

如何在大模型时代找到 科研切入点？

姓名：张绍磊

导师：冯洋 研究员

中国科学院计算技术研究所



大模型时代前的研究范式

■ 自然语言处理任务

- 判别式：文本分类、命名实体识别、信息抽取、词性标注 ...
- 生成式：**机器翻译**、问答、生成、摘要、阅读理解 ...

■ 每个任务的解决方案涉及

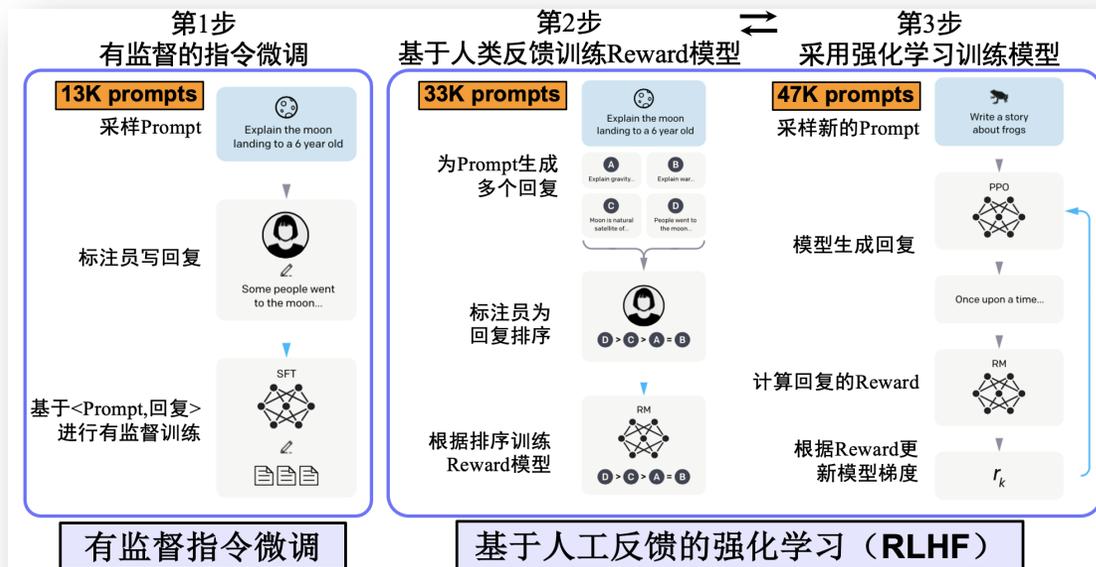
- 模型架构
 - 符号主义、连接主义（统计模型、神经网络）
- 训练数据
 - 高资源、低资源、少样本、零样本
- 学习方法
 - 监督学习、半监督学习、自监督学习、无监督学习

1. 选择一个任务
2. 确定模型架构
3. 确定训练数据
4. 选择学习方法

大模型时代的研究范式

■ 现有范式：基座模型 + 微调

- 预训练大语言模型作为基座模型
- 针对通用领域或垂直领域的微调
 - 全量参数微调
 - 部分参数微调或者添加参数微调
 - 参数不变，in-context learning
- 与人类对齐



如何在新的研究范式中找到切入点?

如何在新的研究范式中找到切入点?

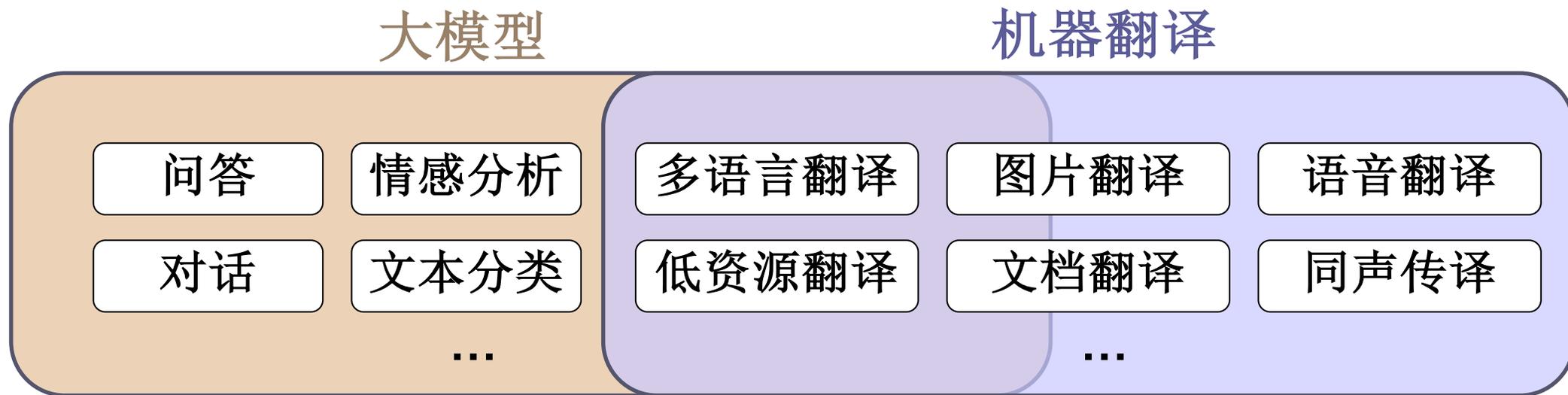
- 大语言模型 \leftrightarrow 机器翻译
 - 大模型**尚未能**完成的任务
 - 大模型**已经能**完成的任务



