

Large Language Model

大语言模型的起源和发展

钱炳州 朱奕杰 赵洛右

Ch.1 深度学习和Transformer

循环神经网络

Recurrent Neural Networks

- 什么是循环神经网络
- 隐藏状态的作用
- 前向传播和反向传播的过程
- RNN的梯度消失问题

长短期记忆

Long Short-Term Memory

- 为什么需要长短期记忆
- 输入门、遗忘门和输出门
- LSTM中的记忆细胞

Transformer和自注意力机制

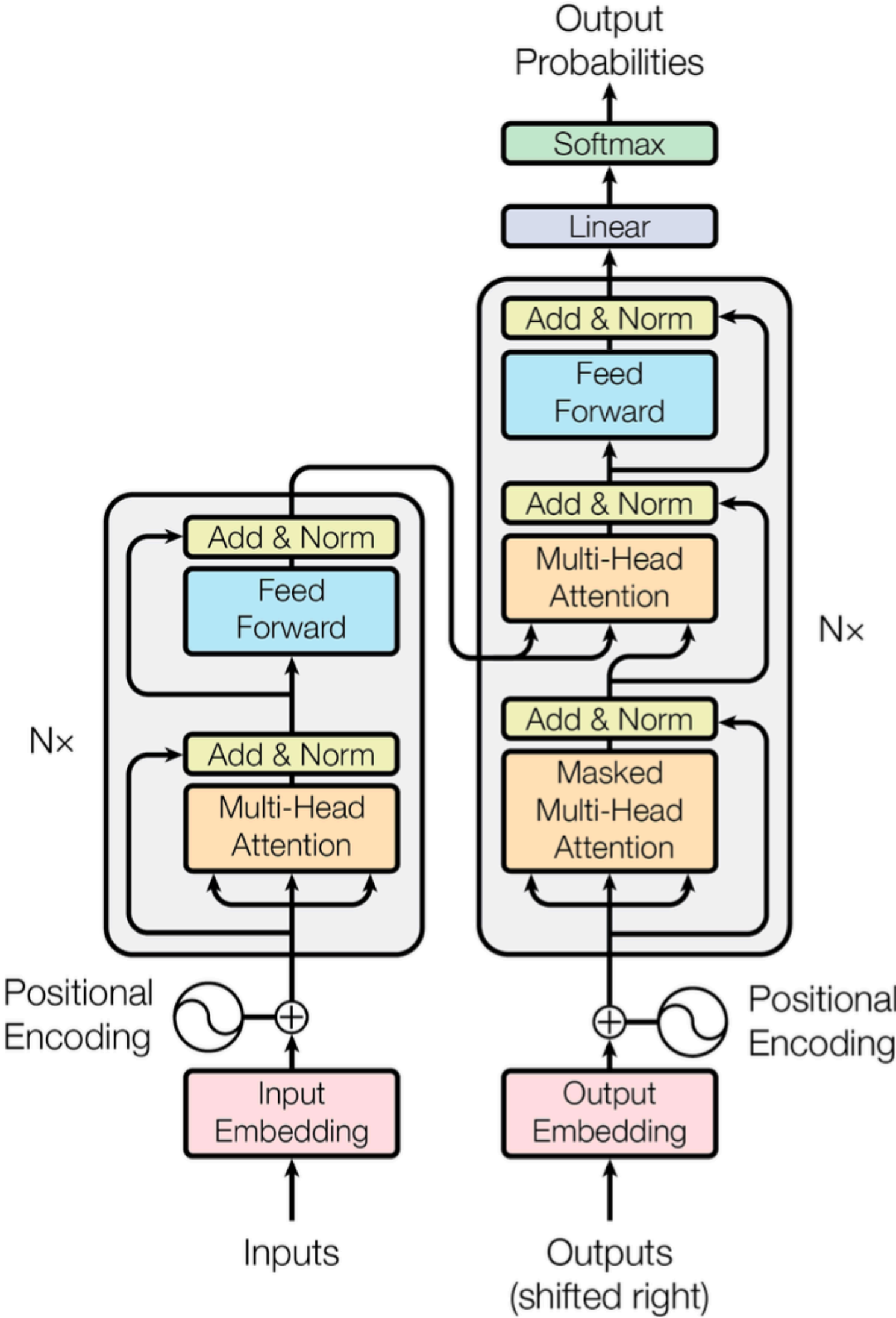
- Self-attention-mechanism
- Query, Key, Value

多头自注意力机制

Multi-Head Self-Attention

编码器和解码器

Encoder & Decoder



Transformer的优势

- 计算效率
- 捕捉长距离依赖
- 全局上下文理解
- 更易于优化

“Transformer is the first transduction model relying entirely on self-attention to compute representations of its input and output without using sequence-aligned RNNs or convolution.”

—Attention Is All You Need

Ch.2 GPT发展史

GPT是什么

- 基于转换器的生成式预训练模型 (Generative pre-trained transformers; GPT)
- 基于 transformer 架构的人工神经网络
- 在无标签的大型数据集上进行预训练
- 最新版本：GPT-4 (2023.03)

模型结构

预训练 (Pre-training) 和微调 (Fine-tuning)

预训练：学习大量未标注数据的共性

微调：学习相关特定领域的少量已标注数据

GPT发展历程

GPT-1

- 发布时间：2018年
- 训练方法：语言建模 (Language modeling)
- 给定已经出现的所有单词，求下一个单词出现的概率分布
- 特点：建立了GPT系列的核心架构，确定了使用上文预测下一单词的基本原则

GPT-2

- 发布时间：2019年
- 训练方法：语言建模 (Language modeling)
- 模型的参数数量提高到15亿个 (1.5B)
- 特点：采用多任务语言模型，进一步提升模型的泛化能力

GPT-3

- 发布时间：2020年
- 去掉了GPT-1的微调步骤，直接输入自然语言当作指示
- 模型的参数数量提高到1750亿个 (175B)

优势

- 模型规模：1750亿个模型参数 (175B)
- 零样本学习 (Zero-Shot Learning)：能够在没有人工标注数据的情况下进行对新任务的学习和执行
- 多样性生成：可以生成多种不同语言风格的文本
- 多语言支持：支持多种不同的语言
- 低资源学习：可以被广泛应用于一些资源有限的场景

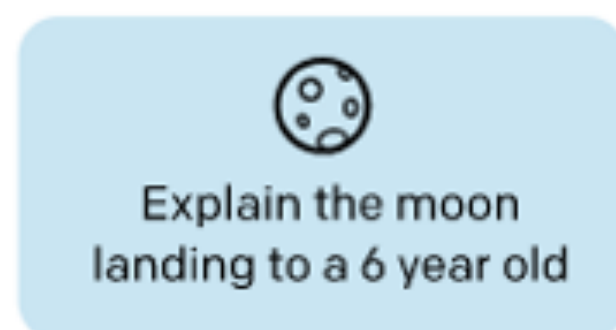
InstructGPT/ChatGPT

- 采用GPT-3的网络结构
- 通过指示学习 (Instruct Learning) 构建训练样本
- 共同点：均基于生成式预训练语言模型
- 区别：训练的数据集和模型的用途不同
- InstructGPT更适用于指令性任务
- ChatGPT更适用于对话生成任务

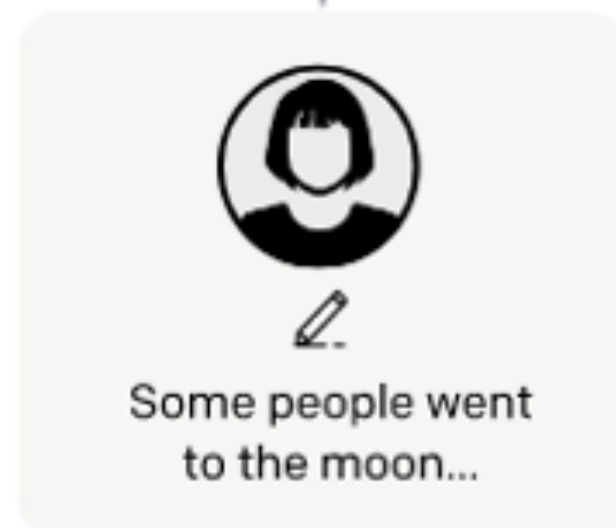
Step 1

Collect demonstration data, and train a supervised policy.

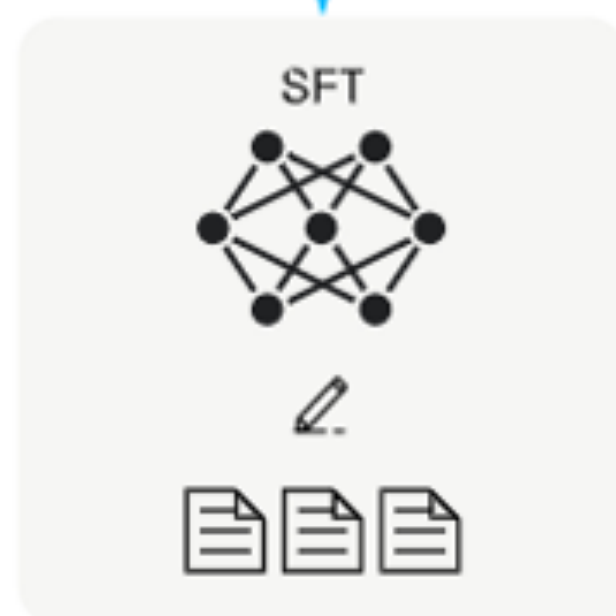
A prompt is
sampled from our
prompt dataset.



A labeler
demonstrates the
desired output
behavior.



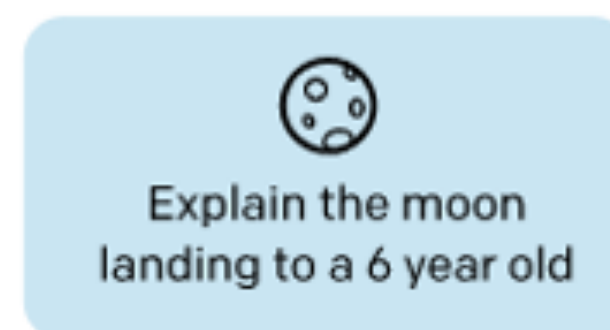
This data is used
to fine-tune GPT-3
with supervised
learning.



Step 2

Collect comparison data, and train a reward model.

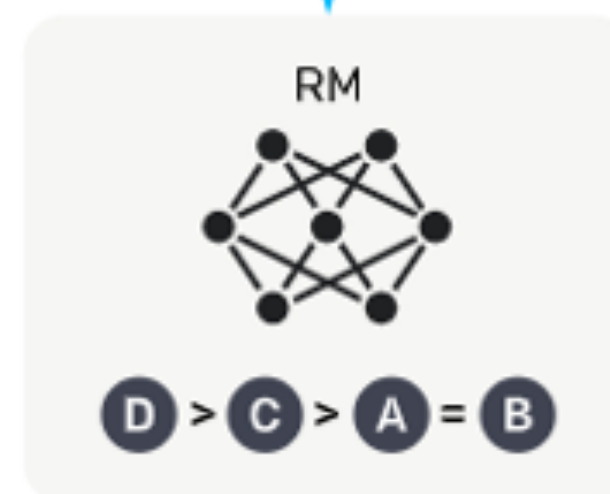
A prompt and
several model
outputs are
sampled.



A labeler ranks
the outputs from
best to worst.



This data is used
to train our
reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

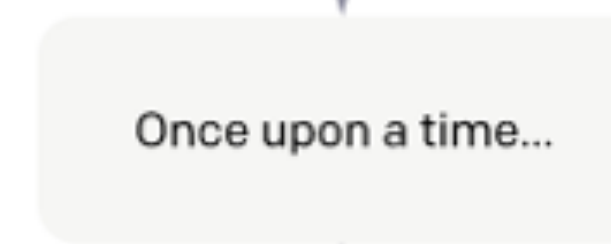
A new prompt
is sampled from
the dataset.



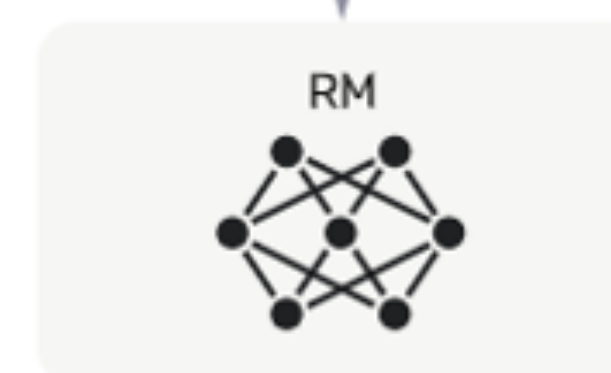
The policy
generates
an output.



