从ChatGPT看大模型的演化

鹏城实验室

王 晖

免责申明：

1. 本附加与原报告无关；
2. 本资料来源互联网公开数据；
3. 本资料在“行业报告资源群”和“知识星球行业与管理资源”均免费获取；
4. 本资料仅限社群内部学习，如需它用请联系版权方

合作与沟通，
请联系客服

客服微信　　客服微信

行业报告资源群

微信扫码 长期有效

1. 进群即领福利《报告与资源合编》，内有近百行业、上万份行研、管理及其他学习资源免费下载；
2. 每日分享学习最新6+份精选行研资料；
3. 群友交流，群主免费提供相关领域行研资料。

微信扫码 行研无忧

知识星球 行业与管理资源 是投资、产业研究、运营管理、价值传播等专业知识库，已成为产业生态圈、企业经营者及数据研究者的智慧工具。

知识星球 行业与管理资源每月更新5000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；

微信扫码加入后无限制搜索下载。

汇报提纲

一 OpenAI的使命任务与机制创新

二 ChatGPT的发展历程

三 ChatGPT的技术创新

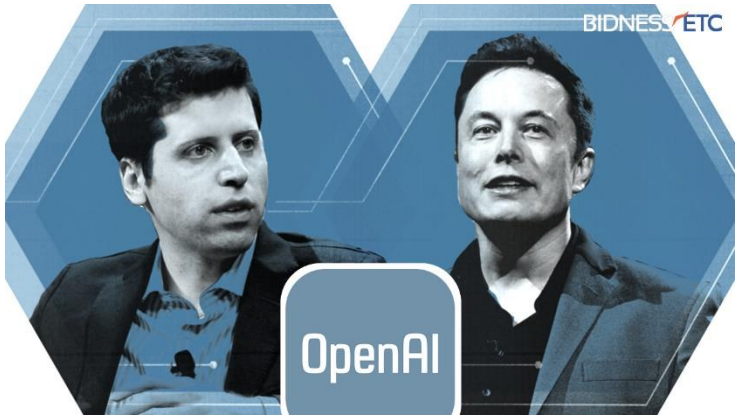
四 基于人机协作的复杂AI软件持续构造与演化

一、OpenAI的使命任务与机制创新

“尼克·波斯托姆的《超级智能》一书非常值得一读。人工智能可能会比核武器还更具危险性，我们一定要谨慎对待它。”

“避免AI潜在威胁的最好方式并不是去限制它，而是透明化地全面普及化。”

——埃隆·马斯克，2014年8月



YC总裁山姆·阿尔特曼Sam Altman

OpenAI

创立于2015年12月
非盈利研究机构

使命任务：突破AGI技术，并普惠人类。

扛开源开放大旗，打破Google和Facebook AI技术壁垒。

使命与愿景：创造一个新的世界（AI for Good）
诗和远方、星辰大海

一、OpenAI的使命任务与机制创新

OpenAI设立的核心技术研发方向：

从强化学习到无监督学习，最终实现AGI。

4款强化学习开源软件平台

第一款：Gym，2016年4月首次发布。

第二款：Universe，2016年12月首次发布。

第三款：Roboschool，2017年5月首次发布。

第四款：Blocksparse，2017年12月首次发布。

一、OpenAI的使命任务与机制创新

发展困境：需要大量资本投入大规模云计算，吸引人才、留住人才，并构建AI超级计算机。

2019年3月11日，OpenAI宣布重组：成立营利性公司OpenAI LP。

山姆·阿尔特曼离任YC，任OpenAI CEO。马斯克离开董事局。

- **OpenAI Nonprofit**：日常工作没有变化，通过开发新的 AI 技术，而非商业产品来创造出最大的价值。
- **OpenAI LP**：被称为 “有限盈利”（capped-profit) 公司，提高筹集资金的能力，增加对计算和人才方面的投资，确保通用人工智能（AGI）有益于全人类。

OpenAI 非营利部门负责管理 OpenAI LP，主持学者和研究人员等教育计划，并负责政策实施。

道路是曲折的：

不忘初心，牢记使命。
一司两制。

一、OpenAI的使命任务与机制创新

山姆·阿尔特曼：2019年3月任OpenAI CEO

提出著名的“万物摩尔定律” