



中国科学院新疆理化技术研究所

Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences

浅谈大模型时代下低资源 机器翻译研究进展

米尔阿迪力江 麦麦提

(伊犁, 2023.11.22)



Outline

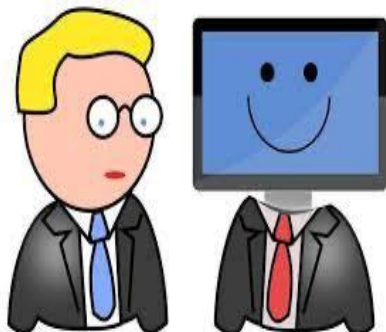
- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

Introduction of LLM

The Three Paths to Achieving AI

输入经验

将知识通过规则等方式教给计算机，进行符号处理



实现人脑机制

解明人脑机制，基于相同原理实现人类智能

从数据中学习

通过数据驱动，ML方法模仿人类智能



Learn From Data

LLM = 通过数据学习 + 受人脑启发 + 输入经验知识

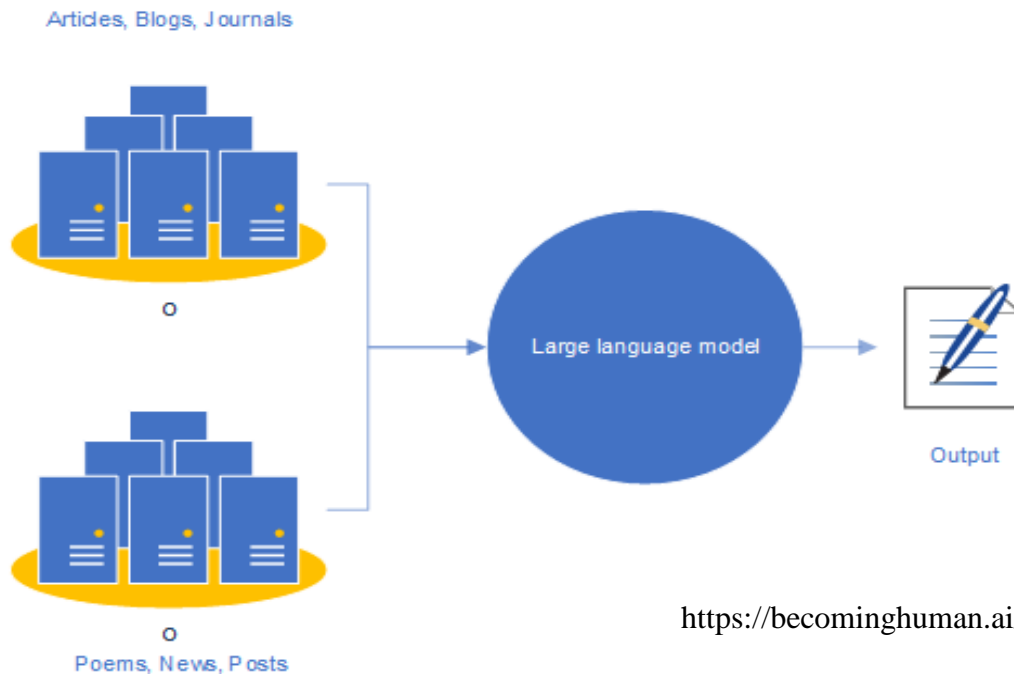
What Are Large Language Models?

大规模**语言模型** (LLM) 是基于**Transformer** 架构的模型，它们通过**自监督或半监督学习**在大量**未标记的文本**上进行广泛训练。这些模型旨在生成听起来自然且与上下文相关的文本，适用于各种风格和格式。

大模型

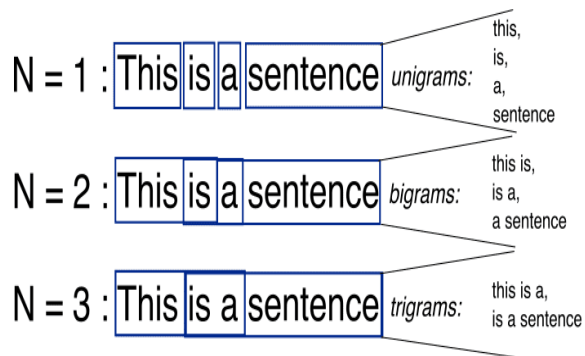


大规模 预训练 语言模型

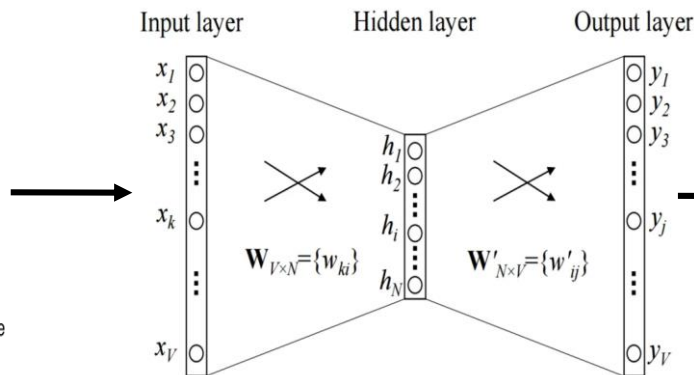


<https://becominghuman.ai/>

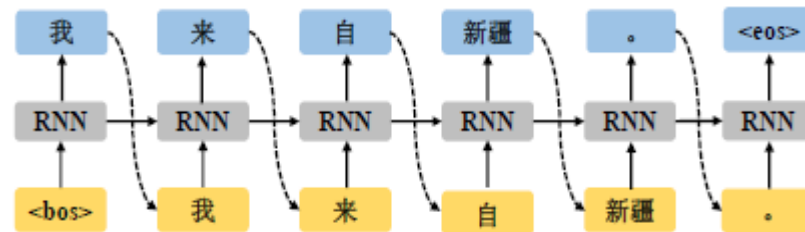
Language Models



N-gram LM



Word2Vec

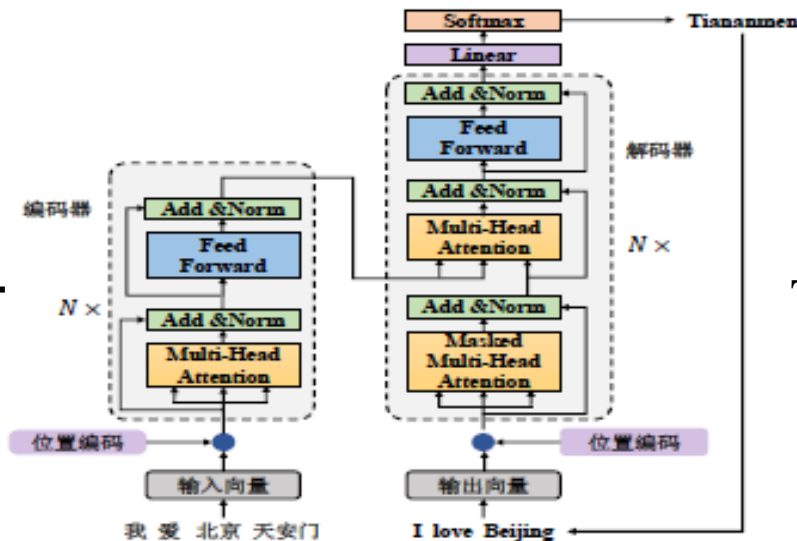


RNNLM

LLM



PLM



Transformer

Large-scale PLMs become the foundational infrastructure of Modern NLP



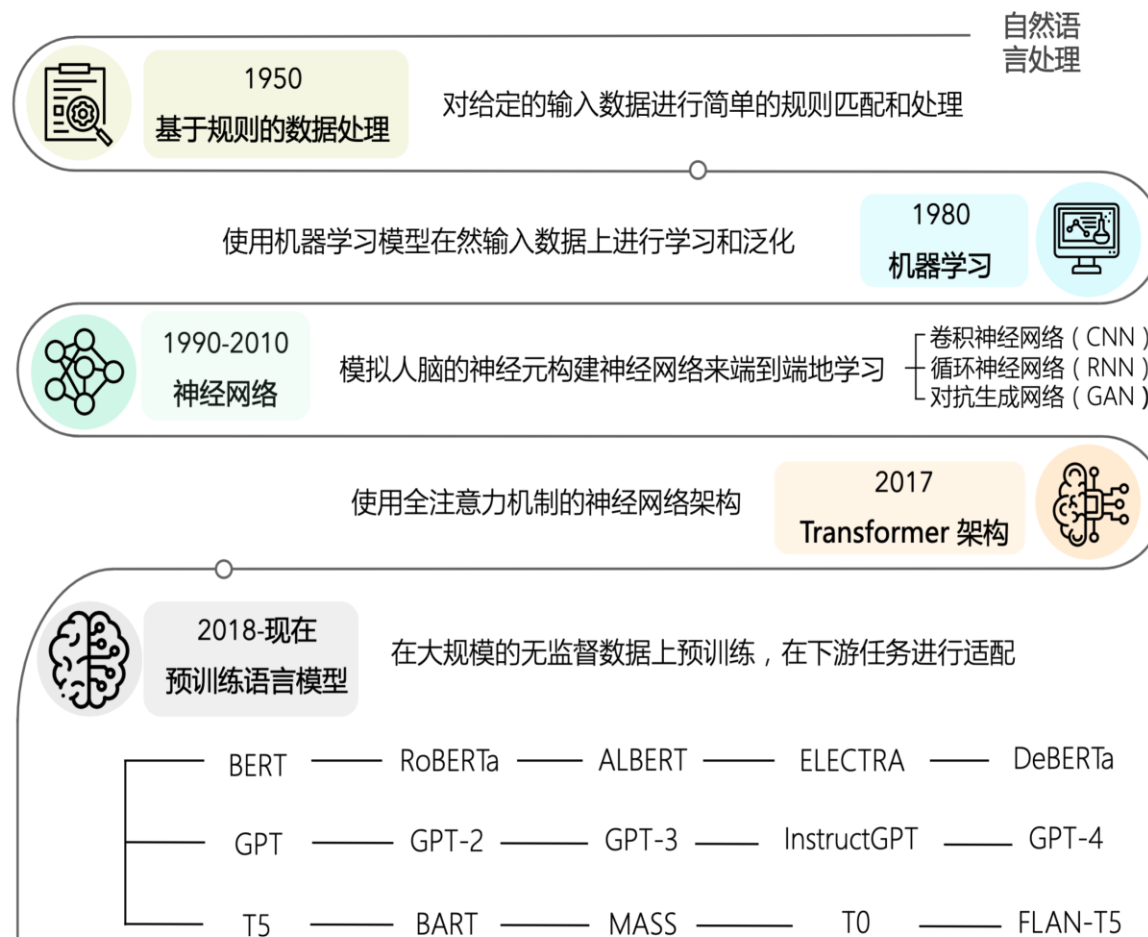
Geoffrey Hinton
(图灵奖获得者)

将GPT-3的壮观表现推断到未来，表明生命、宇宙和万物的答案只是4.398万亿个参数。



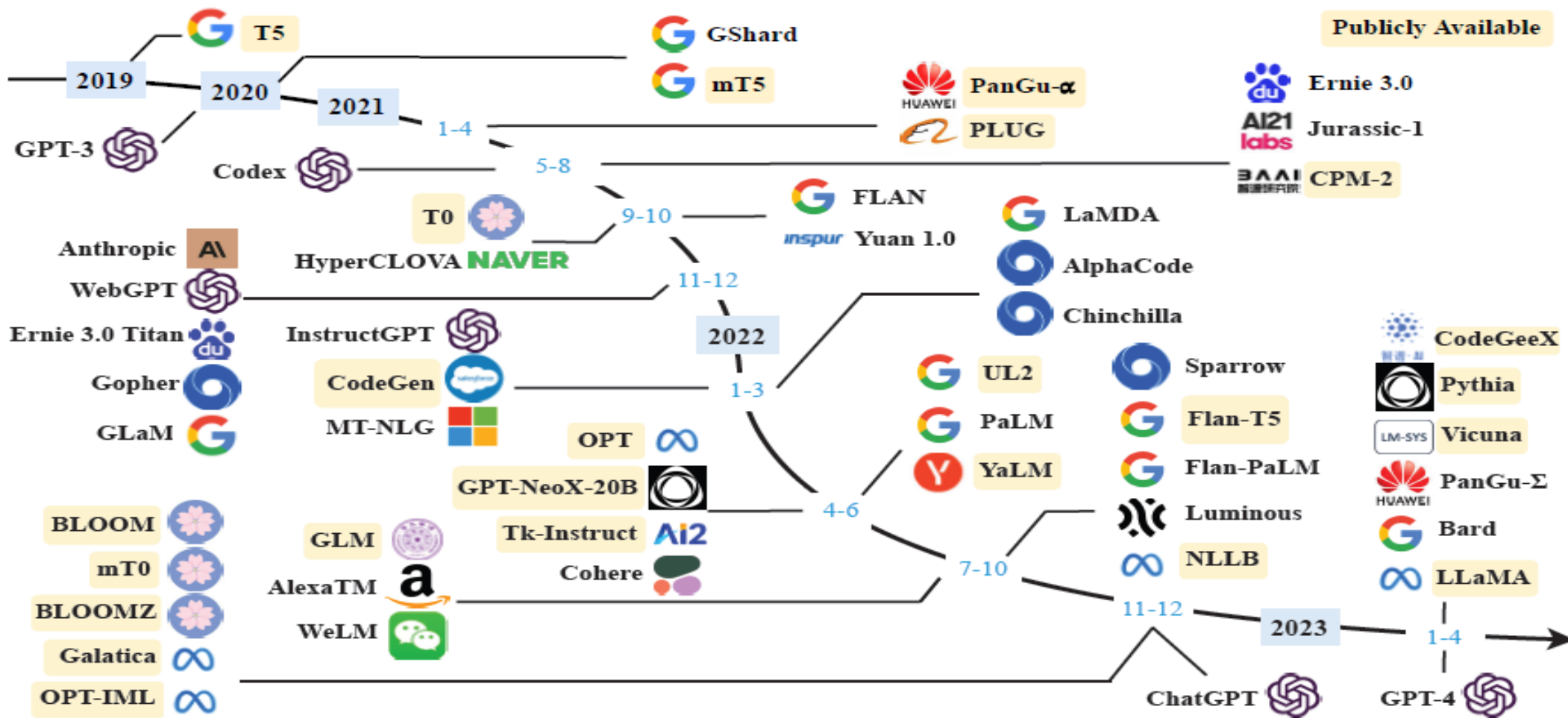
Ilya Sutskever
(OpenAI首席科学家)

将语言模型在学习预测下一个字符的过程中来学习对整个宇宙的规律。



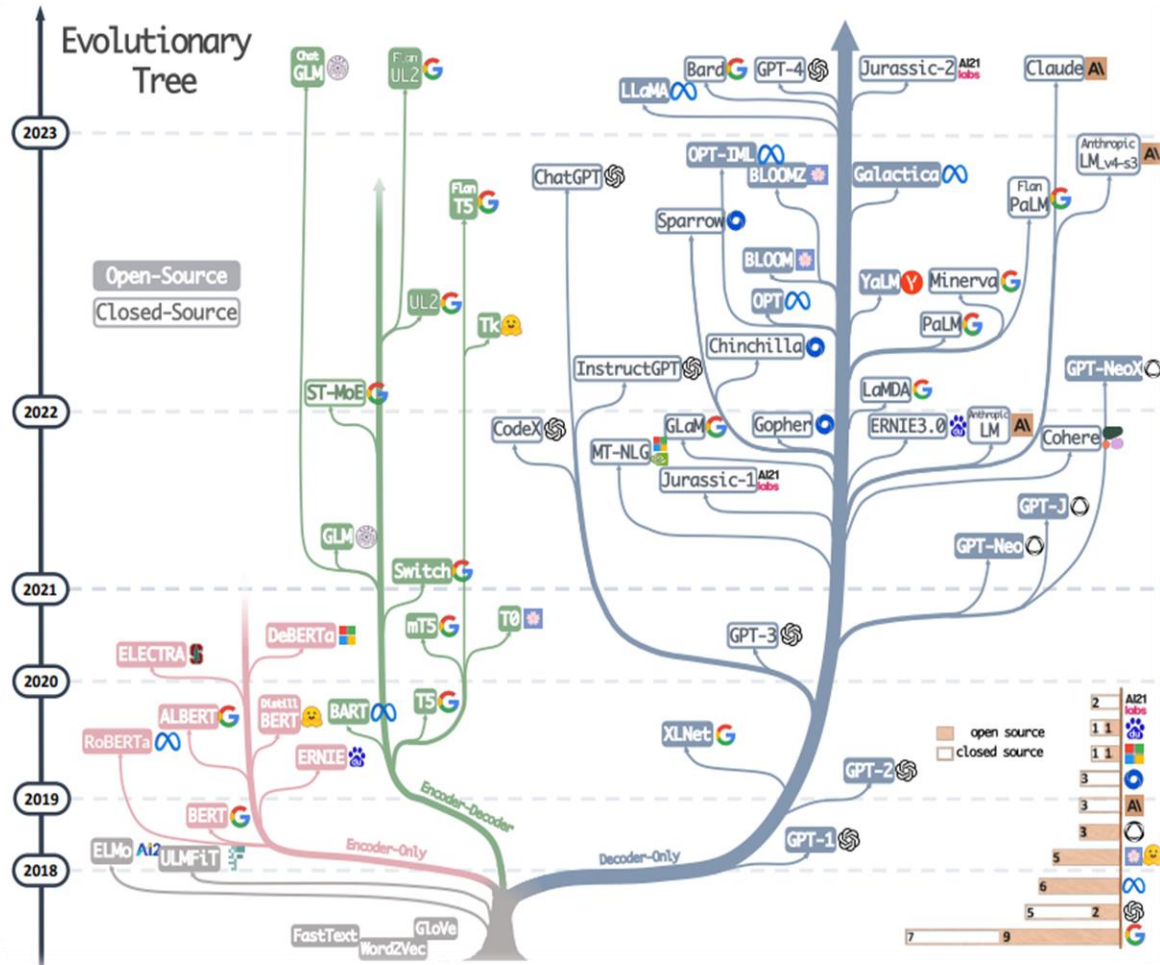
Ding, CCL2023

The timeline of existing LLM



(Xin Zhao et al., 2023)

The Evolutionary of LLM



(Yang et al., 2023)

- **GPT** Improving Language Understanding by Generative Pre-Training. 2018
- **GPT-2** Language Models are Unsupervised Multitask Learners. 2018.
- **GPT-3** "Language Models are Few-Shot Learners". NeurIPS 2020.
- **OPT** "OPT: Open Pre-trained Transformer Language Models". 2022.
- **PaLM** "PaLM: Scaling Language Modeling with Pathways". Aakanksha Chowdhery et al arXiv 2022.
- **BLOOM** "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model". 2022.
- **MT-NLG** "Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model". 2021.
- **GLaM** "GLaM: Efficient Scaling of Language Models with Mixture-of-Experts". ICML 2022.
- **Gopher** "Scaling Language Models: Methods, Analysis & Insights from Training Gopher", 2021.
- **chinchilla** "Training Compute-Optimal Large Language Models". 2022.
- **LaMDA** "LaMDA: Language Models for Dialog Applications". 2021.
- **LLaMA** "LLaMA: Open and Efficient Foundation Language Models". 2023.
- **GPT-4** "GPT-4 Technical Report". 2023.
- **BloombergGPT** BloombergGPT: A Large Language Model for Finance, 2023,
- **GPT-NeoX-20B**: "GPT-NeoX-20B: An Open-Source Autoregressive Language Model". 2022.

The Development History of GPT (Generative Pre-Trained Transformer)



Existing Models

- 近年来大规模语言模型（在本次调查中**规模超过100亿**的模型）的统计数据，包括容量评估、预训练数据规模（以令牌数量或存储大小计算）以及硬件资源成本。
- 在这张表格中，我们仅包括了那些有关**技术细节的公开论文**的大规模语言模型。

	Model	Release Time	Size (B)	Base Model	Adaptation IT RLHF	Pre-train Data Scale	Latest Data Timestamp	Hardware (GPUs / TPUs)	Training Time	Evaluation ICL	CoT
Publicly Available	T5 [73]	Oct-2019	11	-	-	1T tokens	Apr-2019	1024 TPU v3	-	✓	-
	mT5 [74]	Oct-2020	13	-	-	1T tokens	-	-	-	✓	-
	PanGu-α [75]	Apr-2021	13*	-	-	1.1TB	-	2048 Ascend 910	-	✓	-
	CPM-2 [76]	Jun-2021	198	-	-	2.6TB	-	-	-	-	-
	T0 [28]	Oct-2021	11	T5	✓	-	-	512 TPU v3	27 h	✓	-
	CodeGen [77]	Mar-2022	16	-	-	577B tokens	-	-	-	✓	-
	GPT-NeoX-20B [78]	Apr-2022	20	-	-	825GB	-	96 40G A100	-	✓	-
	Tk-Instruct [79]	Apr-2022	11	T5	✓	-	-	256 TPU v3	4 h	✓	-
	UL2 [80]	May-2022	20	-	-	1T tokens	Apr-2019	512 TPU v4	-	✓	✓
	OPT [81]	May-2022	175	-	-	180B tokens	-	992 80G A100	-	✓	-
	NLLB [82]	Jul-2022	54.5	-	-	-	-	-	-	✓	-
	GLM [83]	Oct-2022	130	-	-	400B tokens	-	768 40G A100	60 d	✓	-
	Flan-T5 [64]	Oct-2022	11	T5	✓	-	-	-	-	✓	✓
	BLOOM [69]	Nov-2022	176	-	-	366B tokens	-	384 80G A100	105 d	✓	-
	mT0 [84]	Nov-2022	13	mT5	✓	-	-	-	-	✓	-
	Galactica [35]	Nov-2022	120	-	-	106B tokens	-	-	-	✓	✓
	BLOOMZ [84]	Nov-2022	176	BLOOM	✓	-	-	-	-	✓	-
	OPT-IML [85]	Dec-2022	175	OPT	✓	-	-	128 40G A100	-	✓	✓
	LLaMA [57]	Feb-2023	65	-	-	1.4T tokens	-	2048 80G A100	21 d	✓	-
	CodeGeeX [86]	Sep-2022	13	-	-	850B tokens	-	1536 Ascend 910	60 d	✓	-
Pythia [87]	Apr-2023	12	-	-	300B tokens	-	256 40G A100	-	✓	-	
Closed Source	GPT-3 [55]	May-2020	175	-	-	300B tokens	-	-	-	✓	-
	GShard [88]	Jun-2020	600	-	-	1T tokens	-	2048 TPU v3	4 d	✓	-
	Codex [89]	Jul-2021	12	GPT-3	-	100B tokens	May-2020	-	-	✓	-
	ERNIE 3.0 [90]	Jul-2021	10	-	-	375B tokens	-	384 V100	-	✓	-
	Jurassic-1 [91]	Aug-2021	178	-	-	300B tokens	-	800 GPU	-	✓	-
	HyperCLOVA [92]	Sep-2021	82	-	-	300B tokens	-	1024 A100	13.4 d	✓	-
	FLAN [62]	Sep-2021	137	LaMDA-PT	✓	-	-	128 TPU v3	60 h	✓	-
	Yuan 1.0 [93]	Oct-2021	245	-	-	180B tokens	-	2128 GPU	-	✓	-
	Anthropic [94]	Dec-2021	52	-	-	400B tokens	-	-	-	✓	-
	WebGPT [72]	Dec-2021	175	GPT-3	-	✓	-	-	-	✓	-
	Gopher [59]	Dec-2021	280	-	-	300B tokens	-	4096 TPU v3	920 h	✓	-
	ERNIE 3.0 Titan [95]	Dec-2021	260	-	-	-	-	-	-	✓	-
	GLaM [96]	Dec-2021	1200	-	-	280B tokens	-	1024 TPU v4	574 h	✓	-
	LaMDA [63]	Jan-2022	137	-	-	768B tokens	-	1024 TPU v3	57.7 d	-	-
	MT-NLG [97]	Jan-2022	530	-	-	270B tokens	-	4480 80G A100	-	✓	-
	AlphaCode [98]	Feb-2022	41	-	-	967B tokens	Jul-2021	-	-	✓	-
	InstructGPT [61]	Mar-2022	175	GPT-3	✓	✓	-	-	-	✓	-
	Chinchilla [34]	Mar-2022	70	-	-	1.4T tokens	-	-	-	✓	-
	PaLM [56]	Apr-2022	540	-	-	780B tokens	-	6144 TPU v4	-	✓	✓
	AlexaTM [99]	Aug-2022	20	-	-	1.3T tokens	-	128 A100	120 d	✓	✓
	Sparrow [100]	Sep-2022	70	-	-	✓	-	64 TPU v3	-	✓	-
	WeLM [101]	Sep-2022	10	-	-	300B tokens	-	128 A100 40G	24 d	✓	-
	U-PaLM [102]	Oct-2022	540	PaLM	-	-	-	512 TPU v4	5 d	✓	✓
Flan-PaLM [64]	Oct-2022	540	PaLM	✓	-	-	512 TPU v4	37 h	✓	✓	
Flan-U-PaLM [64]	Oct-2022	540	U-PaLM	✓	-	-	-	-	✓	✓	
GPT-4 [46]	Mar-2023	-	-	✓	✓	-	-	-	✓	✓	
PanGu-Σ [103]	Mar-2023	1085	PanGu-α	-	-	329B tokens	-	512 Ascend 910	100 d	✓	-

(Xin Zhao et al., 2023)

Dataset Statistics

- 信息量估计:
 - 人一生要说的话 5TB
 - 所有语言的wiki百科 5PB (1PB=1024TB)
 - 所有记录在纸上的内容 200PB
 - YouTube所有视频 400PB, Facebook总信息量 500PB
 - 2000年互联网上的信息量 1000PB, 2010年互联网信息量 450EB; 目前互联网总信息量 3600EB (1EB=1024PB)
 - 世界上所有的电子数据约 100ZB (1ZB=1024EB)
- LLM训练 (ChatGPT) 用了**320TB**文字信息, 约3000亿单词/字

Statistics of commonly-used data sources.