Once upon a time...



The reward model
calculates a
reward for
the output.



The reward is
used to update
the policy
using PPO.

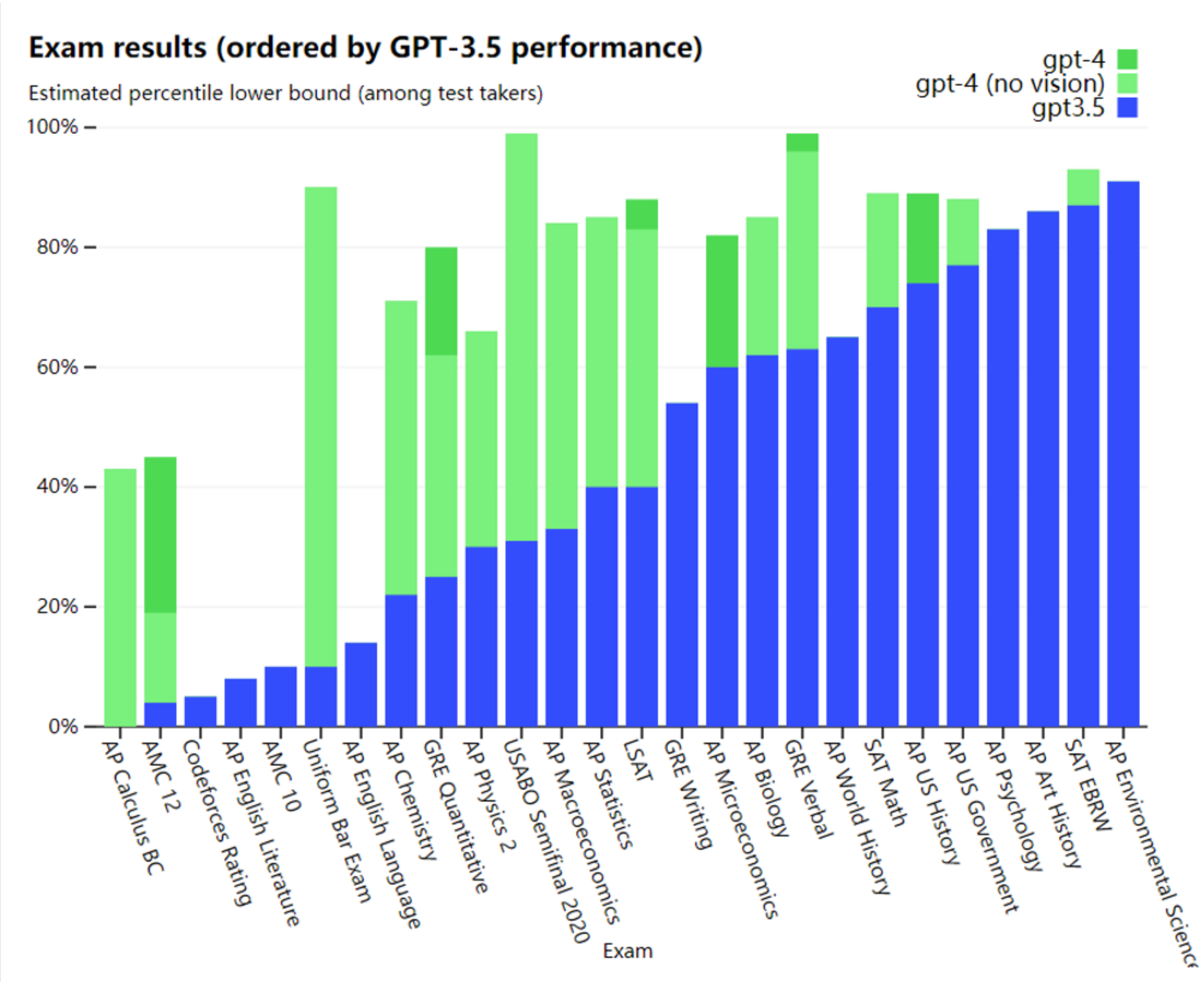


GPT-4

- 发布时间：2023年
- 多模态模型：可同时接收图像和文本输入
- 更可靠，更具创造性，能够处理更细微的指令

基本性能

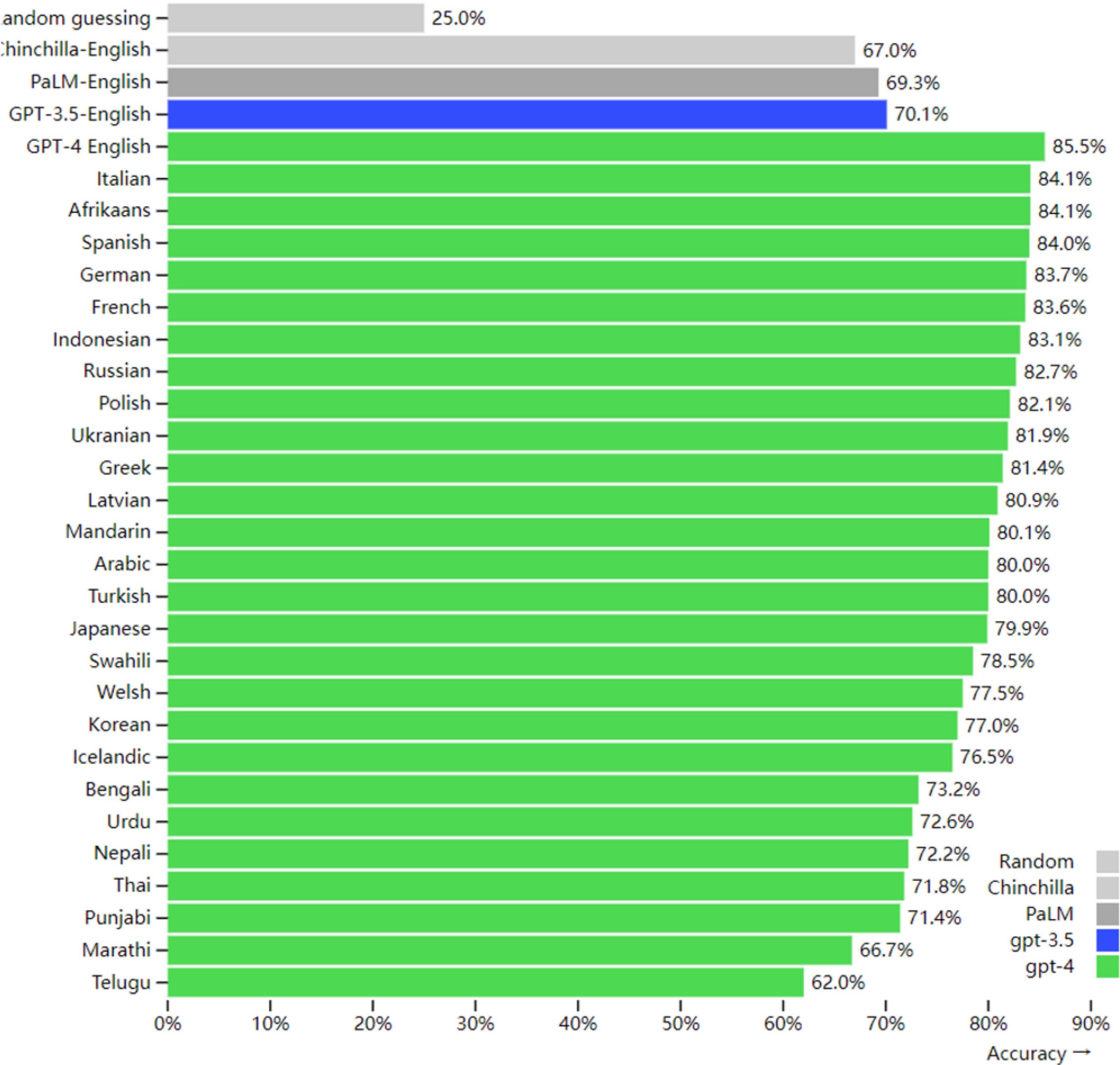
考试得分



基本性能

跨语言能力

GPT-4 3-shot accuracy on MMLU across languages



技术架构

- Transformer架构
- 目前最大规模的Transformer模型
- 拥有超过1万亿个参数
- 更先进的预训练技术
- 更多的人机交互功能

模型名称	训练方法	特点	论文
GPT-1	语言建模（预训练+微调）	建立了GPT系列的核心架构；确定了使用上文预测下一单词的基本原则	Improving Language Understanding Generative Pre-Training
GPT-2	语言建模，扩大模型参数到1.5B	多任务语言模型	Language Models are Unsupervised Multitask Learners
GPT-3	语言建模，扩大模型参数到175B	提示补全，上下文学习， Few-shot Learning	Language Models are Few-Shot Learners
Instruct-GPT	有监督微调，RLHF	指令回答， Zero-shot Learning	Training language models to follow instructions with human feedback
Codex	使用代码+文本训练	代码生成、 chain-of-thought (CoT)	Evaluating Large Language Models Trained on Code
ChatGPT	使用对话数据进行强化学习指令微调	对话历史建模	
