjun Cheng[2], Jun Rao[1], Tengfei Yu[1] , Hexuan Deng[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]Huawei Cloud Computing Technologies Co., Ltd.

EMNLP 2024

CCL框架图

- **课程学习：让模型学习符合当前模型能力的知识**
  - 让模型从简单的样本开始训练，逐步引入更复杂的样本和知识点。
- **问题：如何定义大模型训练样本的难易度？如何评估大模型的当前能力？**

- **样本数据难度评估**
  - 设计样本数据难度衡量指标
- **模型专业度评估**
  - 设计模型专业度衡量指标
- **模型更新**
  - 动态适配更新模型



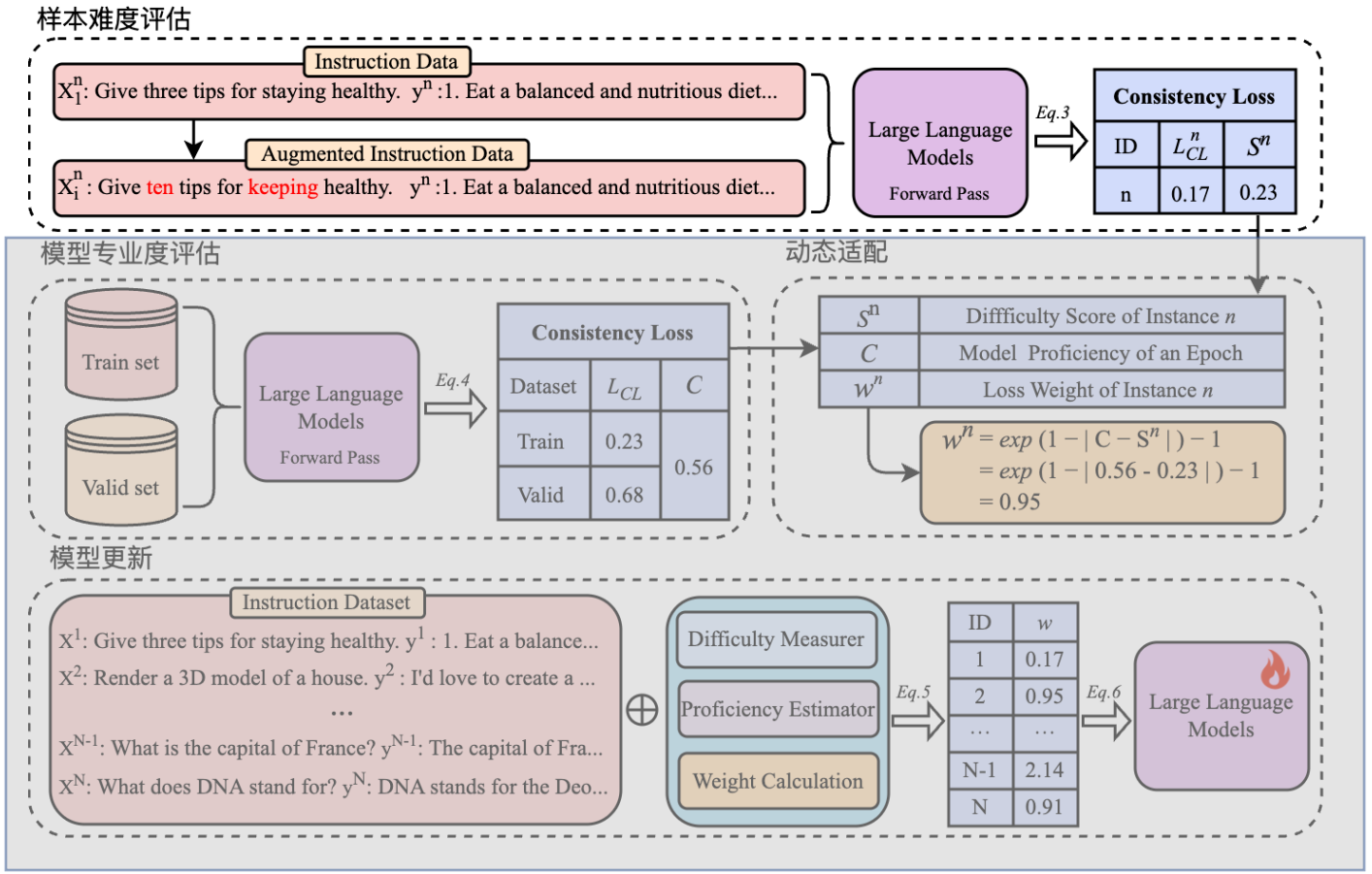
自动化指标：样本数据难度

- **样本数据难度评估技术**
  - **（引入一致性学习理论）**
    - 如果一个特定的输入数据x在经过多种扰动之后，模型预测仍然维持高度一致性，那么可以推断该样本数据具有较低的难度，即其L_{cl}值较低。

$$\begin{aligned}\mathcal{L}_{\mathrm{CL}}^{n} &= \alpha \mathrm{Divergence}(\{P(y^n|x^n,\theta_i)\}_{i=1}^{2}) \\ &= \alpha(\mathrm{KL}(P(y^n|x^n,\theta_1)||P(y^n|x^n,\theta_2)) + \\ &\quad \mathrm{KL}(P(y^n|x^n,\theta_2)||P(y^n|x^n,\theta_1)))\end{aligned}$$

- 相反，一致性损失越大，意味着该样本数据的难度也越大。

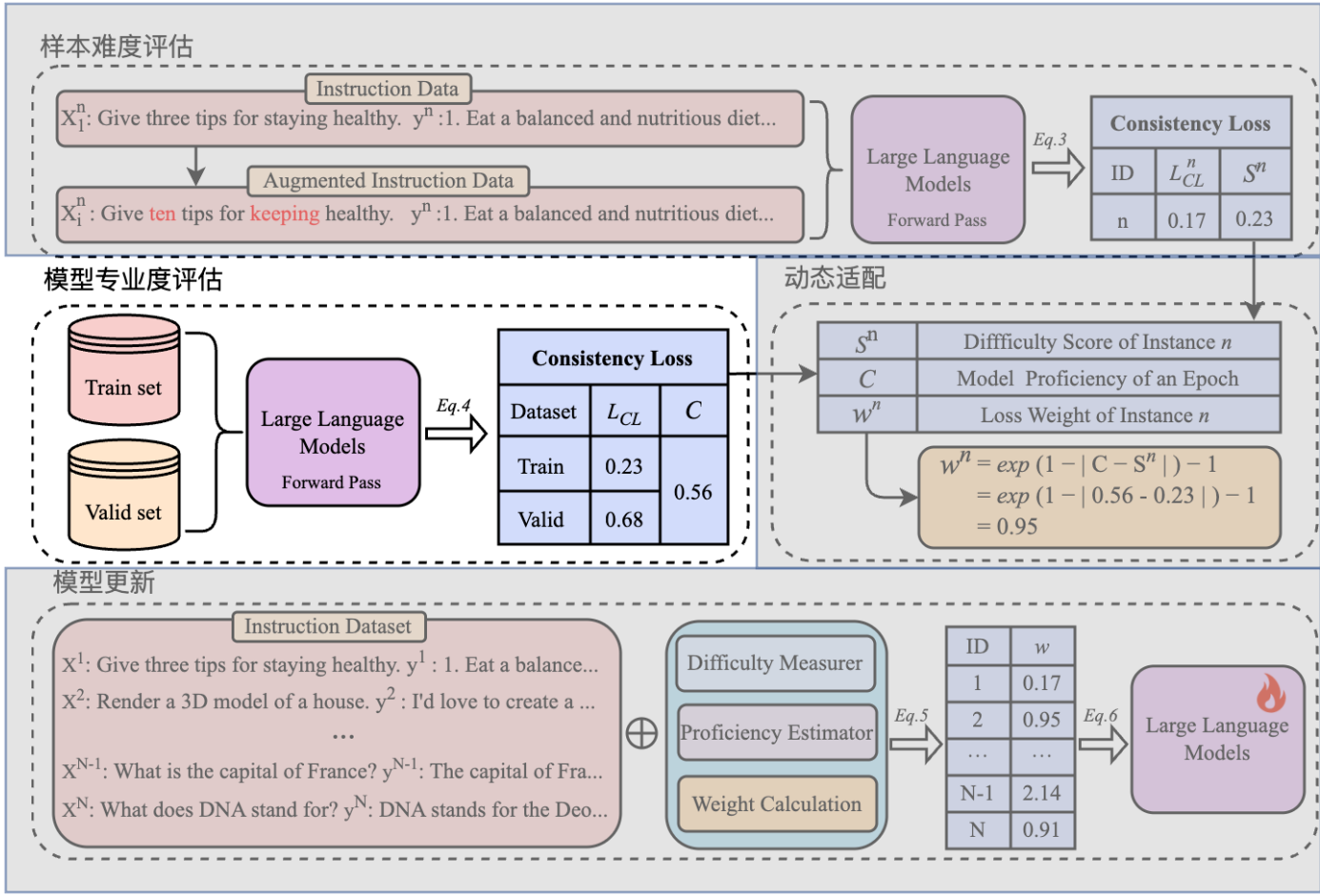
$$S^n \in [0,1) = \frac{\mathcal{L}_{\mathrm{CL}}^{n} - \min\{\mathcal{L}_{\mathrm{CL}}^{n}\}_{n=1}^{N}}{\max\{\mathcal{L}_{\mathrm{CL}}^{n}\}_{n=1}^{N}}$$



自动化指标：模型专业度

- **模型专业度评估技术**
  - 采用经过训练的指令数据集一致性损失$L_{dist}^{训练集}$和未经过训练的验证集的一致性损失值$L_{dist}^{验证集}$集的差异描述模型能力。

$$C = min\left(1, \mid L_{dist}^{验证集} - L_{dist}^{训练集} \mid /\lambda\right)$$



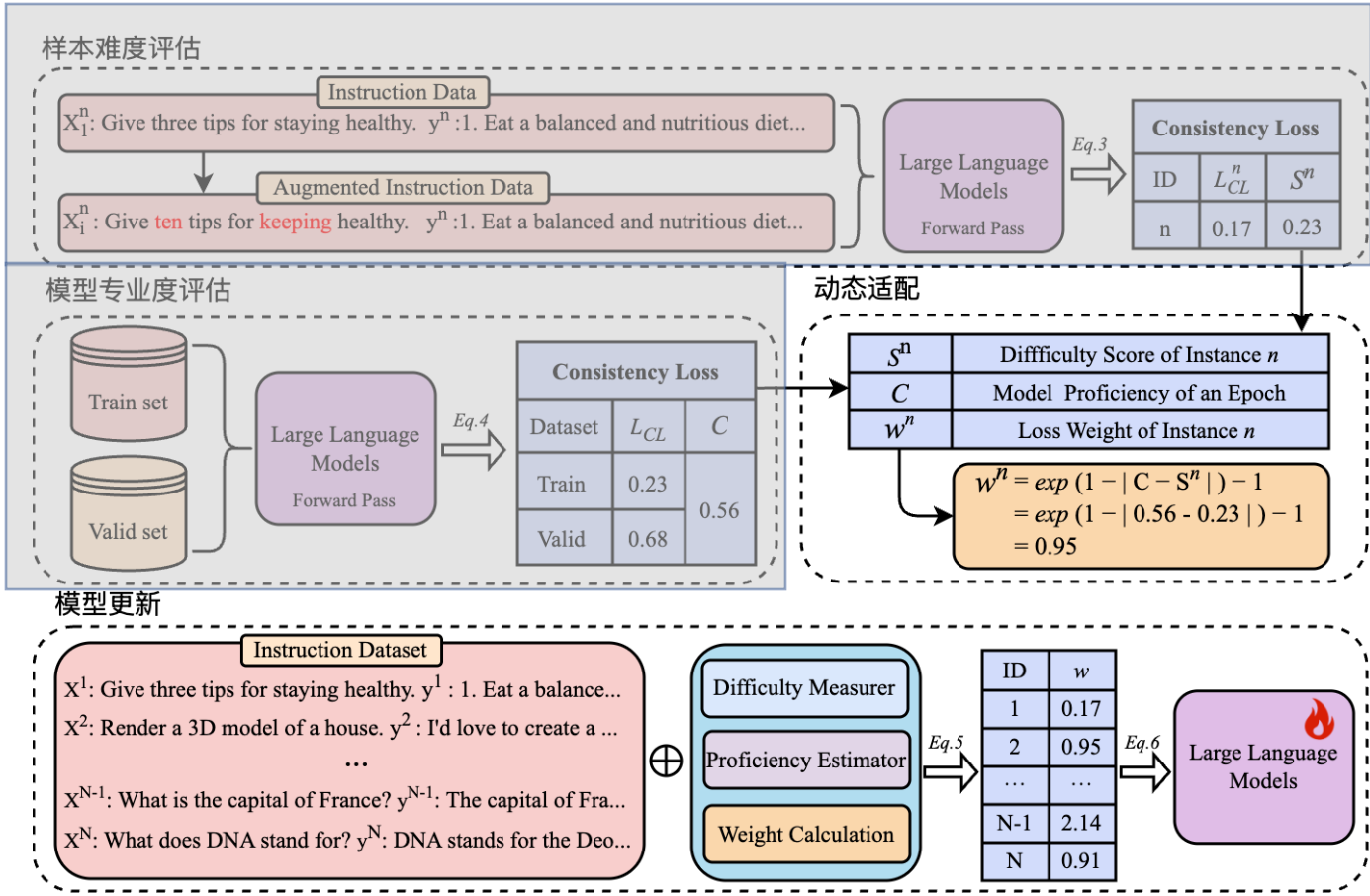
- **动态适配和模型更新**
  - 通过将样本数据难度和模型能力整合，实现了基于课程学习的目标域能力导向的大模型训练与微调。