Corpora	Size	Source	Latest Update Time
BookCorpus [122]	5GB	Books	Dec-2015
Gutenberg [123]	-	Books	Dec-2021
C4 [73]	800GB	CommonCrawl	Apr-2019
CC-Stories-R [124]	31GB	CommonCrawl	Sep-2019
CC-NEWS [27]	78GB	CommonCrawl	Feb-2019
REALNEWS [125]	120GB	CommonCrawl	Apr-2019
OpenWebText [126]	38GB	Reddit links	Mar-2023
Pushift.io [127]	2TB	Reddit links	Mar-2023
Wikipedia [128]	21GB	Wikipedia	Mar-2023
BigQuery [129]	-	Codes	Mar-2023
the Pile [130]	800GB	Other	Dec-2020
ROOTS [131]	1.6TB	Other	Jun-2022

(Xin Zhao et al., 2023)

<https://www.zhihu.com/question/365920362>

Dataset Distribution

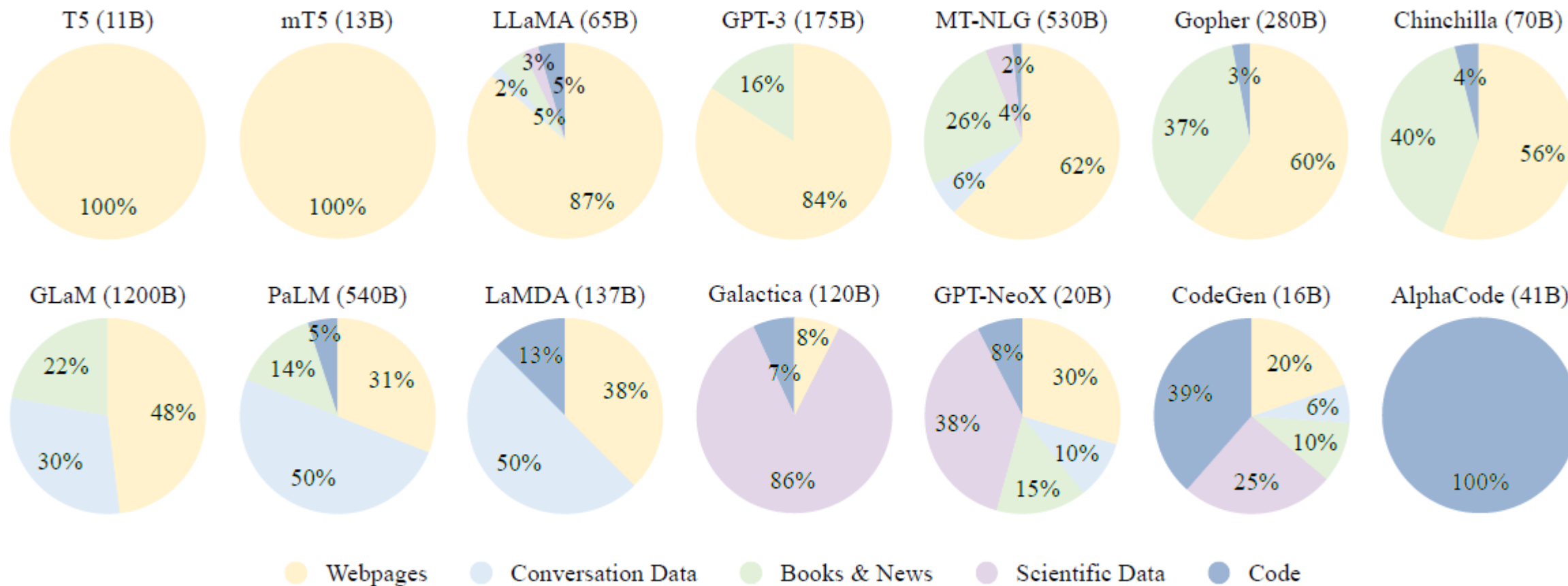
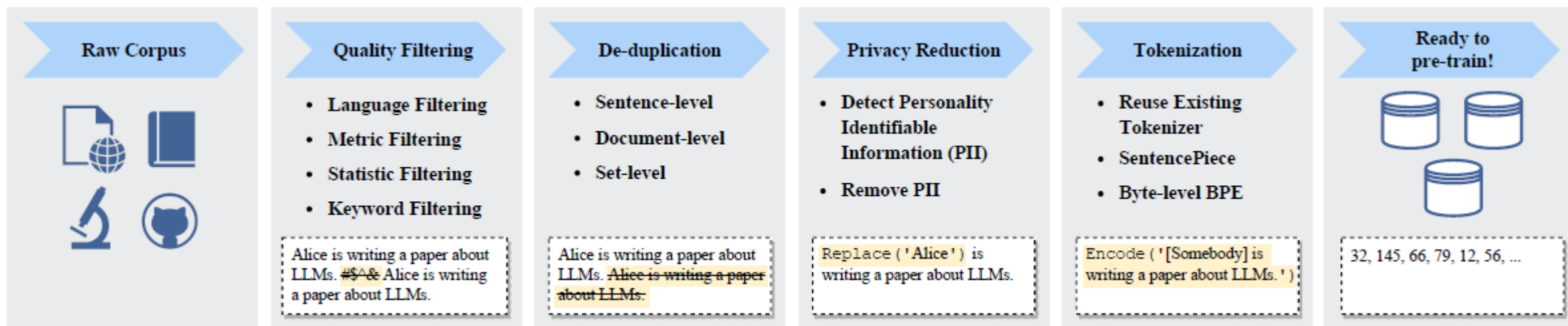


Fig. 2. Ratios of various data sources in the pre-training data for existing LLMs.

(Xin Zhao et al., 2023)

Dataset Preprocessing

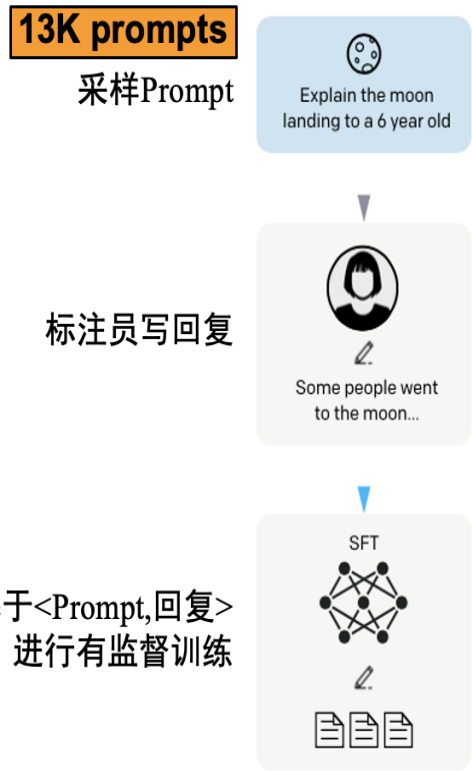
- InstructGPT: 这是GPT-3的一个变体，专门针对用户指令进行了优化。它在2021年底左右被公开介绍。
- WebGPT: 截至2022年1月，WebGPT还处于研究和开发阶段。具体的发布时间可能在2022年之后。
- ChatGPT 3.5: 2022年年底发布



(Xin Zhao et al., 2023)

The Main Architecture of GPT

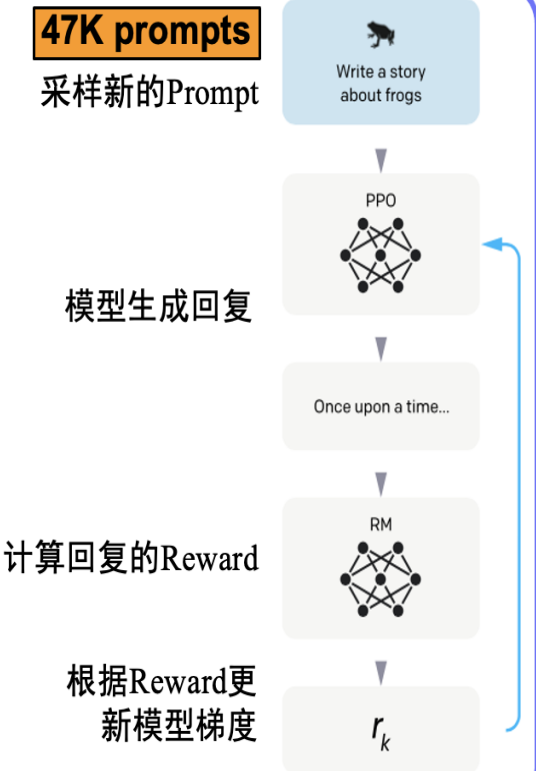
第1步 有监督的指令微调



第2步 基于人类反馈训练Reward模型



第3步 采用强化学习训练模型



有监督指令微调

基于人工反馈的强化学习 (RLHF)

2月28日 晚20:00

ChatGPT类算法的壁垒、 开源和算力私有化

分享内容

1. 从数据、算法、算力、环境等方面讨论国内企业研发ChatGPT的壁垒。
2. 探究ChatGPT开源的意义、可能性、所需要的生态环境、难点以及如何克服这些难点。
3. 围绕算力私有化讨论ChatGPT带给芯片、算法、投资人、云公司的机会是哪些。

主讲嘉宾

北京航空航天大学
自动化科学与电气工程学院教授

秦曾昌教授



千芯科技董事长
前华为系自然语言处理 (NLP) 企业
首席科学家

陈巍博士



阿里巴巴达摩院
语言技术实验室自然语言处理 (NLP)
高级算法工程师

木拉丁博士





2023年2月28日
晚20:00-21:30

名额有限! 扫码报名

创业芯 ChipStar 工作室

创业芯 2023

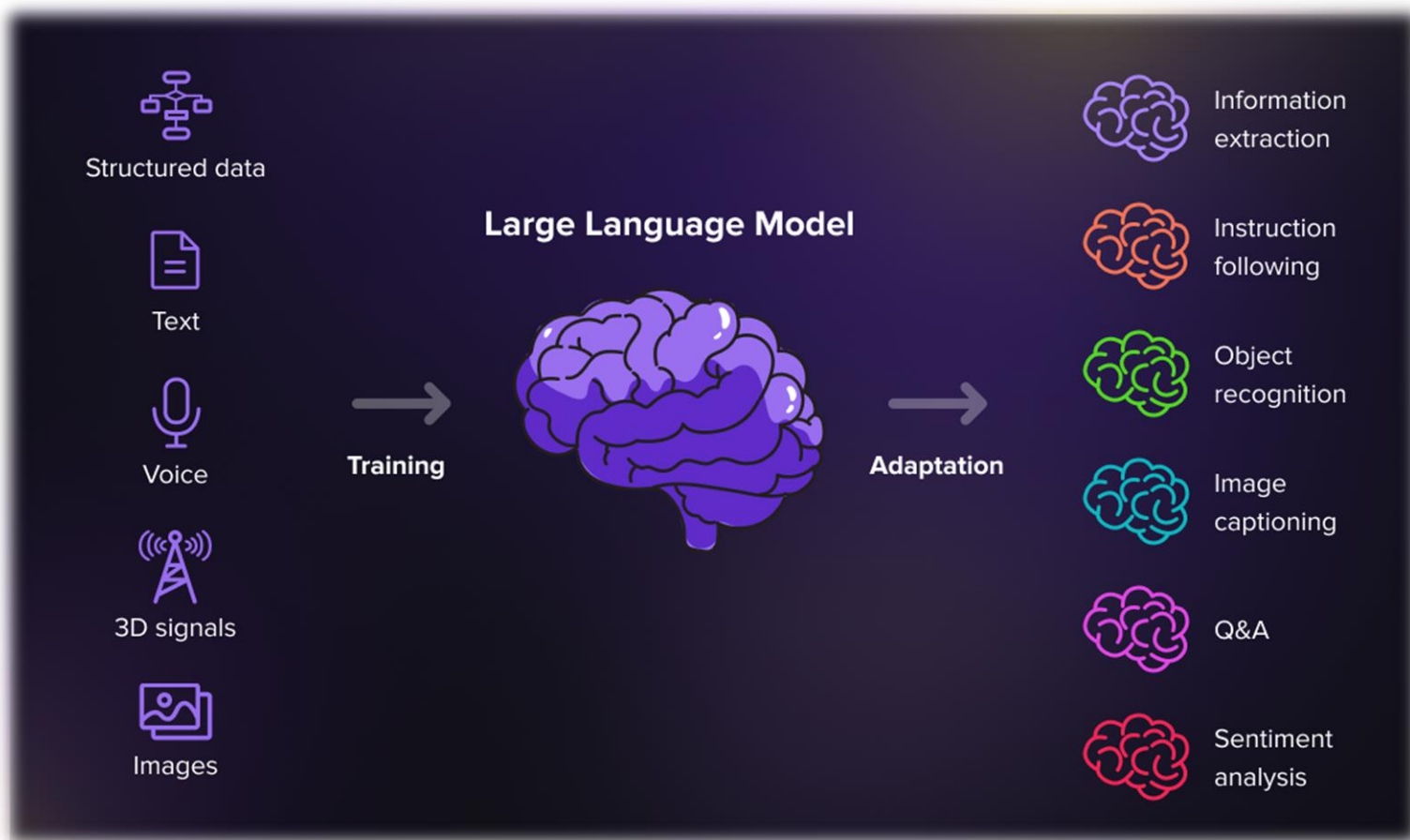


Outline

- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

LLM performance in NLP

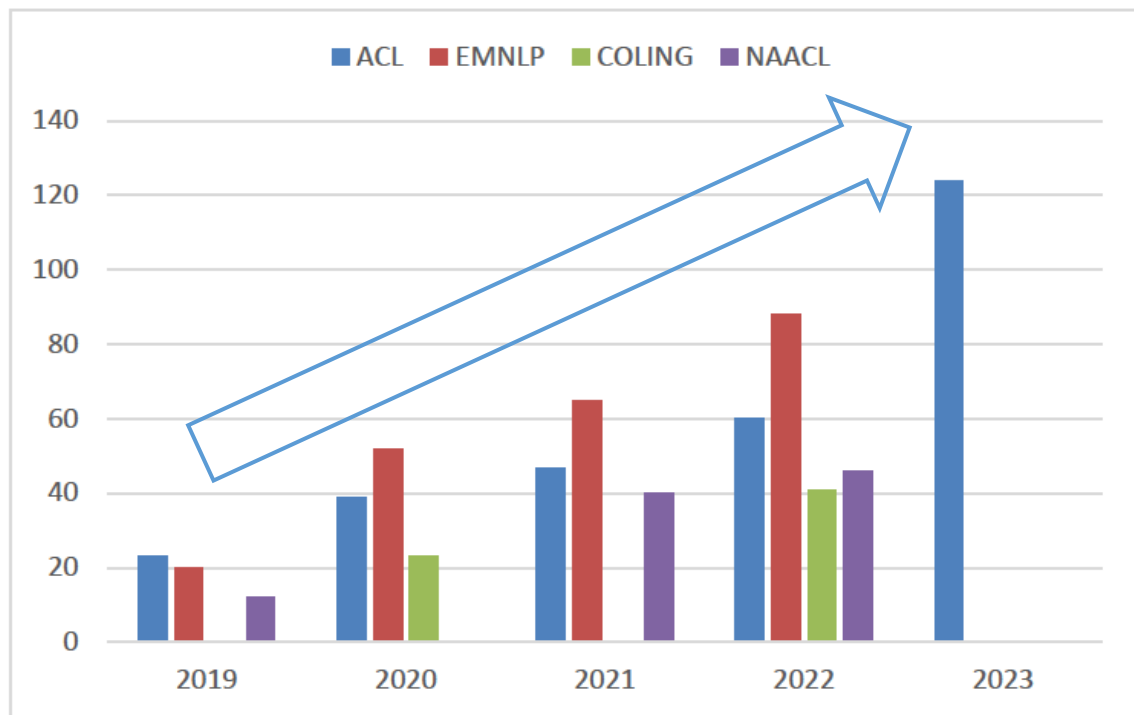
LLM performance on NLP down-stream tasks



<https://blog.gopenai.com/>

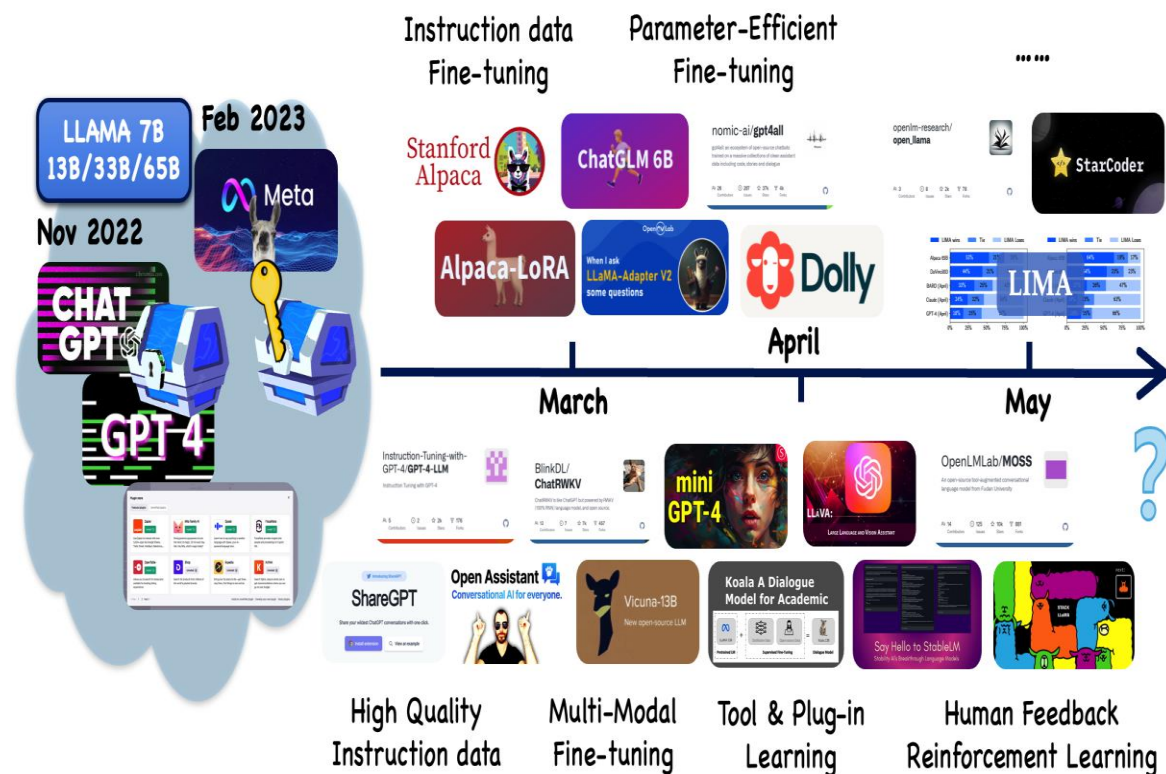
Basic Fact - foundational models is rapidly evolving

语言模型论文粗略统计



相关信息统计自ACL Anthology: <https://www.aclweb.org/anthology/>

一堆指令微调模型层出不穷



(Ding et al., CCL2023)

LLM Everywhere --- Conversation

You
能跟我聊一会儿吗?

ChatGPT
当然可以,很高兴与您聊天。您有什么想谈论的话题吗?无论是科技、文化、旅行还是日常生活的点滴,我都乐于倾听和参与讨论。您可以随时开始。

You
你问我一个问题吧

ChatGPT
好的,让我来问一个问题:您最近读过的一本书是什么?您对它的看法如何?

You
创业有关但是名字忘了

ChatGPT
关于创业的书籍通常包含很多有价值的见解和经验。即使忘记了书名,您能分享一下从这本书中学到的最重要的一两个点吗?比如关于创业的策略、创新思维或者是成功和失败的教训等。这些内容往往能够启发人们对于创业或职业生涯有更深度的思考。



You
新疆伊犁师范大学您怎么看?

ChatGPT
新疆伊犁师范大学是中国新疆维吾尔自治区的一所高等学府。作为一所地区性大学,它在教育和培养当地人才方面扮演着重要的角色。大学通常提供多样化的课程和专业,包括师范教育、艺术、科学等领域,这对于提升区域教育水平和促进地方文化和经济发展都非常重要。

从更广泛的视角来看,位于如此独特地理和文化背景的大学,对于学生而言,不仅提供了学术上的学习机会,也是一个了解和体验多元文化的宝贵平台。学生在这里可以接触到不同的文化和观点,这对于培养全球视角和跨文化交流能力是非常有益的。

您对新疆伊犁师范大学有什么特别的看法或经历吗?您的观点或体验可能会给其他人提供独特的见解。



You
我不太了解这个学校~不过觉得风景不错

ChatGPT
新疆伊犁地区以其壮丽的自然风光和独特的文化风情而闻名,这也为在那里的学校,如新疆伊犁师范大学,提供了一个非常特殊的背景。伊犁的自然景观,包括广阔的草原、雄伟的山脉和丰富的民族文化,都为学生和教职工提供了一个美丽而充满活力的环境。

即使对于不是特别了解该学校的人来说,这样的地理位置和环境也能激发人们对于探索 and 了解这一地区的兴趣。对于那些对自然景观、不同文化以及新疆地区的历史和社会感兴趣的人来说,伊犁无疑是一个值得探访和学习的地方。您对新疆或其他地区的自然风光有兴趣吗?您最喜欢的旅行目的地是哪里?