GPT-4	使用多模态数据	更强的性能、 接受图像输入	GPT-4 Technical Report

仍然需要Fine-tuning transfer

初步具备zero-shot能力引入prompt

强化了zero-shot/one-shot/few-shot的能力，引入In-context Learning的概念

社会化，提出要与人类对齐。

更加社会化，与人类更加对齐。

GPT

GPT-2

GPT-3

instructGPT

ChatGPT

- 1.更大的模型
- 2.更多的数据

1. 更大的模型
2. 更多的数据

1. Instruct learning
2. RLHF

1. Instruct learning采用更多的标注指令
2. 将GPT3.5作为Initial LM

Ch.3 LLM的优越性

为什么GPT-3没有引起人们普遍的关注？

- 从实际性能上看，最初版本的GPT-3哪怕是其中最大的175B模型，其表现也未能超越经过微调的预训练语言模型（PLM）
- 而实验观测得到的比例定律让模型性能与规模挂钩，这意味着即使更改模型架构，也无法突破参数规模的限制，令模型难以取得重大的性能突破。

	SuperGLUE Average	BoolQ Accuracy	CB Accuracy	CB F1	COPA Accuracy	RTE Accuracy
Fine-tuned SOTA	89.0	91.0	96.9	93.9	94.8	92.5
Fine-tuned BERT-Large	69.0	77.4	83.6	75.7	70.6	71.7
GPT-3 Few-Shot	71.8	76.4	75.6	52.0	92.0	69.0

	WiC Accuracy	WSC Accuracy	MultiRC Accuracy	MultiRC F1a	ReCoRD Accuracy	ReCoRD F1
Fine-tuned SOTA	76.1	93.8	62.3	88.2	92.5	93.3
Fine-tuned BERT-Large	69.6	64.6	24.1	70.0	71.3	72.0
GPT-3 Few-Shot	49.4	80.1	30.5	75.4	90.2	91.1

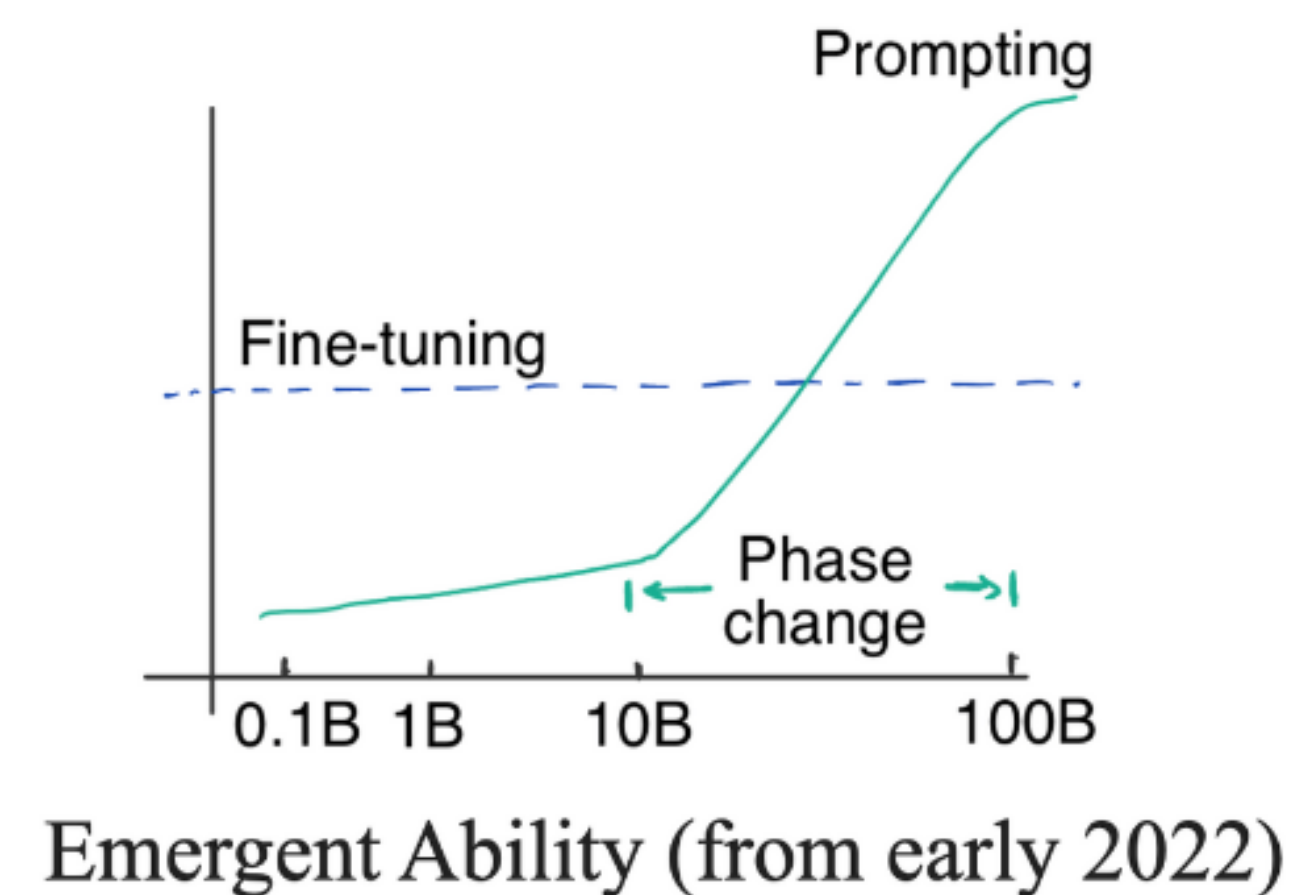
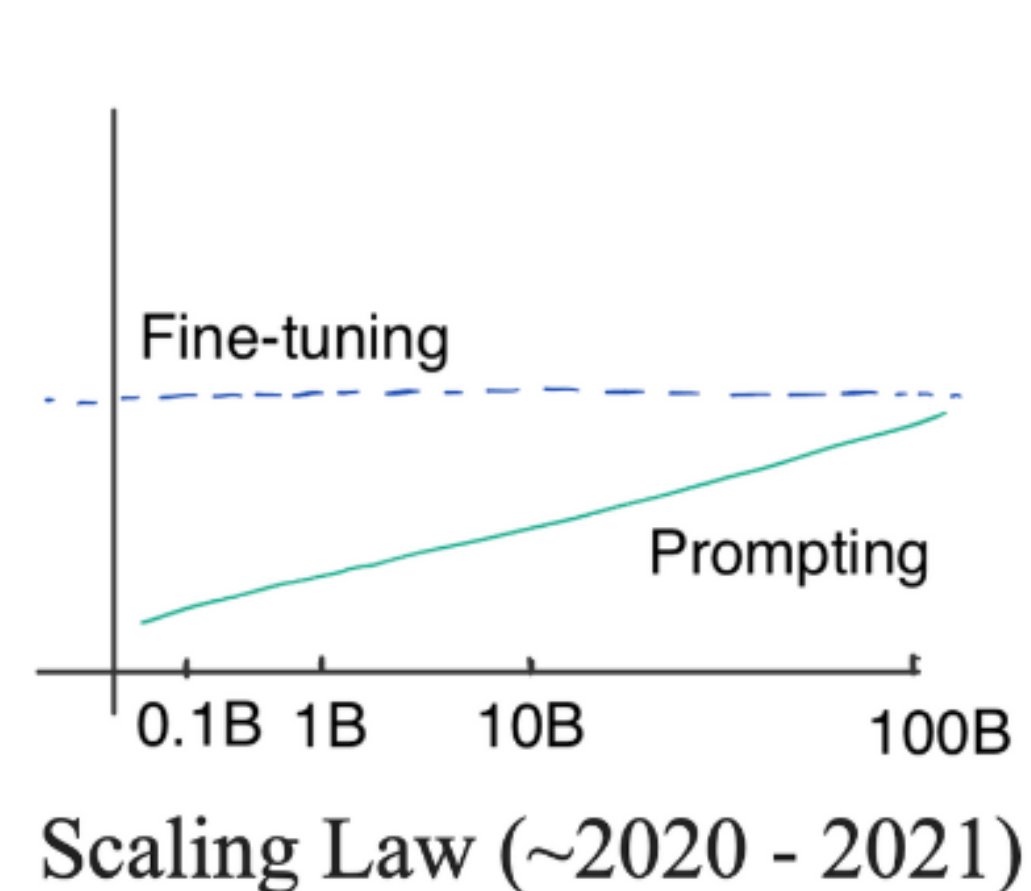
2022

- Google — Chain of Thought (CoT)

激发模型的推理能力，提高了prompting的效果

- OpenAI — instruction tuning

提高模型性能，增强执行任务的能力



指令微调

instruction tuning

- 定义：使用指令数据集（包含指令形式的任务描述、输入输出对、几组示例）对模型进行fine-tune
- 现有大量的开源指令集可供微调
- RLHF对应的微调被称为alignment tuning

TABLE 6
A detailed list of available task collections for instruction tuning. Note that OIG is a large collection consisting of existing collections.

