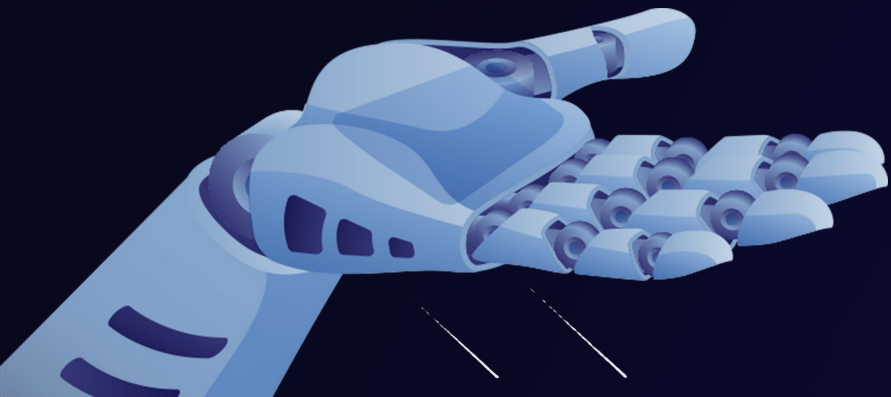


第六章：GPT模型系列理论介绍





CONTENTS



6.1 GPT简介：基于 Transformer 解码器的预训练生成模型

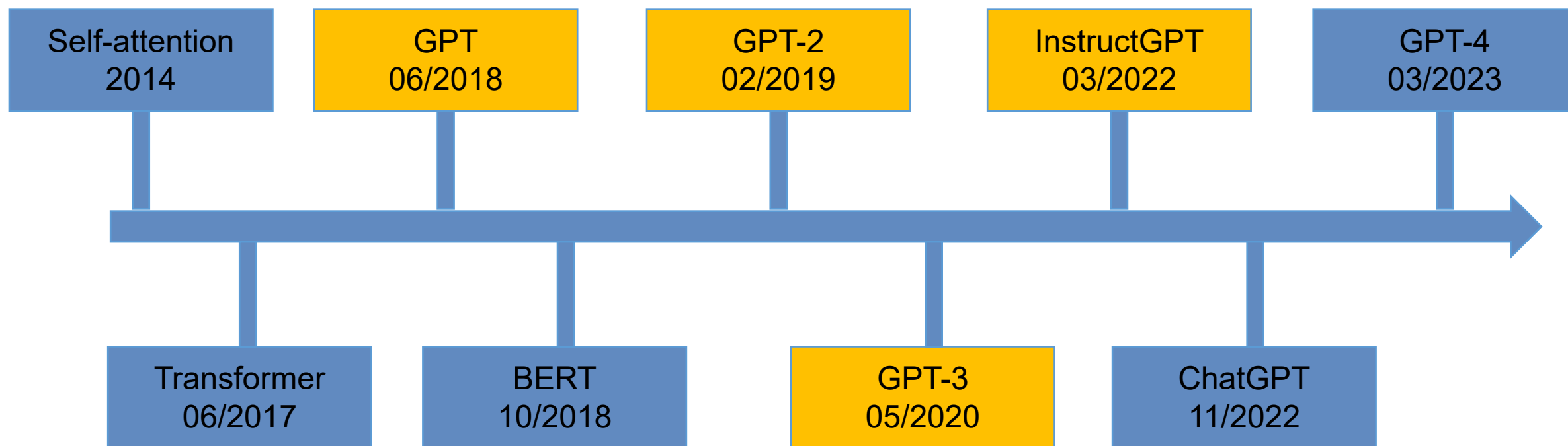
6.2 GPT-1：预训练+微调

6.3 GPT-2：预训练+Zero Shot

6.4 GPT-3：预训练+Few Shot

6.5 InstructGPT：向人类对齐

6.1 GPT简介



6.1 GPT简介

1、GPT1

通常也叫GPT，是openai公司开发的大模型，全称为：Generative Pre-trained Transformer。首次在NLP领域提出了预训练+微调的思想。

论文名称：Improving Language Understanding by Generative Pre-Training。

论文地址：https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

2、GPT2

扩大了数据量和模型容量，并首次提出了ZERO SHOT的思想。

论文名称：Language Models are Unsupervised Multitask Learners。

论文地址：<https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf>

3、GPT3

继续扩大了数据量和模型容量，将模型提升了175B规模，并在FEW SHOT设置下取得了惊艳的效果。

论文名称：Language Models are Few-Shot Learners。

论文地址：<https://splab.sdu.edu.cn/GPT3.pdf>

4、InstructGPT

为解决GPT3会产生不真实、无帮助、有毒的内容，InstructGPT提出了基于人类反馈进行对齐的思想，并取得了很好的效果。

论文名称：Training language models to follow instructions with human feedback。

论文地址：
https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

6.2 GPT-1: 预训练+微调

简介: NLP大多数任务需要人工标注数据, 这既耗时, 又昂贵。openai团队提出了基于未标注的文本和transformer的解码器进行自监督学习, 得到一个预训练的语言模型, 然后仅需将输入进行一些改动, 就可以将预训练模型用于各种NLP任务上, 称为微调。在12项任务评估中, 有9项有显著改进, 比如在常识推理 (故事完形填空测试) 中有8.9%的绝对改进, 在问答任务 (RACE) 中有5.7%的改进, 在文本蕴含 (MultiNLI) 中1.5%的改进。

6.2 GPT-1: 预训练+微调

预训练：基于文本串进行自监督训练，使用标准语言模型来最大下面的似然函数 L_1 ，这里 k 是内容窗口，条件概率 P 是带参数 Θ 的神经网络模型。

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$

$P(u)$ 的计算过程见下公式：首先将输入映射到embedding层的向量，并加上位置向量，得到 h_0 ，然后经过一个多层的transformer解码器到 h_l ，最后进行unembedding变换和softmax变换得到下一个词 u 的预测概率分布。

$$\begin{aligned} h_0 &= UW_e + W_p \\ h_l &= \text{transformer_block}(h_{l-1}) \forall i \in [1, n] \\ P(u) &= \text{softmax}(h_n W_e^T) \end{aligned}$$

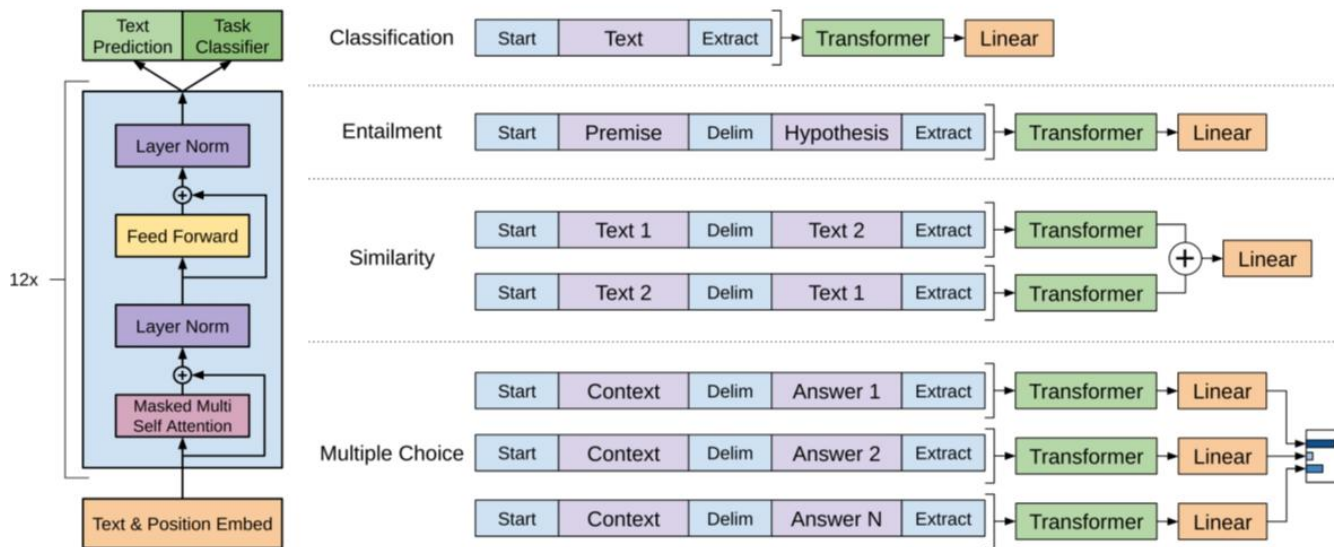
6.2 GPT-1: 预训练+微调

微调: 可用于等多种任务。假如输入是 (x_1, x_2, \dots, x_m) ，输出是 y ，则输出 y 的概率可用下左公式表示，最大化下右公式进行模型微调。

$$P(y|x^1, \dots, x^m) = \text{softmax}(h_l^m W_y).$$

$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \dots, x^m).$$

对不同的任务，输入格式有所不同，但Transformer部分是相同的。下图是分类、蕴含、相似度和多项选择任务的输入形式，与模型架构。



6.2 GPT-1: 预训练+微调

数据:

1) **BookCorpus**: 由多伦多大学和麻省理工学院 (MIT) 的研究团队首次创建和引入的数据集, 包含7000本未出版的书籍, 有探险、幻想和浪漫等多种类型, 包含约9.85亿个词语。

数据地址: <https://huggingface.co/datasets/bookcorpus/bookcorpus>。

2) **1B Word Benchmark**: 由 Google Research 创建的大规模文本数据集, 包含10亿词语, 主要用于训练和评估自然语言处理模型。数据地址: <https://github.com/ciprian-chelba/1-billion-word-language-modeling-benchmark>。

训练:

1) **解码器**: 层数为12层、模型维度为768维、头数为12个、前馈神经网络内部为3072维。

2) **优化算法**: Adam优化器。

3) **学习率**: 在前2000次更新中从0开始线性增加, 最大学习率为 $2.5e-4$, 然后使用余弦调度退火到0。

4) **其他参数**: batchsize: 64, max_seq_len: 512, epochs: 100, 权重初始化: 使用一个正态分布 $N(0, 0.02)$ 生成, 分词方法: 字节对编码 (BPE), 词表长度: 40000, 激活函数: 高斯误差线性单元 (GELU), 位置编码: 可学习编码, 原始数据清洗: 使用ftfy库来清理BookCorpus的原始文本, 标准化一些标点符号和空格。

5) **微调参数**: dropout: 0.1, batchsize: 32, epochs: 3。

6.3 GPT-2: 预训练+Zero Shot

简介：GPT2在模型架构上沿用了GPT1，使用了更大的预训练数据和模型尺寸（最大的GPT-2模型参数量为15亿），而且提出了ZERO SHOT思想，即不需要微调就可以执行特定任务。在论文提到的8个测试中，零样本设置下，有7个达到了最先进的水平。

6.3 GPT-2: 预训练+Zero Shot

方案：首先在一个大的数据集上训练一个Transformer语言模型，然后基于预训练中学习到的知识用于ZERO SHOT预测。因为预训练数据集上集包含了大量多样化的文本，其中就有各种任务的描述和格式，比如问答、翻译等，因此要执行一个翻译任务，可以输入（translate to french, english text, french text），要执行一个文本摘要任务，可以输入（answer the question, document, question, answer），预训练语言模型将根据学习到的知识和自注意力机制理解这些指令上下文，并预测指令的输出。

数据：数据集WebText为openai团队创建的，首先抓取Reddit上2017年12月之前的网页链接，这些链接至少获得了3个karma分数，这种方法可以判断其他用户是否认为该链接有趣、教育性或搞笑。总共抓取了4500万个链接，从中清洗挑选了约800万个文档，总共约40GB的文本数据。所有Wikipedia文档都被移除，因为它们是所有其他数据集中常见的数据源，在做测试评估时会和训练数据冲突。

6.3 GPT-2: 预训练+Zero Shot

训练:

- 1) 输入: 使用字节对编码 (BPE), 词库大小50527。
- 2) 参数: 相对于GPT1的改进, model_size继续扩大, batchsize用了512。
- 3) 尺寸: 尺寸见下图, 117M与GPT1一样, 345M与BERT LARGE一样, 1542M称为GPT-2, 比GPT1超了一个数量级的参数量。

BPE编码方法

- 1) 初始化词库: 词库初始包含所有单个字符。
- 2) 统计字符对频率: 计算文本中每对相邻字符的出现频率。
- 3) 合并高频字符对: 将最常见的字符对合并为一个新的符号, 并将其添加到词库中。
- 4) 重复步骤2和3: 不断重复, 直到词库达到预定大小。

Parameters	Layers	d_{model}
117M	12	768
345M	24	1024
762M	36	1280
1542M	48	1600

评估: 在零样本迁移任务上, 8个有7个都达到了优秀的结果。在小数据集Penn Treebank和WikiText-2上取得了大的提升, 这些数据集只有1到2百万个训练tokens。还有长依赖数据集上也取得了大的进步, 比如LAMBADA和Children's Book Test。在One Billion Word Benchmark上效果比之前要差很多, 可能有两个原因, 一是它是最大的数据集, 二是它进行了破坏性的预处理-句子级的洗牌, 这去掉了长文本结构。

Language Models are Unsupervised Multitask Learners

	LAMBADA (PPL)	LAMBADA (ACC)	CBT-CN (ACC)	CBT-NE (ACC)	WikiText2 (PPL)	PTB (PPL)	enwik8 (BPB)	text8 (BPC)	WikiText103 (PPL)	1BW (PPL)
SOTA	99.8	59.23	85.7	82.3	39.14	46.54	0.99	1.08	18.3	21.8
117M	35.13	45.99	87.65	83.4	29.41	65.85	1.16	1.17	37.50	75.20
345M	15.60	55.48	92.35	87.1	22.76	47.33	1.01	1.06	26.37	55.72
762M	10.87	60.12	93.45	88.0	19.93	40.31	0.97	1.02	22.05	44.575
1542M	8.63	63.24	93.30	89.05	18.34	35.76	0.93	0.98	17.48	42.16

6.4 GPT-3: 预训练+Few Shot

简介：GPT-3在模型架构上继续沿用GPT-2，但将模型参数扩大到了1750亿，同时测试模型在in-context学习，尤其是few-shot设置下的表现，因为人类通过一少部分样例，或简单地提示就可以完成一个新的语言任务。结果，GPT-3在很多NLP数据上表现良好，包括翻译、问答和完形填空，还有一些即时推理或领域适应的场景，比如拼词、句子中使用新词、3位数数学运算。

6.4 GPT-3: 预训练+Few Shot

方案：基本预训练方案，包括模型、数据和训练和GPT-2很相似，只是简单地增大了模型大小、数据的大小和多样性，以及训练时长。上下文学习和GPT-2也很相似，只是在GPT-3中，系统地探索了上下文学习不同的设置，包括zero-shot、one-shot和few-shot，这些设置可以看作依赖特定任务数据的程度，下图展示了不同设置的区别，同时还包含微调作为对比。

The three settings we explore for in-context learning

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 cheese => ..... ← prompt
```

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← example
3 cheese => ..... ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée ←
4 plush giraffe => girafe peluche ←
5 cheese => ..... ← prompt
```

Traditional fine-tuning (not used for GPT-3)

Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



6.4 GPT-3: 预训练+Few Shot

数据：包含四部分：CommonCrawl、WebText2、两个互联网书籍语料、英语维基百科，第一个数据量大，后三个数据质量高，采样权重根据其数据质量而定。

训练：

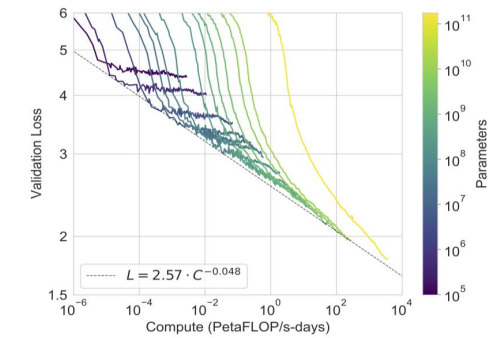
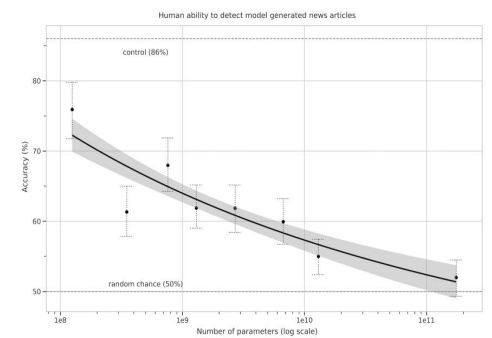
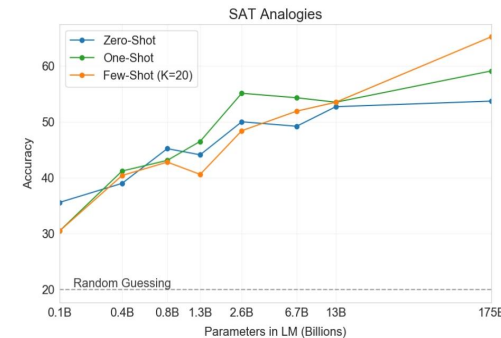
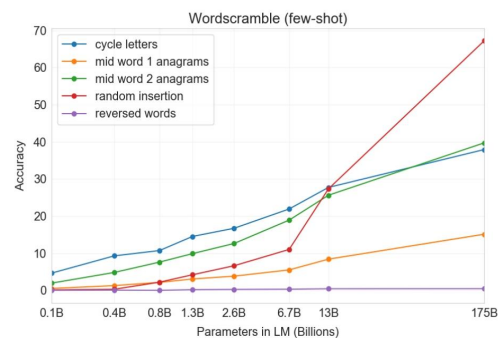
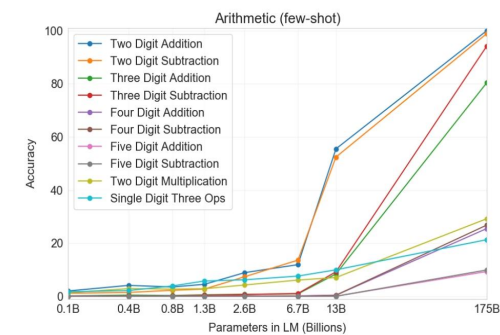
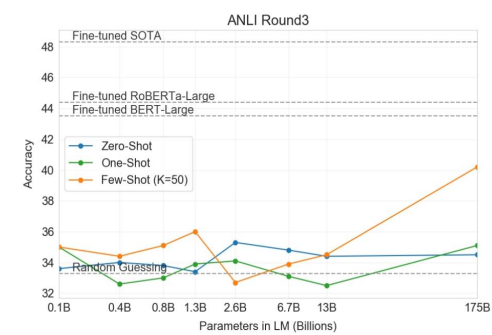
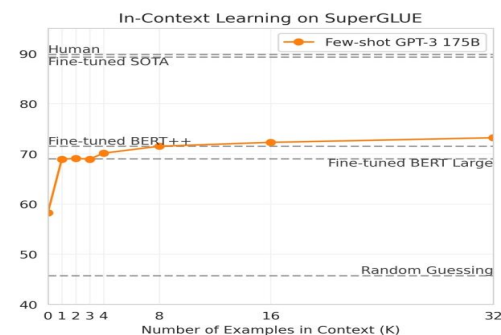
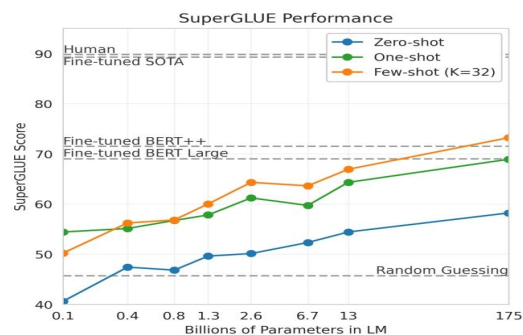
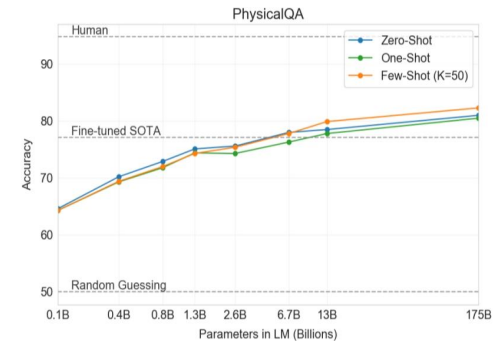
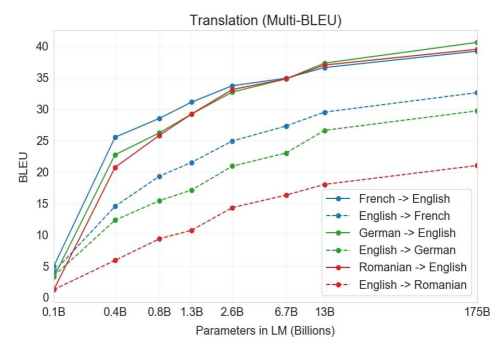
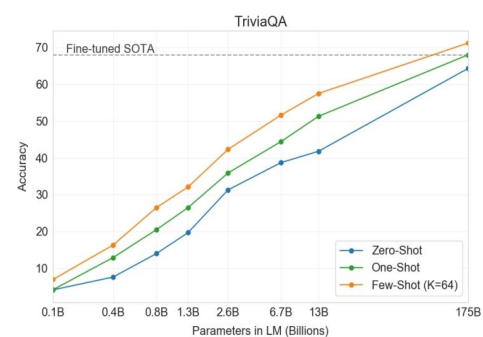
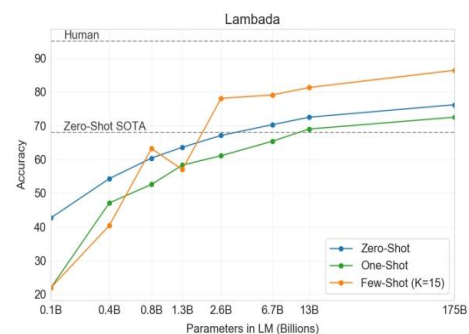
1) **transformer参数：**使用了和GPT-2一样的模型和架构，训练了8种不同尺寸的模型，从1.25亿到1750亿，跨越了三个数量级，1750亿也就是我们称之为GPT-3的模型。

2) **其它参数：**模型越大一般需要更大的batch size，但是需要更小的学习率。下表列出了主要参数设置。为了训练这个大模型时不至于内存溢出，使用了并行策略，这包括矩阵计算方面和网络层级方面。所有的模型都在V100GPU上训练，这是由微软提供的高带宽集群。

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

6.4 GPT-3: 预训练+Few Shot



6.5 InstructGPT: 向人类对齐

简介：模型变得更大并不会更好地理解用户意图，比如，大模型会产生不可信的、有毒的、无用的输出给用户。为了解决这个问题，InstructGPT提出了通过人类反馈来使语言模型适应用户意图，确切地说使模型输出是有帮助的，诚实的和无害的。结果表明，在人类评估结果中，1.3B的InstructGPT比175B的GPT-3还要好，尽管参数量小了100倍。

6.5 InstructGPT: 向人类对齐

标注者编写prompts

包含三种类型:

- **普通版**: 标注员提出任意任务, 同时保证任务具有多样性。
- **few-shot版**: 标注者提出一个指令, 和基于指令的一些查询/响应对。
- **用户版**: 在OpenAI API的申请列表中列出一些用例, 然后让标注者针对这些用例提出一些相应的prompts。

提交至playground版本API的prompts

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

1) **SFT数据集**: 标注员标注输入响应对, 格式为 (prompts, response), 共1.3万个prompts。

2) **RM数据集**: 标注员标注prompts对应输出的排名, 格式为 (prompts, response1, response2), 其中response1优于response2, 共3.3万个prompts。

3) **PPO数据集**: 仅输入prompts, 用于RLHF微调, 共3.1万个prompts。

6.5 InstructGPT: 向人类对齐

方案：使用来自人类反馈的强化学习 (RLHF) 来微调GPT-3，包含三步：

1) **训练监督微调模型 (SFT)：** 收集一个数据集（大多数是英语），输入是 prompts，输出是人工编写的输出，使用这个数据集训练一个监督学习基线模型。

2) **训练奖励模型 (RM)：** 收集另外一个数据集，格式为(prompts, response1, response2)，其中response是SFT模型生成，基于这个数据训练奖励模型 (RM)，奖励模型就是预测哪个输出是人类更偏好的一种模型。

3) **基于RM继续微调SFT模型：** 使用此RM作为奖励函数，并使用PPO算法微调学习基线模型，以最大化此奖励。

Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

Explain the moon landing to a 6 year old

A labeler demonstrates the desired output behavior.

Some people went to the moon...

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

SFT

Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

Explain the moon landing to a 6 year old

A Explain gravity... B Explain war... C Moon is natural satellite of... D People went to the moon...

A labeler ranks the outputs from best to worst.

D > C > A = B

This data is used to train our reward model.

RM

Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

Write a story about frogs

The policy generates an output.

PPO

Once upon a time...

The reward model calculates a reward for the output.