LLM Everywhere --- Conversation, QA



ChatGPT 3.5 ▾

You
新疆伊犁师范大学什么时候改名为“大学”？

ChatGPT
截止到我的知识截止日期为2022年1月，新疆伊犁师范大学尚未改名为“大学”。如果在此之后发生了名称变更，我无法提供相关信息。如果您对该校是否发生了名称变更有疑问，建议您查阅新闻或联系学校的官方网站以获取最新信息。



ChatGPT 4 ▾

You
新疆伊犁师范大学什么时候改名为“大学”？

ChatGPT
新疆伊犁师范大学，原名伊犁师范学院，于2018年12月经教育部批准更名为伊犁师范大学^[1]。

LLM Everywhere --- Conversation, QA, Outline

You
给我提供一个大模型概要有关的提纲

ChatGPT
当然，这里是一个关于大型机器学习模型（如GPT、BERT等）的概要提纲。这个提纲将帮助您深入了解大型模型的基本概念、架构、应用场景以及它们在未来发展中可能面临的挑战和机遇。

1. 引言
 - 大型模型在人工智能领域的兴起背景
 - 大型模型的定义和重要性
2. 大型模型的基本原理和架构
 - 深度学习和神经网络基础
 - 大型模型的核心组件（如变换器架构）
 - 数据和算法在模型训练中的作用
3. 主要大型模型的发展历程
 - 不同代的大型模型（如GPT系列、BERT、T5等）的发展和比较
 - 关键技术突破和创新点
4. 大型模型的应用场景
 - 自然语言处理（NLP）
 - 图像识别和处理
 - 其他领域的应用（如医疗、金融、游戏等）
5. 大型模型的优势和挑战
 - 处理复杂问题和大数据集的能力
 - 模型偏见和伦理问题
 - 计算资源和环境影响
 - 数据隐私和安全性问题
6. 未来发展趋势
 - 技术创新和模型优化



You
帮我写一个slides提纲，有关机器学习基础与实践。

ChatGPT
当然，以下是一个关于“机器学习基础与实践”的幻灯片提纲。这个提纲将帮助您设计一个包含机器学习的基本概念、关键技术和方法、实际应用案例，以及未来发展趋势的完整演示。

1. 幻灯片 1: 标题页
 - 标题：机器学习基础与实践
 - 副标题：从理论到应用
 - 演讲者姓名和机构
2. 幻灯片 2: 介绍
 - 机器学习的定义和重要性
 - 本次演讲的目的和结构
3. 幻灯片 3: 机器学习的历史简介
 - 从最初的概念到现代的发展
 - 关键技术突破和里程碑事件
4. 幻灯片 4: 机器学习的类型
 - 监督学习、非监督学习、半监督学习、强化学习
 - 每种类型的基本概念和应用示例
5. 幻灯片 5: 关键算法和模型
 - 决策树、支持向量机、神经网络、深度学习等
 - 每种算法的基本原理和适用场景
6. 幻灯片 6: 数据处理和特征工程
 - 数据预处理的重要性
 - 特征选择和特征提取
 - 实例展示数据处理流程
7. 幻灯片 7: 机器学习的实际应用案例



LLM Everywhere --- Conversation, QA, Outline, Recommend

You 如果想从事人工智能开发行业的话，对一个新人你有什么建议呀？

ChatGPT 从事人工智能（AI）开发是一个充满挑战和机遇的领域。对于新人来说，以下是一些建议：

- 基础教育：**学习计算机科学、数学和统计学的基础知识。对于人工智能特别是机器学习而言，线性代数、概率论、微积分和算法是非常重要的。
- 编程技能：**精通至少一种编程语言，Python是最受欢迎的语言之一，因为它拥有大量的AI和机器学习库，例如TensorFlow和PyTorch。
- 在线课程和资源：**利用在线平台如Coursera、edX、Udacity等，参加AI和机器学习的课程。
- 实际项目：**通过实际项目来应用你的知识。GitHub上有很多开源项目，你可以参与其中来提高你的技能。
- 理解业务：**了解AI技术如何解决实际问题，并对其可能带来的业务和社会影响有所认识。
- 读原始论文：**阅读领域内的经典和最新论文，如在arXiv上发布的论文，以保持对最新研究动态的了解。
- 网络和社区：**参加Meetups、Conferences、Workshops等，建立职业网络，与同行交流。
- 批判性思维：**培养批判性思维能力，对数据、模型和结果进行仔细分析。
- 伦理和法律知识：**理解与人工智能相关的伦理和法律问题，例如数据隐私、算法偏见等。
- 持续学习：**AI领域在快速发展，持续学习新技术、新算法是非常重要的。

记住，建立一个坚实的基础，然后不断地学习和实践是关键。每个人的学习路径都是独特的，找到最适合你的方式去探索和成长。



You 是非这个行业的，想转行做人工智能，有啥建议的学习方向吗？

ChatGPT 如果您想从其他行业转入人工智能（AI），以下是一些建议的学习方向和步骤：

- 学习基础：**首先，您需要建立在数学（特别是线性代数、概率论和统计学）和编程（推荐Python）方面的坚实基础。这些是深入理解AI的前提。
- 了解AI的子领域：**人工智能包含了机器学习、深度学习、自然语言处理、计算机视觉等子领域。初步了解这些子领域，确定您最感兴趣的领域。
- 参加在线课程：**参加Coursera、edX、Udacity等平台上的入门课程，例如Andrew Ng的《机器学习》课程，或者Google的机器学习速成课程。
- 项目实践：**通过小项目开始实践，比如使用Python编写一个简单的数据分析脚本，或者使用机器学习库如scikit-learn来构建简单模型。
- 深化专业知识：**在有了基础之后，可以通过专门的书籍、课程和项目来深化特定领域的知识，例如深度学习、强化学习或者自然语言处理。
- 参与开源项目：**GitHub上有许多AI相关的开源项目，参与这些项目能够帮助您理解实际应用中的问题，并提高解决问题的能力。
- 建立网络：**加入AI相关的论坛和社区，如Reddit的r/MachineLearning、Kaggle等，与其他从业者交流心得。
- 了解伦理和社会影响：**AI技术的使用会带来伦理和社会问题，理解这些问题对于负责任地开发和应用AI技术很重要。
- 求知若渴：**AI是一个不断发展的领域，保持好奇心和学习的渴望，不断更新您的知识和技能。
- 思考如何结合自己的背景：**思考您之前的行业经验和知识如何与AI结合。许多AI应用是跨领域的，您之前的专业知识可能会成为宝贵的资产。

转行到AI领域可能既充满挑战也富有成就感。重要的是保持耐心，不断学习，并乐于解决问题。



Message ChatGPT...



ChatGPT can make mistakes. Consider checking important information.

LLM Everywhere --- Conversation, QA, Outline, Recommend, MT



You

"我喜欢新疆北疆，景色非常迷人，特别美丽，不过从没去过。" \n\n麻烦把以上这几句话翻译成英语，德语，法语，阿拉伯语，土耳其语，泰语等几个多语种。



ChatGPT

当然，以下是您提供的句子翻译成不同语言的版本：

英语:

"I really like the northern part of Xinjiang. The scenery is very enchanting and incredibly beautiful, though I've never been there."

德语:

"Ich mag den Norden von Xinjiang sehr. Die Landschaft ist sehr bezaubernd und unglaublich schön, obwohl ich noch nie dort war."

法语:

"J'aime beaucoup le nord du Xinjiang. Le paysage est très charmant et incroyablement beau, bien que je n'y sois jamais allé."

阿拉伯语:

"أحب شمال شينجيانغ كثيراً. المناظر الطبيعية ساحرة وجميلة للغاية، على الرغم من أنني لم أذهب إلى هناك قط."

土耳其语:

"Xinjiang'ın kuzeyini çok seviyorum. Manzara çok büyüleyici ve inanılmaz güzel, ama hiç oraya gitmedim."

泰语:

"ฉันชอบภาคเหนือของซินเจียงมาก ทิวทัศน์ที่นั่นมีเสน่ห์และสวยงามมาก แม้ว่าฉันจะไม่เคยไปที่นั่น."

这些翻译应该准确传达了原文的意思和情感 ↓



You

把以上句子可以翻译成东南亚地区几个国家的语言吗?



ChatGPT

当然，我可以将您提供的句子翻译成几种东南亚地区国家的语言：

印度尼西亚语:

"Saya sangat suka bagian utara Xinjiang. Pemandangannya sangat memukau dan sangat indah, meskipun saya belum pernah ke sana."

马来语:

"Saya sangat suka bahagian utara Xinjiang. Pemandangannya sangat menarik dan sangat cantik, walaupun saya belum pernah ke sana."

越南语:

"Tôi rất thích phần phía bắc của Tân Cương. Cảnh quan ở đó rất hấp dẫn và cực kỳ đẹp, mặc dù tôi chưa từng đến đó."

泰语(重复):

"ฉันชอบภาคเหนือของซินเจียงมาก ทิวทัศน์ที่นั่นมีเสน่ห์และสวยงามมาก แม้ว่าฉันจะไม่เคยไปที่นั่น."

菲律宾语:

"Gusto ko talaga ang hilagang bahagi ng Xinjiang. Ang tanawin ay napaka-kaakit-akit at lubos na maganda, bagaman hindi pa ako nakapunta doon."

这些翻译尽可能准确地保持了原文的含义和情感。



LLM Everywhere --- Conversation, QA, Outline, Recommend, MT, Coding

You
麻烦给我写一个python 脚本，需要把tsv 格式的文件转换成json 格式的文件。

ChatGPT
当然，我可以为您提供一个Python脚本来将TSV格式的文件转换成JSON格式的文件。以下是脚本的一个基本示例：

```
python
import csv
import json

# 将TSV文件转换为JSON文件的函数
def tsv_to_json(tsv_file_path, json_file_path):
    # 读取TSV文件
    with open(tsv_file_path, 'r', encoding='utf-8') as tsvfile:
        reader = csv.DictReader(tsvfile, delimiter='\t')

        # 创建一个列表来保存所有行的数据
        data = []
        for row in reader:
            data.append(row)

    # 写入JSON文件
    with open(json_file_path, 'w', encoding='utf-8') as jsonfile:
        json.dump(data, jsonfile, indent=4, ensure_ascii=False)

# 使用示例
tsv_file = 'example.tsv' # TSV文件的路径
json_file = 'output.json' # 输出JSON文件的路径
tsv_to_json(tsv_file, json_file)
```

You
我不喜欢用with open 格式来文件操作，重写一下！

ChatGPT
当然，我可以按照您的要求修改脚本，不使用`with open`来进行文件操作。请注意，虽然`with open`是推荐的方式，因为它可以自动处理文件的关闭，但是如果您不想使用它，可以直接使用`open`和`close`来手动管理文件的打开和关闭。下面是修改后的脚本：

```
python
import csv
import json

def tsv_to_json(tsv_file_path, json_file_path):
    # 打开TSV文件
    tsvfile = open(tsv_file_path, 'r', encoding='utf-8')
    reader = csv.DictReader(tsvfile, delimiter='\t')

    # 创建一个列表来保存所有行的数据
    data = []
    for row in reader:
        data.append(row)

    # 关闭TSV文件
    tsvfile.close()

    # 打开JSON文件
    jsonfile = open(json_file_path, 'w', encoding='utf-8')
    json.dump(data, jsonfile, indent=4, ensure_ascii=False)

    # 关闭JSON文件
    jsonfile.close()

# 使用示例
tsv_file = 'example.tsv' # TSV文件的路径
json_file = 'output.json' # 输出JSON文件的路径
tsv_to_json(tsv_file, json_file)
```

LLM Everywhere --- Conversation, QA, ... , MT, Coding, **Reading Paper**

ChatGPT Plugins ▾



You

"PARROT: Translating During Chat Using Large Language Models", could please summarize what this paper is about?



ChatGPT

Using txyz.ai... ✨ ▾

ChatGPT 4 ▾



You



PARROT_Translating During Chat Usin...
PDF

Please summarize the main idea of what this paper about is!



ChatGPT



LLM Everywhere --- Conversation, QA, ... , MT, Coding, Reading Paper



ChatGPT

The paper "PARROT: Translating During Chat Using Large Language Models" presents a framework designed to enhance and regulate the translation abilities of large language models (LLMs) during chat. The key aspects of the PARROT framework are:

1. **Translation Data Reformulation:** It reformulates translation data into an instruction-following style, making it compatible with chat interactions.
2. **Introduction of a "Hint" Field:** This field incorporates extra requirements or constraints to guide the translation process.
3. **Three Instruction Types for Finetuning:** These include Translation Instruction, Contrastive Instruction, and Error-Guided Instruction. Translation Instruction asks LLMs to generate translations based on source sentences. Contrastive Instruction involves generating translations of two different systems with the preferred one first. Error-Guided Instruction involves generating translations with human-annotated errors as hints.
4. **Experiments and Findings:** The PARROT models were evaluated on Flores subsets and WMT22 test sets. The findings suggest that translation instruction significantly improves the translation performance of vanilla LLMs. The Error-Guided Instruction leads to further improvement, demonstrating the importance of learning from low-quality translations annotated by humans. Additionally, PARROT models can preserve general task abilities with the involvement of the Alpaca multi-task dataset in finetuning.
5. **Future Prospects:** The paper suggests that various aspects can be explored to improve PARROT, like taking advantage of other hints like entity alignments, extending the coverage of translation directions, and exploring parameter-efficient finetuning techniques for larger LLMs.

In summary, PARROT aims to enhance the translation capabilities of LLMs in a chat setting by using human-written translation and evaluation data, with an emphasis on learning from both high-quality and low-quality translations.



You

please summarize shortly.



ChatGPT

The paper "PARROT: Translating During Chat Using Large Language Models" introduces a framework to improve translation abilities of large language models (LLMs) in chat interactions. It focuses on reformulating translation data for chat, using a "hint" field for additional translation guidance, and employing three types of instructions for finetuning the LLMs. The PARROT models show improved translation performance through this method, especially when learning from both high-quality and error-annotated translations. The paper also suggests future enhancements, like exploring different types of hints and extending translation coverage.



You

What is the main problem solved, what is the main proposed idea, what is the contribution, what are the Pros and cons & of this paper? please generate them respectively.



ChatGPT

Main Problem Solved:

The main problem addressed by this paper is the limited ability of large language models (LLMs) to perform accurate and context-sensitive translations during chat interactions. Traditional LLMs often struggle with maintaining the quality of translations in an interactive chat setting, where nuances and context

LLM Everywhere --- Conversation, QA, ... , MT, Coding, Reading Paper

ChatGPT 4 ▾

You
将以上几段的会议能不能用中文描述一下？！

ChatGPT
当然可以，以下是中文版的描述：

解决的主要问题：

本文解决的主要问题是大型语言模型（LLM）在聊天互动中进行准确和符合语境的翻译的能力有限。传统的LLM在聊天环境中经常难以保持翻译质量，因为聊天中的细微差别和上下文非常关键。

主要提出的想法：

论文提出了PARROT框架，旨在提升LLM在聊天场景中的翻译能力。其关键特性包括将翻译数据改造成适合聊天的格式，引入“提示”字段以提供额外的翻译上下文，以及使用三种类型的指令（翻译指令、对比指令和错误引导指令）对模型进行微调，以提高其性能。

贡献：

PARROT框架的主要贡献是：

1. 通过创新的数据改造和基于指令的微调方法，在聊天设置中提高了翻译性能。
2. 引入“提示”字段，为翻译添加了上下文和约束，以更有效地指导翻译。
3. 实证证明了从高质量翻译和人类标注的错误中学习，可以显著提高LLM的翻译能力。

优点和缺点：

优点：

1. **提高翻译准确性：**通过专注于聊天特定的翻译和使用不同类型的指令，PARROT提高了翻译的准确性和对上下文的敏感性。
2. **从错误中学习：**使用错误引导指令帮助模型从错误中学习，从而产生更强大的翻译输出。
3. **多功能性：**该框架具有多功能性，可以适应不同的语言对和聊天环境。

缺点：



Training Procedure of LLM (GPT)



(Prof. Huang., CCL2023)

Knowledge Representation

	信息存储方式	表示方式的精确度	知识调用方式	调用方式的自然度	研究领域	代表应用	代表公司
数据库	结构化	高	SQL语句	低	数据库	DBMS	Oracle Microsoft
互联网	非结构化	中	关键词	中	信息检索	搜索引擎	Google Microsoft
大模型	参数化	低	自然语言	高	自然语言处理	ChatGPT	OpenAI Microsoft Google

(Prof. Liu., CCL 2023)

LLM --- Image Generation (DALL-E)

You
生成一张厚重大雪下的新疆伊犁那拉提草原的照片!

ChatGPT



Here is an image depicting the Nalati Grassland in Xinjiang, Yili, under a heavy snowfall. This serene winter landscape showcases the vast, snow-covered grassland with gentle snowflakes falling from the sky, rolling hills, sparse trees, and distant snow-covered mountains under a cloudy sky.



You
不需要大雪，需要最美丽的那拉提草原

ChatGPT



Here is the image depicting the Nalati Grassland in Xinjiang, Yili, on a beautiful sunny day. This scene showcases the vibrant and colorful landscape, with lush green grass, a variety of wildflowers, clear blue skies, and the majestic Tianshan Mountains in the background.



LLM --- Image Generation (DALL-E)

You
在以上生成的图上放几匹黑马，再生成一下

ChatGPT



Here is the updated image of the Nalati Grassland in Xinjiang, Yili, now featuring **several black horses** gracefully roaming the vibrant and colorful landscape. This scene captures the peaceful and picturesque beauty of the grassland, with the majestic Tianshan Mountains in the background.



You
I do not want you to chage the background image everytime, please regenerate the picture what I need!

ChatGPT



Here is the revised image of the Nalati Grassland in Xinjiang, Yili, now featuring both black and white horses gracefully roaming across the landscape. The original background with the Tianshan Mountains remains unchanged.





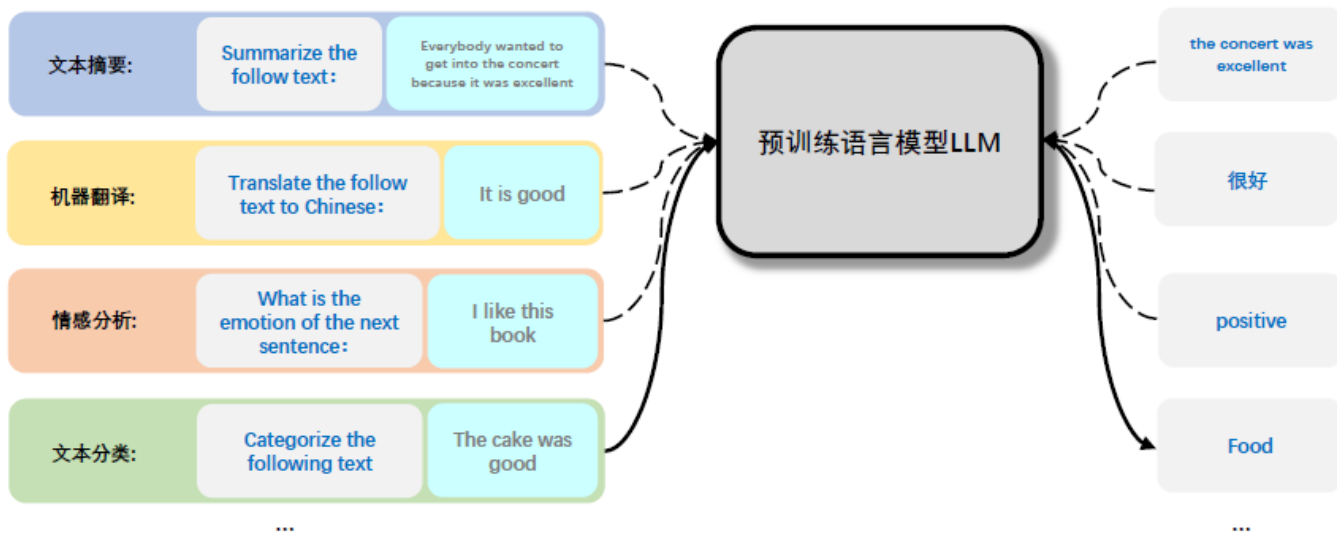
Outline

- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

Fine-Tuning Strategies & Challenges for LLM

What Is the Instruction Tuning?

- Instruct tuning指在**多个任务**上，使用自然语言的形式构建指令数据集，并对模型进行**微调**
- Instruct通常由**人类的自然语言描述**与**任务输入**拼接而成
- 根据人工标注的指令形式，可分为学术性指令和人工自然语言指令



上图为Instruct tuning示例，训练阶段（虚线）在机器翻译、情感分析等多个任务上构造指令微调LLM，推断阶段（实线）可泛化到文本分类任务上

学术性指令

Template 1

```
<premise>
Based on the paragraph
above, can we conclude that
<hypothesis>?
<options>
```

Template 2

```
<premise>
Can we infer the following?
<hypothesis>
<options>
```

Template 3

```
Read the following and
determine if the hypothesis can
be inferred from the premise:
Premise: <premise>
Hypothesis: <hypothesis>
<options>
```

Template 4, ...

人工自然语言指令

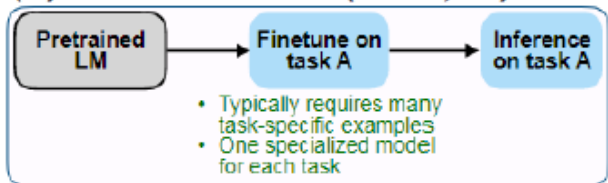
Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: "" {summary} "" This is the outline of the commercial for that play: ""

内容来源：小牛翻译-生成式LLM系列直播，5期，刘新宇

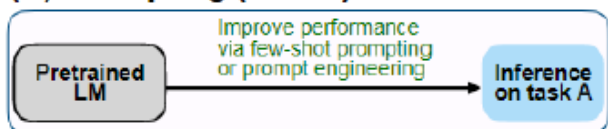
Why Instruction Tuning?

- 使大模型的输入与人类习惯“对齐”
- 指令微调可以提高模型的指令理解能力，使得模型能根据输入的Instruction**自动识别任务类型**，并执行对应的操作
- Instruct tuning区别于fine-tuning和prompting，需要标注数据更少，可在未见任务上**指令泛化**

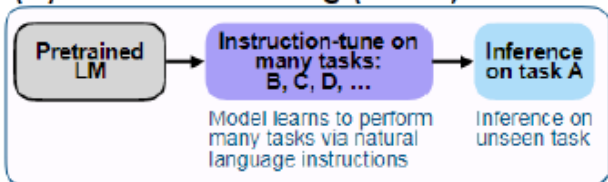
(A) Pretrain-finetune (BERT, T5)



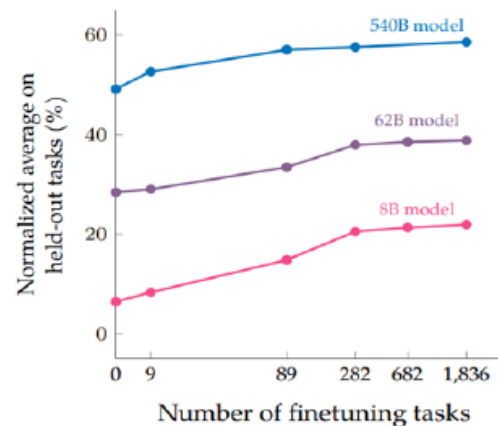
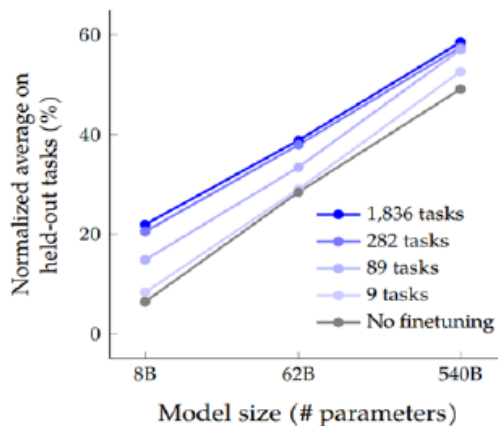
(B) Prompting (GPT-3)



(C) Instruction tuning (FLAN)



研究表明，当模型**规模越大**，微调时见过的指令**任务种类越多**，其泛化能力越强。



Flan-PaLM模型在不同scale下进行instruct tuning性能对比

Instruction tuning 与 **fine-tuning** 和 **prompt** 对比

内容来源：小牛翻译-生成式LLM系列直播, 5期, 刘新宇

What Is the RLHF (Reinforcement Learning with Human Feedback)?