Collections	Time	#Task types	#Tasks	#Examples
Nat. Inst. [208]	Apr-2021	6	61	193K
CrossFit [209]	Apr-2021	13	160	7.1M
FLAN [62]	Sep-2021	12	62	4.4M
P3 [210]	Oct-2021	13	267	12.1M
ExMix [211]	Nov-2021	11	107	18M
UnifiedSKG [212]	Jan-2022	6	21	812K
Super Nat. Inst. [79]	Apr-2022	76	1616	5M
MVPCorpus [213]	Jun-2022	11	77	41M
xP3 [84]	Nov-2022	17	85	81M
OIG ²³	Mar-2023	-	-	43M

Emergent Ability

- 定义：小模型没有，只有大模型才有的能力。
 - 突现能力体现之处很多，我们重点关注具有以下特点的涌现能力：
 1. NLP领域长期努力但未能实现的
 2. 其他NLP模型难以解决的
 3. 关系到自然语言本质的选
- (不同的综述对重要的突现能力的认识存在微妙差异，但总体方向相似)

上下文学习

In-context learning

定义：假设现在有一个新的任务，只需要给语言模型提供任务描述或几个示例，它可以产生预期的输出，而不需要额外的训练和梯度更新。

- Few-Shot：输入10~100个包含上文和回复的示例
- One-Shot：输入1个包含上文和回复的示例
- Zero-Shot：只输入目标任务的自然语言描述

Setting	LAMBADA (acc)	LAMBADA (ppl)	StoryCloze (acc)	HellaSwag (acc)
SOTA	68.0 ^a	8.63 ^b	91.8^c	85.6^d
GPT-3 Zero-Shot	76.2	3.00	83.2	78.9
GPT-3 One-Shot	72.5	3.35	84.7	78.1
GPT-3 Few-Shot	86.4	1.92	87.7	79.3

指令服从

Instruction following

定义：模型可以只根据指令形式的任务描述执行新任务，而不需要给出具体示例。

前提：

- 经过一组使用自然语言描述的多任务数据集微调（instruction tuning）
- 模型参数规模至少62B

值得注意的是，对较小的模型，指令微调仍然可能提高模型的性能。

Params	Model	Norm. avg.	MMLU		BBH		TyDiQA	MGSM
			Direct	CoT	Direct	CoT	Direct	CoT
80M	T5-Small	-9.2	26.7	5.6	27.0	7.2	0.0	0.4
	Flan-T5-Small	-3.1 (+6.1)	28.7	12.1	29.1	19.2	1.1	0.2
250M	T5-Base	-5.1	25.7	14.5	27.8	14.6	0.0	0.5
	Flan-T5-Base	6.5 (+11.6)	35.9	33.7	31.3	27.9	4.1	0.4
780M	T5-Large	-5.0	25.1	15.0	27.7	16.1	0.0	0.3
	Flan-T5-Large	13.8 (+18.8)	45.1	40.5	37.5	31.5	12.3	0.7
3B	T5-XL	-4.1	25.7	14.5	27.4	19.2	0.0	0.8
	Flan-T5-XL	19.1 (+23.2)	52.4	45.5	41.0	35.2	16.6	1.9
11B	T5-XXL	-2.9	25.9	18.7	29.5	19.3	0.0	1.0
	Flan-T5-XXL	23.7 (+26.6)	55.1	48.6	45.3	41.4	19.0	4.9
8B	PaLM	6.4	24.3	24.1	30.8	30.1	25.0	3.4
	Flan-PaLM	21.9 (+15.5)	49.3	41.3	36.4	31.1	47.5	8.2
62B	PaLM	28.4	55.1	49.0	37.4	43.0	40.5	18.2
	Flan-PaLM	38.8 (+10.4)	59.6	56.9	47.5	44.9	58.7	28.5
540B	PaLM	49.1	71.3	62.9	49.1	63.7	52.9	45.9
	Flan-PaLM	58.4 (+9.3)	73.5	70.9	57.9	66.3	67.8	57.0
62B	cont-PaLM	38.1	61.2	57.6	41.7	53.1	45.7	32.0
	Flan-cont-PaLM	46.7 (+8.6)	66.1	62.0	51.0	53.3	62.7	40.3
540B	U-PaLM	50.2	71.5	64.0	49.2	62.4	54.6	49.9
	Flan-U-PaLM	59.1 (+8.9)	74.1	69.8	59.3	64.9	68.3	60.4

人类对模型输出的评价

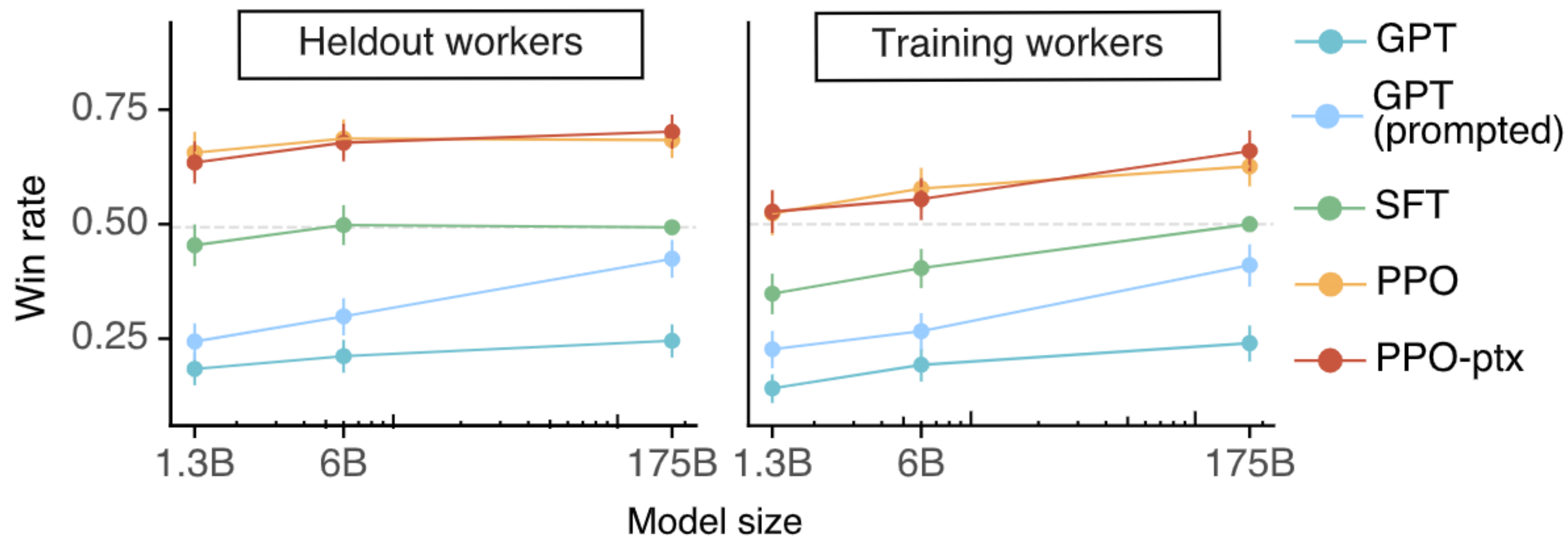


Figure 3: Preference results of our models, measured by winrate against the 175B SFT model.