如何在新的研究范式中找到切入点?

■ 大语言模型 \leftrightarrow 机器翻译

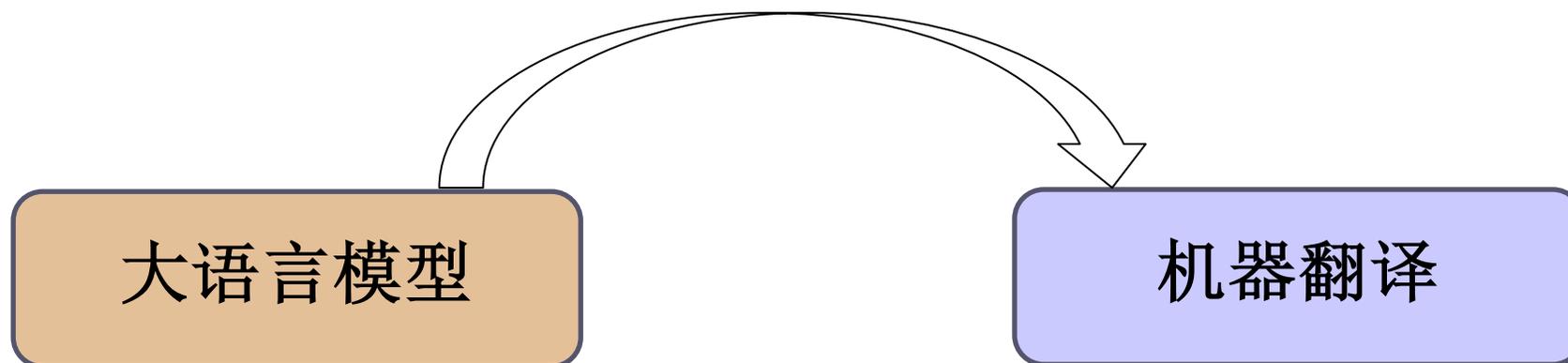
- 大模型**尚未能**完成的任务：设计新的架构将大模型适配特定任务
- 大模型**已经能**完成的任务：空间更小、挑战更大



如何在新的研究范式中找到切入点?