- RLHF是一种**基于人类反馈**的强化学习方法，使用基于Agent的强化学习框架进行模型参数的更新
- 通过交互界面进行打分，使得模型可以利用人类反馈进行学习
- **评分模型**训好后，可与**生成模型**形成迭代闭环，完成LLM的训练

PPO强化学习方法架构图



内容来源：小牛翻译-生成式LLM系列直播, 5期, 刘新宇

Why RLHF?

- 使大模型的输出与人类期望“对齐”
- RLHF可以提高模型的鲁棒性与安全性，使得生成结果更加遵循指令和事实，并大幅度减少有毒输出
- RLHF使得模型具备了Say no的能力

Prompt: Explain the moon landing to a 6 year old in a few sentences.

Completion: GPT-3

Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.

InstructGPT

People went to the moon, and they took pictures of what they saw, and sent them back to the earth so we could all see them.

使用了RLHF的InstructGPT比GPT-3更加遵循指令

XM 你知道2023年中国春节联欢晚会的节目单吗?

很抱歉，我无法预测未来的事件和信息。2023年中国春节联欢晚会的节目单还没有发布，一般会在春节前夕或者节目录制前公布。您可以关注相关的新闻和娱乐网站，以获取最新的节目单信息。

可以拒绝回答

Dataset	Model	Score
RealToxicity	GPT	0.233
	Supervised Fine-Tuning	0.199
	InstructGPT	0.196
TruthfulQA	GPT	0.224
	Supervised Fine-Tuning	0.206
	InstructGPT	0.413
API Dataset Hallucinations	GPT	0.414
	Supervised Fine-Tuning	0.078
	InstructGPT	0.172
API Dataset Customer Assistant Appropriate	GPT	0.811
	Supervised Fine-Tuning	0.880
	InstructGPT	0.902

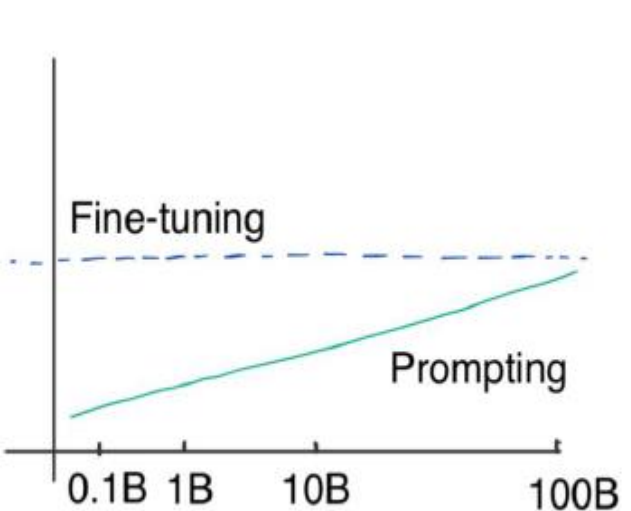
使用了RLHF的InstructGPT相对于GPT-3产生的谎言更少(TruthfulQA)，毒性更少(RealToxicity)，同时不会频繁的编造事实(hallucinates)

内容来源：小牛翻译-生成式LLM系列直播，5期，刘新宇

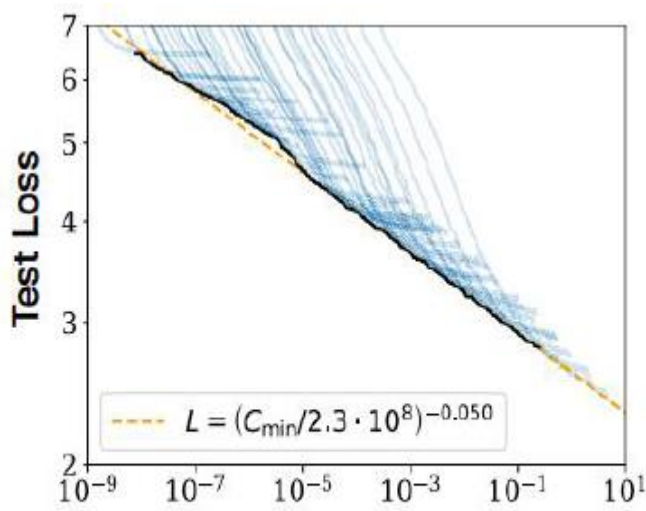
What Is the Scaling Law?

- 语言模型的Scaling law是一条log-linear曲线，指模型性能随着**模型规模呈指数级增长**而线性增加
- 模型规模包括**计算规模**、**数据规模**以及**参数规模**

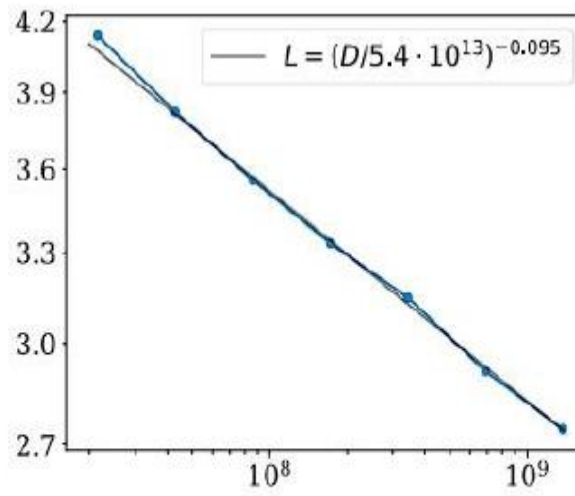
Scaling law的恐怖之处在于，模型的性能似乎存在一个与规模有关的上界。模型被“锁定”在相应的标度内，无法明显的突破。



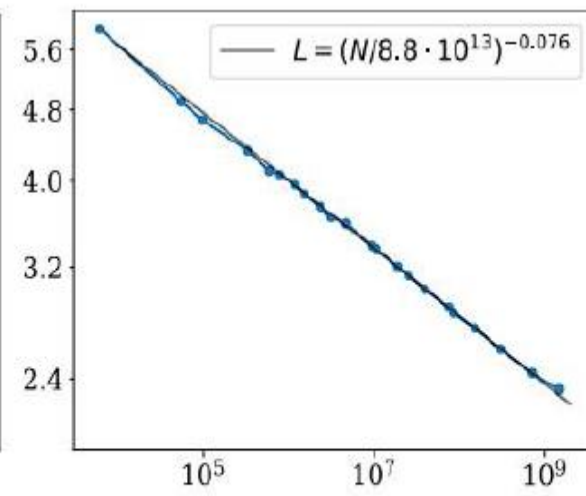
Scaling Law (~2020 - 2021)



Compute
PF-days, non-embedding



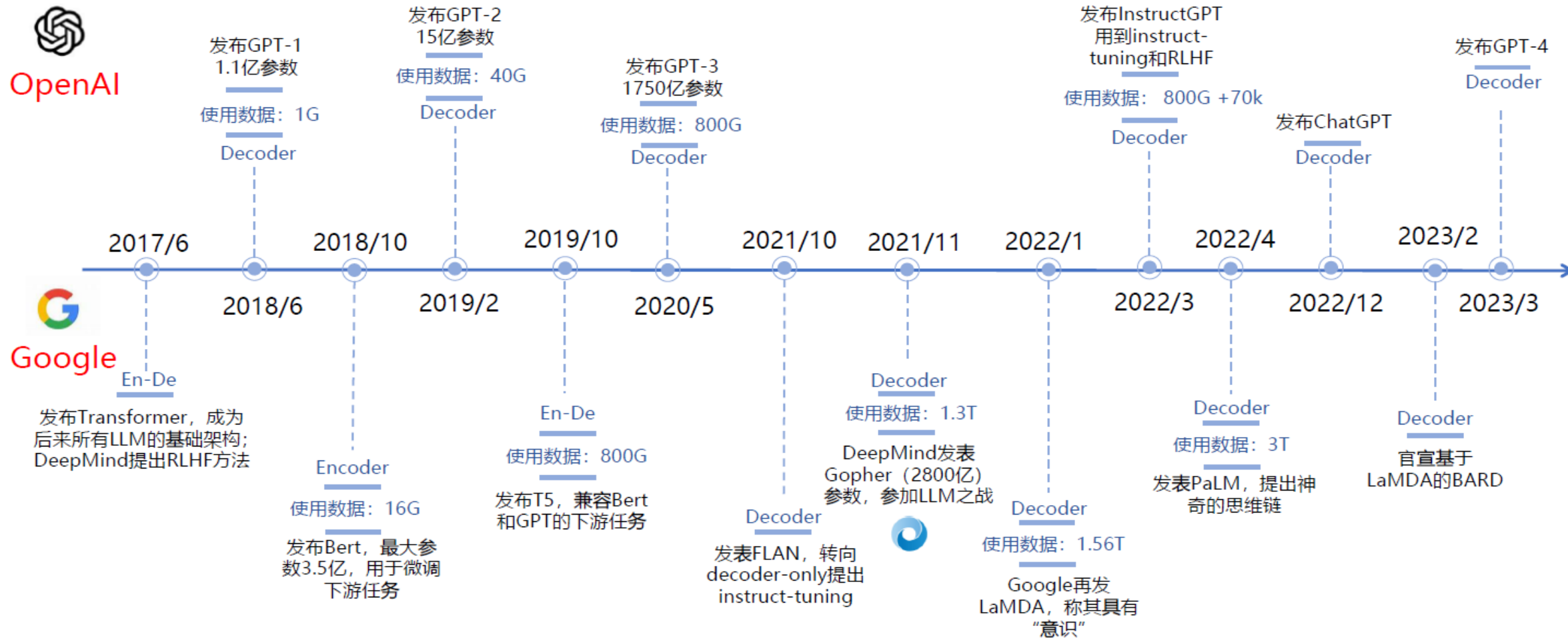
Dataset Size
tokens



Parameters
non-embedding

内容来源：小牛翻译-生成式LLM系列直播, 5期, 刘新

Making the LLM getting Larger

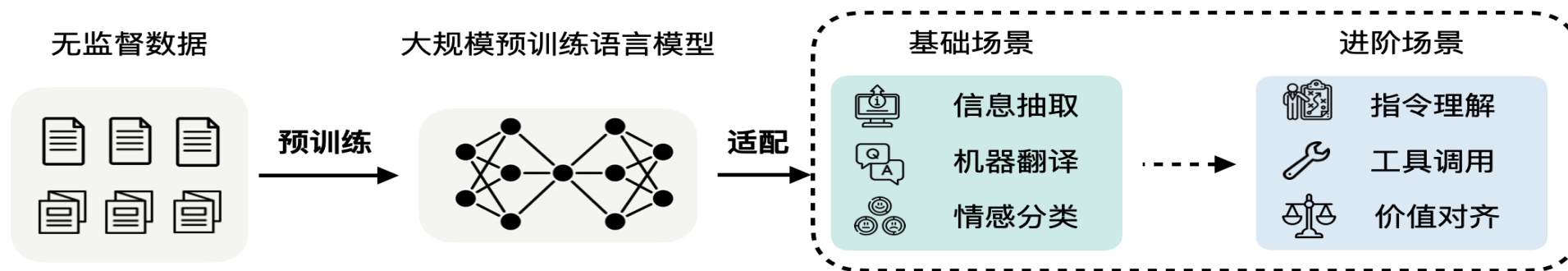


Indications of Scaling Law?

- **Scaling Law** 昭示了更大的计算量一定会带来更好的效果 (更低的**预训练损失**)

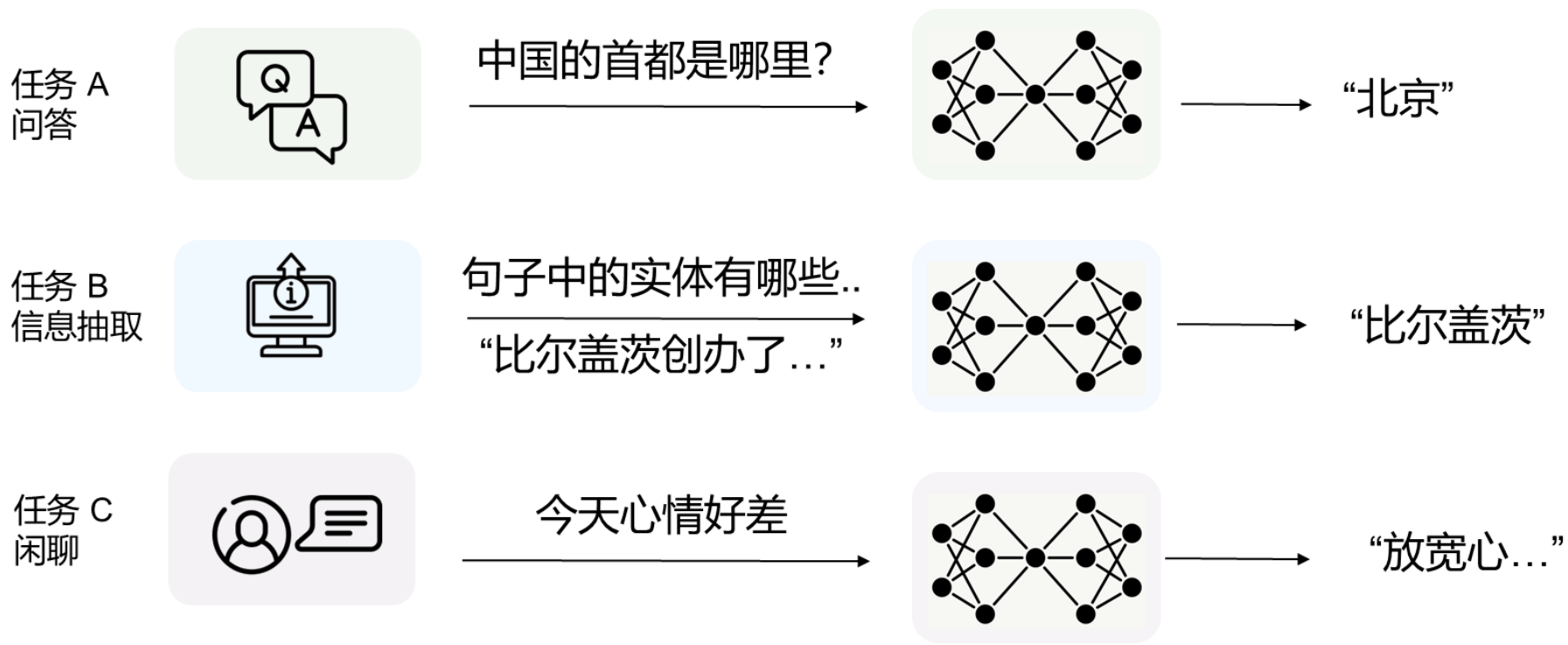


- 更大的计算量也会带来更好的**下游任务表现**



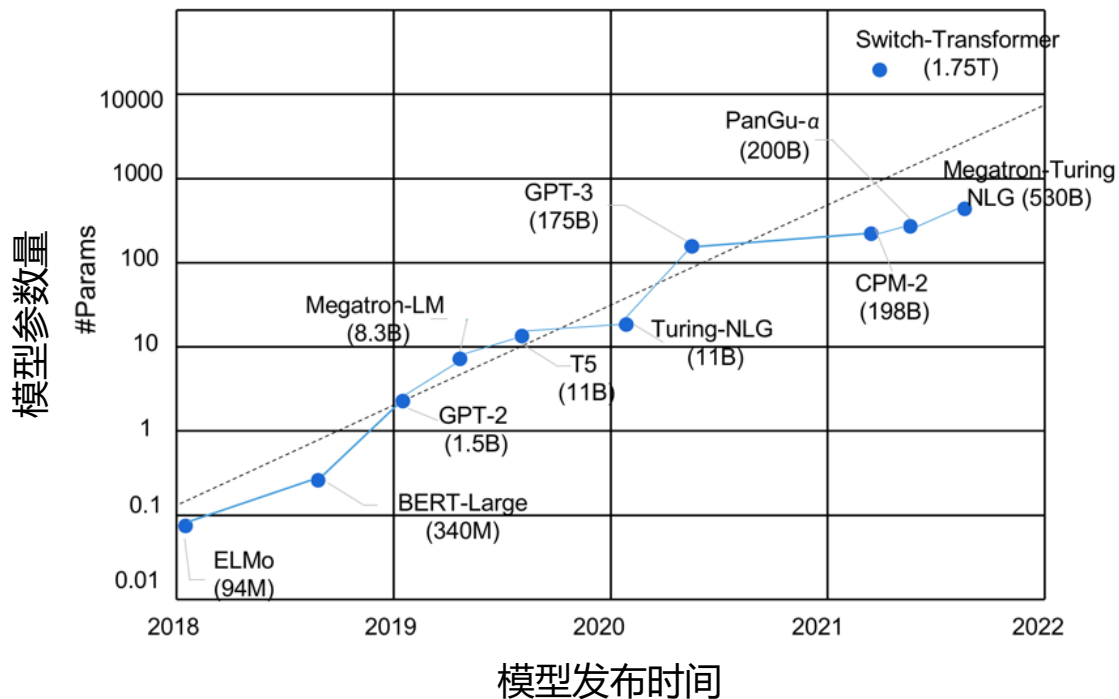
Big Model Fine-tuning Task

- BERT时代的模型微调：对预训练模型在**特定任务的全量数据**上进行**全参数微调**
- **缺少多任务之间的泛化能力**

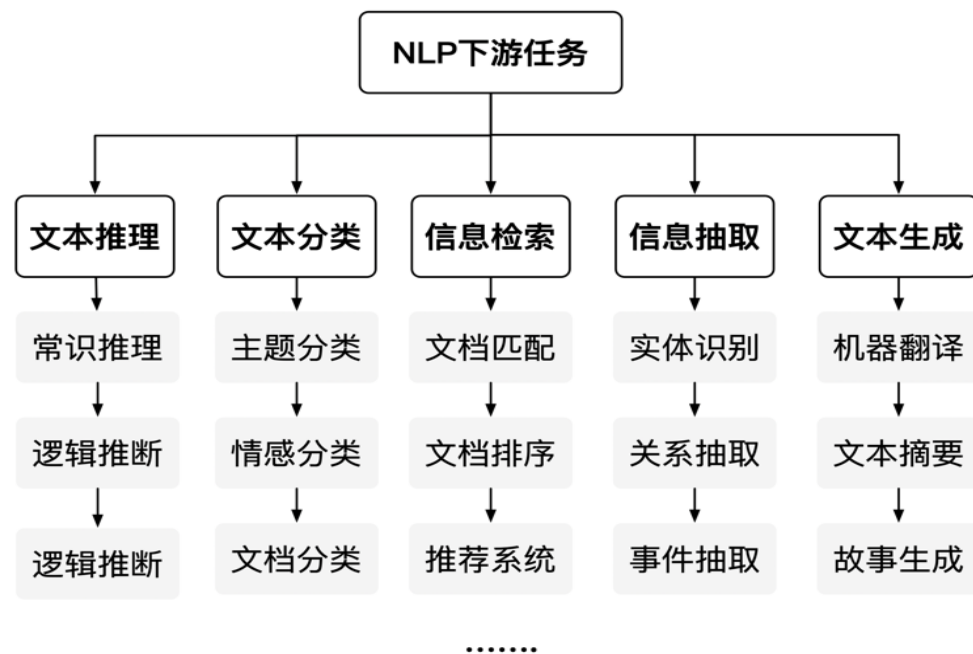


Main Challenge --- Difficult to fine-tune LLM

- 大模型在微调中更新所有参数，需要在**计算**中占用庞大的显存
- 大模型在微调中对每一个任务存储一个大模型，需要占用庞大的**存储空间**



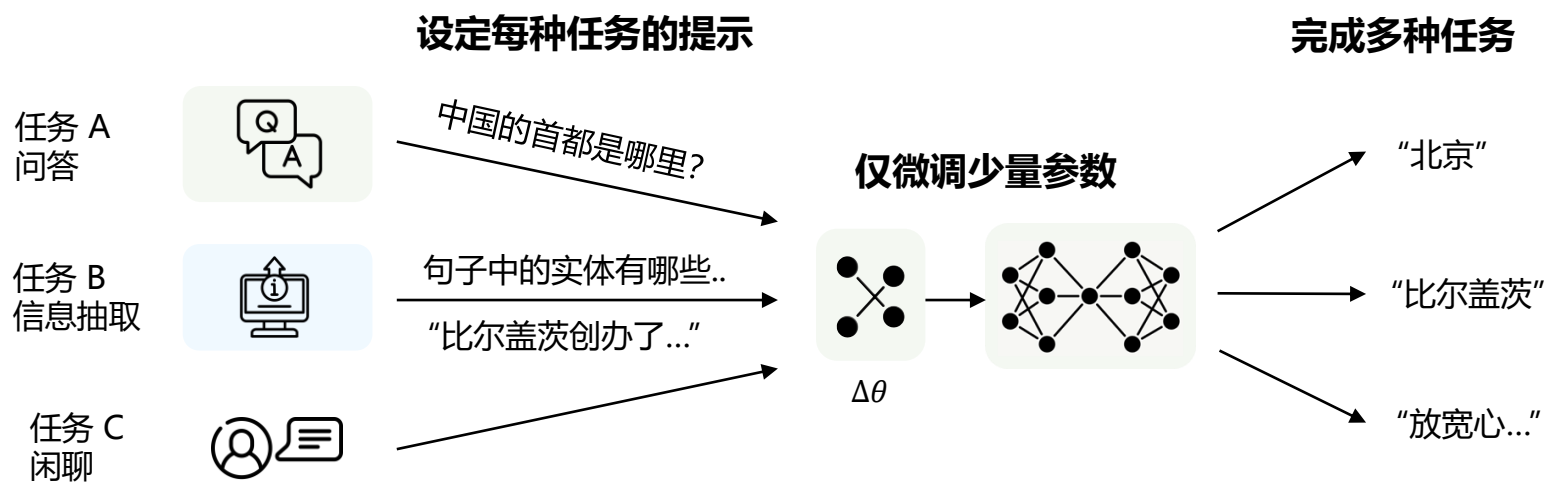
模型的参数规模越来越大，使之更加难以微调



任务种类繁多，难以对所有任务都进行全参数微调

Key Points for Fine-tuning the LLM

- 大模型时代的模型微调：具备少数据、少参数、强任务的泛化能力
 - 少量数据上的**提示微调**
 - 少量参数上的**增量微调**
 - 多种任务上的**指令微调**



Fine-tuning the LLM

- 微调是大模型适配到各种场景的关键技术
- **数据高效**：使用提示微调组织训练和任务格式
 - 统一预训练与下游任务形式
 - 仅需少量样本就可适配新任务
- **参数高效**：使用增量微调激活大模型能力
 - 计算 & 存储高效
 - 仅需少量参数就可与全参数微调相当
- **任务泛化**：使用指令微调打破任务壁垒
 - 极大增强了大模型的意图理解能力
 - 构建ChatGPT的关键一步