$$\omega^n = exp(1 - |C - S^n| - 1)$$

  - 当$|C - S^n|$越小时，损失权重应较大，即ω值越大，这意味着模型将更加关注这些合适的指令数据，反之亦然。

  - 基于课程学习的大模型微调策略的最终损失函数可以表示为：

$$L^n_{total} = \omega^n(L^n_{lm} + \alpha \cdot L^n_{cl})$$

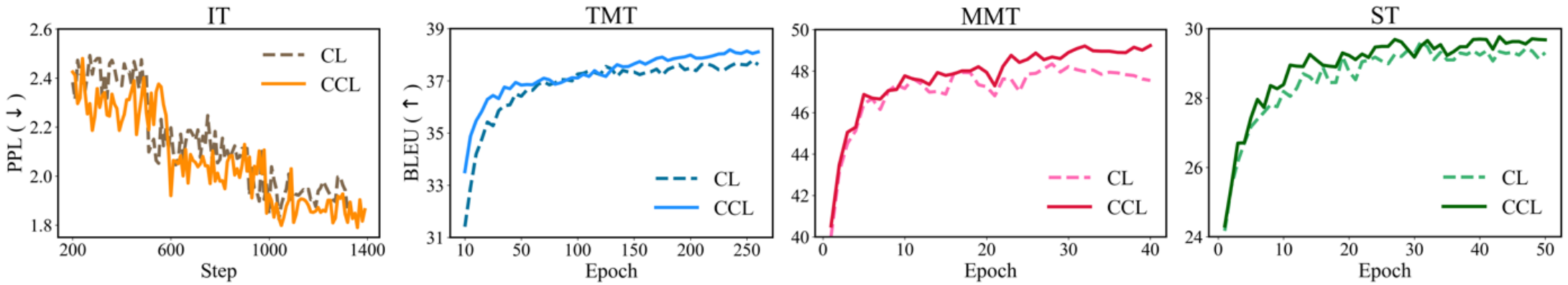


指令微调实验结果

- **CCL方法与主流方法在LLaMA-2-7B和LLaMA-2-13B上面进行对比验证**

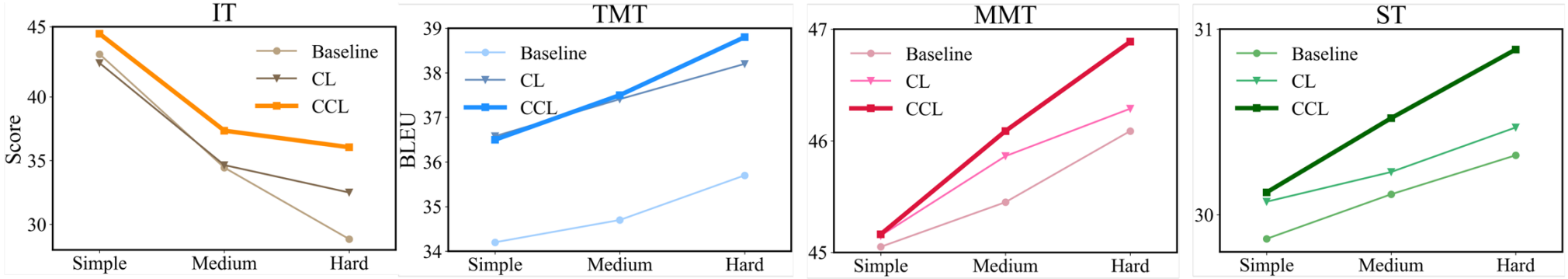
ID	System	External Data	External Model	MMLU	BBH	GSM	TydiQA	CodeX	AE	Overall AVG	Overall Δ (↑)
Base Model: LLaMA-2-7B			*Implemented Existing Method*								
1	Alpaca-GPT4			46.2	39.2	15.0	43.3	27.8	33.7	34.3	-
2	1 + AlpaGasus	✗	✓	46.8	39.2	14.5	**48.4**	26.5	34.6	35.0	+0.8
3	1 + Q2Q	✗	✓	46.7	39.8	**16.5**	45.5	28.1	35.1	35.3	+1.1
4	1 + Instruction Mining	✓	✓	47.0	40.0	**16.5**	47.8	29.6	34.4	35.9	+1.7
5	1 + Data CL	✗	✗	47.3	39.0	14.5	43.7	29.5	35.1	29.5	+0.6
				Our Method							
6	5 + CCL	✗	✗	**47.6**	**40.8**	15.5	**48.4**	**30.7**	**35.6**	**36.4**	**+2.1**
Base Model: LLaMA-2-13B				*Implemented Existing Method*							
7	Alpaca-GPT4			**55.7**	47.3	31.0	49.1	41.8	46.5	45.2	-
8	7 + AlpaGasus	✗	✓	54.1	**49.3**	32.0	52.6	39.3	47.5	45.8	+0.6
9	7 + Q2Q	✗	✓	55.3	48.5	**34.0**	50.8	42.3	**47.7**	46.4	+1.2
10	7 + Instruction Mining	✓	✓	55.5	49.7	33.0	51.2	41.5	46.1	46.2	+1.0
11	7 + Data CL	✗	✗	55.2	47.2	33.0	51.2	40.8	46.2	45.6	+0.4
				Our Method							
12	11 + CCL	✗	✗	55.6	**49.3**	**34.0**	**53.6**	**42.7**	47.6	**47.1**	**+1.9**

CCL通过充分利用大语言模型自身的能力，结合课程学习的优化方法，显著提升了指令微调性能，比传统一致性学习方法提高了约2.1分。



- **探究**模型在不同任务上的学习曲线

CCL拥有更快的学习收敛速度



- **探究**模型在不同难度数据集上面的性能表现

CCL能够更好处理挑战性数据

CCL 总结

- **CCL是一种动态数据优化方法**
  - **样本难度评估 → 通过一致性学习识别样本难度数据**
  - **模型专业度评估 → 通过一致性学习建模模型当前专业度**
  - **动态适配模型更新 → 结合样本数据难度及模型能力动态更新**

参考文献

[1] Xiaobo Liang, Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, Tie-Yan Liu. R-Drop: Regularized Dropout for Neural Networks. NeurIPS 2021.

[2] Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, Jianfeng Gao. Instruction Tuning with GPT-4. arXiv 2023.

[3] Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, Omer Levy. LIMA: Less Is More for Alignment. NeurIPS 2023.

[4] Lichang Chen, Shiyang Li, Jun Yan, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Srinivasan, Tianyi Zhou, Heng Huang, Hongxia Jin. AlpaGasus: Training a Better Alpaca with Fewer Data. ICLR 2024.

[5] Ming Li, Yong Zhang, Zhitao Li, Jiuhai Chen, Lichang Chen, Ning Cheng, Jianzong Wang, Tianyi Zhou, Jing Xiao. From Quantity to Quality: Boosting LLM Performance with Self-Guided Data Selection for Instruction Tuning. NAACL 2024.

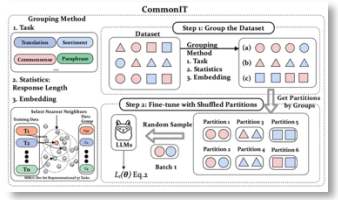
报告总览

- 3种方式帮助模型**更好地学习数据**

模型性能优化

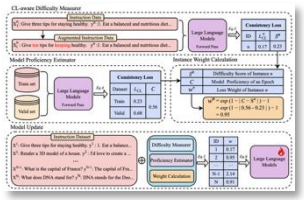
计算效率优化

捕获数据共性



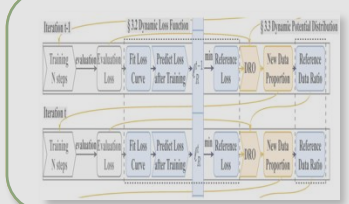
CommonIT：基于数据划分的共性感知指令微调方法

捕获模型反馈



CCL：面向条件句生成的课程一致性学习

捕获下游领域特性



DRPruning：基于数据分布鲁棒的模型剪枝

DRPruning: Efficient Large Language Model Pruning through Distributionally Robust Optimization

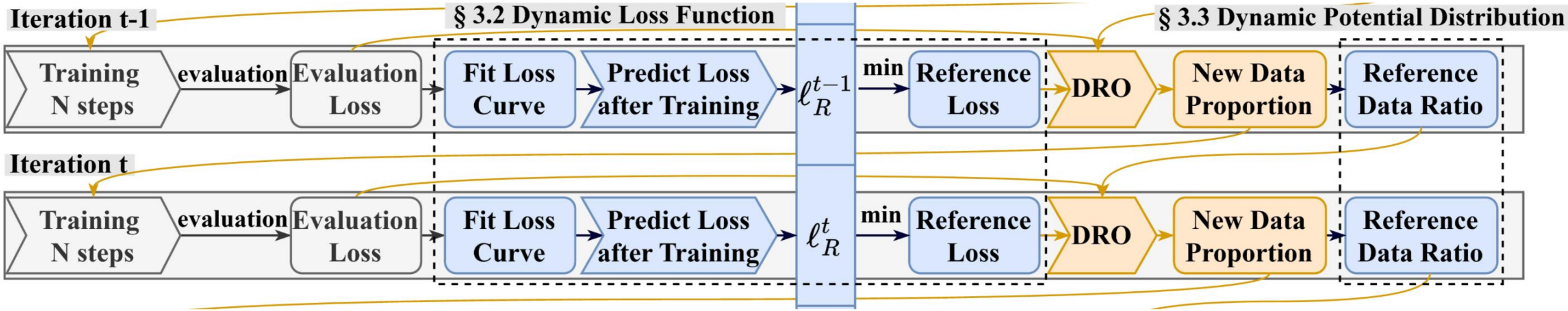
Hexuan Deng[1], Wenxiang Jiao, Xuebo Liu[1], Min Zhang[1], Zhaopeng Tu
[1]Harbin Institute of Technology, Shenzhen

ACL 2025

DRPruning框架图

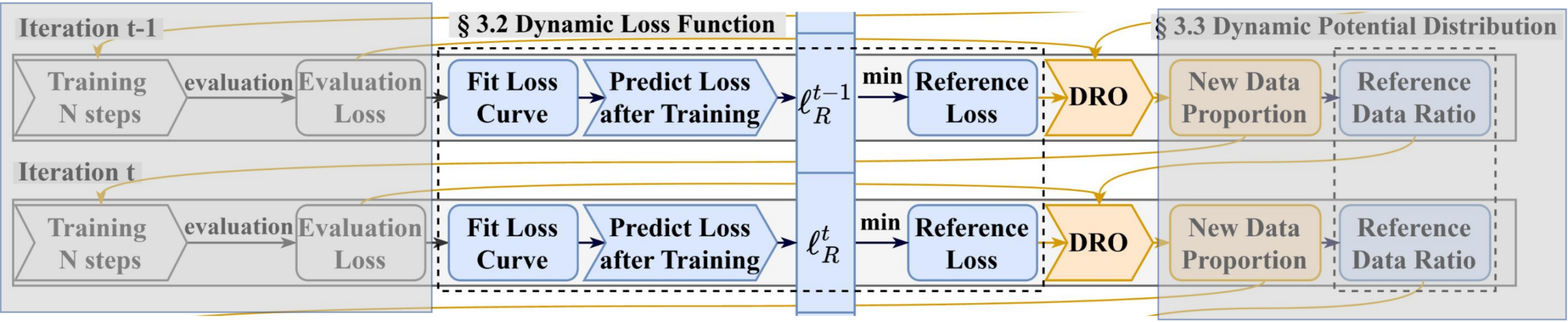
- **捕获下游领域特性：让模型更好的掌握难领域知识**
  - 根据训练动态感知难度较大的领域，在训练中为其分配更高权重
  - **问题**：**如何评估大模型的各领域能力？如何定义大模型训练领域的难易度？**

- **动态决定期望Loss**
  - 为各个领域设定更合理的训练目标
- **动态决定期望数据配比**
  - 动态偏向性能较差领域



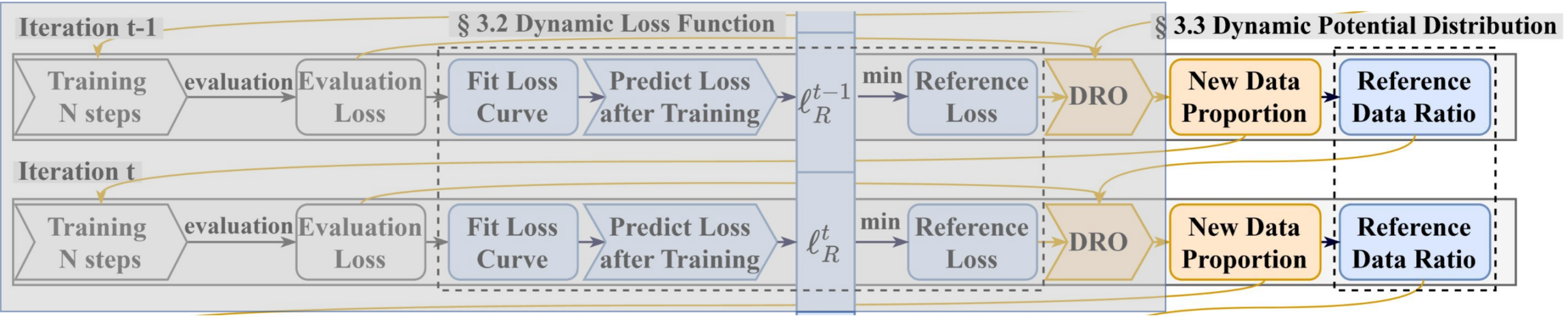
- **传统训练过程**
  - 使用无标签文本进行结构化剪枝与继续预训练恢复能力
  - **两阶段训练**：少量数据（0.4B）用于剪枝，中量数据（50B）用于增训以恢复能力

- **领域训练动态感知的期望Loss动态优化**
  - 目的：计算对于分布迁移鲁棒的期望Loss，以设定更合理的训练目标
  - 拟合Loss，数据量P与模型参数量T间的函数，并预测训练结束时的Loss作为期望Loss
  - 训练10%的步数后则收集足够的点用于拟合，随后每1%的步数更新一次



$$\hat{\ell}(P,T) = A \cdot \frac{1}{P^{\alpha}} \cdot \frac{1}{T^{\beta}} + E$$

- **领域难度感知的期望数据配比动态优化**
  - 目的：让模型对于更困难的异质性数据分布鲁棒
  - 将期望数据配比p_R向更难的方向q移动，提高对于困难领域的训练权重
  - 训练20%的步数，在当前分布充分训练后才启用，随后每1%的步数更新一次

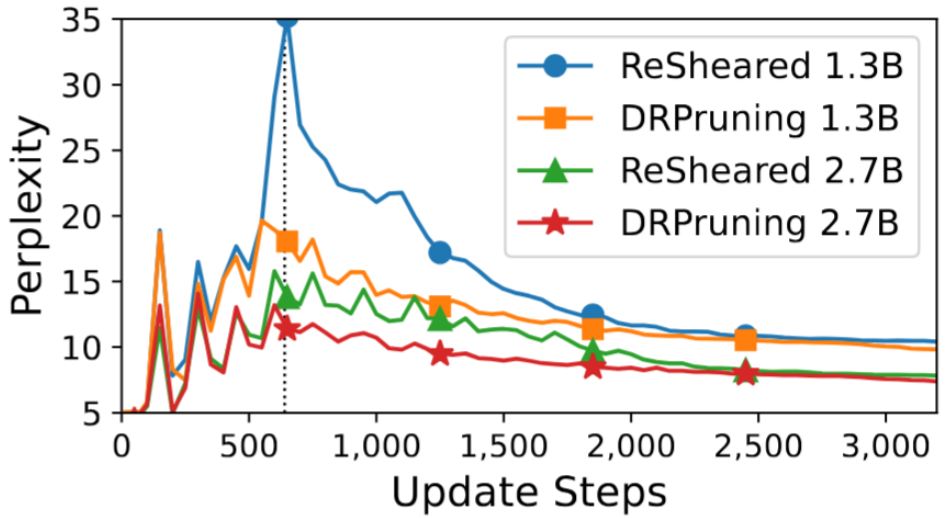


$$\mathbf{p}_R^{t+1} = \delta \cdot \mathbf{q}^t + (1-\delta) \cdot \mathbf{p}_R^t$$

DRPruning垂域任务实验效果

- **剪枝后模型测试：**
- 使用Llama2-7B作为基模型，剪枝到1.3B和2.7B

Method	From	To	PPL ↓	Task ↑
Sheared Llama	7B	1.3B	10.05	34.89
ReSheared	7B	1.3B	10.42	34.85
DRPruning	7B	1.3B	**9.83**	**35.60**
Sheared Llama	7B	2.7B	7.64	39.75
ReSheared	7B	2.7B	7.83	39.98
DRPruning	7B	2.7B	**7.40**	**40.18**