- 大模型已经能高质量完成的任务

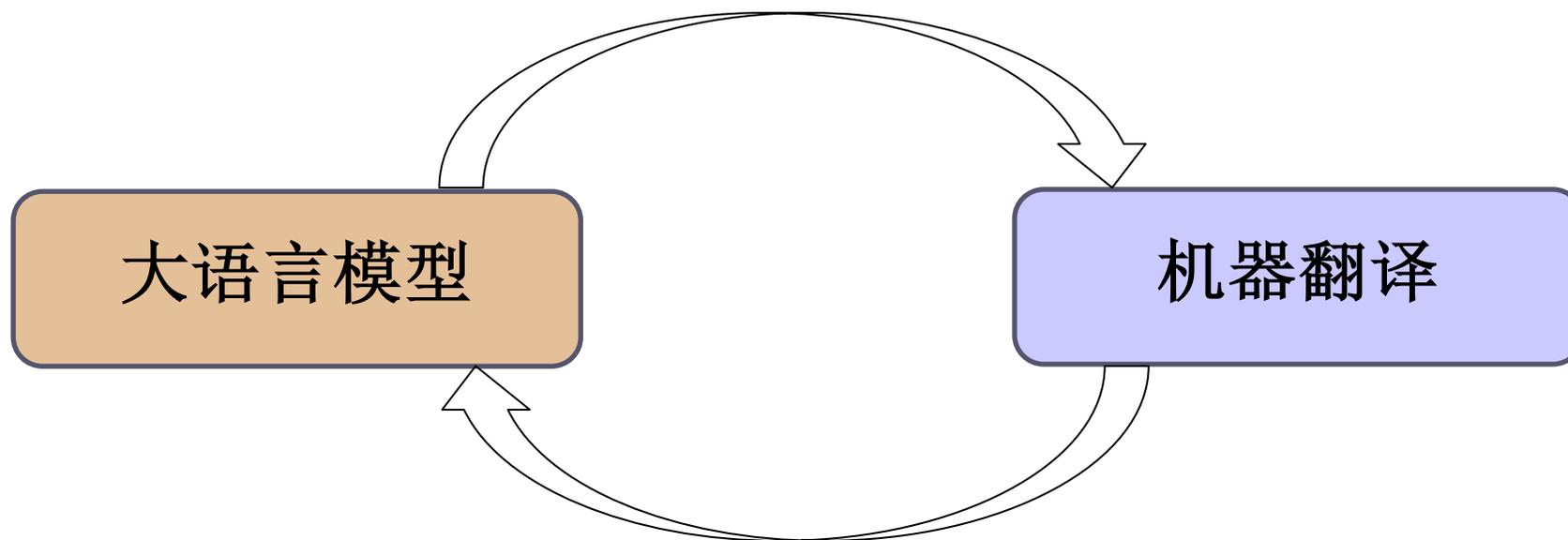
利用大模型完成翻译任务



如何在新的研究范式中找到切入点?