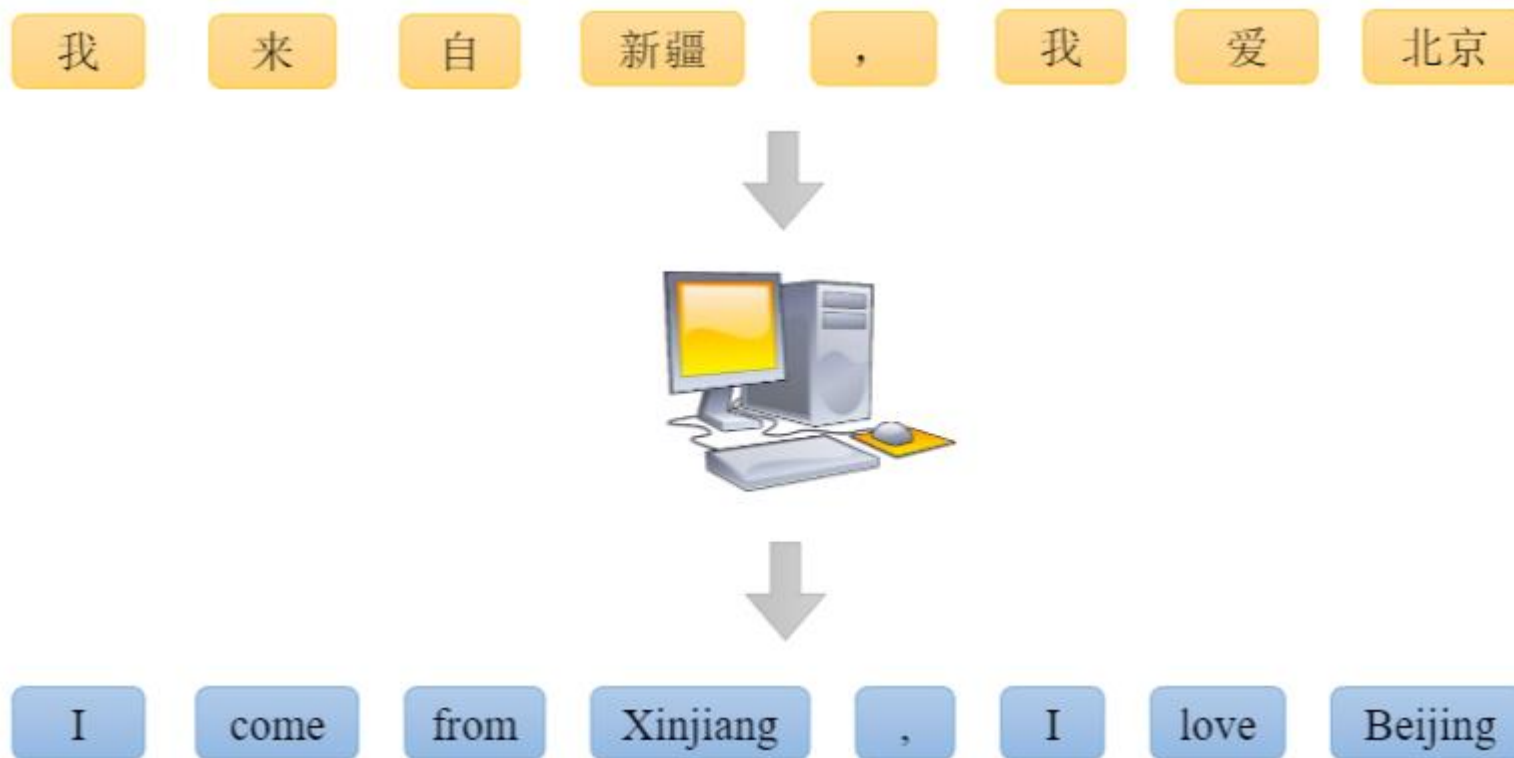
Outline

- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

Machine Translation with LLM

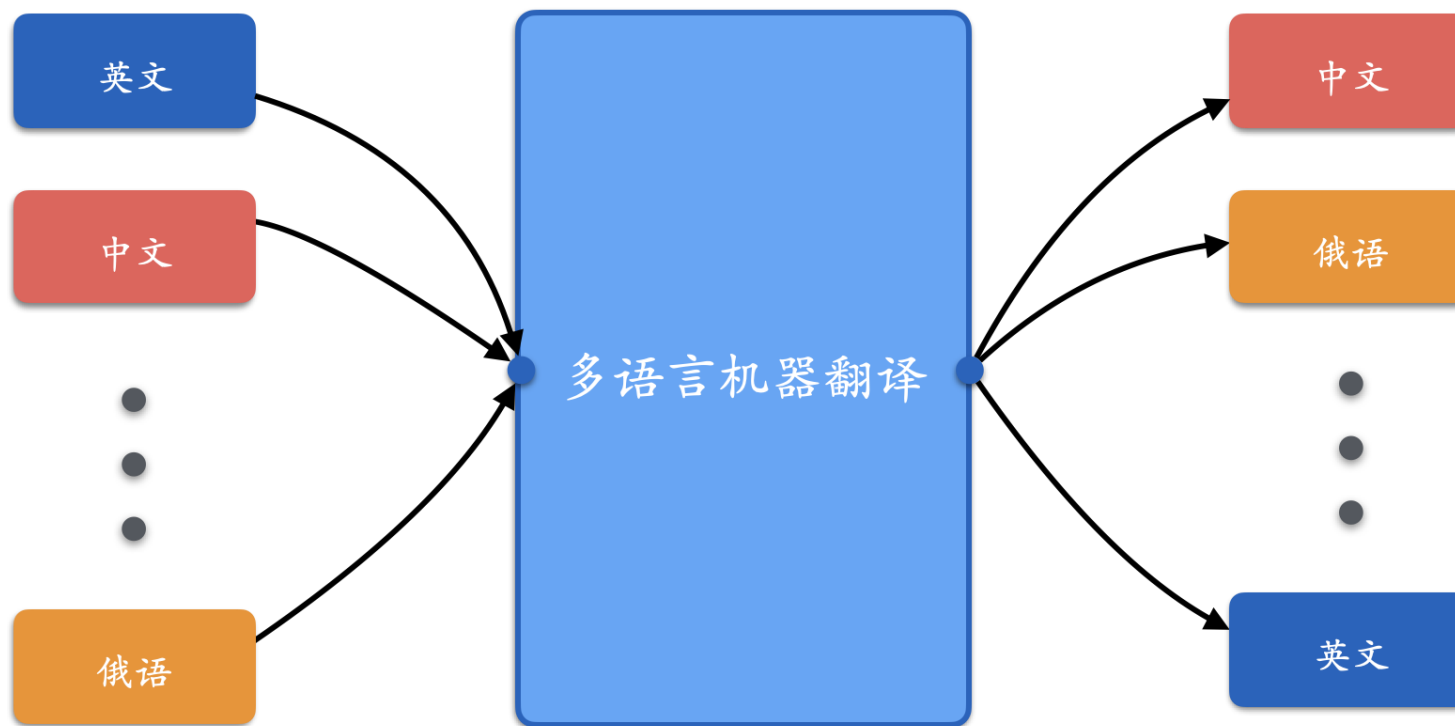
Machine Translation

- 用计算机将一种序列转化为另一种序列



Multi-lingual Machine Translation

- **多语言机器翻译**旨在使用一个统一的翻译模型实现**多种语言之间的互译**，从而显著提升低资源语言的翻译性能，大幅减少部署成本。



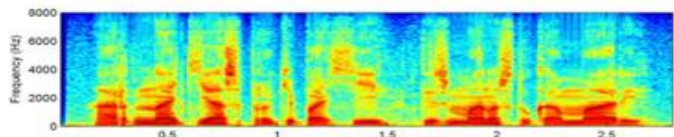
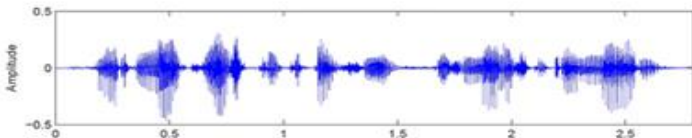
Multi-Modal Translation --- Speech

我 来自 中国



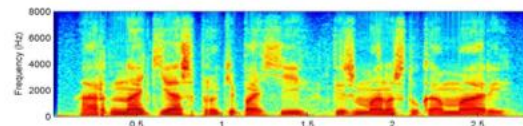
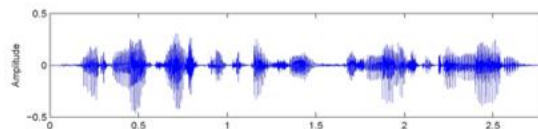
MT

I am from China

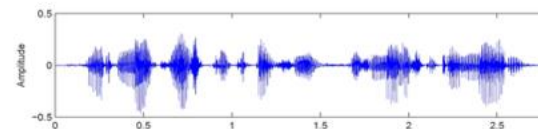


I am from China

ASR TTS



I am from China



9	16	52	...	84
...
71	12	38	...	67
12	43	8	...	59

I am from China

ASR TTS

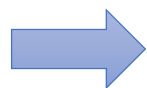
Speech Processing vs NLP

From Continuous Signals to Discrete Tokens

Multi-Modal Translation --- Image



Image Captioning
Model

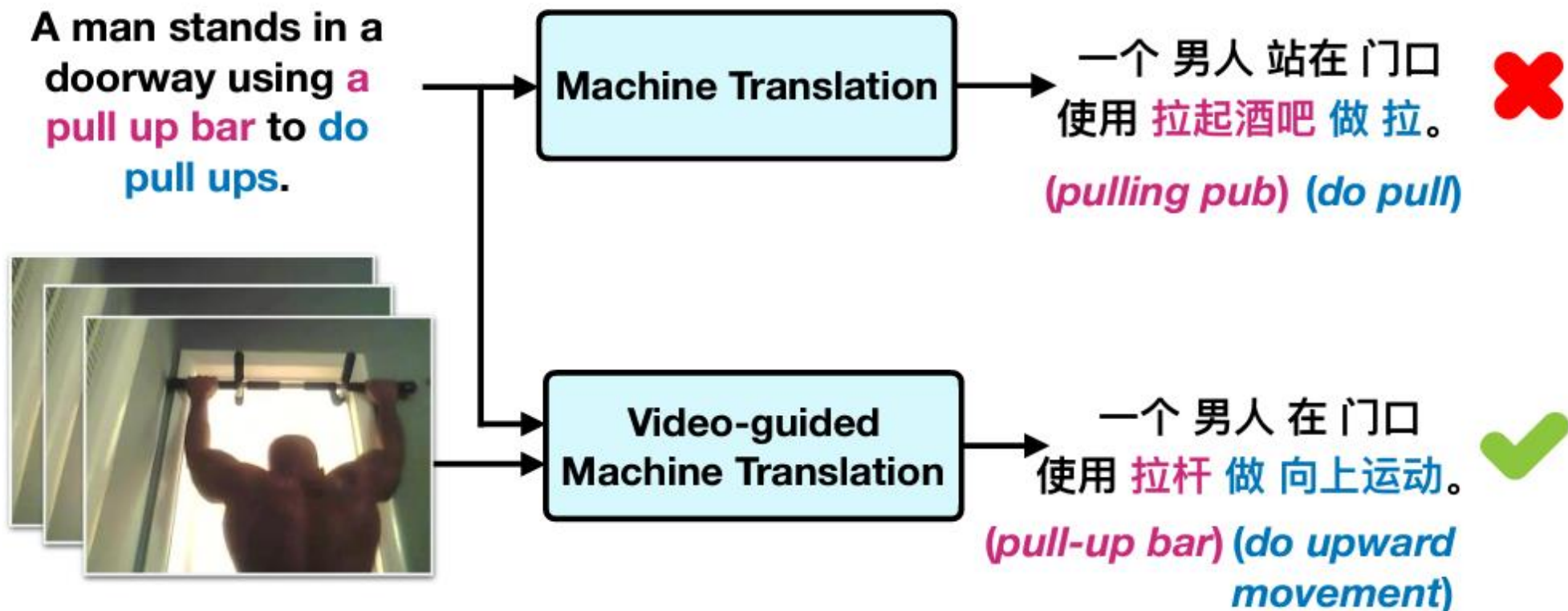


There is a teacher teaching his
students in the classroom.



有一个老师在教室里教他的学生。

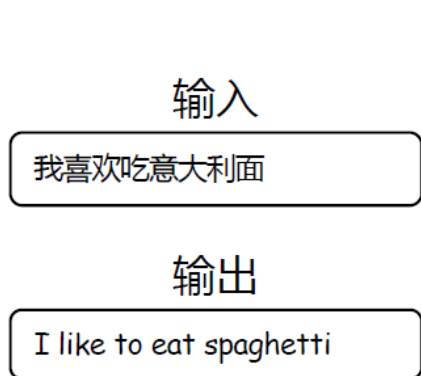
Multi-Modal Translation --- Video



Machine Translation with LLM

• 通用模型 v.s. 执行特定任务 (翻译)

- Instruction Following 通过指令指定模型行为
- In-context Learning (ICL) 从上下文中学习

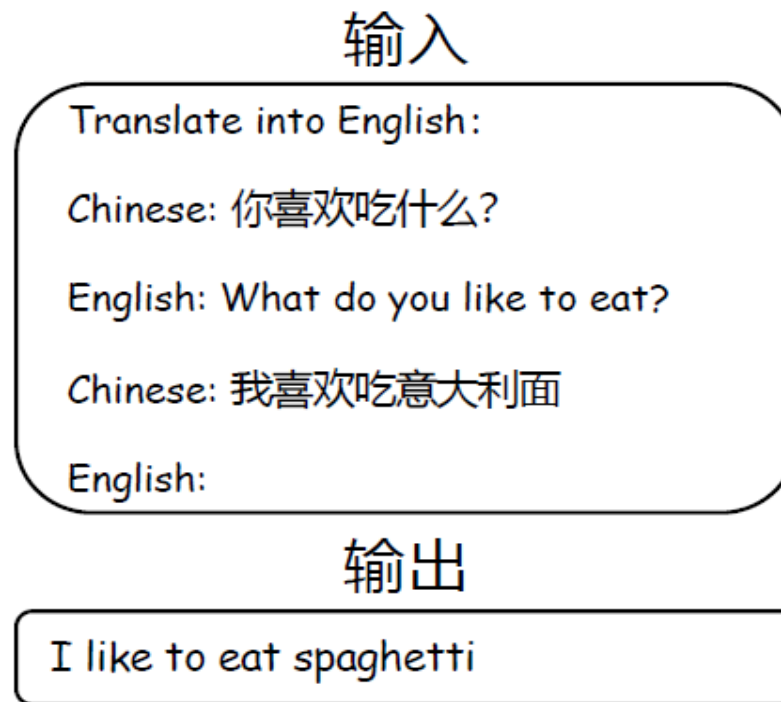


Seq2Seq



Instruction

通过指令学习获得
直接完成指令



In-context Learning

预训练中
获得少量的
示例

Machine Translation with LLM

• 翻译（多语言）能力还可以表现在更多方面：

- 理解不同语言的输入
- 生成不同语言的回复
-

You
“我喜欢新疆北疆，景色非常迷人，特别美丽，不过从没去过。” \n\n 麻烦把以上这几句话翻译成英语，德语，法语，阿拉伯语，土耳其语，泰语等几个多语种。

ChatGPT
当然，以下是您提供的句子翻译成不同语言版本：

英语:
"I really like the northern part of Xinjiang. The scenery is very enchanting and incredibly beautiful, though I've never been there."

德语:
"Ich mag den Norden von Xinjiang sehr. Die Landschaft ist sehr bezaubernd und unglaublich schön, obwohl ich noch nie dort war."

法语:
"J'aime beaucoup le nord du Xinjiang. Le paysage est très charmant et incroyablement beau, bien que je n'y sois jamais allé."

阿拉伯语:
"أحب شمال شينجيانغ كثيراً. المناظر الطبيعية ساحرة وجميلة للغاية، على الرغم من أنني لم أذهب إلى هناك قط."

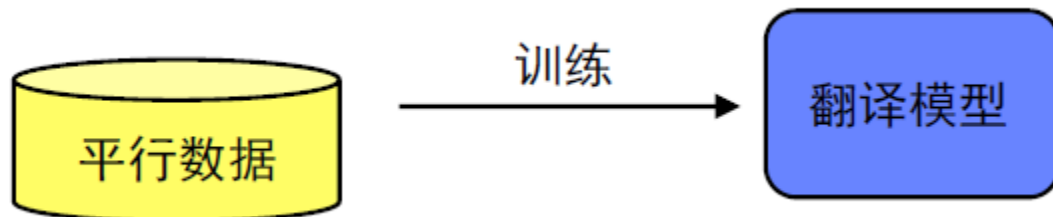
土耳其语:
"Xinjiang'ın kuzeyini çok seviyorum. Manzara çok büyüleyici ve inanılmaz güzel, ama hiç oraya gitmedim."

泰语:
"ฉันชอบภาคเหนือของซินเจียงมาก ทิวทัศน์ที่นั่นมีเสน่ห์และสวยงามมาก แม้ว่าฉันจะไม่เคยไปที่นั่น."

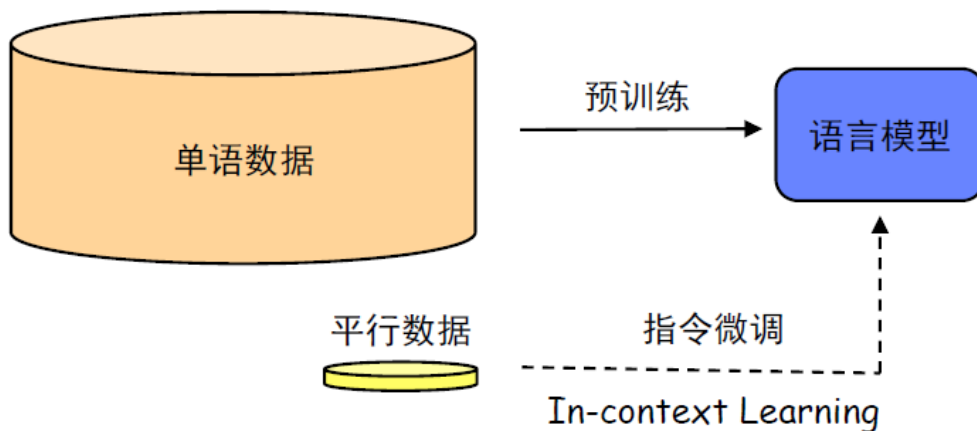
这些翻译应该准确传达了原文的意思和信息 ↓

The paradigm of learning is shifting

- **神经机器翻译: 主要从平行数据中学习翻译知识**



- **大语言模型: 主要从单语数据中学习通用知识 (包含翻译知识)**



New research questions from new learning paradigms

- **当前状况评估**

- 大语言模型翻译表现的好不好?
- 不同的表现方式有什么区别?

- **探索模型潜力**

- 大语言模型翻译潜力如何?
- 如何能够激发模型的翻译能力?

- **产生更大价值**

- 模型如何在不同语言上展现能力?

当前状况评估



探索模型潜力



产生更大价值

Understanding the translation capability of LLM

- 经过预训练的模型大多具有ICL能力
- 可通过ICL进行翻译任务从而进行评估

输入

Translate into English:

Chinese: 你喜欢吃什么?

English: What do you like to eat?

Chinese: 我喜欢吃意大利面

English:

输出

I like to eat spaghetti

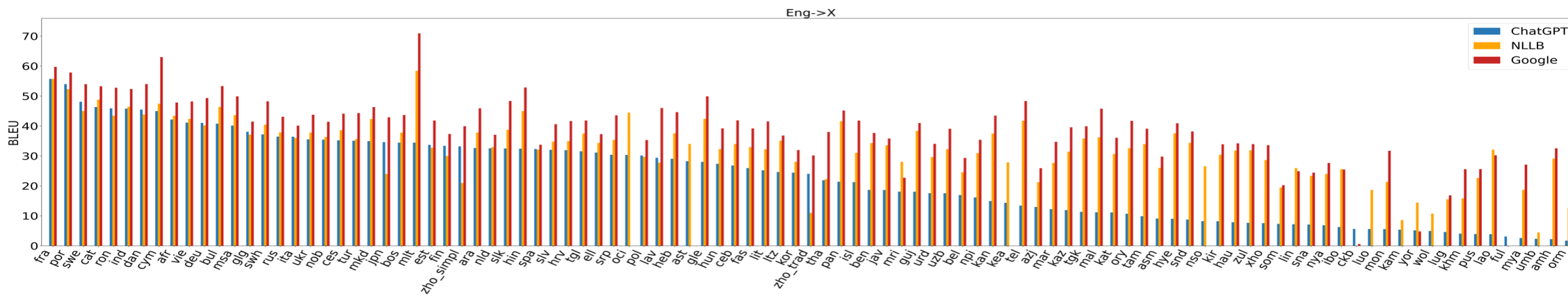
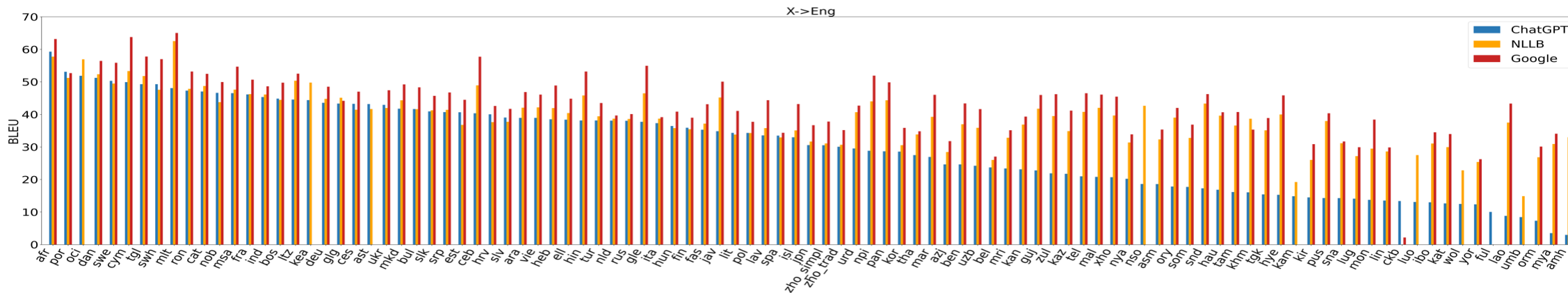
In-context Learning

预训练中获得
少量的示例

LLM MT vs NMT



- **大语言模型仍然落后于强大有监督基线模型：ChatGPT (16.67%) v.s. NLLB (83.33%)**
- **尤其在低资源语言上远远落后于商用翻译系统Google Translate**



Privileges of LLM MT

1. 语境理解能力

提升翻译的准确性和流畅程度

2. 交互式翻译功能

用户能够为翻译提供实时反馈或指导

3. 多功能文本处理

能够翻译正式文本、口语和俚语

4. 文档级翻译功能

在长文本和复杂文档中保持一致性和语境信息

5. 多语言功能

能同时进行多种语言的翻译

...

语境

GPT-4



AD

Please translate the following sentence into Chinese:
During lunch, bank employee John told Mary they'd meet at the river bank to celebrate her promotion. "Break a leg with your presentation," he said.



在午餐期间，银行员工约翰告诉玛丽他们会在河岸见面，以庆祝她的晋升。"祝你的演讲一切顺利，"他说。

交互式

AD

Please translate it into English:

大型语言模型已被证明在使用少量样本学习的情况下，能在各种自然语言任务中表现出色。为了进一步了解规模对少量样本学习的影响，我们训练了一个具有5400亿参数、密集激活的Transformer语言模型，我们称之为Pathways语言模型。



The translation of the provided text into English is:

"Large language models have been shown to perform well in various natural language tasks with limited sample learning. To further understand the impact of scale on limited sample learning, we trained a Transformer language model with 540 billion parameters and dense activations, which we call the Pathways language model."