DRPruning是更优的剪枝算法，收敛速度显著更高

DRPruning垂域任务实验效果

- **多领域任务数据集测试：**
- 使用Llama2-7B作为基模型，剪枝到1.3B和2.7B，并经过50B增训

Tasks	7B	2.7B				1.3B			
	Llama2†	Pythia†	Sheared†	ReSheared	DRPrun.	Pythia†	Sheared†	ReSheared	DRPrun.
WSC	36.54	38.46	**48.08**	36.54	46.15	36.54	36.54	40.38	**50.00**
TriQA (5)	64.16	27.17	42.92	40.14	**43.33**	18.19	26.03	24.98	**28.10**
NQ (5)	25.98	7.12	14.85	13.49	**15.82**	4.79	8.75	8.39	**10.44**
TruthQA	32.09	28.79	**30.21**	28.41	30.13	**30.75**	29.12	28.09	29.68
LogiQA	30.11	28.11	28.26	26.27	**28.73**	27.50	27.50	28.11	**28.88**
BoolQ	77.71	64.50	**65.99**	64.92	65.08	63.30	62.05	61.01	**63.36**
LAMB	73.90	64.76	**68.21**	66.18	66.91	**61.67**	61.09	58.84	60.28
MMLU (5)	44.18	**27.09**	26.63	25.70	26.99	26.75	25.70	26.60	**27.28**
SciQ	94.00	88.50	**91.10**	90.10	89.80	86.70	87.00	86.40	**87.70**
ARCE	76.35	64.27	67.34	**67.72**	67.13	60.40	**60.90**	60.35	**60.90**
ARCC (25)	52.65	36.35	**42.66**	40.10	40.53	33.02	33.96	**34.30**	33.62
PIQA	78.07	73.88	76.12	**76.71**	75.19	70.84	73.50	**74.59**	72.69
WinoG	69.06	59.83	**65.04**	63.38	64.72	57.38	57.85	**60.06**	58.01
SQuAD	40.02	26.81	**49.26**	49.17	44.69	22.66	29.57	**37.59**	35.06
HelS (10)	78.95	60.81	71.24	**72.03**	69.22	53.49	61.05	**63.06**	58.88
Average	58.25	46.43	**52.53**	50.72	51.63	43.60	45.37	46.18	**46.99**

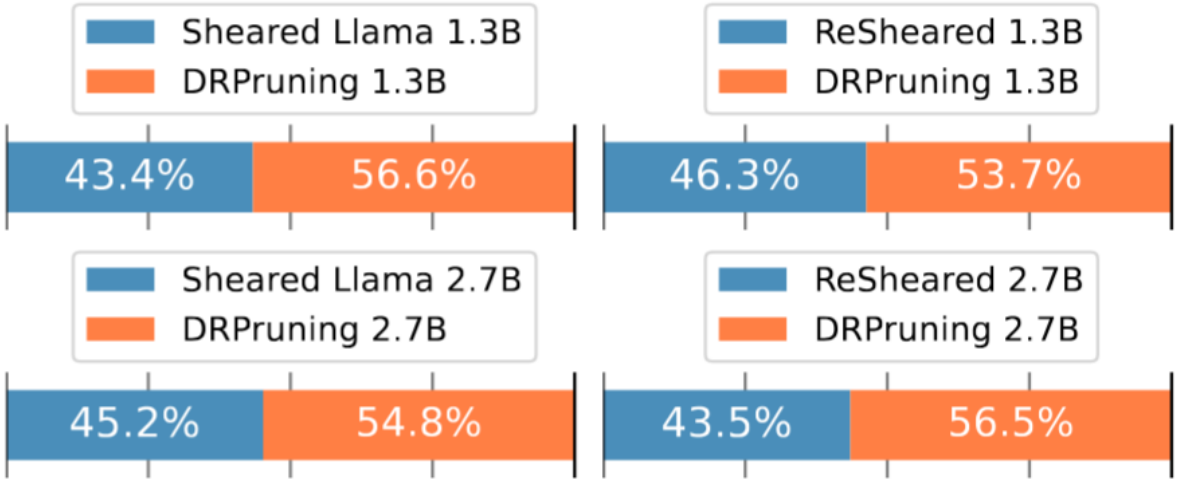
在使用相同数据的公平对比下（ReSheared vs. DRPruning），显著更优

DRPruning分析实验

Base Model	Prune	PT	Method	EN	RU	ZH	JA	AR	TR	KO	TH	Average
XGLM-1.7B	X	X	-	55.06	52.97	51.02	51.00	42.89	37.99	49.00	38.63	47.32
Qwen1.5-1.8B	X	X	-	60.89	52.30	**56.13**	53.30	42.17	34.98	48.25	36.75	48.10
Qwen2-1.5B	X	X	-	61.58	57.83	55.72	55.30	43.31	35.98	49.25	36.02	49.37
Qwen2-1.5B	X	✓	ReSheared	**62.16**	58.95	54.93	**55.60**	43.91	37.27	**54.05**	39.96	50.85
Qwen2-1.5B	X	✓	DRPruning	61.67	**59.09**	54.01	54.95	45.14	**46.91**	52.65	**44.42**	**52.35**
Qwen2-7B	✓	✓	DRPruning	60.43	56.80	55.72	55.05	**45.69**	43.82	53.95	43.53	51.87

- **多语种场景任务测试**

在分布偏移严重的场景表现更佳



- **垂域高效微调模型测试**

训练的模型是更优的基座模型

DRPruning总结

- DRPruning是一种数据异质性感知的模型剪枝方法
  - 分布鲁棒训练引入剪枝 → 缓解异质性领域数据上不均衡性能损失
  - 训练动态感知的期望Loss优化 → 解决数据与模型间的分布偏移
  - 难度感知的期望数据配比优化 → 让模型对更难的数据分布鲁棒

高效数据学习与利用总览

- 3种面向数据高效利用的建模方法
  - 静态数据共性利用
    - CommonIT：挖掘数据的共性进行指令微调，适用性广、泛化性强、效率更高
  - 动态数据适配利用
    - CCL：评估样本数据难度与模型能力，动态调整数据训练策略，优化学习效果
    - DRPruning：采用分布鲁棒训练方法剪枝，动态调整损失和数据配比，增强模型在异质数据上的鲁棒性
- 高效数据学习与利用瓶颈
  - 当前的方法仍主要基于传统的深度学习方法
  - 未针对大模型范式充分优化

01 大模型与数据合成背景

02 基础：通用与垂域数据合成

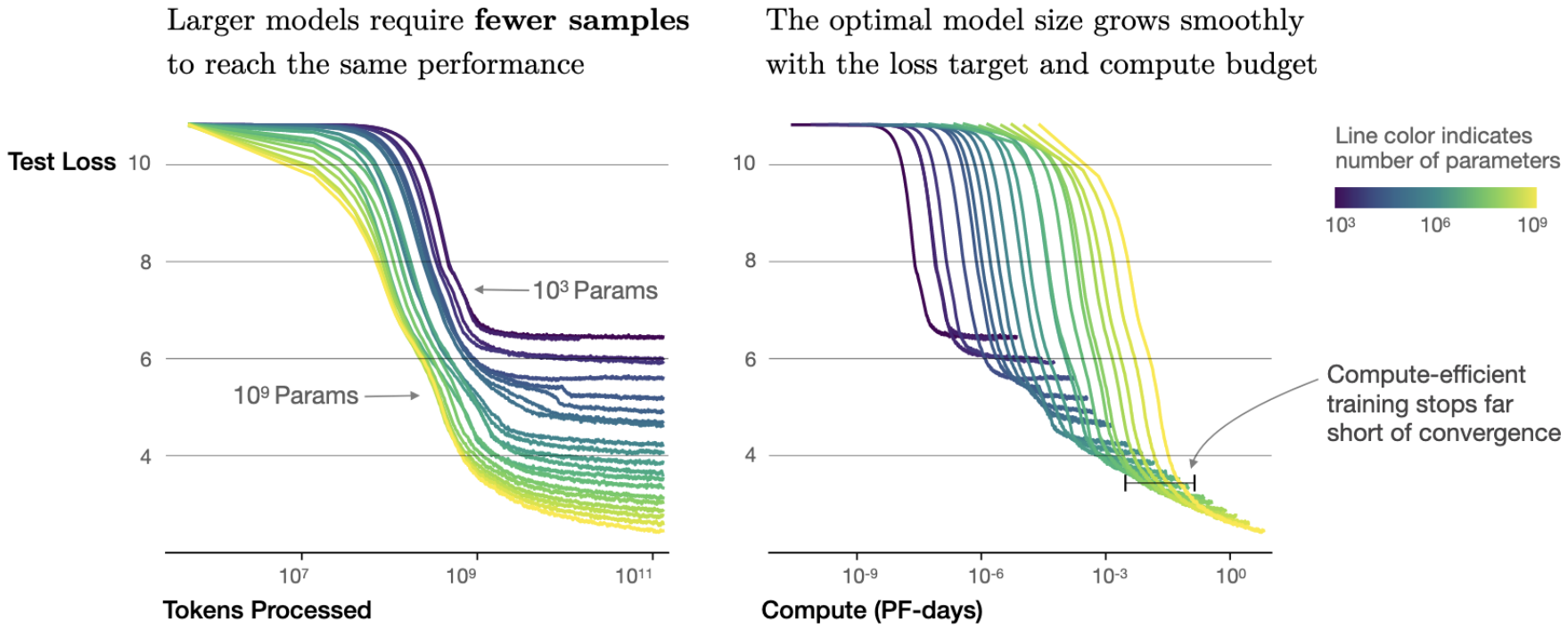
03 核心：高效数据学习与利用

04 进阶：“数据-模型”能力对齐

05 领域瓶颈与未来展望

背景：“数据-模型”能力对齐

- **大模型时代的核心矛盾：数据投入与能力提升失衡**
  - **现状：**业界长期依赖 “数据越多越好” 的范式，然而根据 Scaling Law，模型训练成本随数据规模呈指数级增长。例如，GPT-4 训练数据量是 GPT-3 的数十倍，训练成本超过 1 亿美元，而性能提升却无法匹配投入 。



背景：“数据-模型”能力对齐

- **能力因果链断裂：从数据到能力的黑箱映射**
  - 缺乏“数据属性→模型内部表征→目标能力”的可追溯分析框架，导致优化手段停留在“试错调参”。如盲目增加数据多样性却意外降低任务准确率。

“难题揭榜”第九十七期--四野会战难题第五期

欢迎大家毛遂自荐、踊跃揭榜。 对于解决难题或提供重大思路的，会给予及时激励！并张榜公布。 如有任何问题，请直接与接口专家联系； 如有其它建议，可与总负责金颖 jinying@hisilicon.com、樊玉伟fanyuwei2@huawei.com联系。

难题1	大模型课程学习技术
难题2	大模型运行时提示压缩技术
难题3	个性化 LLM-based Agent 难题
难题4	拓扑亲和的最优集合通信算法生成
难题5	面向AI流量的高性能路由与调度

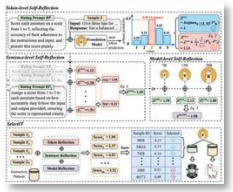
“难题揭榜”第100期-华为云难题第五期

欢迎大家毛遂自荐、踊跃揭榜。 对于解决难题或提供重大思路的，会给予及时激励！并张榜公布。 如有任何问题，请直接与接口专家联系； 如有其它建议，可与总架构师顾炯炯 dennis.gu@huawei.com 联系。

难题1	[AI平台-高可靠] AI集群中的任务调度和碎片卡整理技术
难题2	[LLM SFT] 行业大模型SFT数据动态配比技术
难题3	[数据]如何利用生成数据提升行业场景下的视觉理解能力
难题4	[行业模型] 基于图数据的大模型知识增强
难题5	无微调适配多领域的NL2SQL技术

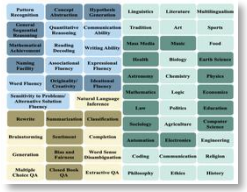
报告总览

- 攻克 “数据属性与模型能力映射关系不清” 的挑战
  - 模型自身决定能力（黑盒）：指令数据筛选
  - 人类已有理论决定能力（白盒）：基于认知理论的能力分类



SelectIT：不确定性感知自反思引导的选择性指令微调

黑盒

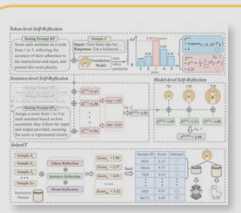


CDT：多维度的数据驱动大语言模型能力框架

白盒

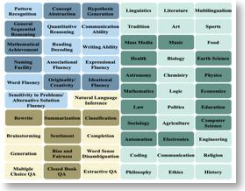
报告总览

- 攻克 “数据属性与模型能力映射关系不清” 的挑战
  - 模型自身决定能力（黑盒）：指令数据筛选
  - 人类已有理论决定能力（白盒）：基于认知理论的能力分类



SelectIT：不确定性感知自反思引导的选择性指令微调

黑盒

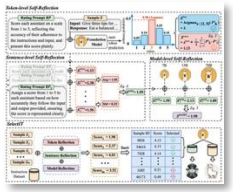


CDT：多维度的数据驱动大语言模型能力框架

白盒

报告总览

- 攻克 “数据属性与模型能力映射关系不清” 的挑战
  - 模型自身决定能力（黑盒）：指令数据筛选
  - 人类已有理论决定能力（白盒）：基于认知理论的能力分类



SelectIT：不确定性感知自反思引导的选择性指令微调

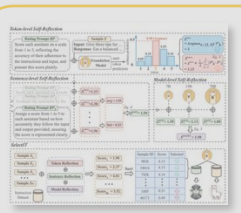
黑盒

CDT：多维度的数据驱动大语言模型能力框架

白盒

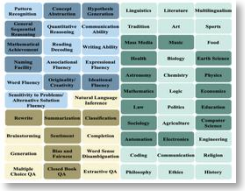
报告总览

- 攻克 “数据属性与模型能力映射关系不清” 的挑战
  - 模型自身决定能力（黑盒）：指令数据筛选
  - 人类已有理论决定能力（白盒）：基于认知理论的能力分类



SelectIT：不确定性感知自反思引导的选择性指令微调

黑盒



CDT：多维度的数据驱动大语言模型能力框架

白盒

SelectIT: Selective Instruction Tuning for LLMs via Uncertainty-Aware Self-Reflection

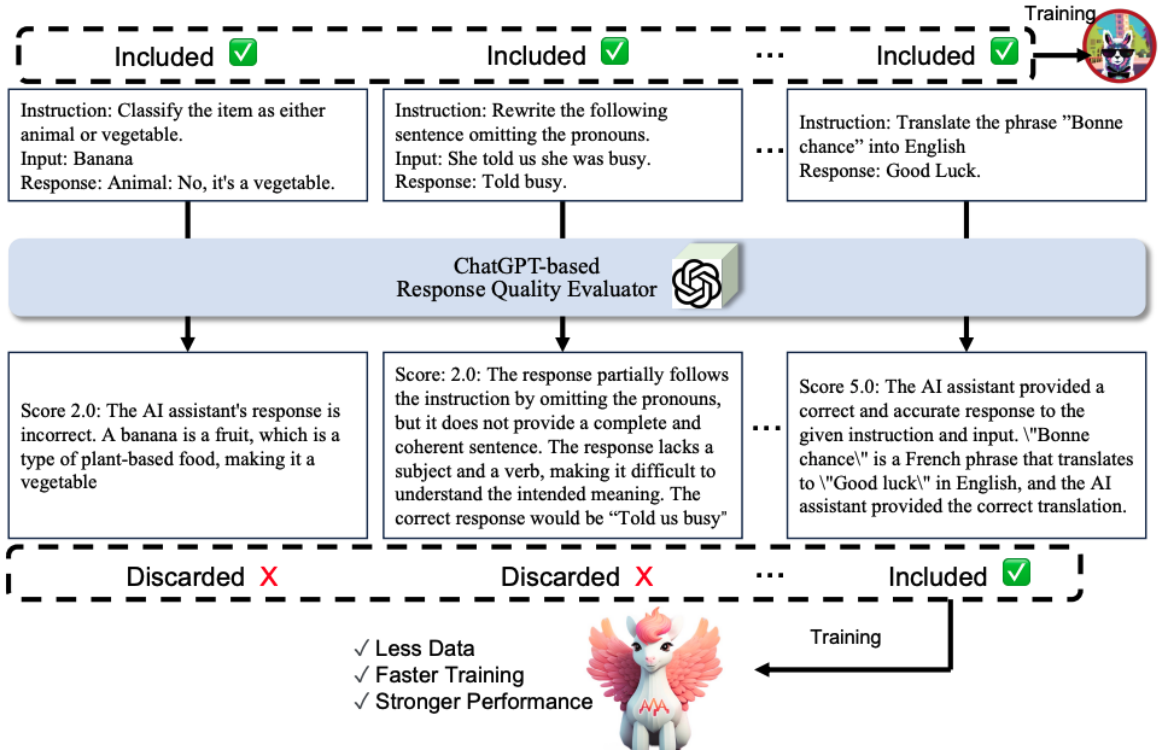
Liangxin Liu[1], Xuebo Liu[1*], Derek F. Wong[2], Dongfang Li[1], Ziyi Wang[1], Baotian Hu[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]University of Macau