- 大模型已经能高质量完成的任务

利用大模型完成翻译任务



利用机器翻译增强大模型的语言能力

Llama等大模型英语强、其他语言弱

百聆：利用交互式翻译增强LLM的跨语言对齐

- 英语为主的大模型 → 增强其他语言生成/指令遵循能力



百聆：利用交互式翻译增强LLM的跨语言对齐

- 英语为主的大模型 → 增强其他语言生成/指令遵循能力

成本高、效率低、持续预训练效果无法保证

收集大规模多语言语料，重新预训练+微调

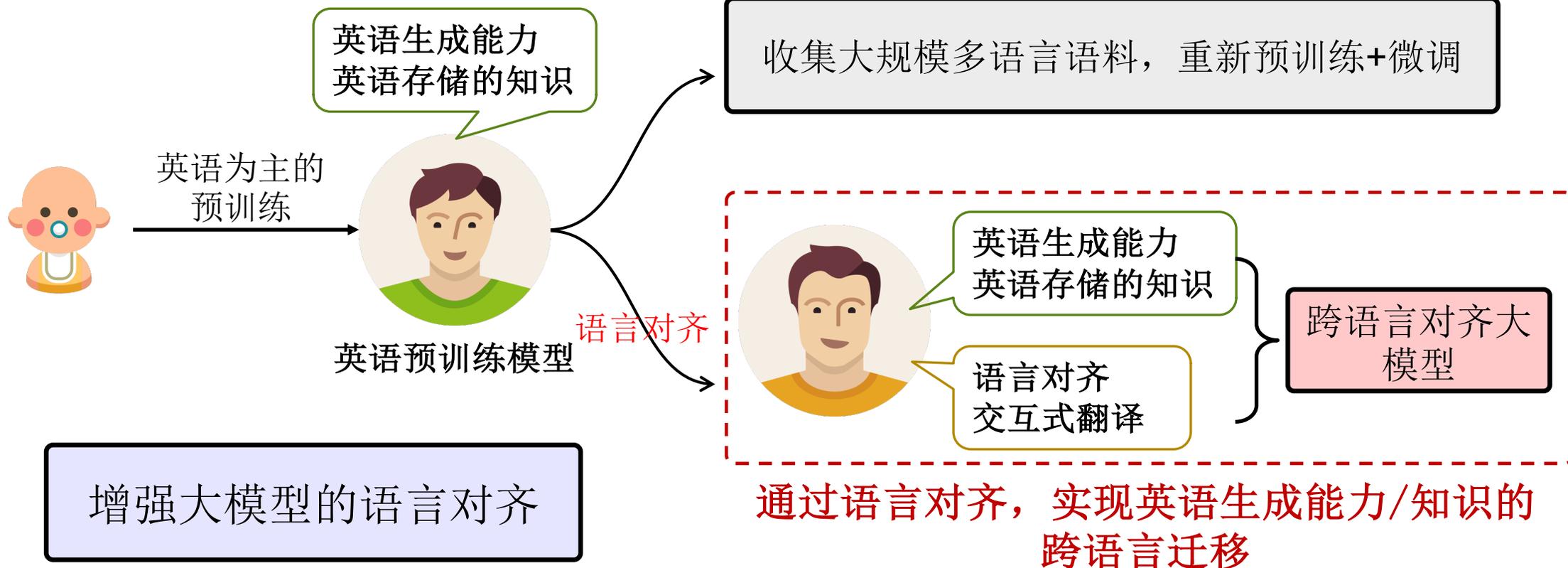


百聆：利用交互式翻译增强LLM的跨语言对齐

- 英语为主的大模型 → 增强其他语言生成/指令遵循能力

成本高、效率低、持续预训练效果无法保证

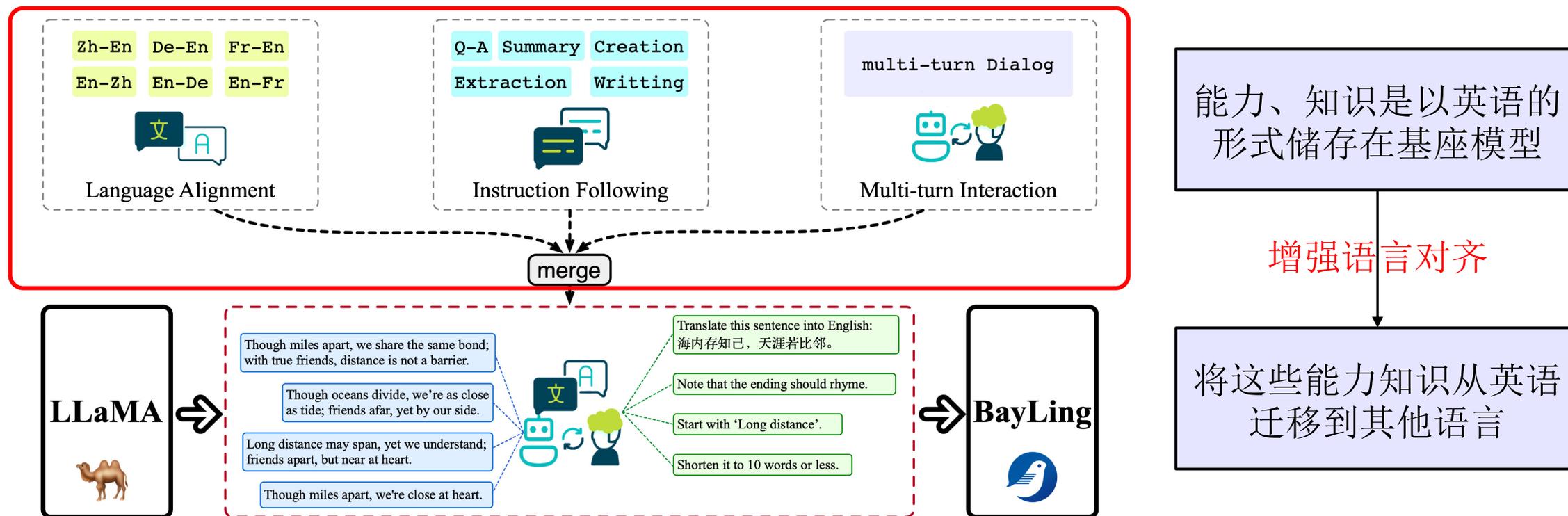
收集大规模多语言语料，重新预训练+微调