多功能文本处理

AD

Please translate "远水救不了近火" into English.

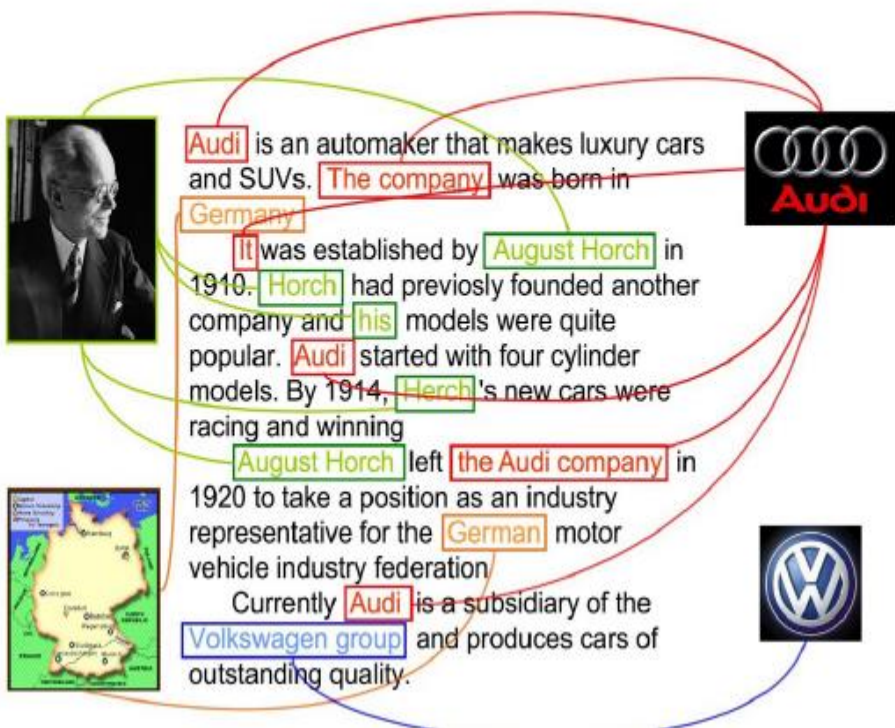


The Chinese phrase "远水救不了近火" literally translates to "distant water cannot put out a nearby fire." This idiom is used to express the idea that help that is too far away is of no use in an emergency or urgent situation. In other words, when you're facing an immediate problem, solutions that are too far off in time or space won't be of any help.

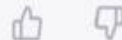
文档级别翻译



Translate this document from English to Chinese:



奥迪是一家生产豪华汽车和SUV的汽车制造商。该公司起源于德国。它由奥古斯特·霍希（August Horch）于1910年创立。霍希此前曾创立过另一家公司，其车型相当受欢迎。奥迪起初生产四缸车型。到1914年，霍希的新车开始参加比赛并获胜。奥古斯特·霍希于1920年离开奥迪公司，担任德国机动车行业联合会的行业代表。目前，奥迪是大众集团的子公司，生产质量卓越的汽车。



Cons of LLM MT --- Drawbacks

1. 缺乏领域内的专业知识，难以进行知识迁移

无法准确捕捉特定领域的术语或细节

在适配新数据时会出现严重的灾难性遗忘

2. 在低资源语言上翻译性能欠佳

少见语言对的翻译效果仍有提升空间

3. 自动评估遇到瓶颈

翻译的质量太高，传统的评估指标能力有限

4. 额外的算力开销

需要大量的计算资源来进行监督微调和适配

...

LLM MT --- New paradigm

机器翻译：可以被视作一种具有代表性的大模型微调任务

关键研究技术在机器翻译领域的体现

大模型场景

大模型垂域适配



实时自适应: Test-time Adaptation for Machine Translation Evaluation by Uncertainty Minimization (ACL 2023 Oral)

大模型多语性



语言相似性: Clustering Pseudo Language Family in Multilingual Translation Models with Fisher Information Matrix (EMNLP 2023)

大模型检索鲁棒性



模版一致性: TemplateGEC: Improving Grammatical Error Correction with Detection Template (ACL 2023 Oral)

大模型能力评估



常识推理能力: Revisiting Commonsense Reasoning in Machine Translation: Training, Evaluation and Challenge (ACL 2023)



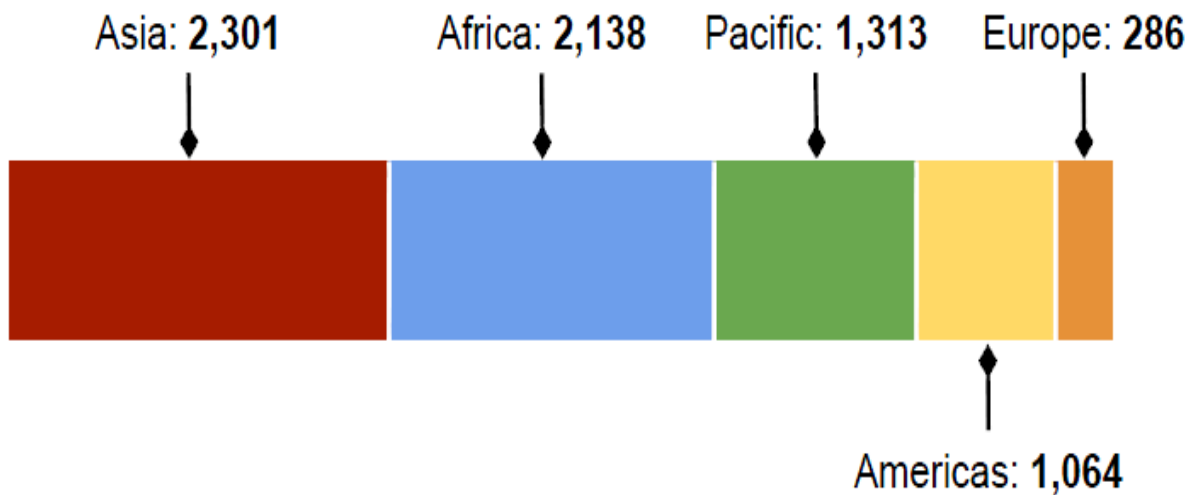
Outline

- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

LRLs Machine Translation with LLM

Low-Resource Languages (LRLs)

- 学术界与企业界都关心的热点。
- 建设“一带一路”工作中跨语言的顺畅沟通成为越发迫切的需求。
- “一带一路”沿线国家特别是东盟、中亚国家，这些国家的语言几乎都是属于低资源语言。



图片来源：Washington Post Article mentioning “Ethnologue language of the world”, 8th ed



图片来源：<http://www.mrcjcn.com/n/224527.html>

LRLs MT

- 目前以机器翻译为代表的跨语言任务一大挑战：低资源语言翻译任务
 - 由于低资源语料的限制，模型常在小语种上表现较差
 - 数据增强、平行语料挖掘以及多语言模型为代表的多种方法被应用
- LLM与传统NMT的差异
 - LLM多数使用Decoder-only的结构，而传统NMT常使用Encoder-Decoder结构
 - LLM的训练语料以单语为主，而传统NMT需要大量平行语料
- LLM低资源翻译语言以及多语言翻译能力
 - 多数LLM是以英文为中心的LLM
 - 低资源语种和极低资源语种训练数据少、翻译表现差
 - 与目前的SOTA翻译系统存在较大差距

Main topics for LRLs MT with LLM

- 基于LLM的翻译表现评估
- LLM低资源翻译提示方法
- LLM低资源翻译训练方法
- 小结

Sample



“我来自我国最西部的地理位置上比较偏僻的，经济不发达的，教育水平和条件很低的，以及医疗和医学落后的新疆，但是我一直以来坚持不懈地努力而向往美好的未来。”这句话翻译成维吾尔语。



مەن مەملىكىتىمىزنىڭ ئەلتە تەرەپتىكى، ئېكېمى يېتەكچىلىك ۋە ئېكونومىكىسى ئېشەلمەپتىكى، تەلىمەن تەڭلىما ۋە ھالەتى تۆۋەن بولغان شىنجاڭدىن كېلىدىم، بۇنداق تىلەن مەن داۋاملىق نەسەمسىزلىك بىلەن كۆتۈرۈشنى ئىككىلەپ، تورۇپ كۆتۈكچىلىك كۈندەلىك نىڭ چەكسىزلىقىغا ئېغىزلىنىپ چىكىپ، ئالدىنقى تەسەمسىزلىكتىكى ئېلىم بىلەن مۇتلەق ھەمكار بولۇپ تونۇمدىن يۇقىرى قارايداشقان ئۇنتۇقى ئاۋازىنىڭ ئىچكىمىگە ئاڭلىدۇرماققا سۈردۈم.



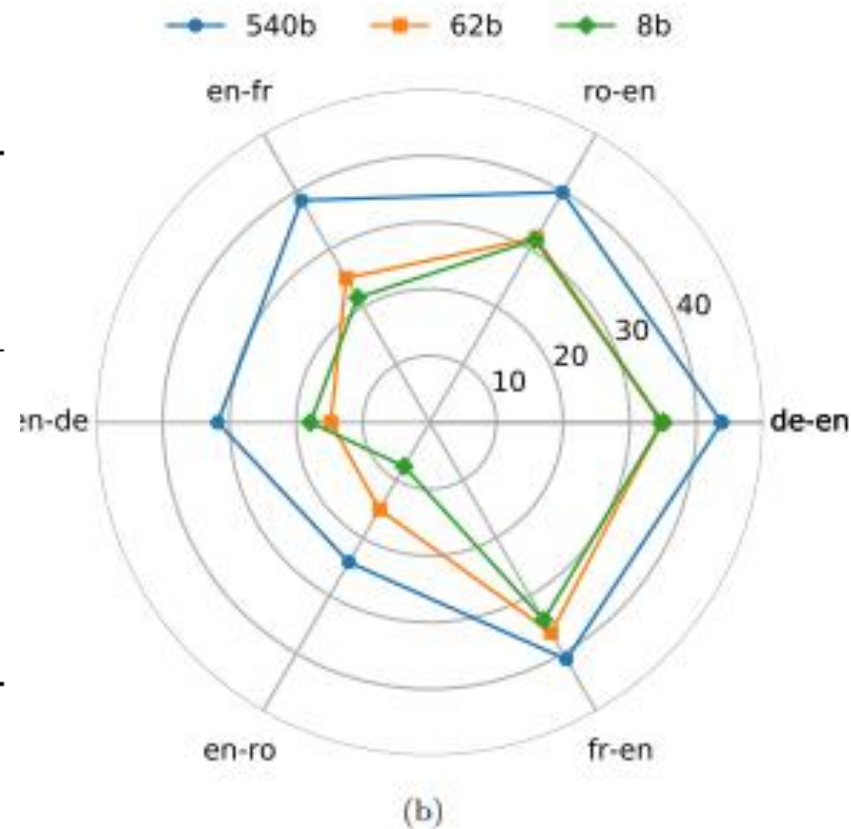
Evaluation --- HRLs

- 高、低资源MT性能表现
- LLM的机器翻译高资源语言对表现

Src	Tgt	0-shot		1-shot		Few-shot		Supervised
		Prior SOTA	PaLM 540B	Prior SOTA	PaLM 540B	Prior SOTA	PaLM 540B	Finetuned SOTA
en	fr	32.9 ^a	38.5	28.3 ^b	37.5	33.9 ^a (9)	44.0	<u>45.6^c</u>
en	de	25.4 ^a	31.8	26.2 ^b	31.8	26.8 ^a (11)	37.4	<u>41.2^d</u>
en	ro	16.7 ^a	24.2	20.6 ^b	28.2	20.5 ^a (9)	28.7	<u>33.4^e</u>
fr	en	35.5 ^a	41.1	33.7 ^b	37.4	38.0 ^a (9)	42.8	<u>45.4^f</u>
de	en	38.9 ^a	43.8	30.4 ^b	43.9	40.6 ^a (11)	47.5	41.2 ^g
ro	en	36.8 ^a	39.9	38.6 ^b	42.1	37.3 ^a (9)	43.8	39.1 ^h

- PaLM目前已经在所有few-shot模型方法上达到了SOTA
- 部分语言对的表现已经超越Supervised Model
- Model的size大小对于翻译的性能有较大差异

0-shot BLEU Scores for PaLM Model Scales



Evaluation --- HRLs

- 其他语言到英语的方向表现
 - Zero-Shot表现已经超越或者匹配WMT和商用翻译系统
 - Few-Shot不会带来可见的提升
- 英语到其他语言的方向表现
 - 在Token级别的Metric会有表现退化的现象
 - Few-Shot会带来一定的提升
 - Few-Shot中样例质量是带来提升的关键因素

System	COMET-22	COMETkiwi	ChrF	BLEU	COMET-22	COMETkiwi	ChrF	BLEU
	DE-EN				EN-DE			
WMT-Best	85.0	81.4	58.5	33.4	87.2	83.6	64.6	38.4
MS-Translator	84.7	81.0	58.5	33.5	86.8	83.4	64.2	37.3
GPT Zeroshot	84.8	81.2	56.8	30.9	85.6	82.8	60.2	31.8
GPT 1-Shot RR	84.9	81.3	56.1	30.4	86.1	83.0	60.7	31.9
GPT 1-Shot QR	84.9	81.3	56.7	31.1	85.8	82.8	60.7	32.4
GPT 5-Shot RR	85.2	81.5	56.5	31.2	86.5*	83.2 *	61.0	32.4
GPT 5-Shot QR	85.4*	81.5*	57.7	32.4	86.4	83.1	61.3*	33.2*
GPT 5-Shot QS	85.0	81.3	57.8*	32.5*	85.9	82.9	60.8	32.7

Evaluation --- LRLs

LLM的机器翻译低资源语言对和非英语语言对表现

Src	Tgt	0-shot	1-shot	Few-shot	Supervised
		PaLM 540B	PaLM 540B	PaLM 540B	Finetuned SOTA
en	kk	1.8	4.2	5.1	15.5^a
de	fr	28.6	20.9	25.7	31.5^b
kk	en	18.0	20.3	20.8	30.5^c
fr	de	25.2	9.5	17.4	24.9 ^b

- 非英语语言对中, PaLM比较擅长目标语言为德语的翻译
- 低资源语言翻译时, 大模型更加擅长其他语言到英语方向的翻译
- 低资源语言翻译表现和当前的SOTA翻译模型差距较大

Evaluation --- LRLs

LLM的机器翻译低资源语言对和非英语语言对表现 -- GPT

System	COMET-22	COMETkiwi	ChrF	BLEU	COMET-22	COMETkiwi	ChrF	BLEU
	IS-EN				EN-IS			
WMT-Best	87.0	81.4	62.3	41.7	86.8	81.8	59.6	33.3
MS-Translator	85.9	80.3	62.8	40.5	84.3	80.2	56.8	28.7
GPT Zeroshot	82.1	78.7	55.6	31.9	76.3	74.0	43.5	15.9
GPT 1-Shot RR	84.1	80.2	57.8	34.7	77.0	74.6	43.7	15.3
GPT 1-Shot QR	83.5	79.7	56.7	33.3	77.4	75.1	44.5	16.2
GPT 5-Shot RR	84.4*	80.4*	58.1*	35.0*	77.9*	75.2*	45.1*	16.8*
GPT 5-Shot QR	84.2	80.2	58.0	35.2	76.0	74.1	44.1	16.3
	HA-EN				EN-HA			
WMT-Best	80.0	74.5	48.7	21.0	79.8	61.5	51.1	20.1
MS-Translator	73.3	68.5	43.4	16.2	72.5	57.2	38.4	10.3
GPT Zeroshot	76.1	73.1	45.5	17.3	73.3	58.6	38.4*	9.4*
GPT 1-Shot RR	75.7	72.7	45.7	17.3	74.0	59.0	38.4*	8.8
GPT 1-Shot QR	78.1	74.4	47.5*	19.1*	74.1*	59.7*	37.8	8.9
GPT 5-Shot RR	75.5	72.2	45.9	17.8	72.1	57.7	36.0	8.0
GPT 5-Shot QR	78.2*	74.5*	47.5*	18.9	72.6	58.5	36.9	8.5

- LLM的低资源翻译表现和目前的有监督模型和商用模型有较大差距
- Few-shot会产生特定的收益，在翻译方向为其他语言翻译到英语时，收益更加明显
- Few-Shot时，示例质量十分关键

Evaluation --- Instruction

- LLM的机器翻译提示策略(Prompt Strategy)评估
 - 模版(Template)
 - 语种选择: 英语模版会带来更好的翻译表现
 - 具体形式: 不同形式对于表现差异较大
 - Few-shot的示例(Demonstration):
 - 示例选择方法: 翻译质量、相关性、随机采样
 - 示例翻译质量十分重要: 高质量样例会为In-Context Learning提供一个“更加密集的信号”, 其影响力大于相关性

Instruction for LRLs --- DiPMT

Dictionary-based Phrase-level Prompting

- Dictionary-based Phrase-level Prompting*
 - 针对LLMs在稀缺词翻译方面表现差的问题
 - 低资源翻译
 - Domain Adaption
- 尝试将现有的词典或者可进行抽取的词典注入到Prompt中
 - 已有的词典: MUSE
 - 需要抽取的词典: BLI

Translate the following sentence to English:
Pada dasarnya, hal tersebut terbagi ke dalam dua kategori: Anda bekerja sambil mengadakan perjalanan atau mencoba mencoba atau membatasi pengeluaran Anda. Artikel ini berfokus pada hal yang terakhir.

In this context, the word "sambil" means "while"; the word "membatasi" means "limiting", "restrict", "limit".

The full translation to English is: *Basically, they fall into two categories: Either work while you travel or try and limit your expenses. This article is focused on the latter.*

Translate the following sentence to English:
Ia melakukan pembuatan bel pintu dengan teknologi WiFi, katanya.

In this context, the word "pembuatan" means "creation"; the word "bel" means "buzzer", "bell"; the word "pintu" means "door", "doors".

The full translation to English is:

* Ghazvininejad M, Gonen H, Zettlemoyer L. Dictionary-based phrase-level prompting of large language models for machine translation,

Instruction for LRLs --- DiPMT

Language	Baseline	DiPMT on OPT	Delta	Baseline	DiPMT on Bloom	Delta
Cat-Eng	37.80	37.61	-0.19	46.16	45.88	-0.28
Eng-Cat	17.86	19.08	1.22	40.79	40.79	0.00
Hrv-Eng	30.72	31.59	0.87	23.72	25.11	1.39
Eng-Hrv	11.55	12.56	1.01	7.94	9.21	1.27
Dan-Eng	42.39	42.82	0.43	34.30	36.87	2.57
Eng-Dan	26.63	27.04	0.41	16.55	19.58	3.03
Nld-Eng	27.18	27.49	0.31	25.16	25.99	0.83
Eng-Nld	15.98	16.66	0.68	12.49	13.40	0.91
Tgl-Eng	31.21	32.12	0.91	16.95	20.17	3.22
Eng-Tgl	14.05	14.78	0.73	8.13	9.57	1.44
Ind-Eng	31.30	32.20	0.90	42.55	42.18	-0.37
Eng-Ind	16.03	18.18	2.15	42.26	41.98	-0.28
Ita-Eng	30.03	29.84	-0.19	30.60	30.95	0.35
Eng-Ita	18.99	19.49	0.50	19.40	19.85	0.45
Msa-Eng	27.08	28.80	1.72	42.31	42.47	0.16
Eng-Msa	10.95	12.98	2.03	30.92	31.30	0.38
Nob-Eng	38.45	39.85	1.40	30.68	32.73	2.05
Eng-Nob	20.99	22.34	1.35	13.86	15.38	1.52
Slk-Eng	24.53	27.12	2.59	20.15	22.33	2.18
Eng-Slk	6.28	5.71	-0.57	6.87	8.21	1.34

Low-resource Benchmark: FLORES-101实验表现

Method	MED.	LAW	IT	KORAN	Avg.
Baseline	37.33	35.30	23.14	16.89	28.16
Repp	44.18	45.87	32.44	14.32	34.20
DiPMT	50.38	45.92	33.58	20.34	37.56
KNN-MT	54.54	61.11	48.63	19.22	45.87

Out-of-Domain数据集实验表现

Instruction for LRLs --- XLT

Cross-Lingual-Thought Prompting

- 过去的大量Prompt设计都是基于单语任务的
- 提出一种通用in-context learning方法来提高LLM的多语言能力，命名为cross-lingual-thought prompting (XLT)
- 通过利用NLP任务的原始信息和模型的英文能力实现对于多语言任务的充分理解
- 用一种单轮对话的形式实现任务
- few-shot learning：实现对于LLM输出的利用

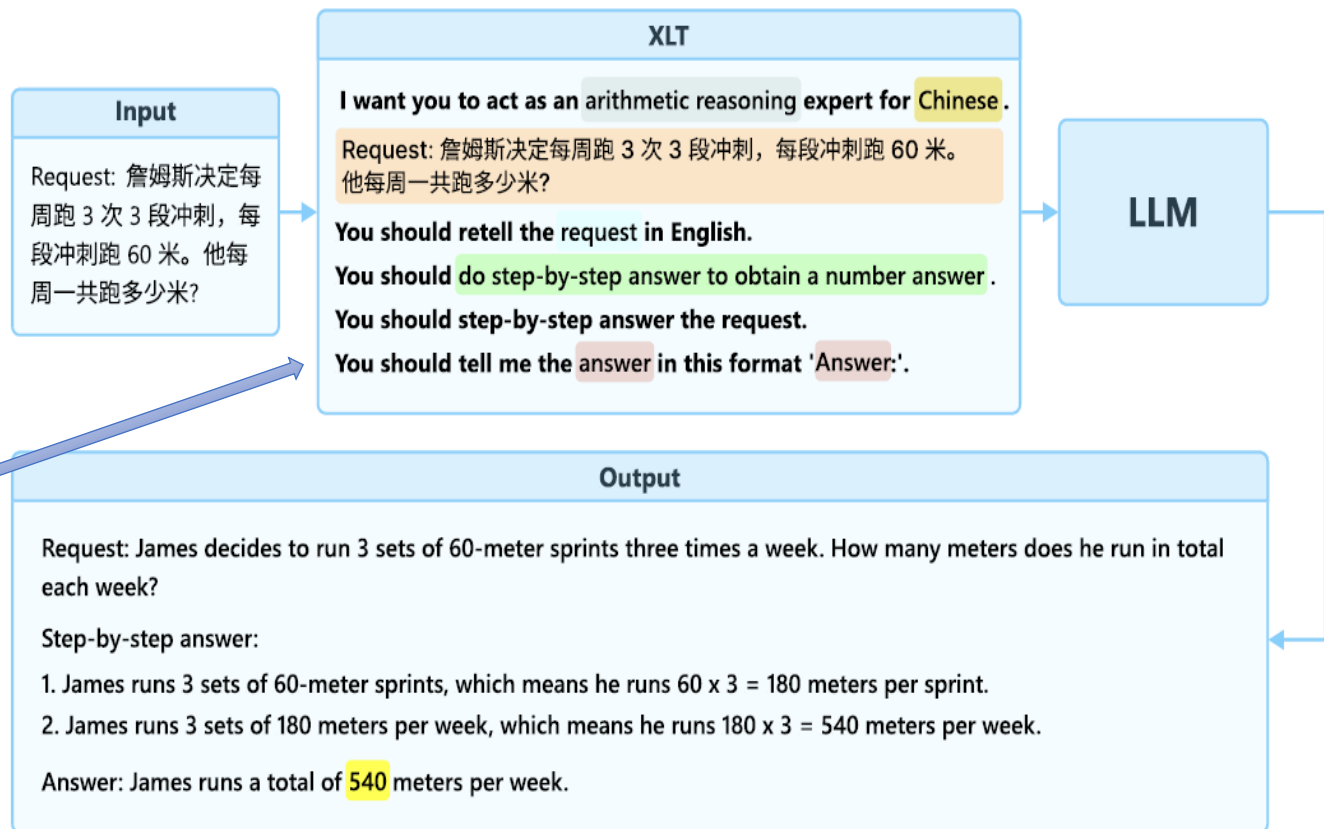
Instruction for LRLs --- XLT

Provide NLP task Meta Data



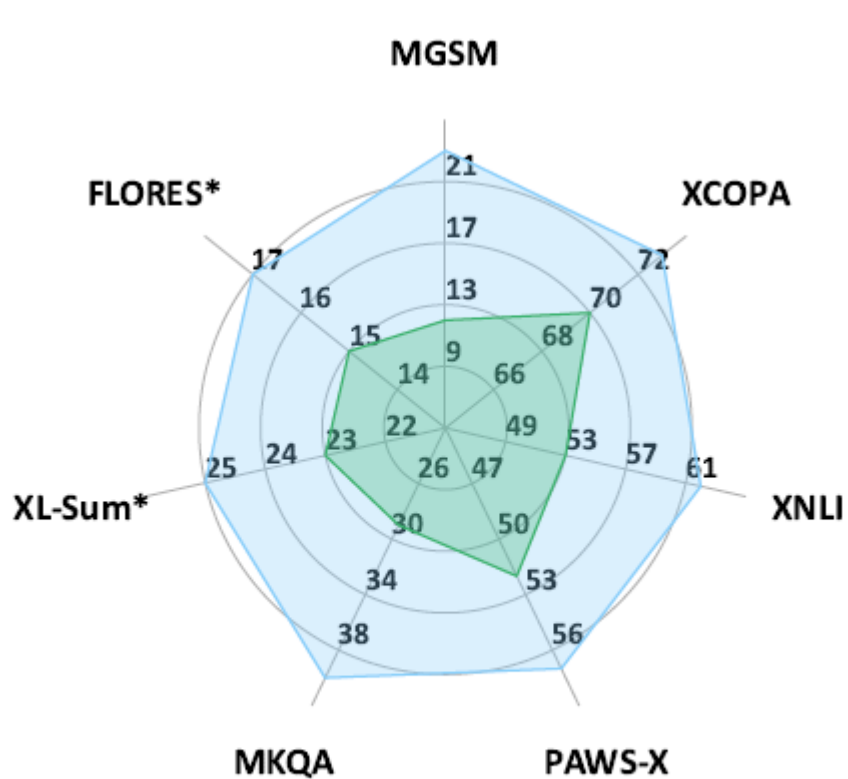
I want you to act as a `task_name` expert for `task_language` .
`task_input`
 You should retell/repeat the `input_tag` in English.
 You should `task_goal` .
 You should step-by-step answer the request.
 You should tell me the `output_type` (`output_constraint`) in this format '`output_type`' .