强化学习 (PPO)、监督微调 (SFT)

逐步推理

Step-by-step reasoning

- 核心：Chain of Thought (CoT)

示例：

- 问题：小明有5个乒乓球，又买了2罐子乒乓球，每个罐子有3个乒乓球。请问小明现在有多少个乒乓球？
- 回答：小明从5个乒乓球开始，2个罐子，每罐3个，总共6个乒乓球。5+6=11。
答案是11。

Zero-shot CoT

输入：

- 问题：小明有5个乒乓球，又买了2罐子乒乓球，每个罐子有3个乒乓球。请问小明现在有多少个乒乓球？
- 回答： **Let's think step by step**

输出：

- **Let's think step by step** 把木棒锯成5段，需要4次。锯一次2分钟，4次需要4*2=8 分钟。答案是8分钟。

人工CoT

输入：

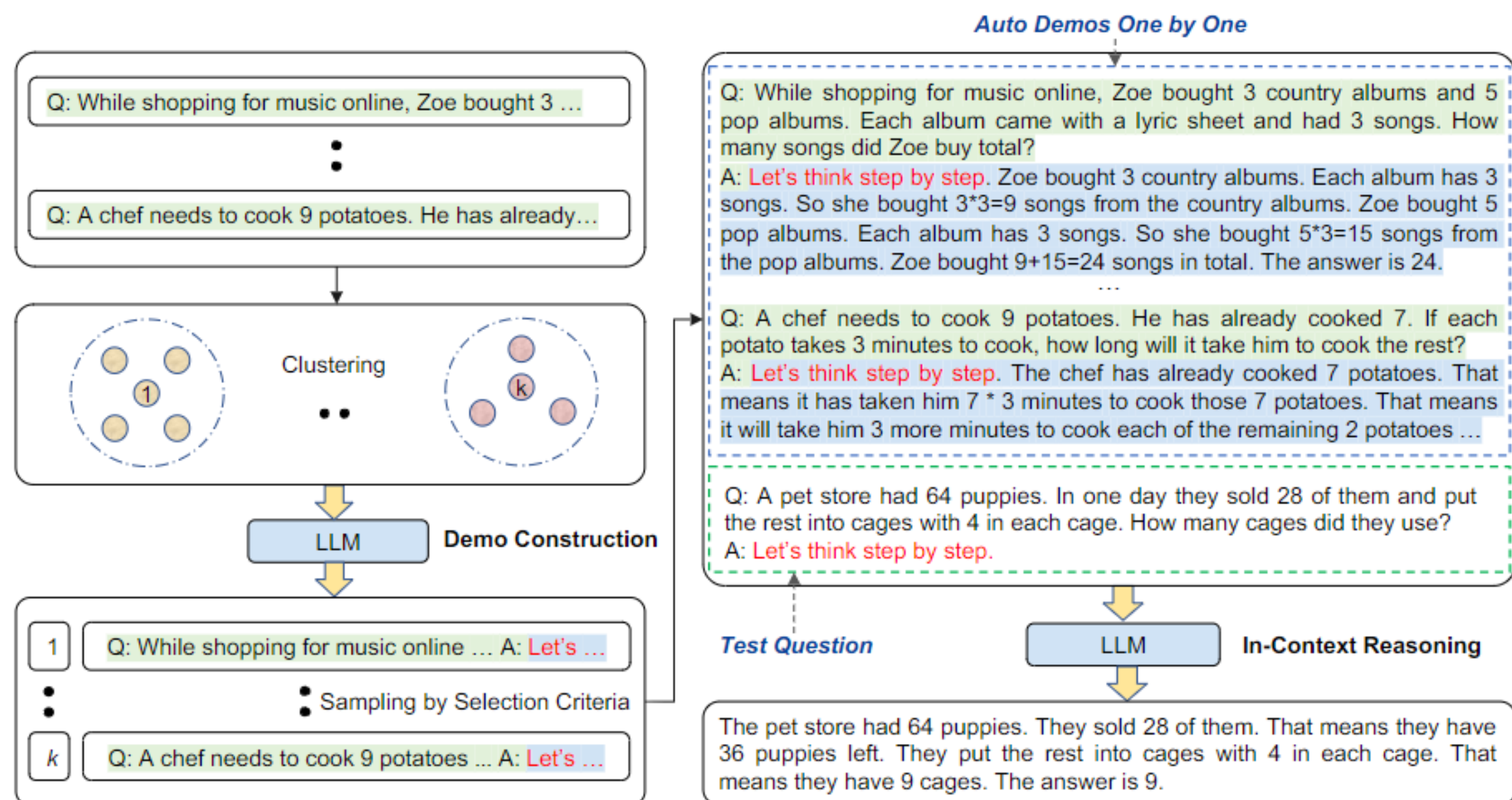
- 问题：小明有5个乒乓球，又买了2罐子乒乓球，每个罐子有3个乒乓球。请问小明现在有多少个乒乓球？
- 回答：小明从5个乒乓球开始，2个罐子，每罐3个，总共6个乒乓球。 $5+6=11$ 。答案是11。
- 问题：锯一根10米长的木棒，每锯一段要2分钟。如果把这根木棒锯成相等的5段，请问需要多长时间？

输出：

- 把木棒锯成5段，需要4次。锯一次2分钟，4次需要 $4*2=8$ 分钟。答案是8分钟。

Auto-CoT

1. 聚类选取有代表性的问题
2. 对于每一个采样的问题拼接上“Let's think step by step”（类似于 Zero-Shot-CoT）
3. 输入到语言模型，让语言模型生成中间推理步骤和答案，然后把所有采样的问题以及语言模型生成的中间推理步骤和答案全部拼接在一起
4. 使用上面生成的内容作为问题的样例，拼接上需要求解的问题一起输入到语言模型