NeurIPS 2024

研究背景：精简数据对指令微调的重要性

- **在指令微调中，指令数据的质量远比数量更为重要**

- 目前数据挑选方法普遍依赖额外资源
  - 外部模型：使用闭源的ChatGPT进行样本挑选
  - 外部数据：使用额外的数据集



ALPAGASUS框架图

- **SelectIT 利用基座大模型固有的不确定性来评估和挑选指令数据，无需额外微调。**

- ALPAGASUS 通过 ChatGPT 评估指令-响应对的质量，筛选出高分样本用于训练。
  - 闭源模型，可扩展性较差

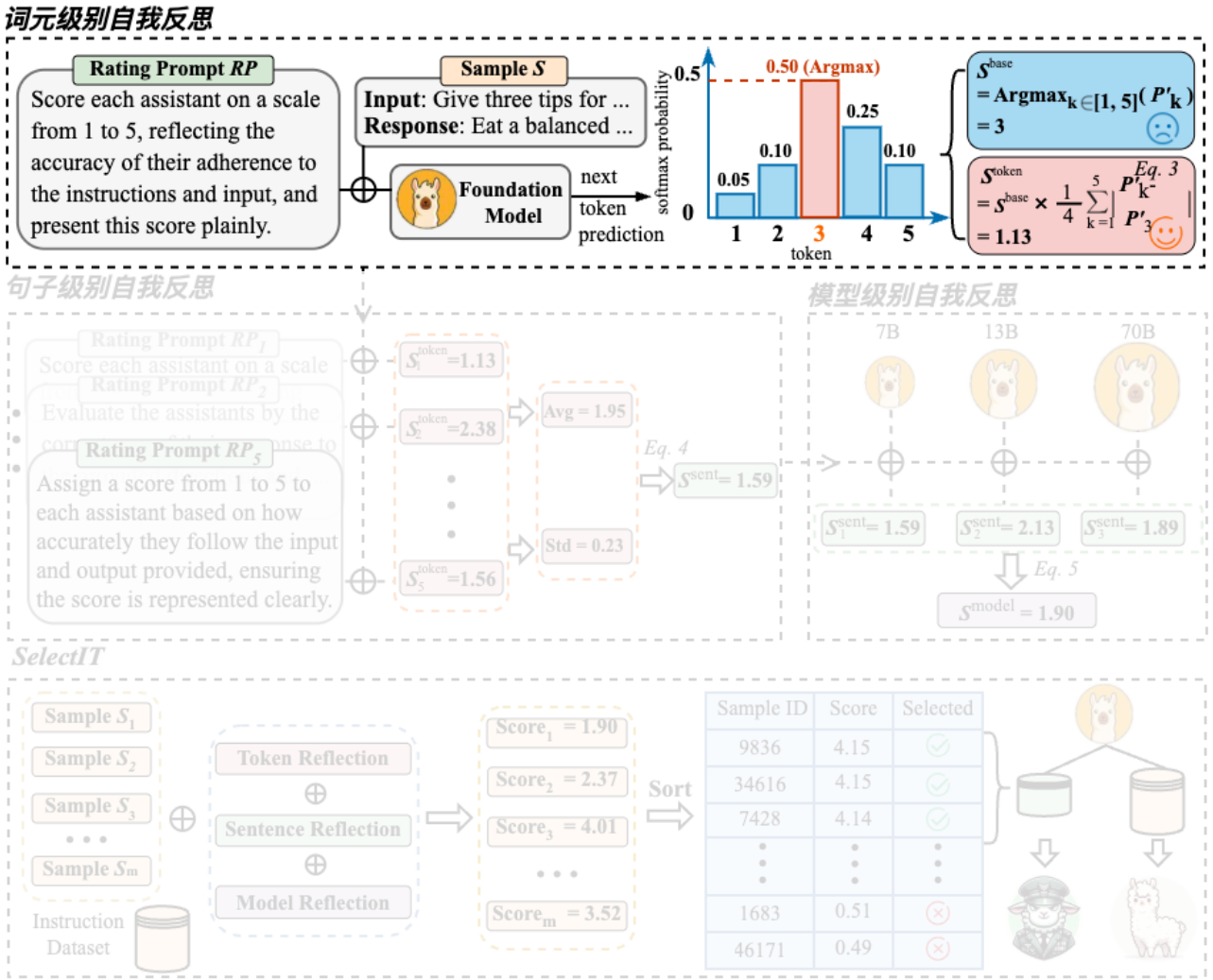
- **SelectIT框架图（词元级评分）**

▶ 词元级评分：计算基座模型输出每个评分词元的概率P_k。概率最高的评分词元被视为该样本的质量得分。

$$S^{base} = \arg\max_{k \in \{1,\dots,K\}} P'_k, P'_k = \left(\frac{P_k}{\sum_{j=1}^{K} P_j} \right)$$

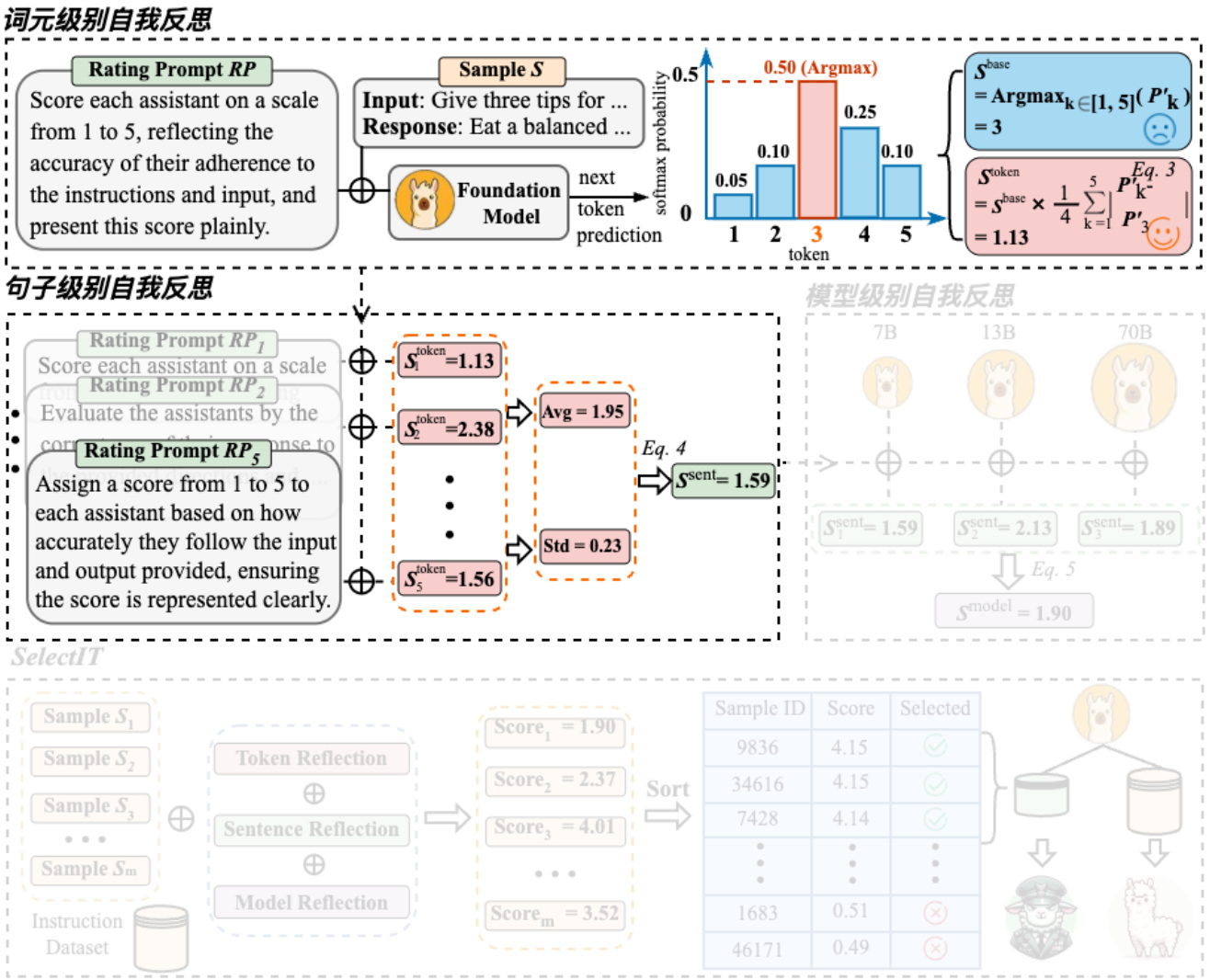
▶ 词元自反思机制：用于衡量模型对评分词元的确信程度，最终的词元级评分为：

$$S^{token} = S^{base} \times \underbrace{\frac{1}{K-1} \sum_{i=1}^{K} |P'_i - P'_{S^{base}}|}_{Uncertainty}$$



- **SelectIT框架图（句子级评分）**
  - 句子级评分：引入多样化提示和计算标准差，量化 LLMs 对评分提示变化的敏感性，从而更准确评估数据质量。K 表示评分提示的数量。

$$S^{sent} = \frac{\mathbf{Avg}\{S_i^{token}\}_{i=1}^{K}}{1 + \alpha \times \underbrace{\mathbf{Std}\{S_i^{token}\}_{i=1}^{K}}_{Uncertainty}}$$

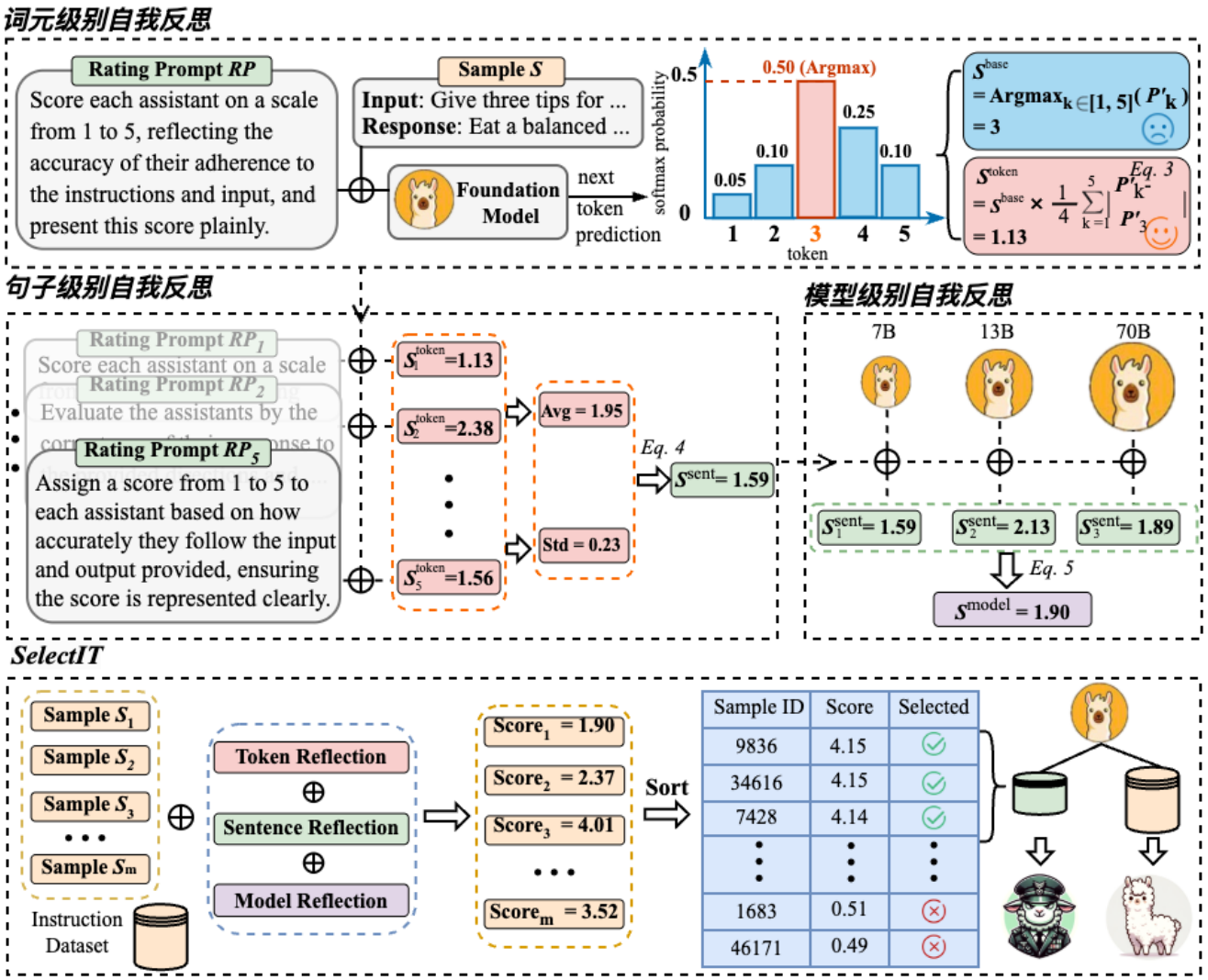


- **SelectIT框架图（模型级评分）**

▶ 模型级评分：利用不同基座模型的评估结果以减少模型级不确定性，根据模型参数量加权整合评估结果，提供样本质量的综合评分。

$$Quality \propto S^{model} = \sum_{i=1}^{N} \left(\frac{\theta_i}{\sum_{j=1}^{N} \theta_j} \times S_i^{sent} \right)$$