百聆：利用交互式翻译增强LLM的跨语言对齐

■ 通过交互式机器翻译同时提升语言生成和指令遵循能力

- **语言对齐**：完成语言生成能力从英语到其他语言的迁移。
- **复合任务**：同时提升多语言、指令理解、多轮交互等能力。



百聆：利用交互式翻译增强LLM的跨语言对齐

■ 通过交互式机器翻译同时提升语言生成和指令遵循能力

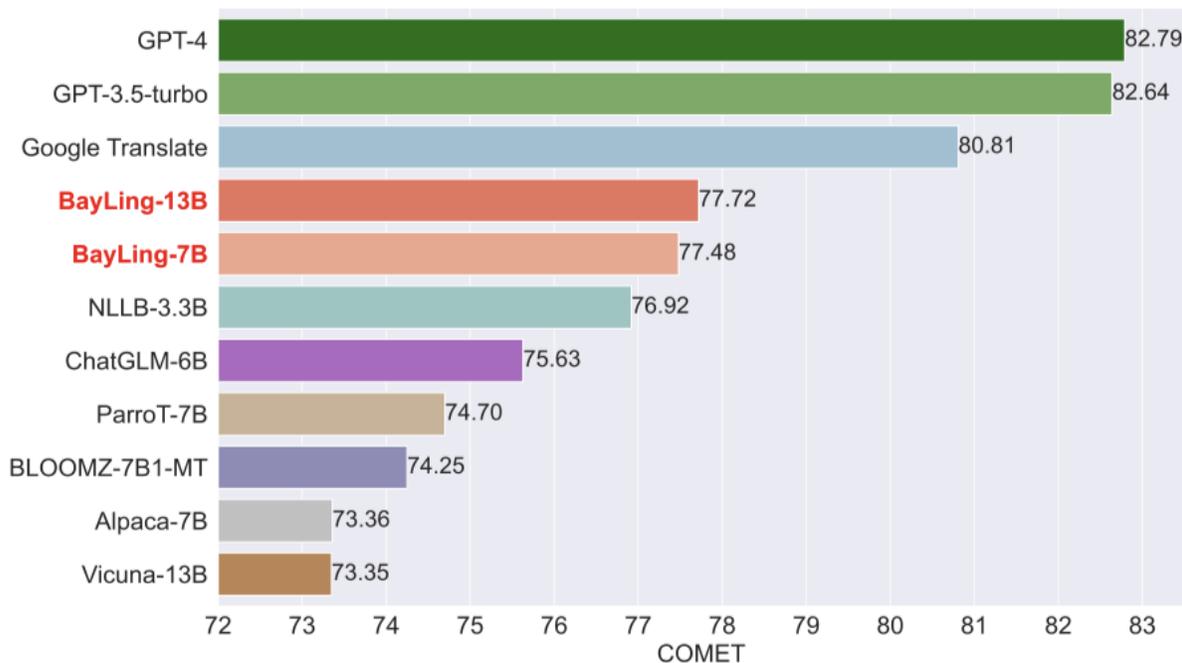
- **语言对齐**：完成语言生成能力从英语到其他语言的迁移。
- **复合任务**：同时提升多语言、指令理解、多轮交互等能力。
- **数据质量高**：无毒性、偏见。

Source	Interactive	Languages		#Instances
Alpaca	Single-turn	English		52K