CoT Emergent Scale

- 当模型规模大于62B（根据模型和任务的不同，这个临界值有差异），CoT才比直接回答问题存在优势
- 如果模型规模较小，CoT反而会导致性能降低。

Params	Model	Norm. avg.	MMLU		BBH		TyDiQA	MGSM
			Direct	CoT	Direct	CoT	Direct	CoT
80M	T5-Small	-9.2	26.7	5.6	27.0	7.2	0.0	0.4
	Flan-T5-Small	-3.1 (+6.1)	28.7	12.1	29.1	19.2	1.1	0.2
250M	T5-Base	-5.1	25.7	14.5	27.8	14.6	0.0	0.5
	Flan-T5-Base	6.5 (+11.6)	35.9	33.7	31.3	27.9	4.1	0.4
780M	T5-Large	-5.0	25.1	15.0	27.7	16.1	0.0	0.3
	Flan-T5-Large	13.8 (+18.8)	45.1	40.5	37.5	31.5	12.3	0.7
3B	T5-XL	-4.1	25.7	14.5	27.4	19.2	0.0	0.8
	Flan-T5-XL	19.1 (+23.2)	52.4	45.5	41.0	35.2	16.6	1.9
11B	T5-XXL	-2.9	25.9	18.7	29.5	19.3	0.0	1.0
	Flan-T5-XXL	23.7 (+26.6)	55.1	48.6	45.3	41.4	19.0	4.9
8B	PaLM	6.4	24.3	24.1	30.8	30.1	25.0	3.4
	Flan-PaLM	21.9 (+15.5)	49.3	41.3	36.4	31.1	47.5	8.2
62B	PaLM	28.4	55.1	49.0	37.4	43.0	40.5	18.2
	Flan-PaLM	38.8 (+10.4)	59.6	56.9	47.5	44.9	58.7	28.5
540B	PaLM	49.1	71.3	62.9	49.1	63.7	52.9	45.9
	Flan-PaLM	58.4 (+9.3)	73.5	70.9	57.9	66.3	67.8	57.0
62B	cont-PaLM	38.1	61.2	57.6	41.7	53.1	45.7	32.0
	Flan-cont-PaLM	46.7 (+8.6)	66.1	62.0	51.0	53.3	62.7	40.3
540B	U-PaLM	50.2	71.5	64.0	49.2	62.4	54.6	49.9
	Flan-U-PaLM	59.1 (+8.9)	74.1	69.8	59.3	64.9	68.3	60.4

CoT来源分析

- CoT出现的必要条件是模型达到一定规模，但规模并不是CoT的充分条件。

在175B的初始GPT-3、OPT、BLOOM等模型中都没有表现出CoT能力。（表现出CoT指采用CoT prompting后，模型性能比普通prompting和微调T5-11B更高）

- CoT与训练集中的代码数据相关。

使用代码数据作为训练集，大小为GPT-3数据集28%的CodeX，展现了较强的链式推理能力。

- CoT与指令微调相关

GPT-3版本text-davinci-001链式推理能力差，而经过强化学习指令微调后的版本text-davinci-002，模型链式推理能力显著增强。

其他突现能力

- **Few-Shot Prompted Tasks** 在模型达到一定规模后，能通过few-shot prompting策略在benchmark上取得突破性的准确率增长
- **Augmented Prompting Strategies** 能够增强LLM的prompting技巧。如果某些技巧在模型达到一定程度才能起作用，在此之前都没有效果甚至有害，那也称其为突现能力

注意，一个LLM并不一定包含所有突现能力，包含突现能力也不说明LLM一定优于PLM。

	Emergent scale		Model	Reference
	Train. FLOPs	Params.		
Few-shot prompting abilities				
• Addition/subtraction (3 digit)	2.3E+22	13B	GPT-3	Brown et al. (2020)
• Addition/subtraction (4-5 digit)	3.1E+23	175B	GPT-3	Hendrycks et al. (2021a)
• MMLU Benchmark (57 topic avg.)	3.1E+23	175B		
• Toxicity classification (CivilComments)	1.3E+22	7.1B		
• Truthfulness (Truthful QA)	5.0E+23	280B	Gopher	Rae et al. (2021)
• MMLU Benchmark (26 topics)	5.0E+23	280B	GPT-3	Patel & Pavlick (2022)
• Grounded conceptual mappings	3.1E+23	175B		
• MMLU Benchmark (30 topics)	5.0E+23	70B		
• Word in Context (WiC) benchmark	2.5E+24	540B		
• Many BIG-Bench tasks (see Appendix E)	Many	Many		
Augmented prompting abilities				
• Instruction following (finetuning)	1.3E+23	68B	FLAN	Wei et al. (2022a)
• Scratchpad: 8-digit addition (finetuning)	8.9E+19	40M	LaMDA	Nye et al. (2021)
• Using open-book knowledge for fact checking	1.3E+22	7.1B	Gopher	Rae et al. (2021)
• Chain-of-thought: Math word problems	1.3E+23	68B	LaMDA	Wei et al. (2022b)
• Chain-of-thought: StrategyQA	2.9E+23	62B	PaLM	Chowdhery et al. (2022)
• Differentiable search index	3.3E+22	11B	T5	Tay et al. (2022b)
• Self-consistency decoding	1.3E+23	68B	LaMDA	Wang et al. (2022b)
• Leveraging explanations in prompting	5.0E+23	280B	Gopher	Lampinen et al. (2022)
• Least-to-most prompting	3.1E+23	175B	GPT-3	Zhou et al. (2022)
• Zero-shot chain-of-thought reasoning	3.1E+23	175B	GPT-3	Kojima et al. (2022)
• Calibration via P(True)	2.6E+23	52B	Anthropic	Kadavath et al. (2022)
• Multilingual chain-of-thought reasoning	2.9E+23	62B	PaLM	Shi et al. (2022)
• Ask me anything prompting	1.4E+22	6B	EleutherAI	Arora et al. (2022)

REFERENCE

- Polosukhin, A. V., Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia. (n.d.). *Attention is All you Need*.
- Sutskever**, A. R., Jeffrey Wu*, Rewon Child, David Luan, Dario Amodei**, Ilya. (n.d.). *Language Models are Unsupervised Multitask Learners*.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35. <https://doi.org/10.1145/3560815>
- Harnessing the Power of LLMs in practice: A survey on chatgpt and beyond. (n.d.).
- Adilov, S. (2021). Generative pre-training from molecules.
- Winata, G. I., Madotto, A., Lin, Z., Liu, R., Yosinski, J., & Fung, P. (2021). Language models are few-shot multilingual learners. *Proceedings of the 1st Workshop on Multilingual Representation Learning*.

谢谢大家