其中，N表示基座模型的数量。



Selective Alapca数据集

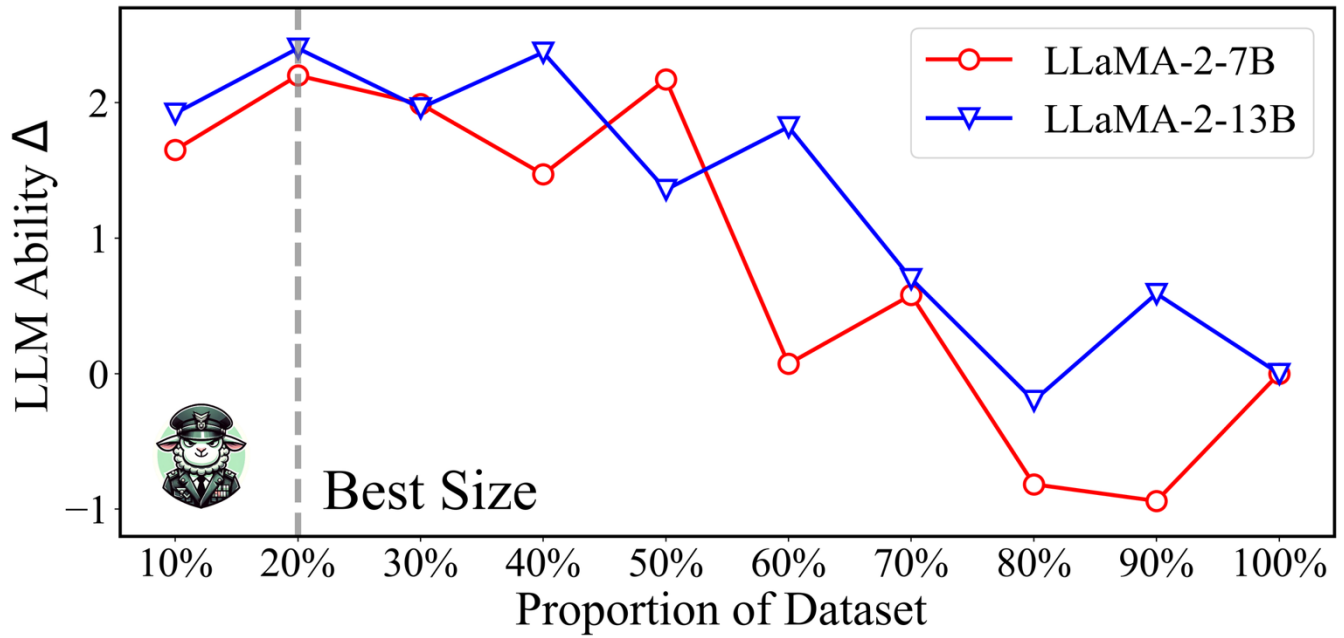


图: 不同 Alpaca 数据比例下大模型能力的比较

- 根据训练资源、训练时间和模型性能的权衡，我们从全量的Alpaca-GPT4选择 20% 的数据，称为**Selective Alpaca数据集**

SelectIT垂域任务实验结果

- 在知识、推理、多语言、代码四大领域的测试集上进行测试
- 对比五种方法：Alpaca-GPT4、LIMA、AlpaGasus、From Quantity to Quality、Instruction Mining

ID	System	External		MMLU	BBH	GSM	TydiQA	CodeX	AE	Overall	
		Model	Data							AVG	Δ (↑)
Base Model: LLaMA-2-7B				*Implemented Existing Method*							
1	Alpaca-GPT4			46.5	38.4	15.0	43.4	26.8	34.2	34.1	-
2	LIMA	✗	✓	45.4	37.5	14.3	45.1	24.6	33.1	33.3	-0.7
3	1 + AlpaGasus	✓	✗	45.9	39.0	14.5	46.4	27.5	35.4	34.8	+0.7
4	1 + Q2Q	✓	✗	46.9	39.4	15.3	46.7	28.2	35.7	35.4	+1.3
5	1 + Instruction Mining	✓	✓	47.0	39.6	16.5	47.1	28.6	34.4	35.5	+1.5
				Our Proposed Method (Individual)							
6	1 + Token-R	✗	✗	46.8	36.5	14.5	44.6	28.9	35.5	34.5	+0.4
7	1 + Sentence-R	✗	✗	46.9	38.1	16.1	**48.4**	26.9	35.3	35.3	+1.2
8	1 + Model-R	✗	✗	47.3	37.4	16.1	45.3	28.4	**35.8**	35.1	+1.0
				Our Proposed Method (All)							
9	SelectIT (6 + 7 + 8)	✗	✗	**47.4**	**40.6**	**16.8**	47.4	**29.4**	35.7	**36.2**	**+2.2**
Base Model: LLaMA-2-13B				*Implemented Existing Method*							
10	Alpaca-GPT4			**55.7**	46.6	30.5	48.1	40.8	46.5	44.7	-
11	LIMA	✗	✓	54.6	45.3	30.5	51.1	34.1	42.6	43.0	-1.7
12	10 + AlpaGasus	✓	✗	54.1	47.3	31.5	50.6	41.3	46.3	45.2	+0.5
13	10 + Q2Q	✓	✗	55.3	48.5	32.0	50.8	41.3	47.3	45.9	+1.2
14	10 + Instruction Mining	✓	✓	54.1	47.3	32.5	52.6	**43.3**	48.3	46.3	+1.6
				Our Proposed Method (Individual)							
15	10 + Token-R	✗	✗	55.3	47.3	30.5	51.3	39.8	46.2	45.1	+0.4
16	10 + Sentence-R	✗	✗	55.2	48.3	31.0	52.2	42.5	46.3	45.9	+1.2
17	10 + Model-R	✗	✗	55.1	47.5	31.5	52.3	40.2	46.1	45.5	+0.8
				Our Proposed Method (All)							
18	SelectIT (15 + 16 + 17)	✗	✗	**55.7**	**48.9**	**33.0**	**54.1**	42.2	**48.8**	**47.1**	**+2.4**

SelectIT显著提升了LLaMA-2模型的性能，在多个领域上优于其他方法

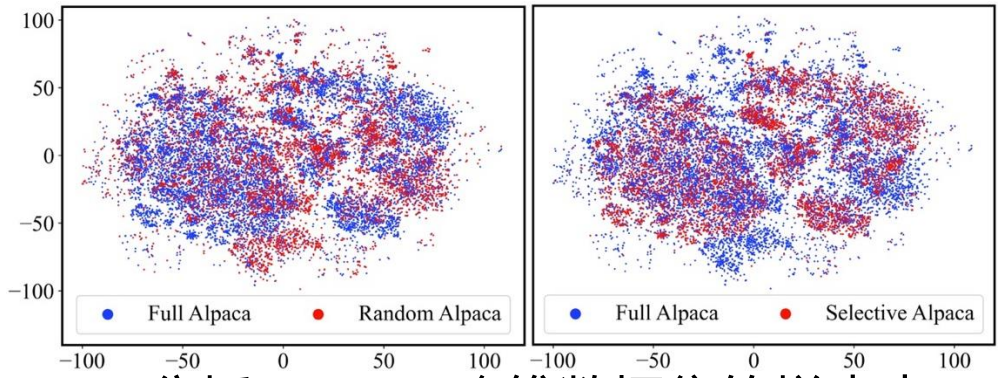
SelectIT垂域任务实验结果

- 对比Alpaca-GPT4数据集和我们的Selective Alpaca数据集
- 使用多种类型模型，包括Llama 2及3系列模型，以及Mistral模型

Base Model	Datasets	MMLU	BBH	GSM	Tydiqa	CodeX	AE	Overall	
								AVG	Δ (↑)
LLaMA-2-7B	Alpaca-GPT4	46.5	38.4	15.0	43.4	26.8	34.2	34.1	-
	Selective Alpaca	**47.4**	**40.6**	**16.8**	**47.4**	**29.4**	**35.7**	**36.2**	**+2.1**
LLaMA-2-13B	Alpaca-GPT4	**55.7**	46.6	30.5	47.1	38.8	46.5	44.2	-
	Selective Alpaca	55.3	**48.5**	**32.5**	**54.1**	**41.2**	**47.8**	**46.6**	**+2.4**
Mistral-7B	Alpaca-GPT4	52.5	51.7	33.5	**51.1**	54.7	43.1	47.8	-
	Selective Alpaca	**56.9**	**53.7**	**36.0**	49.3	**55.3**	**44.3**	**49.3**	**+1.5**
LLaMA-3-8B	Alpaca-GPT4	59.6	52.3	34.5	**43.1**	60.2	**48.2**	49.7	-
	Selective Alpaca	**61.2**	**55.0**	**37.5**	41.1	**65.4**	47.7	**51.3**	**+1.6**

在多个领域上，Selective Alpaca对每个模型均带来了显著提升

分析实验

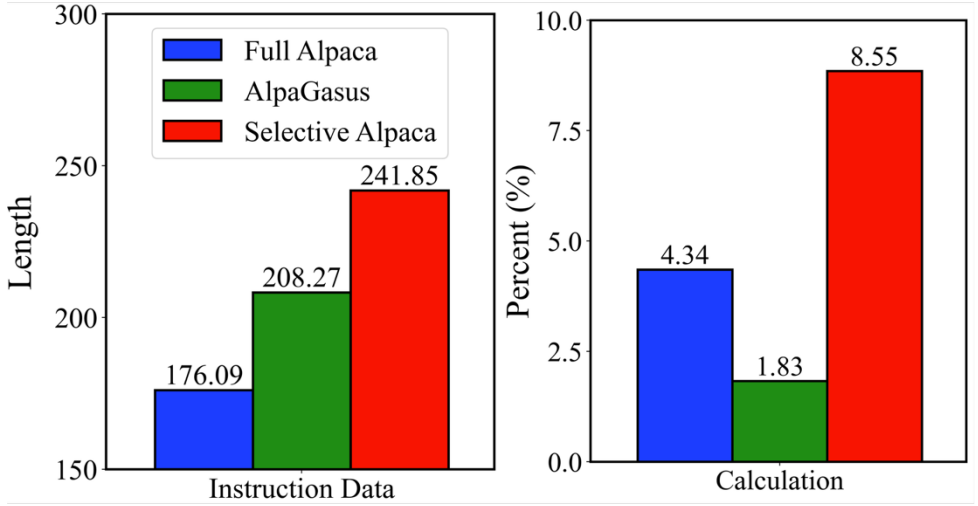


分析：T-SNE降维数据集的样本表示大多集中在中心，表明其主要包含高质量数据并有效剔除异常数据

Method	LLaMA-2		ALMA		Δ (↑)
	7B	13B	7B	13B	
Full Dataset	34.1	44.2	29.7	31.5	-
w/ Random (Full)	34.1	45.1	29.3	31.0	0.0
w/ Random (Unselected)	34.6	44.3	29.1	31.2	-0.4
w/ Length	35.5	47.1	30.1	31.8	+5.0
w/ SelectIT	**36.2**	**47.1**	**30.5**	**32.2**	**+6.5**

SelectIT为最优选择策略

分析：选择的数据中更长的指令和计算问题比例更高，为后续指令数据集构建提供指导

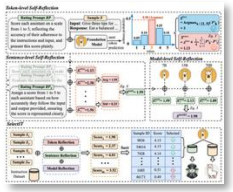


SelectIT 总结

- SelectIT是一种基于模型不确定性的指令数据筛选方法
  - 利用模型内部token概率分布 → 低成本数据质量评估，无需外部模型
  - 多提示语与多模型评分 → 增强数据选择稳定性
  - 选择计算类与长文本数据 → 数据更丰富，提升模型推理能力

报告总览

- 攻克 “数据属性与模型能力映射关系不清” 的挑战
  - 模型自身决定能力（黑盒）：指令数据筛选
  - 人类已有理论决定能力（白盒）：基于认知理论的能力分类



SelectIT：不确定性感知自反思引导的选择性指令微调

黑盒

CDT：多维度的数据驱动大语言模型能力框架

白盒

CDT: A Comprehensive Capability Framework for Large Language Models Across Cognition, Domain, and Task

Haosi Mo[1], Xinyu Ma[1], Xuebo Liu[1], Derek F. Wong[2], Yu Li[1], Jie Liu[1], Min Zhang[1]
[1]Harbin Institute of Technology, Shenzhen
[2]University of Macau

EMNLP 2025 Findings

背景：大模型能力定义

- **大模型能力评估的现状**
  - 当前主流评估方式多依赖任务特定的基准集
  - 现有的框架如 FLASK 、FAC²E 和INSTAG 等工作从不同角度探索了能力定义
- **存在问题**
  - **维度单一**

    多数方法仅关注任务完成性能，忽视模型能力的多维度本质
  - **定义模糊**

    缺乏统一、可解释的能力定义框架
  - **互补性不足**

    不同能力之间的协同效应尚未体现
- **CDT**
  - **横跨认知、领域与任务三个能力维度**
  - **能力组合与能力解耦**

Framework	Open Source Tagging Models	Multiple Dimensions	Capability Decomposition	Cognition Oriented	Domain Oriented	Task Oriented
FLASK	✗	✓	✗	✓	✓	✗
FAC²E	✗	✓	✓	✓	✗	✓
INSTAG	✓	✗	✗	✗	✓	✗
CDT (Ours)	✓	✓	✓	✓	✓	✓

CDT 框架概览

- CDT = Cognition（认知）+ Domain（领域）+ Task（任务）

Cognition

基于 **Cattell-Horn-Carroll (CHC) 认知理论**，定义模型核心认知能力

Task

系统化梳理指令中的任务类型

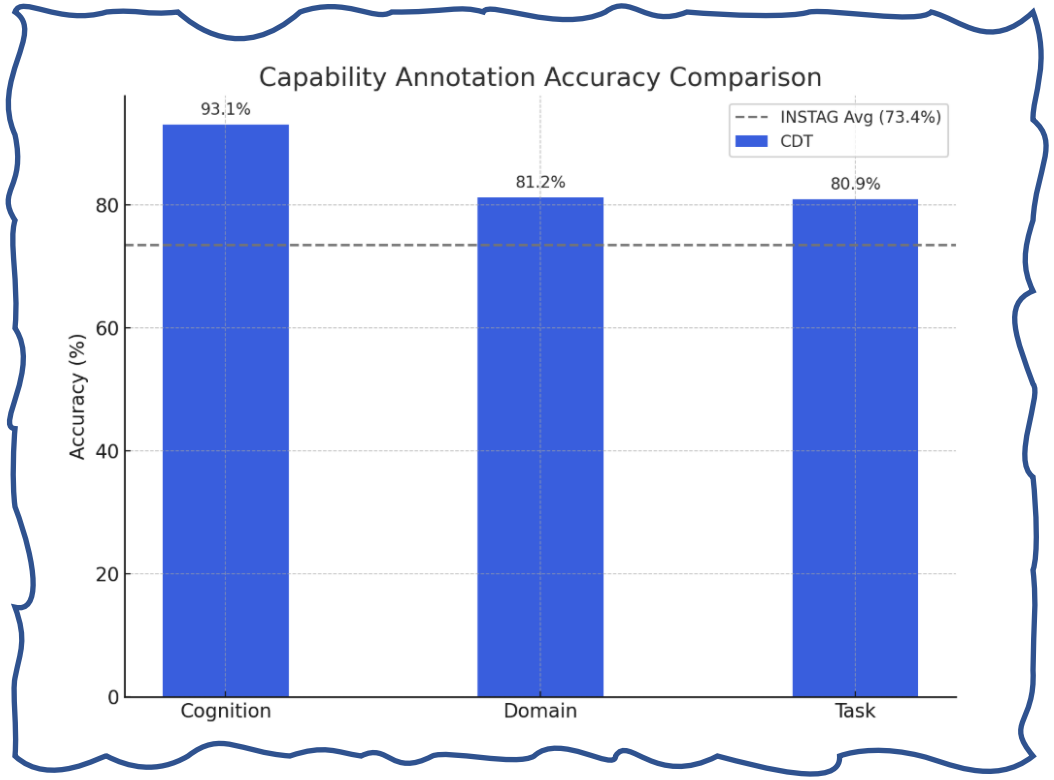
Pattern Recognition	Concept Abstraction	Hypothesis Generation	Linguistics	Literature	Multilingualism	Rewrite	Summarization	Classification
Problem Decomposition	Quantitative Reasoning	Logical Analysis	Tradition	Art	Sports	Extraction	Program Execution	Detection
Number Facility	Reading Decoding	Writing Ability	Mass Media	Music	Food	Brainstorming	Sentiment	Completion
Naming Facility	Associational Fluency	Expressional Fluency	Health	Biology	Earth Science	Generation	Bias and Fairness	Word Sense Disambiguation
Word Fluency	Sensitivity to Problems/ Alternative Solution Fluency		Astronomy	Chemistry	Physics	Multiple Choice QA	Closed QA	Open QA
Abstract Coding Concept	Originality/ Creativity	Ideational Fluency	Mathematics	Logic	Economics			
General Sequential Reasoning	Natural Language Inference		Law	Politics	Education			
			Sociology	Agriculture	Computer Science			
			Automation	Electronics	Engineering			
			Coding	Communication	Religion			
			Philosophy	Ethics	History			

Domain

覆盖大模型常见应用领域

CDT框架多场景应用

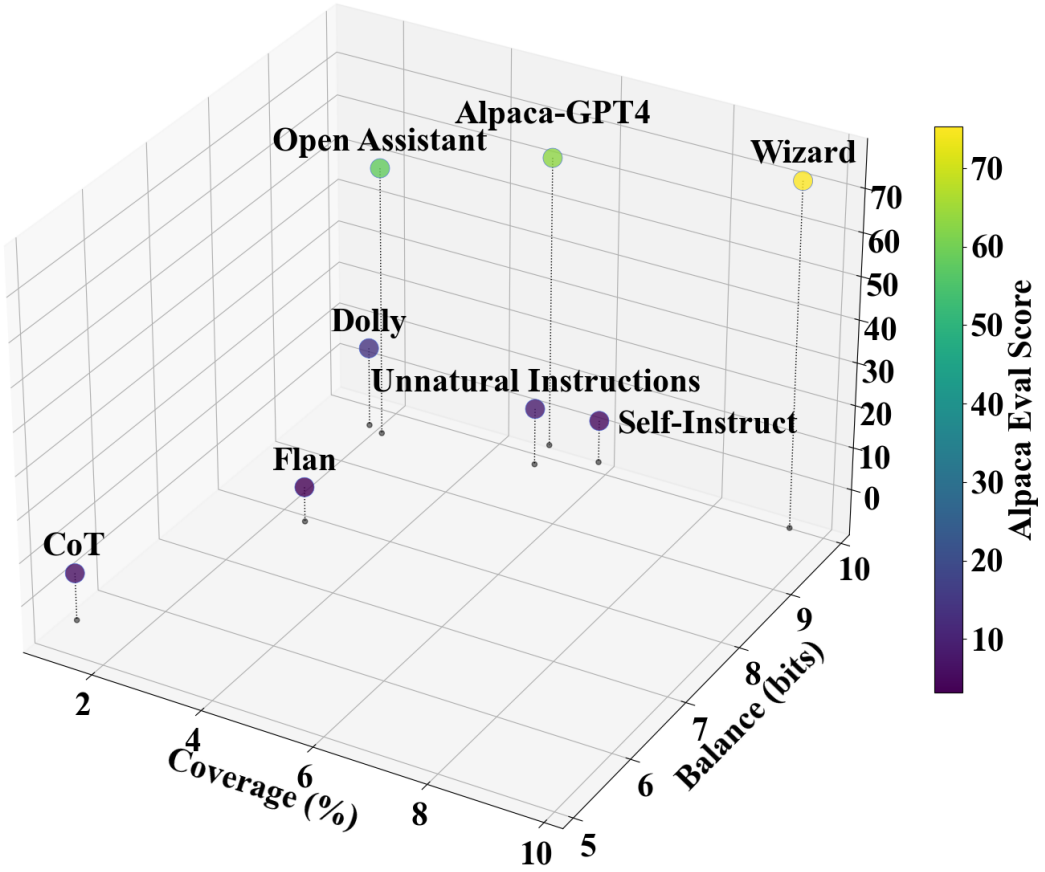
- 为实现框架应用，训练了维度特定的能力标注模型
- 使用 GPT-4o 对种子数据进行三个维度的细粒度能力标注
  - (认知维度≤2个标签，领域/任务维度1个标签)
- 基于 FLASK 数据集训练 Qwen2.5-7B-Base 模型作为标注器



CDT框架多场景应用

可用于指导数据合成与数据集构建

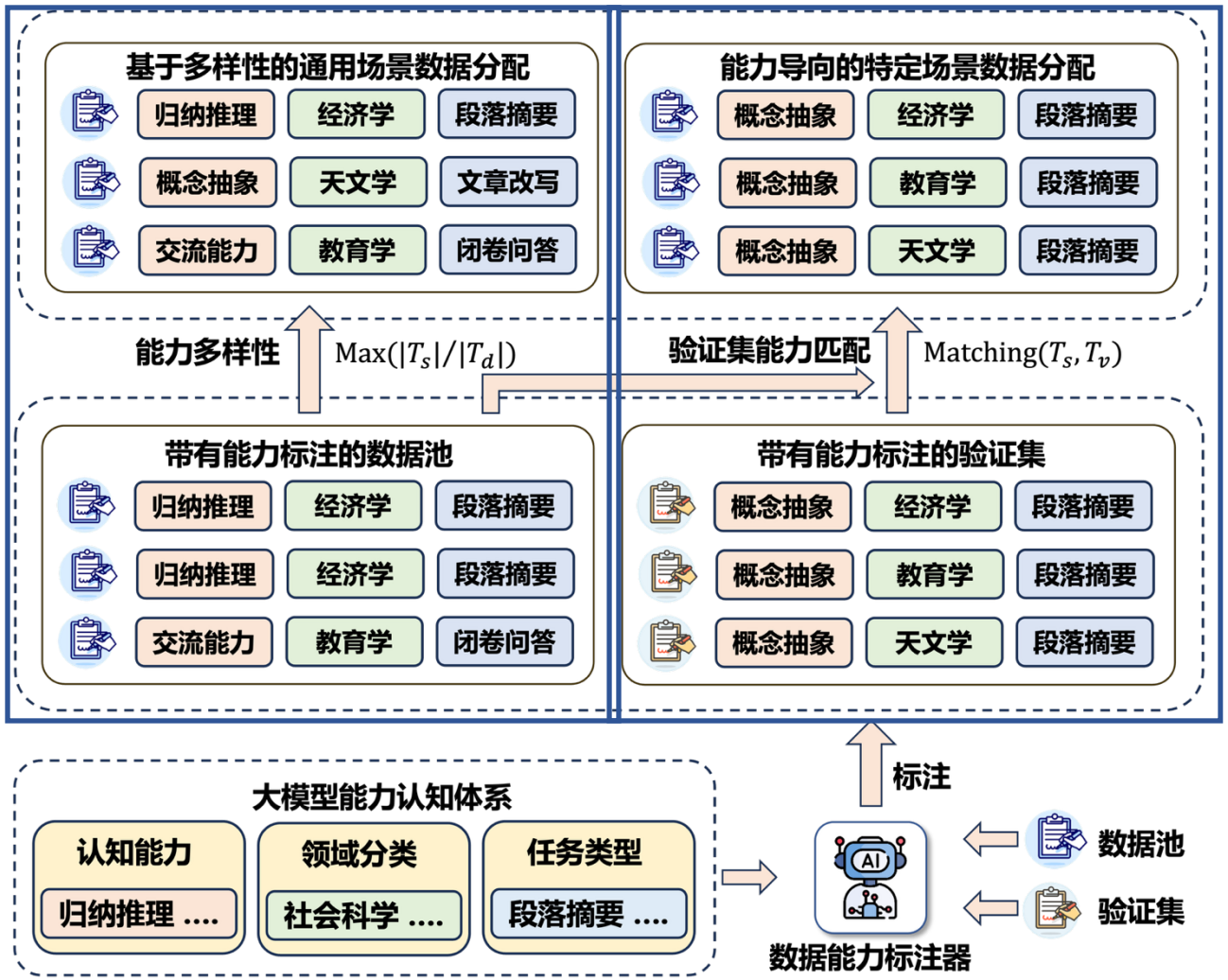
- 对开源数据集进行能力标注
- 分析数据集的能力特征
- Coverage
  - 数据集中的能力标注占整体能力组合数量的比例
  - 评估数据集的能力多样性
- Balance
  - 评估数据集能力分布的均衡程度



CDT框架多场景应用

可用于高质量指令筛选

- **基于多样性的通用场景数据分配**
  - 将三个维度能力组合定义为复合能力
  - 筛选出的训练数据集具有多样化的复合能力
  - 最大化训练数据相对数据池的复合能力覆盖度



- **能力导向的特定场景数据分配**
  - 针对特定测试需求，选择具备特定能力组合的数据
  - 根据测试任务对应的验证集的能力分布筛选训练数据

实验效果：基于多样性的通用场景

- **从数据池中筛选20%数据微调LLAMA2-7B模型**

Methods	ARC-C	MMLU	BBH	CEVAL	TYDIQA	AVG.
Baselines						
Base	43.5	45.2	**41.6**	31.9	47.8	42.0
All	44.5	45.9	39.6	35.6	**53.3**	43.8
Random	45.0	45.5	39.8	32.9	50.4	42.7
InsTag	44.8	45.8	39.3	33.2	51.9	43.0
Our Methods						
CDT_Cognition	45.3	45.3	38.2	36.6	51.9	43.5
CDT_Domain	45.9	46.1	38.5	34.3	52.2	43.4
CDT_Task	45.7	46.1	39.3	35.9	50.5	43.5
CDT	**46.1**	**46.3**	38.8	**36.9**	53.2	**44.3**

即便只考虑单一维度的多样性(Cognition/Domain/Task)，CDT方法仍能取得良好效果
三个维度的协同作用（CDT）取得最优表现

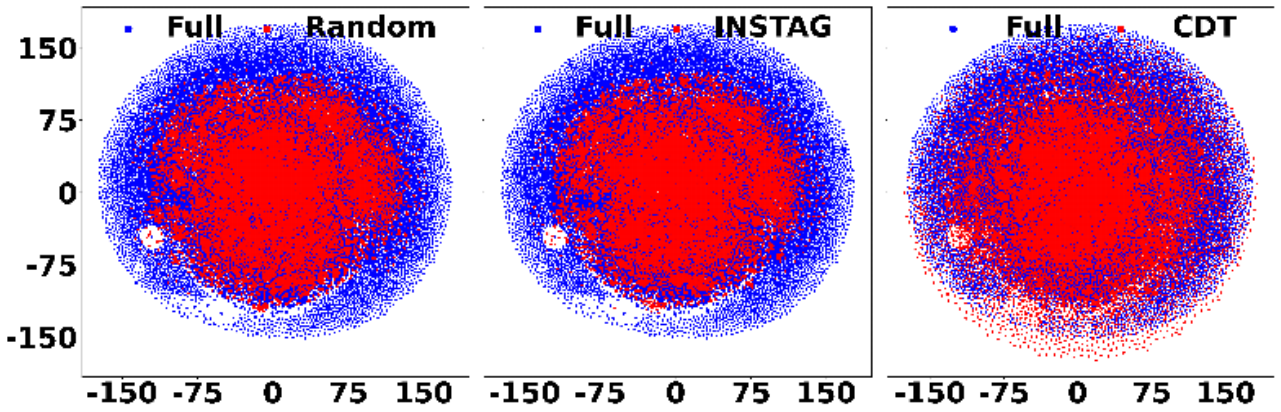
实验效果：基于多样性的通用场景

▶ 不同数据量的影响

Volume	Methods	ARC-C	BBH	MMLU	CEVAL	TYDIQA	AVG.
5%	INSTAG	44.3	38.3	44.4	32.1	49.4	41.7
	CDT	45.6	**39.4**	45.7	32.7	50.1	42.7
20%	INSTAG	44.8	39.3	45.8	33.2	51.9	43.0
	CDT	**46.1**	38.8	46.3	**36.9**	**53.2**	**44.3**
40%	INSTAG	45.2	**39.4**	46.3	33.7	51.5	43.2
	CDT	45.1	38.1	**46.7**	**36.9**	51.6	43.7

CDT 在不同数据量下均优于INSTAG

▶ t-SNE多样性分析



CDT 选择的数据多样性优于 Random 和 INSTAG，解释了性能优势

实验效果：能力导向的特定场景

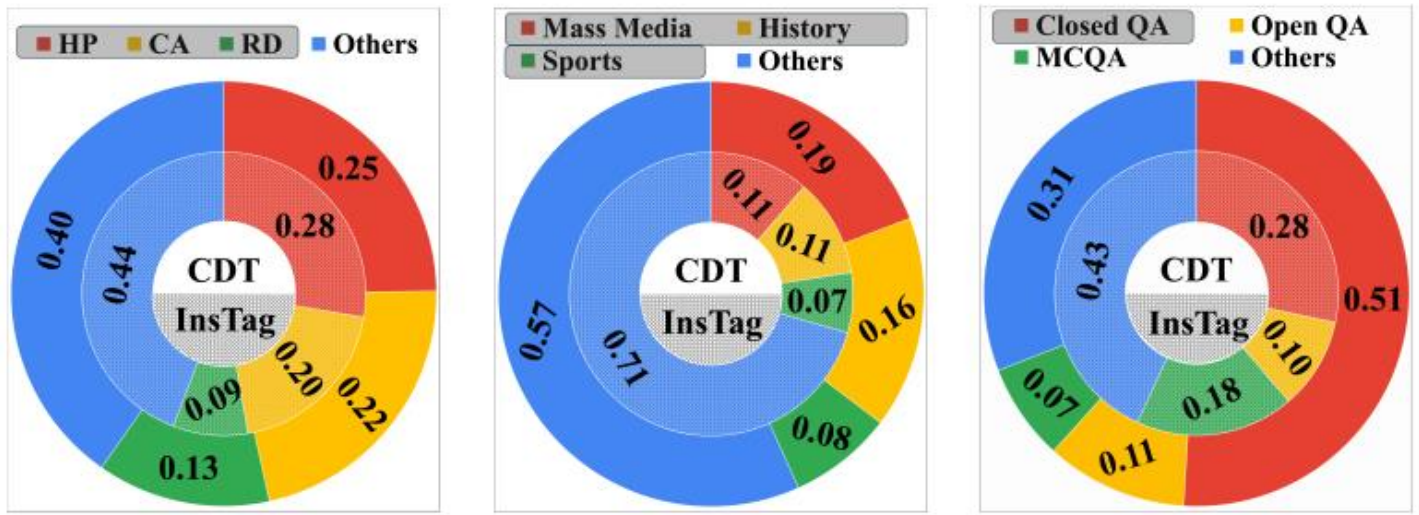
- **面向不同特定场景的测试集**：
  - DROP：阅读理解任务
  - GSM：数学任务
  - HISTORY：从MMLU选取历史领域任务

Methods	DROP		GSM	HISTORY	AVG.
	EM	F1	EM	Acc.	
Base	0.0	1.3	14.5	51.0	16.7
All	49.0	**58.3**	21.0	51.3	44.9
Random	46.7	55.8	19.0	51.2	43.2
InsTag	47.9	57.2	19.0	52.4	44.1
CDT	**49.3**	**58.3**	**21.5**	**52.5**	**45.4**