ShareGPT	Multi-turn	English-dominant		90K
→ Interactive Translation	Multi-turn	Instruction Languages		160K
		English, Chinese	Translation Languages English, Chinese German, French	

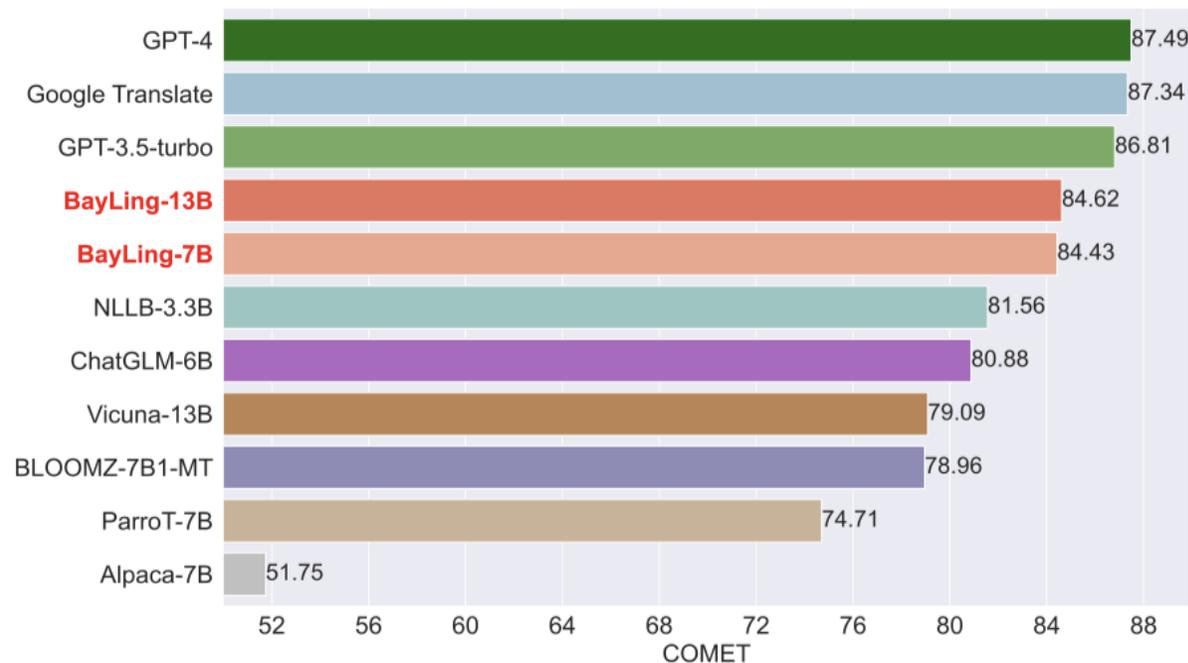
Shaolei Zhang, Qingkai Fang, Zhuocheng Zhang, et.al. BayLing: Bridging Cross-lingual Alignment and Instruction Following through Interactive Translation for Large Language Models.