XLT template

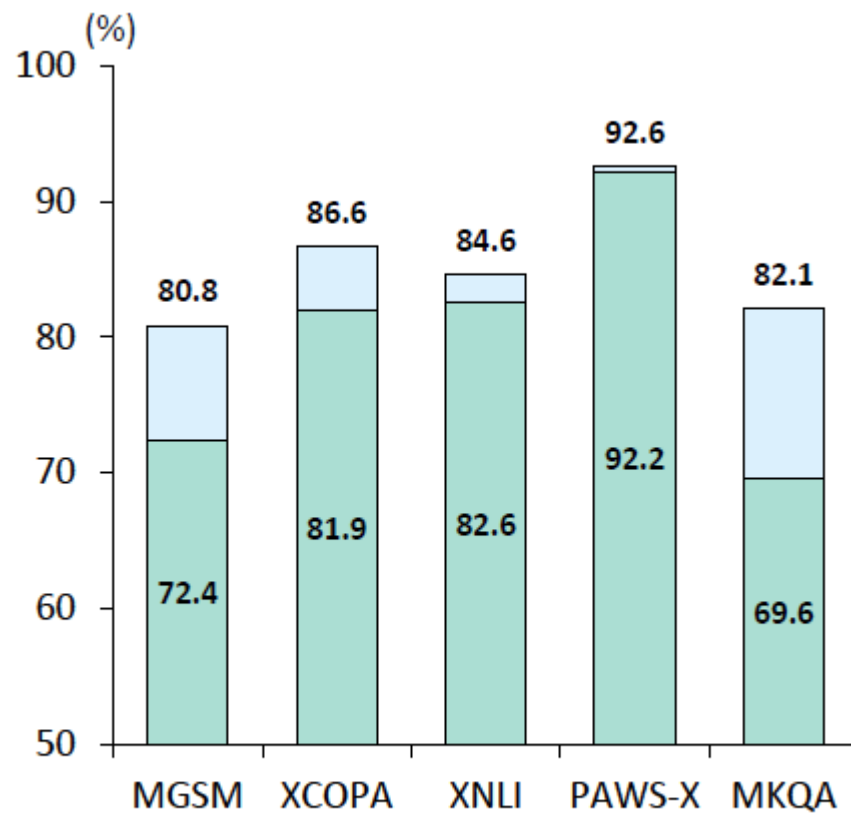


既有翻译, 又有推理

Instruction for LRLs --- XLT



(a)



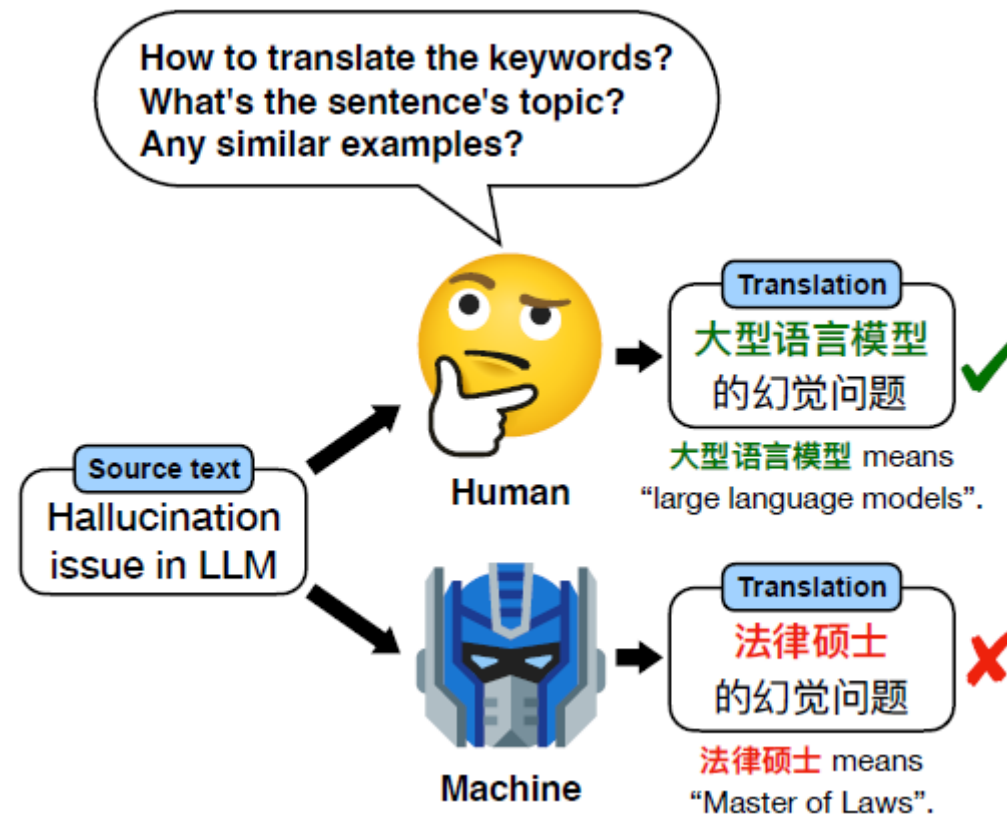
(b)

Cross-LingualPrompt基于text-davinci-003实现了包括多语言翻译任务的一系列任务表现提升

Instruction for LRLs --- MAPS

Exploring Human-Like Translation Strategy

- 模仿人类翻译的思考过程：翻译时会采取很多准备步骤确保高质量翻译
- 提出MAPS(Multi-Aspect Prompting & Selection)Framework:
 - 多方面的提示和结果选择方法来保证翻译质量
 - 提取了三个方面的翻译相关知识：关键词、主题以及相关的示例
 - 基于质量估计方法实现了对于噪声知识带来的翻译结果的过滤



Instruction for LRLs --- MAPS architecture

Step 1: Knowledge Mining

Ask for keyword pairs

Extract the keywords in the following English sentence, and then translate these keywords into Chinese.

English: <source>

Keyword Pairs:

<src_word>₁=<tgt_word>₁,
<src_word>₂=<tgt_word>₂,

Ask for topics

Use a few words to describe the topics of the following input sentence.

Input: <source>

Topics: <topic>₁, <topic>₂,
<topic>₃,

Ask for demonstration

Write an English sentence related to but different from the input English sentence and translate it into Chinese.

English: <source>

Output English-Chinese sentence pair: <src_demo>
<tgt_demo>

Step 2: Knowledge Integration

Keyword Pairs: <src_word>₁=<tgt_word>₁, <src_word>₂=<tgt_word>₂,

Topics: <topic>₁, <topic>₂, <topic>₃,

Related English-Chinese sentence pair: <src_demo> <tgt_demo>

Instruction: Given the above knowledge, translate the following English text into Chinese.

English: <source>

Chinese: <Candidate Demo>

Step 3: Knowledge Selection

Candidate Keyword	Score Keyword
Candidate Topic	Score Topic
Candidate Demo	Score Demo
Candidate Base	Score Base
Quality Estimation	

Best translation

- 知识挖掘:
 - 关键词keyword
 - 主题topic
 - 示例demonstration
- 知识整合
- 翻译结果选择

Instruction for LRLs --- MAPS results

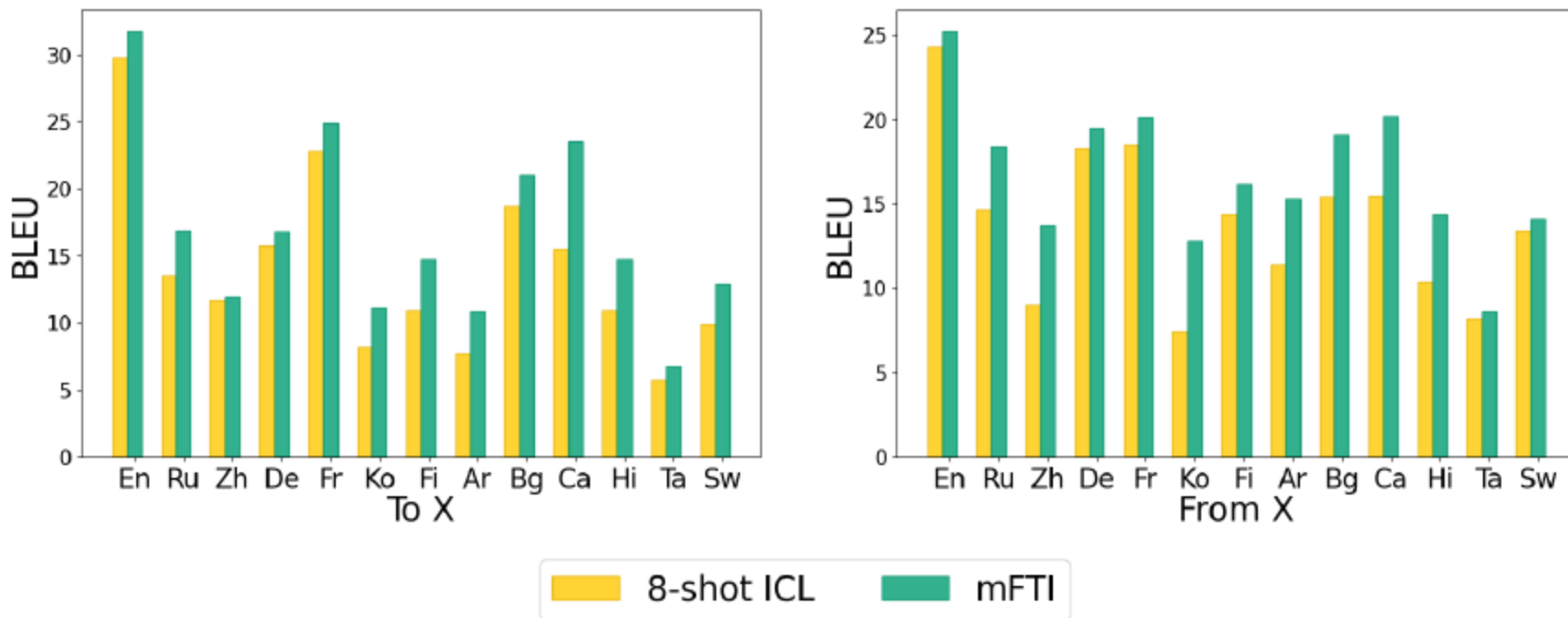
在基于text-davinci-003以及Alpaca的一系列开源模型上翻译表现得到了一定提升

Method	En-Zh	Zh-En	En-De	De-En	En-Ja	Ja-En	De-Fr	Fr-De
Existing works COMET								
WMT22 Best	86.8	81.0	87.4	85.0	89.3	81.6	85.7	89.5
Hendy et al.	85.8	81.6	85.6	84.8	87.5	81.5	78.5	84.6
text-davinci-003 COMET								
Baseline	86.2	81.6	85.8	85.2	87.9	81.8	82.8	86.3
+Keyword	86.2	81.5	85.5	84.9	88.0	81.5	82.6	86.2
+Topic	86.4	81.7	85.6	85.2	88.1	81.9	83.1	86.3
+Demo	86.9	81.8	86.6	85.2	88.5	81.8	83.4	86.7
Rerank	86.9/87.5	82.1/82.6	86.4/86.9	85.5/85.8	88.8/89.3	82.3/83.0	83.4/84.1	86.8/87.4
MAPS	87.6/88.5	82.6/83.8	87.2/88.0	85.7/86.7	89.5/90.3	82.9/84.8	84.1/85.4	87.5/88.7

Instruction for LRLs --- mFTI

- Instruction Tuning来激发LLM的翻译能力:
 - 基座模型: XGLM-7B
 - 数据集: WikiMatrix, Caligned
 - 语言对: 13种语言, 156个语言对
 - 数据数目: 每种语言对—1000个语言对
 - 指令的形式: Translation: [S]: X [T]: Y

Multilingual Fine-tuning with Translation Instructions



BLEU平均值均有增长：mFTI比ICL更好地激发了LLM的翻译能力：无论是En-X还是X-En的翻译方向上，经过instructiontuning的模型的翻译能力比原始8-shotICL的模型都有了提升

Multilingual Fine-tuning with Translation Instructions

	Same Direction	Seen Both Sides Reversed Direction	Unseen Direction	Unseen Src	Unseen Tgt	Unseen Both Sides
8-shot ICL	14.5	14.5	11.2	13.5	13.6	14.6
mFTI-16	15.7(+1.2)	13.7(-0.8)	12.6(+1.4)	14.9(+1.4)	14.5(+0.9)	15.3(+0.7)
mFTI-all	16.7	16.8	14.6	17.6	17.0	18.4

Table 3: Translation performances on different partitions. mFTI-16: XGLM multilingual finetuned with translation instructions on a mixture of 16 language pairs described in 4.1.

激发对未知语言对的翻译能力：此外，实验发现mFIT也对**instruction tuning**中未包含的语言对也激发了对应的翻译能力

Instruction for LRLs --- Big Translate

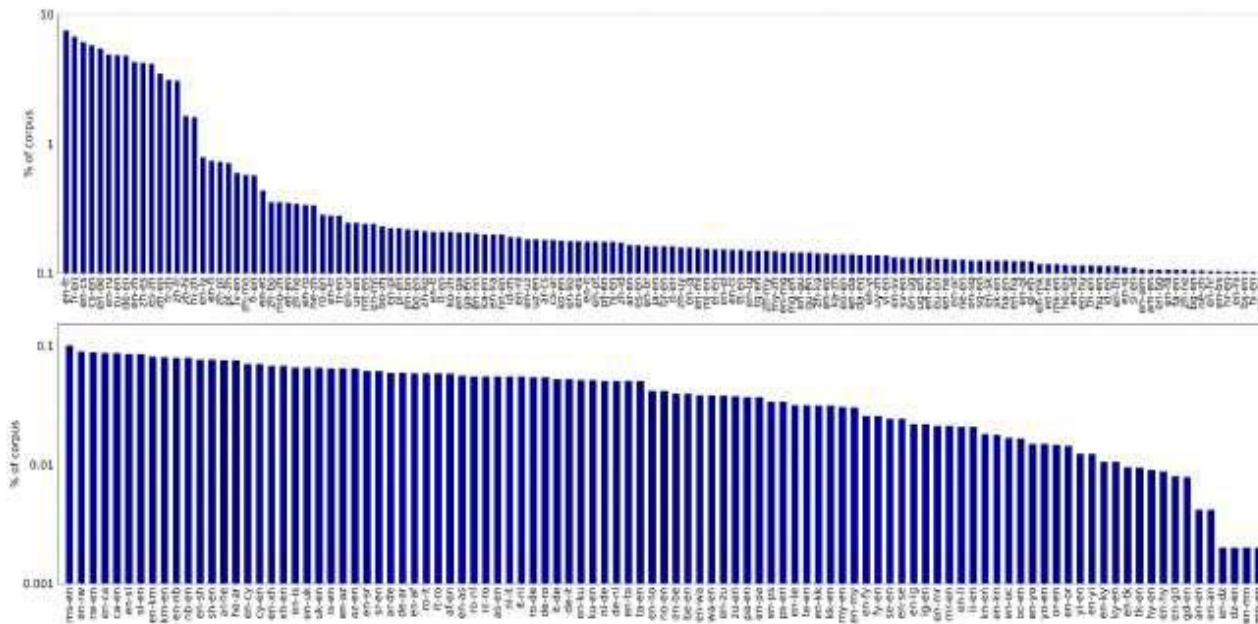
Big Translate

- 为了让LLM具备更多的多语言能力，BigTrans基于LLaMA进行训练，使其能够翻译100多种自然语言
 - 基座模型：LLaMA-13B
 - Step1: 中文预训练
 - Step 2:102种语言平行语料进行训练
 - Step3:InstructionTuning

Instruction for LRLs --- Big Translate

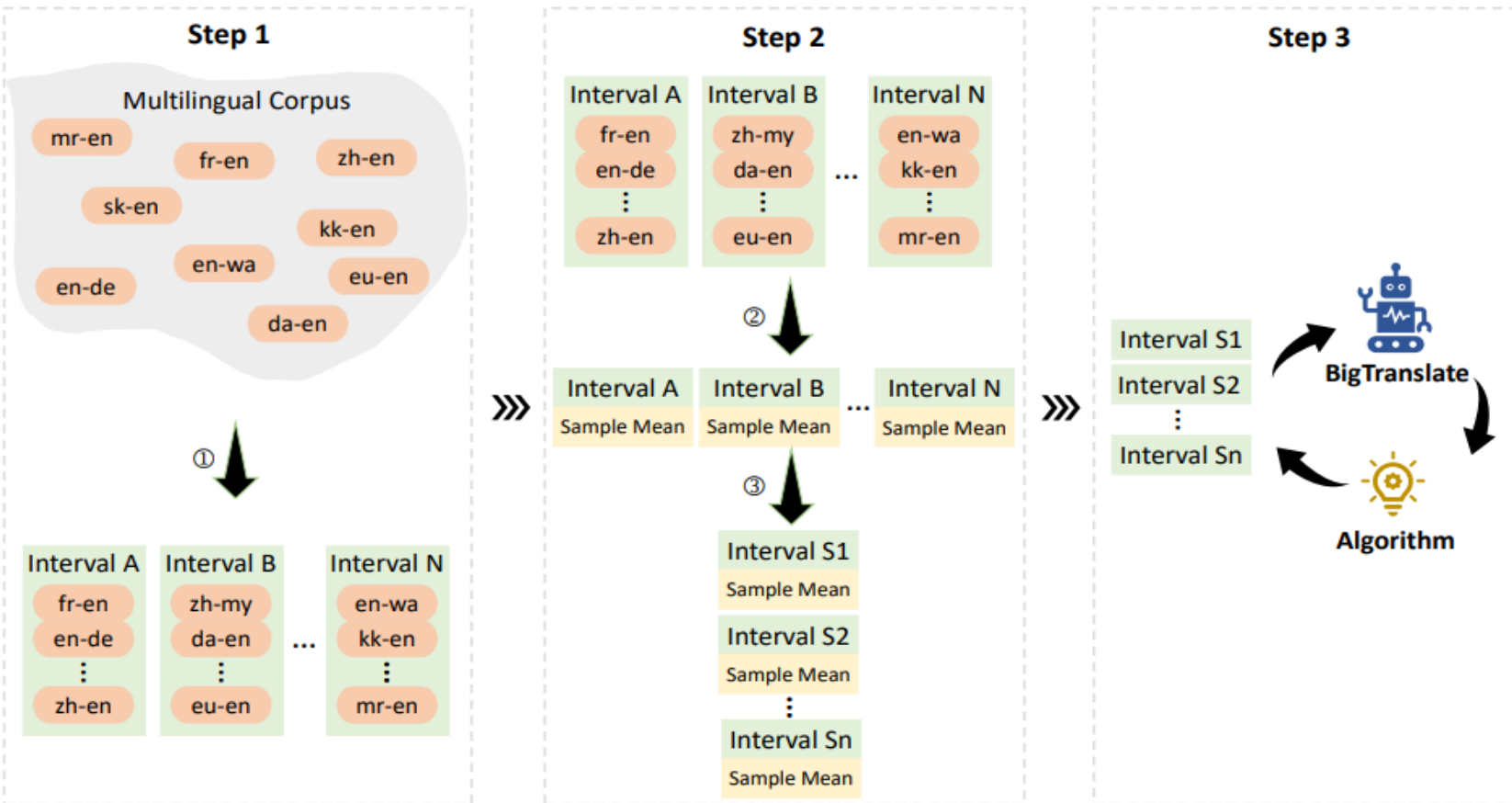
Big Translate—Step2:102种语言的平行语料进行训练

- 使用102种语言的平行语料进行训练
 - 防止中文预训练导致原始语言遗忘
 - 增强模型的跨语言对齐能力
- 对于训练数据进行数据增强平衡数据分布



数据集中语言对分布

Instruction for LRLs --- Big Translate



第1步：划分不同子集

第2步：计算不同子集每个语言对样本数目均值，子集进行降序排序

第3步：动态更新未训练的样本数目均值，与后续子集合并

BigTranslate—Step2:
102种语言的平行语料进行训练
增量多语言预训练：平衡高资源语言和低资源语言的翻译能力

Instruction for LRLs --- Big Translate result

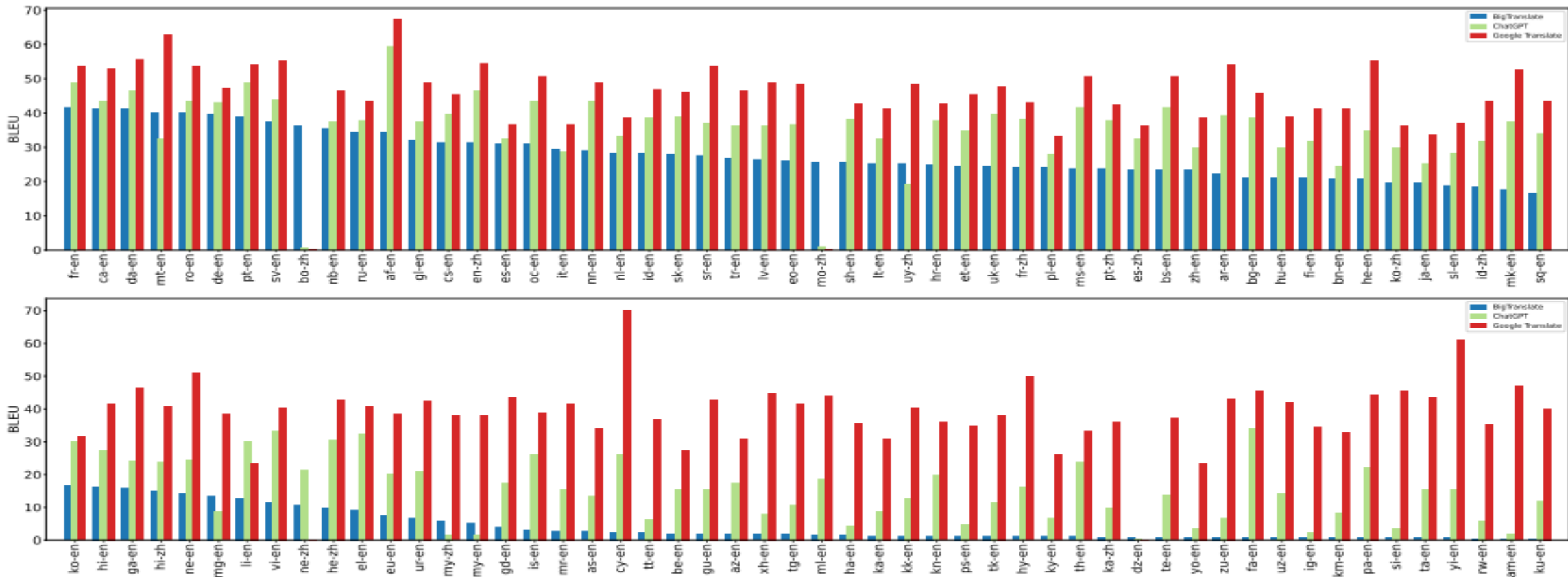


Figure 3: An illustrated comparison of 102 languages from X to English or Chinese between **BigTranslate**, **ChatGPT**, and **Google Translate**. We sort the language scores in BLEU for BigTranslate in descending order.