CDT 在三个特定能力测试中均表现最佳，显著优于所有基线方法

实验效果：能力导向的特定场景

- 以DROP任务为例，筛选出数据的能力分布对比



(a) Cognition (b) Domain (c) Task

CDT**筛选出的数据中，对应测试任务所需能力占比最高**
CDT能准确识别并优化对应测试任务所需的能力

CDT 总结

- CDT 是一种大模型综合能力框架
  - 从认知、领域、任务三维分解能力 → 实现全面、系统的**能力定义**
  - 认知维度融入 CHC 认知理论 → 具有坚实的**理论基础**
  - 为各维度提供标注模型 → 支持细粒度的**能力分析与应用**
- 多模态火花🔥
  - 认知维度跨模态扩展，**定义跨模态多维能力**，构建多模态能力框架
  - 构建多模态的数据-能力图谱，**指导多模态数据的采集与合成**
  - 基于能力感知，规划**从单模态到跨模态组合能力的课程学习路径**，优化训练效率

“数据-模型”能力对齐总览

- 2种理清数据与模型能力映射关系的方法
  - 模型自身决定能力
    - SelectIT：通过利用大语言模型自身的标记级、句子级和模型级不确定性进行自反思，高效选择高质量指令调优数据，提升模型性能。
  - 人类已有理论决定能力
    - CDT：提出认知 - 领域 - 任务框架，构建三维评估体系
      - 高效数据集多样性和覆盖度评估方法，有效保证数据集/数据合成质量
      - 多样性驱动和能力导向的数据选择方法，有效提升模型在通用及特定场景下的性能。

01 大模型与数据合成背景

02 基础：通用与垂域数据合成

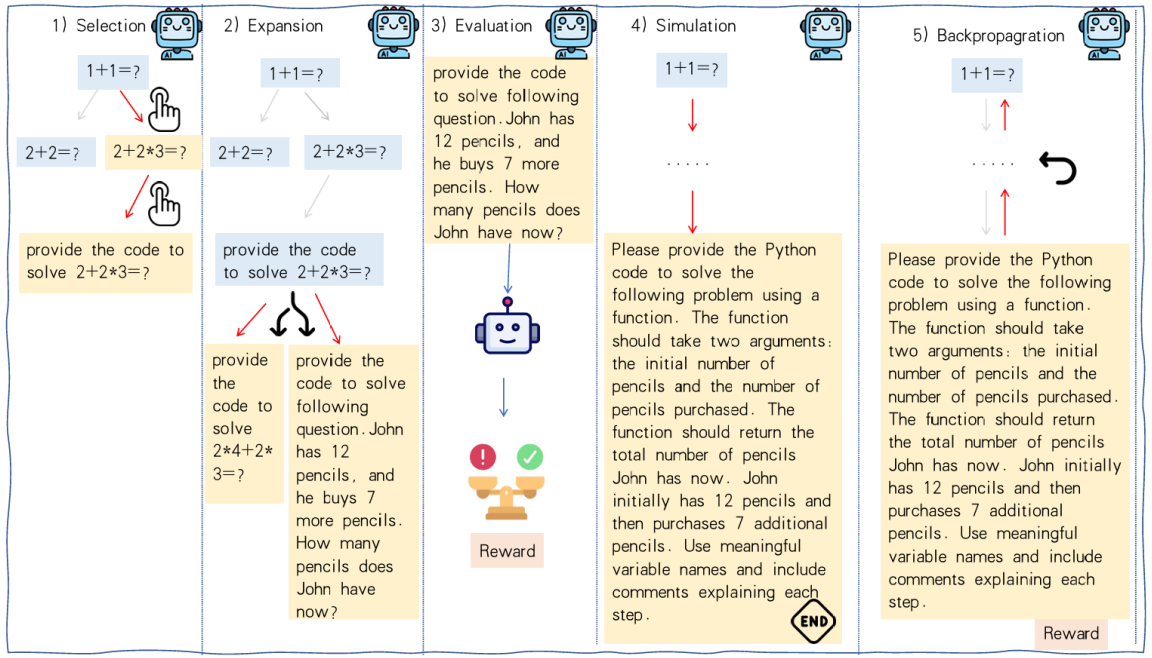
03 核心：高效数据学习与利用

04 进阶：“数据-模型”能力对齐

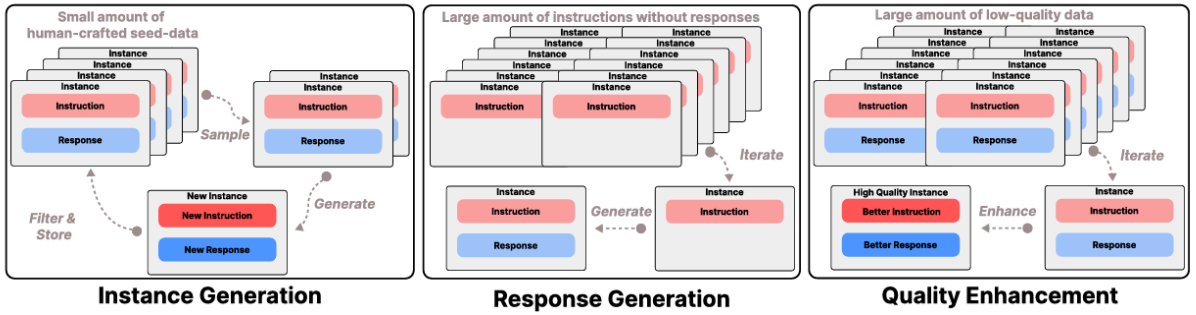
05 领域瓶颈与未来展望

未来展望

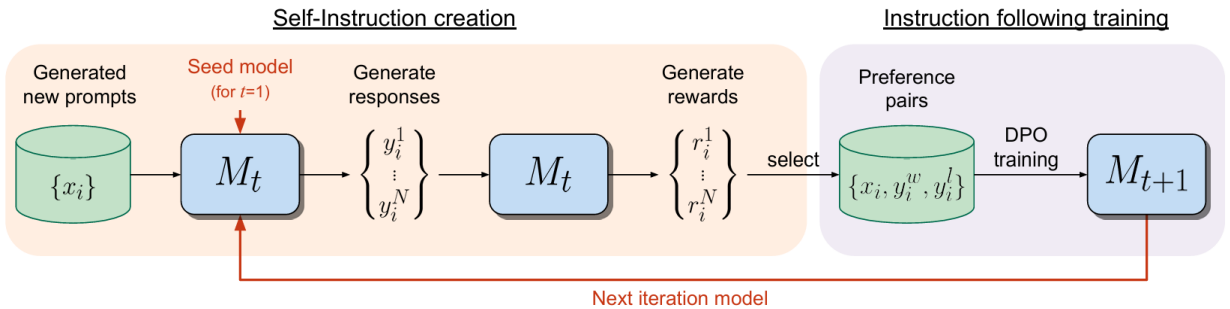
- **具备数据合成能力的通用大模型**[1]
  - 数据合成与能力评估
  - 数据合成长思维链推理与自我评估



- **完善合成数据质量评估体系**[2]：
  - 构建轻量化评估模型，实现生成-评估-优化的端到端流水线



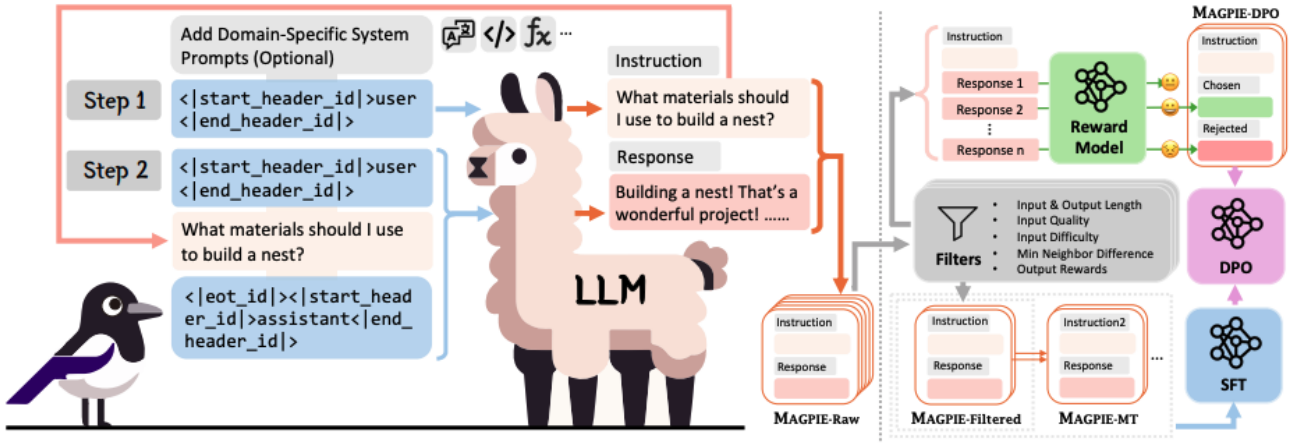
- **Self-Rewarding**[3]：
  - 自监督迭代提升合成指令数据质量



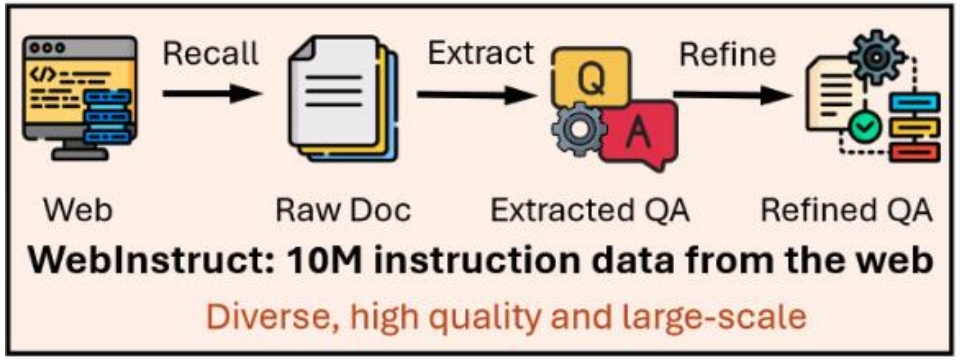
[1] Optimizing Instruction Synthesis: Effective Exploration of Evolutionary Space with Tree Search. EMNLP 2024.
[2] Evaluating Language Models as Synthetic Data Generators. Arxiv 2024.
[3] Self-Rewarding Language Models. ICML 2024.

未来展望

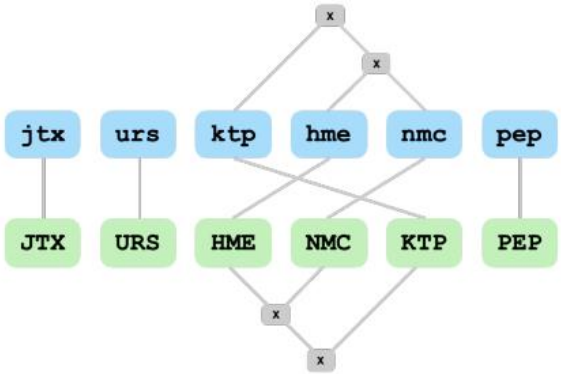
- **预训练数据合成**
  - **数据集蒸馏**[1]：
  - 通过输入该模型的提示词模板，即可引导模型生成训练数据



- **Real-Query 挖掘**[2]：
  - 利用真实数据进一步合成



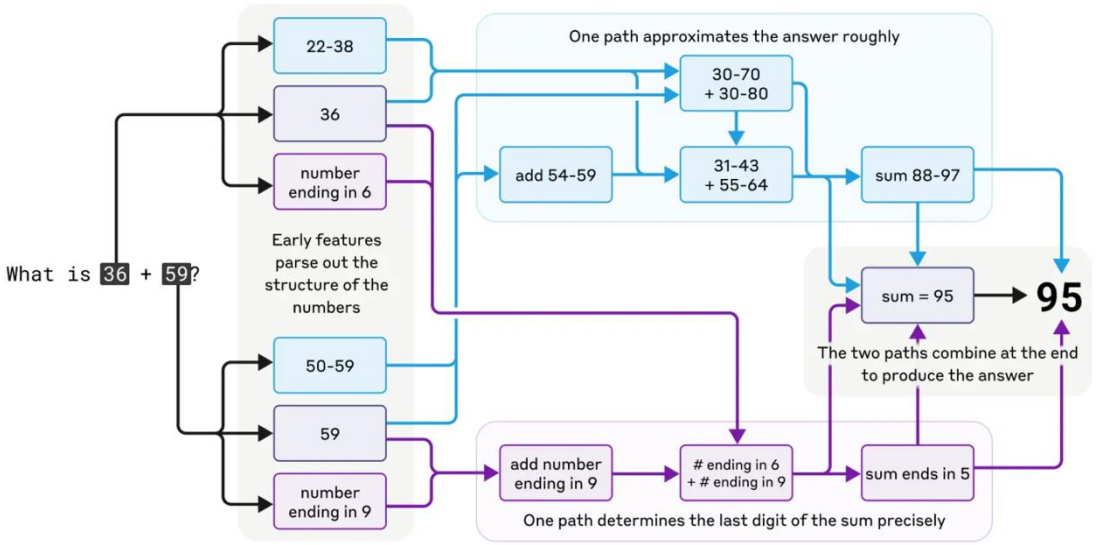
- **合成人类无法理解的抽象数据**[3][4]：
  - 减少毒性、偏见、版权和隐私



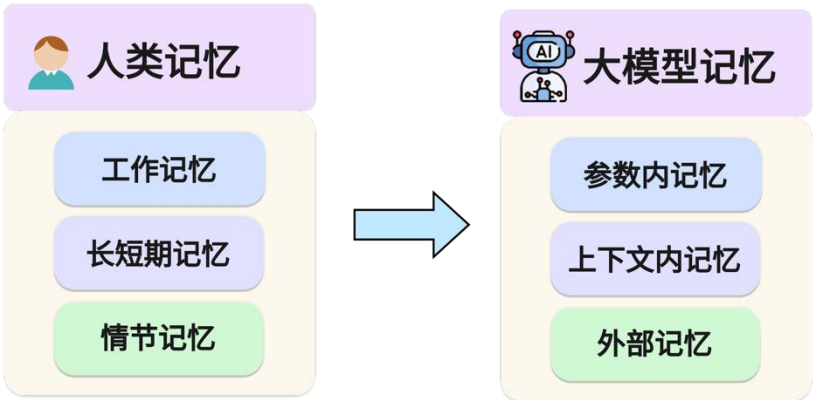
[1] Magpie: Alignment Data Synthesis from Scratch by Prompting Aligned LLMs with Nothing. ICLR 2025.
[2] MAmmoTH2: Scaling Instructions from the Web. NeurIPS 2024.
[3] Synthetic Pre-Training Tasks for Neural Machine Translation. ACL 2023.
[4] Pre-training with Synthetic Data Helps Offline Reinforcement Learning. ICLR 2024.