RM

The reward is used to update the policy using PPO.

r_k

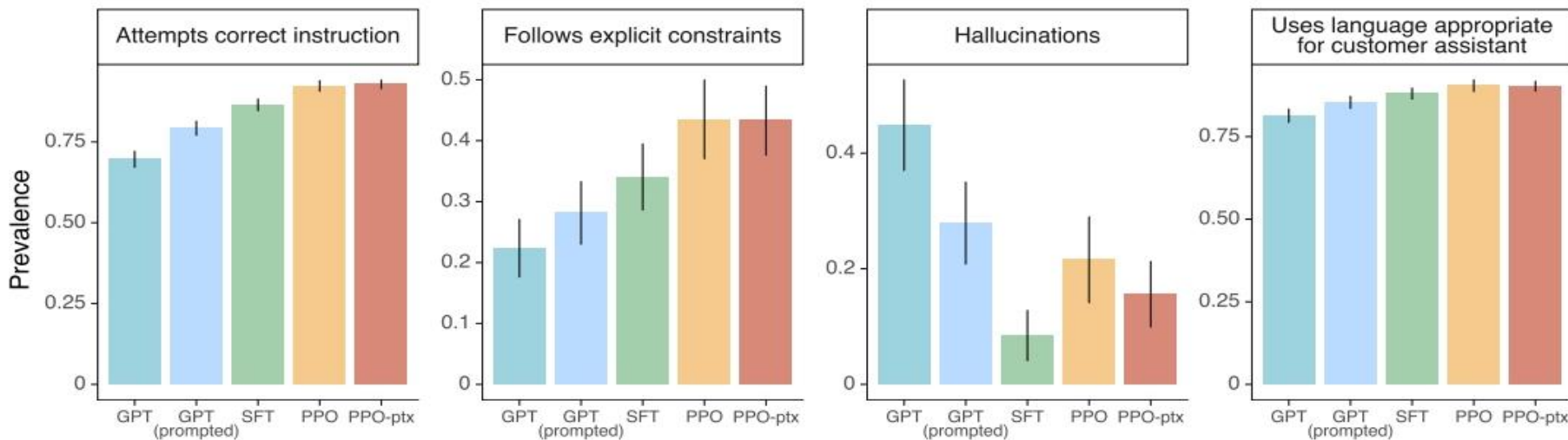
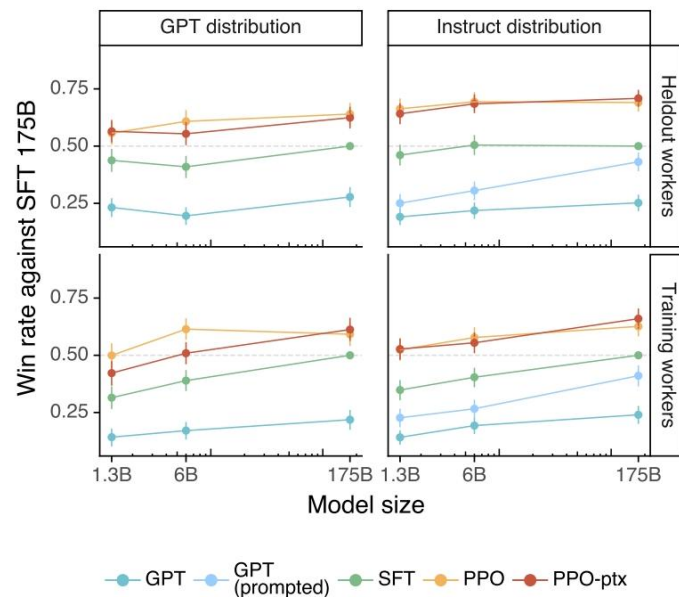
$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} E_{(x, y_w, y_l) \sim D} [\log(\sigma(r_\theta(x, y_w) - r_\theta(x, y_l)))]$$

$$\begin{aligned} \text{objective}(\phi) = & E_{(x, y) \sim D_{\pi_\phi^{\text{RL}}}} [r_\theta(x, y) - \beta \log(\pi_\phi^{\text{RL}}(y | x) / \pi^{\text{SFT}}(y | x))] + \\ & \gamma E_{x \sim D_{\text{pretrain}}} [\log(\pi_\phi^{\text{RL}}(x))] \end{aligned}$$