翻译能力：中英翻译性能

■ BayLing-13B 相比 GPT-4 取得 95% 翻译性能



(a) 中英翻译上的COMET得分



(b) 英中翻译上的COMET得分

图： WMT22 中英翻译测评

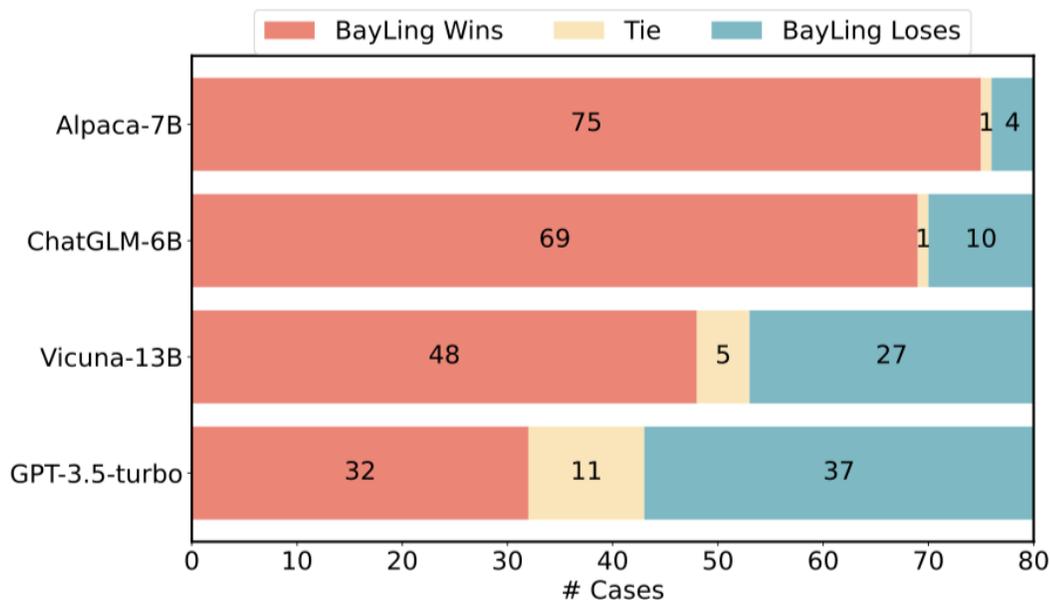
指令遵循能力：通用任务上的GPT-4评测结果

■ 多轮中英指令测试集

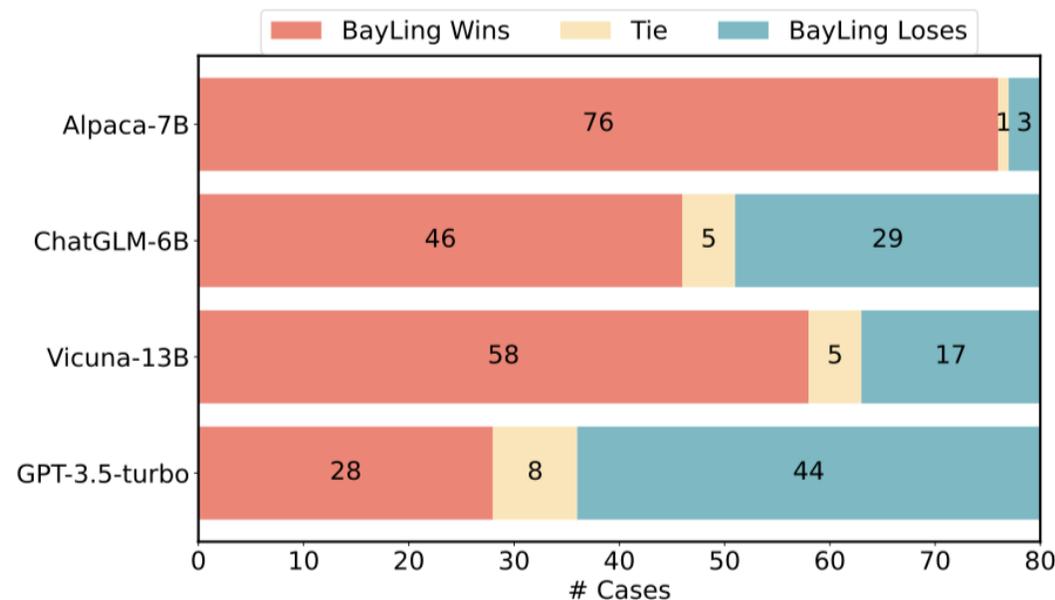
□ 35% 优于 GPT-3.5-turbo

□ 50% 不差于 GPT-3.5-turbo

BayLing表现出更强的
中文能力、多轮交互能力



(a) 英语指令



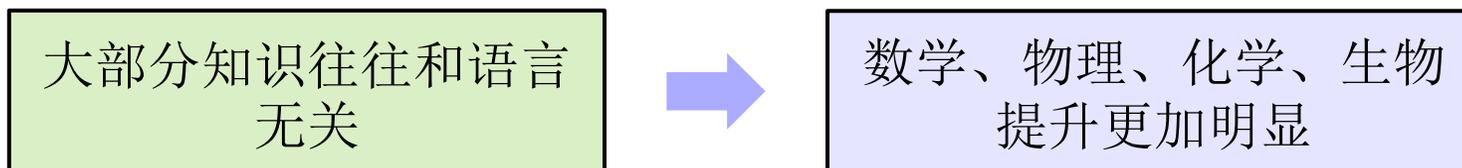
(b) 中文指令

图：多轮交互指令上的GPT-4评估

知识跨语言迁移：AGIEval中文高考评测结果

■ 语言对齐可以将知识从英语迁移到中文

- 基座模型中的知识以英语形式存储
- 通过语言对齐将知识高效地迁移到中文，避免了用大量数据重新注入中文知识



表：百聆在高考测试集的得分

Systems	Avg.	GaoKao (%)								
		chinese	english	mathqa	physics	chemistry	biology	history	geography	mathcloze
GPT-3.5-turbo	43.87	42.68	86.27	30.48	21.00	44.44	46.19	59.57	63.32	0.85
BayLing-13B	32.13	29.27	69.28	29.34	21.50	36.71	30.00	34.04	38.19	0.85
BayLing-7B	28.20	27.64	55.56	26.78	24.50	29.95	29.05	33.19	27.14	0.00
ChatGLM-6B	31.83	31.71	52.29	26.50	16.00	27.54	28.10	54.04	47.74	2.54
Vicuna-13B	29.36	21.14	71.24	21.94	23.00	31.88	27.14	33.19	34.67	0.00
Alpaca-7B	20.03	24.80	36.27	17.95	6.00	20.77	20.95	24.68	27.14	1.69

机器翻译技术 → 提升大模型能力

- Transformer 最早在MT任务中展现出强大的性能，逐渐拓展到其他任务。
- 机器翻译技术具有更好的泛化性
 - 多语言翻译 ⇒ 大模型的低资源语言能力
 - 交互式翻译 ⇒ 同时提升大模型的语言能力/指令遵循能力
 - 多模态翻译 ⇒ 多模态大模型
 - 翻译可解释性 ⇒ 大模型可解释性
 - ...

百聆

发挥传统NLP任务积累的研究经验

谢谢大家!

张绍磊

中国科学院计算技术研究所
zhangshaolei20z@ict.ac.cn