未来展望

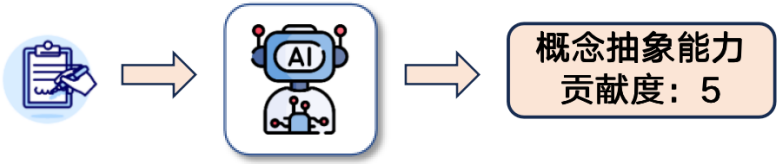
- **面向大模型的认知体系构建：**
  - 现有的模型认知研究几乎都从人类认知科学迁移[1][2]
  - 然而大模型处理信息的方式和思考的方式与人类有显著不同[3]



- **基于大模型特性进行认知体系搭建：**



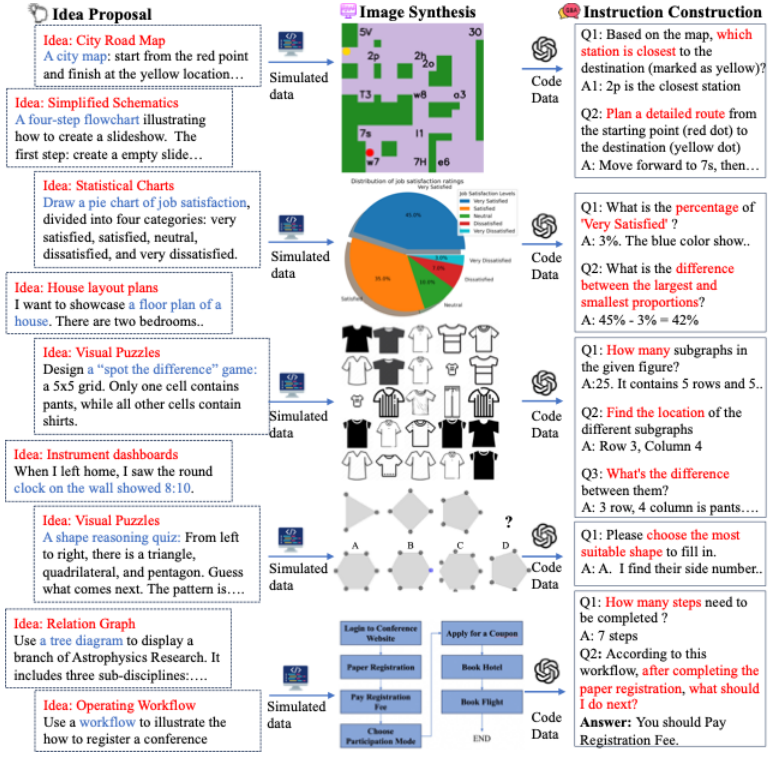
- **动态“数据-能力”标注体系：**
  - **不仅对数据标注能力，还要量化数据对能力的贡献度**
  - **不同场景动态更新认知体系与标注框架**



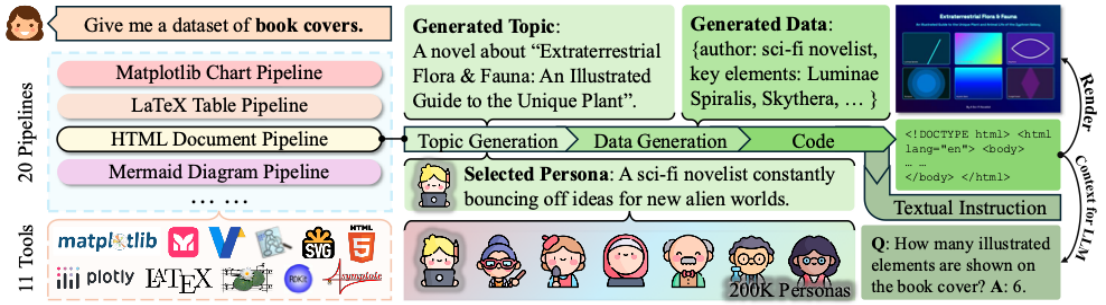
[1] CogBench: a large language model walks into a psychology lab. ICML 2024.
[2] M3GIA: A Cognition Inspired Multilingual and Multimodal General Intelligence Ability Benchmark. Arxiv 2024.
[3] Tracing the thoughts of a large language model. Anthropic 2025.

未来展望

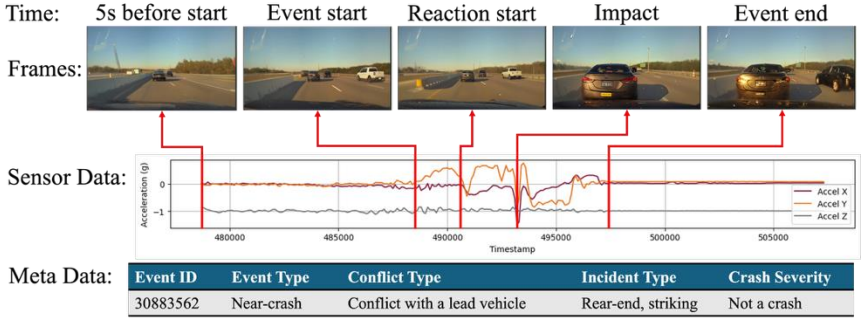
- 多模态指令微调数据合成
  - 以语言为中心的数据合成[1]：



- 多模态大模型的数据合成[2]：
  - 通过大模型代码（如LaTex）作图能力生成多模态高质量数据



- 物理世界/具身智能的数据合成[3]：
  - 通过图像生成技术模型现实场景



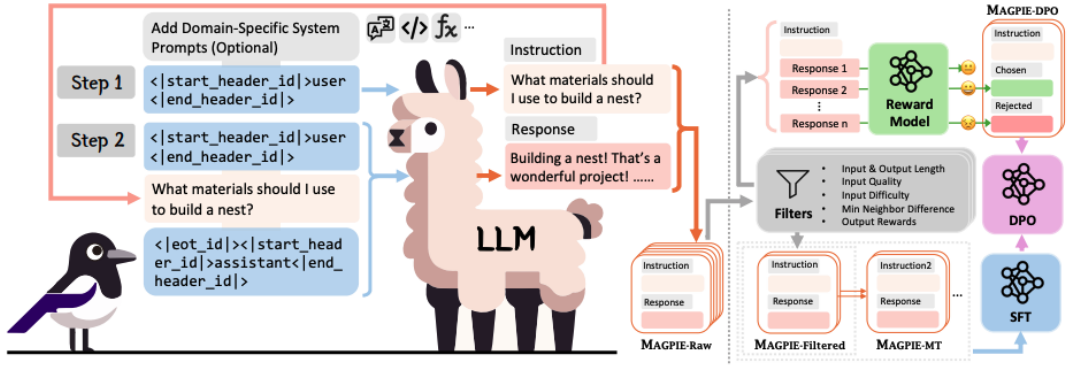
Event ID	Event Type	Conflict Type	Incident Type	Crash Severity
30883562	Near-crash	Conflict with a lead vehicle	Rear-end, striking	Not a crash

[1] Multimodal self-instruct: Synthetic abstract image and visual reasoning instruction using language model. EMNLP 2024.
[2] Scaling Text-Rich Image Understanding via Code-Guided Synthetic Multimodal Data Generation. ACL 2025
[3] SynSHRP2: A Synthetic Multimodal Benchmark for Driving Safety-critical Events Derived from Real-world Driving Data. Arxiv 2025.

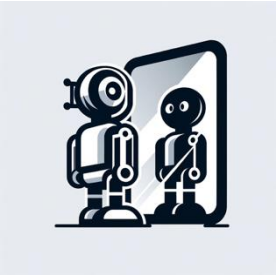
未来展望

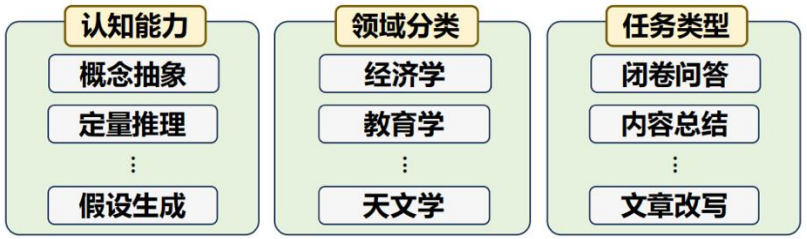
- 数据合成和AGI

▶ 1 训练具备数据合成能力的通用大模型

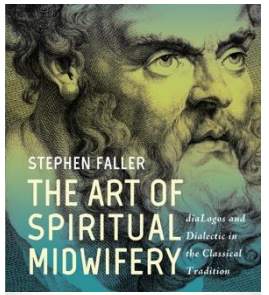


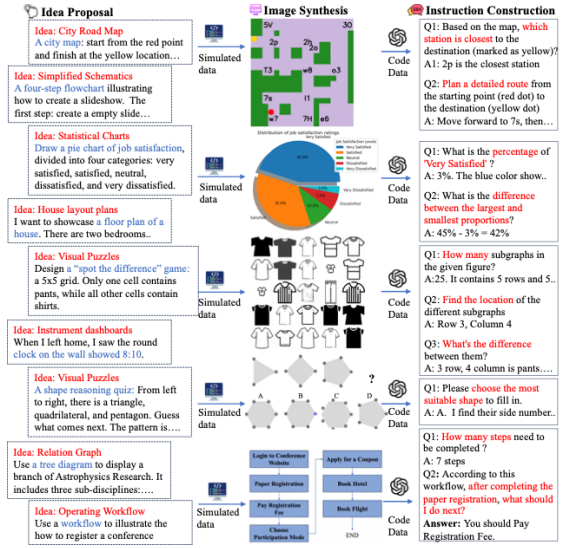
▶ 2 模型根据认知体系自我感知缺失能力



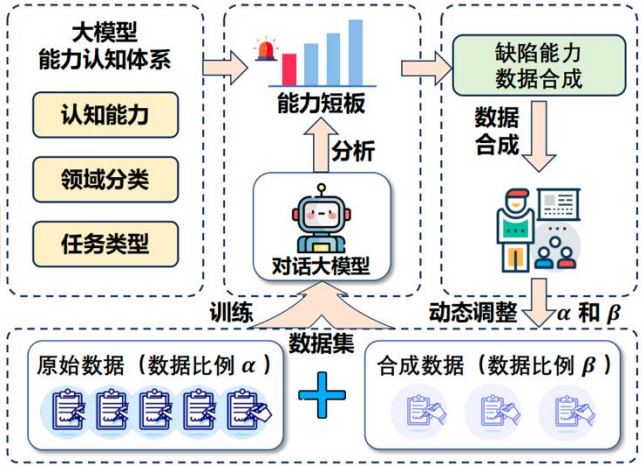


▶ 3 合成面向现实世界的高质量多模态数据





▶ 4 迭代优化模型自身能力



AI开发体系的最后一块拼图

$$\hat{\theta} = \arg\min_{\theta} \left\{ \sum_{(x,y)\in\mathcal{D}} L(f(x;\theta), y) \right\}$$

算力

数据

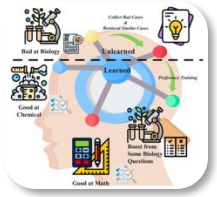
算法（模型）

Ascend 910
华为昇腾910

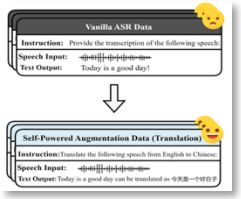
NVIDIA

报告总览

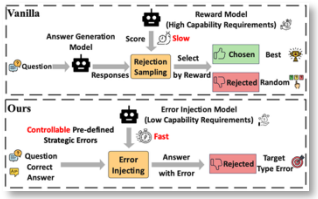
合成高质量数据



APT：弱点数据生成与迭代式能力对齐



Self-Powered LSM：面向语音-文本大模型模态扩展的自驱动数据合成

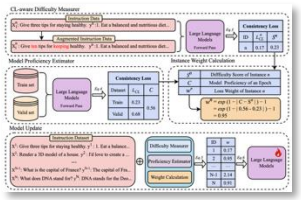


SeaPO：策略性错误放大的偏好数据合成

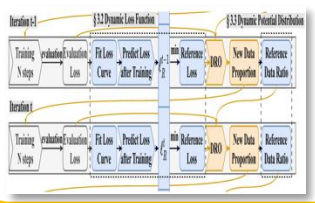
LongMT: CoT偏好数据广域搜索与细粒度策略合成

AQuilt：逻辑与反思增强的指令对齐数据合成

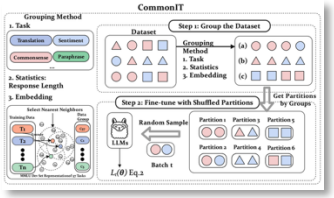
更好地利用数据



CCL：数据驱动的课程一致性学习

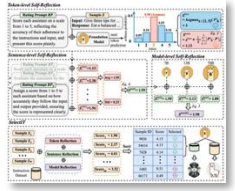


DRPruning：基于数据分布鲁棒的模型剪枝

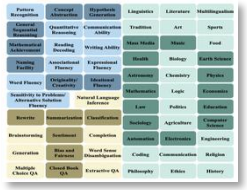


CommonIT：基于数据划分的共性感知指令微调方法

理解数据与模型能力之间的关系



SelectIT：不确定性感知的选择性指令数据微调



CDT：多维度的数据驱动大语言模型能力框架

工作汇总（刘学博为这10篇工作的唯一通讯作者）

[1] Xiaopeng Ke, Hexuan Deng, **Xuebo Liu**, Jun Rao, Lian Lian, Dong Jin, Shengjun Cheng, Jun Yu, Min Zhang. AQuilt: Weaving Logic and Self-Inspection into Low-Cost, High-Relevance Data Synthesis for Specialist LLMs. EMNLP 2025.

[2] Tengfei Yu, **Xuebo Liu**, Zhiyi Hou, Liang Ding, Dacheng Tao, Min Zhang. Self-Powered LLM Modality Expansion for Large Speech-Text Models. EMNLP 2024.

[3] Jun Rao, Zepeng Lin, **Xuebo Liu**, Xiaopeng Ke, Lian Lian, Dong Jin, Shengjun Cheng, Jun Yu, Min Zhang. APT: Improving Specialist LLM Performance with Weakness Case Acquisition and Iterative Preference Training. ACL 2025 Findings.

[4] Jun Rao, Yunjie Liao, **Xuebo Liu**, Zepeng Lin, Lian Lian, Dong Jin, Shengjun Cheng, Jun Yu, Min Zhang. SeaPO: Strategic Error Amplification for Robust Preference Optimization of Large Language Models. EMNLP 2025 Findings.

[5] Yutong Wang, Zepeng Lin, Yunjie Liao, **Xuebo Liu**, Min Zhang. LongMT: Towards Humen-Like Document-Level Translation Agent. Ongoing Work.

[6] Jun Rao, **Xuebo Liu**, Lian Lian, Shengjun Cheng, Yunjie Liao, Min Zhang. CommonIT: Commonality-Aware Instruction Tuning for Large Language Models via Data Partitions. EMNLP 2024.

[7] Liangxin Liu, **Xuebo Liu**, Lian Lian, Dong Jin, Shengjun Cheng, Jun Rao, Tengfei Yu, Hexuan Deng, Min Zhang. Curriculum Consistency Learning for Conditional Sentence Generation. EMNLP 2024.

[8] Hexuan Deng, Wenxiang Jiao, **Xuebo Liu**, Min Zhang, Zhaopeng Tu. DRPruning: Efficient Large Language Model Pruning through Distributionally Robust Optimization. ACL 2025.

[9] Liangxin Liu, **Xuebo Liu**, Derek F. Wong, Dongfang Li, Ziyi Wang, Baotian Hu, Min Zhang. SelectIT: Selective Instruction Tuning for LLMs via Uncertainty-Aware Self-Reflection. NeurIPS 2024.

[10] Haosi Mo, Xinyu Ma, **Xuebo Liu**, Derek F. Wong, Yu Li, Jie Liu, Min Zhang. CDT: A Comprehensive Capability Framework for Large Language Models Across Cognition, Domain, and Task. EMNLP 2025 Findings.