6.5 InstructGPT: 向人类对齐

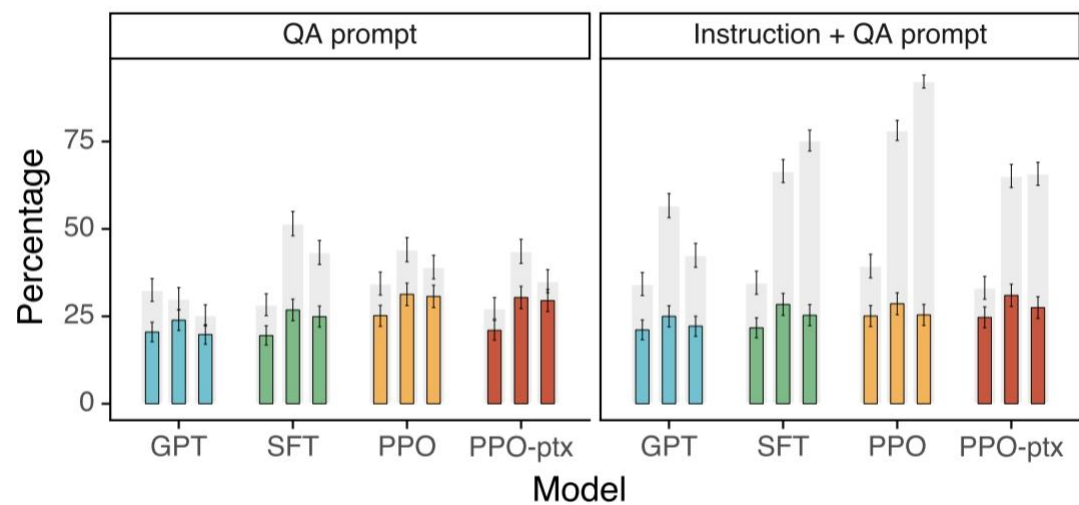
API整体评估结果（右图）：GPT-3是最差的，然后GPT-3（prompted）、SFT、PPO次第增加。增加预训练部分到PPO上并没有带来标注者偏好的显著提升。

API细分维度评估结果（下图）：具体来说，相较于GPT-3，InstructGPT输出在客户助理的内容上更合适，在明确限制的指令上回复得更好（e.g. “Write your answer in 2 paragraphs or less.”），在正确地响应指令上面表现得更好，更少地虚构事实。

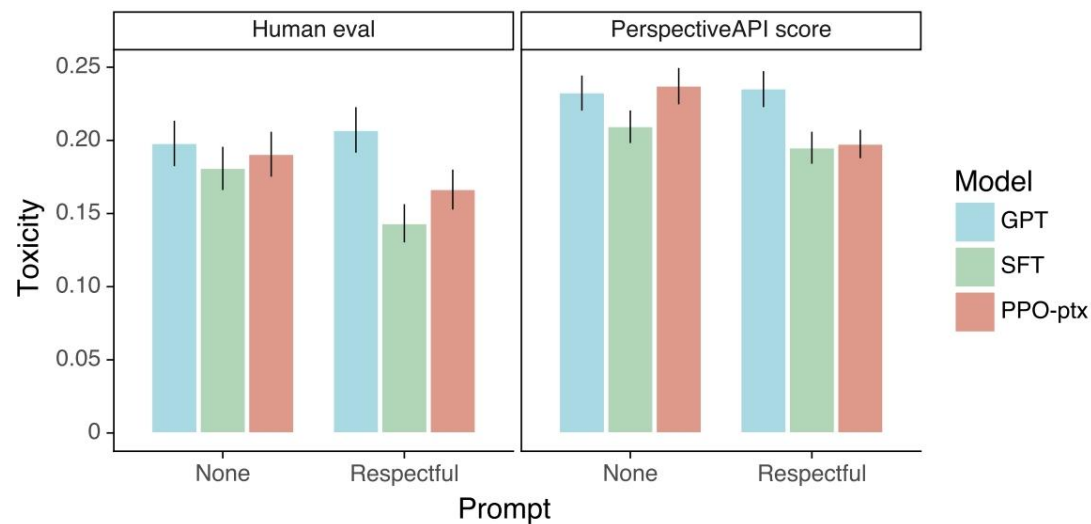


6.5 InstructGPT: 向人类对齐

InstructGPT模型在真实度上超越了GPT-3



InstructGPT在毒性方面相对于GPT-3有小的改善



感谢您的观看