Summary for LRLs MT with LLM

- 当前研究进展：
 - 基于LLM的机器翻译性能评估：PaLM、GPT-3.5等
 - 基于LLM的低资源翻译提示方法：基于Dictionary、利用英文Thought能力等
 - 基于LLM的低资源翻译训练方法：mFIT, BigTranslate等
- 未来研究方向：
 - 如何更好地平衡大模型的多语言低资源翻译能力？
 - 如何建立以中文为核心的机器翻译大模型？
 - 如何依据大模型的生成特点做好的翻译能力评估？
 -



Outline

- Introduction of LLM
- LLM Performance in NLP
- Fine-Tuning Strategies & Challenges for LLM
- Machine Translation with LLMs
- LRLs Machine Translation with LLM
- Future of NLP with LLMs

Future of NLP with LLMs

Future Research Topic for LLM

NLP领域未来科研工作的发展方向掀起广泛的探讨

中国科学：信息科学

SCIENTIA SINICA Informationis

评述

大模型时代的自然语言处理：挑战、机遇与发展

车万翔⁴, 窦志成¹¹, 冯岩松¹, 桂韬³, 韩先培¹⁰, 户保田⁵, 黄民烈⁶, 黄萱菁^{2*}, 刘康⁹, 刘挺⁴, 刘知远^{6*}, 秦兵⁴, 邱锡鹏², 万小军¹, 王宇轩⁸, 文继荣¹¹, 严睿¹¹, 张家俊⁹, 张民^{5,7*}, 张奇², 赵军⁹, 赵鑫¹¹, 赵妍妍⁴

1. 北京大学王选计算机研究所, 北京 100080

2. 复旦大学计算机科学技术学院, 上海 200438

3. 复旦大学现代语言学研究院, 上海 200433

4. 哈尔滨工业大学计算学部, 哈尔滨 150001

5. 哈尔滨工业大学(深圳), 深圳 518055

6. 清华大学计算机科学与技术系, 北京 100084

7. 苏州大学计算机科学与技术学院, 苏州 215006

8. 之江实验室, 杭州 310000

9. 中国科学院自动化研究所, 北京 100190

10. 中国科学院软件研究所, 北京 100190

11. 中国人民大学高瓴人工智能学院, 北京 100872

* 通信作者. E-mail: xjhuang@fudan.edu.cn, liuzy@tsinghua.edu.cn, zhangmin2021@hit.edu.cn

† 作者和单位排名不分先后, 同等贡献, 按拼音排序

- LLM
 - Challenges
 - Opportunities
 - Future Directions
- 43 Pages
- NLP down-stream tasks
 - 文本分类
 - 结构化预测
 - 语义分析
 - 知识图谱与文本信息抽取
 - 情感计算
 - 文本生成
 - 自动文摘
 - 机器翻译
 - 对话系统
 - 信息检索
 - 自动问答

[Towards a comprehensive understanding of the impact of large language models on natural language processing: challenges, opportunities and future directions[†] \(sciengine.com\)](https://www.sciengine.com)

Future Research Topic for LLM

A PhD Student's Perspective on Research in NLP in the Era of Very Large Language Models

Oana Ignat*, Zhijing Jin*, Artem Abzaliev, Laura Biester, Santiago Castro, Naihao Deng, Xinyi Gao, Aylin Gunal, Jacky He, Ashkan Kazemi, Muhammad Khalifa, Namho Koh, Andrew Lee, Siyang Liu, Do June Min, Shinka Mori, Joan Nwatu, Veronica Perez-Rosas, Siqi Shen, Zekun Wang, Winston Wu, Rada Mihalcea

LANGUAGE AND INFORMATION TECHNOLOGIES (LIT)
UNIVERSITY OF MICHIGAN
{oignat, jinzhi, mihalcea}@umich.edu

Abstract

Recent progress in large language models has enabled the deployment of many generative NLP applications. At the same time, it has also led to a misleading public discourse that “it’s all been solved.” Not surprisingly, this has in turn made many NLP researchers – especially those at the beginning of their career – wonder about what NLP research area they should focus on. This document is a compilation of NLP research directions that are rich for exploration, reflecting the views of a diverse group of PhD students in an academic research lab. While we identify many research areas, many others exist; we do not cover those areas that are currently addressed by LLMs but where LLMs lag behind in performance, or those focused on LLM development. We welcome suggestions for other research directions to include: <https://bit.ly/nlp-era-llm>

1 Background

Language models represent one of the fundamental building blocks in NLP, with their roots traced back to 1948 when Claude Shannon introduced Markov chains to model sequences of letters in English text (Shannon, 1948). They were then heavily used in connection to the early research on statistical ma-

idea: given a series of previous words or characters, we can predict what will come next. The new large language models (LLMs) benefit from two main developments: (1) the proliferation of Web 2.0 and user-generated data, which has led to a sharp increase in the availability of data; and (2) the growth in computational capabilities through the introduction of Graphics Processing Units (GPUs). Together, these developments have facilitated the resurgence of neural networks (or deep learning) and the availability of very large training datasets for these models.

Current LLMs have output quality comparable to human performance, with the added benefit of integrating information from enormous data sources, far surpassing what one individual can accumulate in their lifetime. The number of applications that benefit from using LLMs is continuously growing, with many cases where the LLMs are used to replace entire complex pipelines. LLMs becoming “lucrative” has led to a surge in industry interest and funding, alongside a sharp increase in the number of research publications on LLMs. For instance, a search on Google Scholar for “language models” leads to 50,000 publications over the past five years, a third of the roughly 150,000 papers published during the past 25 years.

- 博士研究生关于大语言模型时代自然语言处理研究的视角
- [2305.12544.pdf \(arxiv.org\)](https://arxiv.org/pdf/2305.12544v1)
- Multilinguality and **Low-Resource Languages**

iv:2305.12544v1 [cs.CL] 21 May 2023

Future Research Topic for LLM

Choose Your Weapon: Survival Strategies for Depressed AI Academics

Julian Togelius and Georgios N. Yannakakis*

April 14, 2023

Abstract

Are you an AI researcher at an academic institution? Are you anxious you are not coping with the current pace of AI advancements? Do you feel you have no (or very limited) access to the computational and human resources required for an AI research breakthrough? You are not alone; we feel the same way. A growing number of AI academics can no longer find the means and resources to compete at a global scale. This is a somewhat recent phenomenon, but an accelerating one, with private actors investing enormous compute resources into cutting edge AI research. Here, we discuss what you can do to stay competitive while remaining an academic. We also briefly discuss what universities and the private sector could do improve the situation, if they are so inclined. This is not an exhaustive list of strategies, and you may not agree with all of them, but it serves to start a discussion.

1 Introduction

As someone who does AI research in a university, you develop a complicated relationship to the corporate AI research powerhouses, such as DeepMind, Open AI, Google Brain and Meta AI. Whenever you see one of these papers that train some kind of gigantic neural net model to do something you weren't even sure a neural network could do, unquestionably pushing the state of the art and reconfiguring your ideas of what is possible, you get conflicting emotions. On the one hand: it's very impressive. Good on you for pushing AI forward. On the other hand: how could we possibly keep up? As an AI academic, leading a lab with a few PhD students and (if you're lucky enough) some postdocs, perhaps with a few dozen GPUs in your lab, this kind of research is simply not possible to do.

To be clear, this was not always the case. As recently as ten years ago, if you had a decent desktop computer and an internet connection you had everything you needed to compete with the best of researchers out there. Ground-breaking papers were often written by one or two people who ran all the experiments on their regular workstations. It is useful to point this out particularly for those who've come into the research field within the last decade, and for which the need for giant compute resources is a given.

- Try Scaling Anyway
- Scale Down
- Reuse and Remaster
- Analysis Instead of Synthesis
- **RL! No Data!**
- **Small Models! No Compute!**
- Work on Specialized Application Areas or Domains
- Solve Problems Few Care About (For Now!)
- **Try Things that Shouldn't Work**
- Do Things That Have Bad Optics
- Start it Up; Spin it Out!
- **Collaborate or Jump Ship!**
-

[2304.06035.pdf \(arxiv.org\)](https://arxiv.org/pdf/2304.06035.pdf)

arXiv:2304.06035v1 [cs.OH] 31 Mar 2023

Future Research Topic for LLM

大模型在各个传统任务上的性能测试报告层出不穷

Is ChatGPT a General-Purpose Natural Language Processing Task Solver?

Chengwei Qin[†], Aston Zhang, Zhuosheng Zhang[‡], Jiaao Chen[‡],
Michihiro Yasunaga[‡], Diyi Yang[‡]

[†]Nanyang Technological University, [‡]Shanghai Jiao Tong University
[‡]Georgia Institute of Technology, [‡]Stanford University

Abstract

Spurred by advancements in scale, large language models (LLMs) have demonstrated the ability to perform a variety of natural language processing (NLP) tasks zero-shot—i.e., without adaptation on downstream data. Recently, the debut of ChatGPT¹ has drawn a great deal of attention from the natural language processing (NLP) community due to the fact that it can generate high-quality responses to human input and self-correct previous mistakes based on subsequent conversations. However, it is not yet known whether ChatGPT can serve as a generalist model that can perform many NLP tasks zero-shot. In this work, we empirically analyze the zero-shot learning ability of ChatGPT by evaluating it on 20 popular NLP datasets covering 7 representative task categories. With extensive empirical studies, we demonstrate both the effectiveness and limitations of the current version of ChatGPT. We find that ChatGPT performs well on many tasks favoring reasoning capabilities (e.g., arithmetic reasoning) while

in some cases, current LLMs are still prone to various mistakes in zero-shot learning. In addition, the format of the prompt can have a substantial impact—for example, simply adding “Let’s think step by step” (Kojima et al., 2022) has been shown to significantly improve the performance of InstructGPT (Ouyang et al., 2022) on reasoning tasks. These limitations illustrate that current LLMs are not truly general-purpose language systems.

Recently, the ChatGPT LLM released by OpenAI has attracted a great deal of attention from the NLP community. ChatGPT was created by training a GPT-3.5 series model through reinforcement learning from human feedback (RLHF) (Christiano et al., 2017) (similarly to InstructGPT). RLHF mainly includes three steps: training a language model with supervised learning, collecting comparison data based on human preferences and training a reward model, and optimizing the language model against the reward model using reinforcement learning (Ouyang et al., 2022). Through RLHF training,

[2302.06476.pdf \(arxiv.org\)](https://arxiv.org/abs/2302.06476)

A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity

Yejin Bang^{*} Samuel Cahyawijaya Nayeon Lee Wenliang Dai Dan Su Bryan Wilie
Holy Lovenia Ziwei Ji Tiezheng Yu Willy Chung Quyet V. Do Yan Xu Pascale Fung^{*}
Centre for Artificial Intelligence Research (CAiRE)
The Hong Kong University of Science and Technology
yjbang@connect.ust.hk, pascalle@ece.ust.hk

Abstract

This paper proposes a framework for quantitatively evaluating interactive LLMs such as ChatGPT using publicly available data sets, using 23 data sets covering 8 different common NLP application tasks. We extensively evaluate the multitask, multilingual, and multi-modal aspects of ChatGPT based on these data sets and a newly designed multimodal dataset. We find that ChatGPT outperforms LLMs with zero-shot learning on most tasks and even outperforms fine-tuned models on some tasks. We find that it is better at understanding non-Latin script languages than generating them. It is able to generate multimodal content from textual prompts via an intermediate code generation step. Moreover, we find that ChatGPT is 63.41% accurate on average in 10 different reasoning categories under logical reasoning, non-textual reasoning, and commonsense reasoning, hence making it an unreliable reasoner. ChatGPT suffers from hallucination problems like other LLMs. Finally, the interactive feature of ChatGPT enables human collaboration with the underlying LLM to improve its performance, i.e., 8% ROUGE-1 on summarization and 2% ChrF++ on machine translation, in a multi-turn “prompt engineering” fashion. We release a code for evaluation set extraction.¹

[2302.04023.pdf \(arxiv.org\)](https://arxiv.org/abs/2302.04023)

its scale and via RLHF, has shown impressive abilities in many areas of NLP as well as emergent abilities. Another reason is that its dialog interface allows users to interact with the underlying LLM more effectively and efficiently via interactive chats that are akin to multi-turn prompting.

However, despite its powerful abilities, anecdotal reports on ChatGPT consistently showed remaining challenges - for example, it fails in some elementary mathematical (Gilson et al., 2022; Goldberg, 2023; Frieder et al., 2023; Choi et al., 2023; Davis, 2023b) and commonsense reasoning tasks (Guo et al., 2023; Davis, 2023b); it hallucinates with human-like fluency and eloquence on things that are not based on truth (Shen et al., 2023; Thorp, 2023; Smith, 2023); and as a general-purpose language model trained from everything on the web, its language coverage is questionable (Lu et al., 2022; Jiao et al., 2023). Consequently, it is not clear what people can or cannot use ChatGPT for despite its popularity.

Since OpenAI never published any benchmarking results on ChatGPT at the time, seeing this need, in February 2023, we proposed a comprehensive framework for quantitatively evaluating interactive LLMs such as ChatGPT through standard public

Is ChatGPT Equipped with Emotional Dialogue Capabilities?

Weixiang Zhao, Yanyan Zhao^{*}, Xin Lu, Shilong Wang, Yanpeng Tong, Bing Qin
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{wxzhao, yyzhao}@ir.hit.edu.cn

Abstract

This report presents a study on the emotional dialogue capability of ChatGPT, an advanced language model developed by OpenAI. The study evaluates the performance of ChatGPT on emotional dialogue understanding and generation through a series of experiments on several downstream tasks. Our findings indicate that while ChatGPT’s performance on emotional dialogue understanding may still lag behind that of supervised models, it exhibits promising results in generating emotional responses. Furthermore, the study suggests potential avenues for future research directions.

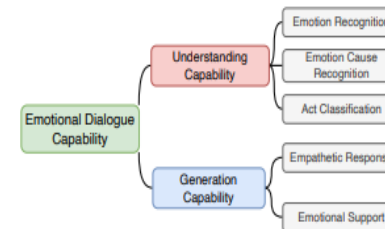


Figure 1: The emotional dialogue capability of a chatbot can be divided into two aspects: understanding and generation capability, with several downstream tasks.

1 Introduction

Emotional dialogue technology is a promising research area that aims to equip chatbots with human-like emotions, enabling them to recognize, understand, and express emotions in their interactions

has brought about unprecedented semantic understanding and response generation capabilities for conversational robots, greatly improving the interaction experience with human users. Considering the significant breakthrough of ChatGPT in basic conversational technology, as well as recent research analyzing its performance in various tradi-

[2304.09582.pdf \(arxiv.org\)](https://arxiv.org/abs/2304.09582)

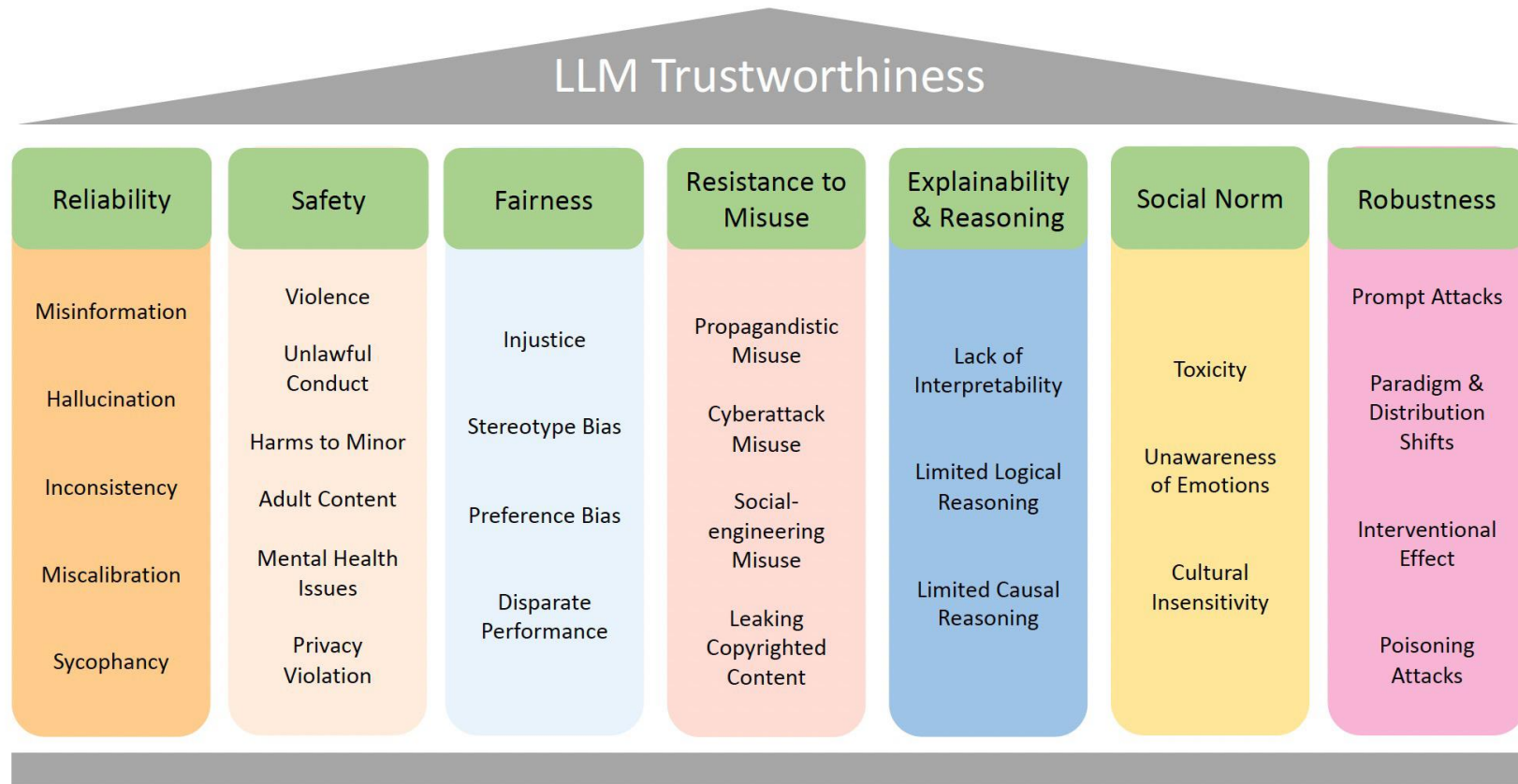
Future Research Topic for LLM

国内外一些优秀的 Survey 文章不能不重视

- [**A Survey of Large Language Models**](#)
- [**Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond**](#)
- [**Tool Learning with Foundation Models**](#)
- [**Evaluating Large Language Models: A Comprehensive Survey**](#)
- [**Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models**](#)
- [**Interactive Natural Language Processing**](#)
- [**A Survey on In-context Learning**](#)
- [**Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and future Prospects**](#)
- [**Aligning Large Language Models with Human: A Survey**](#)
- [**A Comprehensive Survey of AI-Generated Content \(AIGC\)**](#)
- [**Explanation-Based Human Debugging of NLP Models: A Survey**](#)
- [**Augmented Language Models: a Survey**](#)
- [**Training language models to follow instructions with human feedback**](#)
- [**Deep reinforcement learning from human preferences**](#)
- [**A REVIEW ON LARGE LANGUAGE MODELS: ARCHITECTURES, APPLICATIONS, TAXONOMIES, OPEN ISSUES AND CHALLENGES**](#)

Future Research Topic for LLM

- 大模型的优化
- 大模型的真实性和
- 大模型的理论
- **多模态大模型**
- **大模型+逻辑推理**
- 大模型+世界模型?
- 大模型+规划
- **可信赖大模型**



From Zhihu

Thank You!

Contact Me

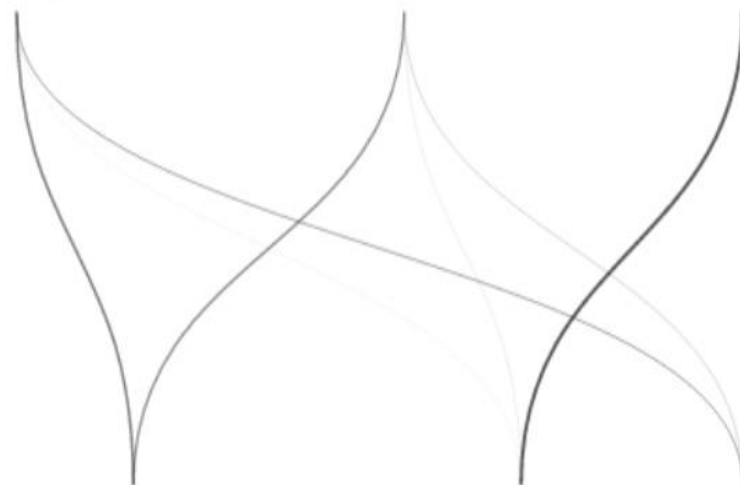
Email: miradel_51@Hotmail.com

I'm looking for **self-motivated master's** and **PhD** students to join our lab.

- 资源:

- A100 40G or 80G
- V100 32Gself-motivated
- 4090 24G

Any Questions ?



Questions diversives ?

This inspiration comes from Dzmitry Bahdanau @ ICLR2014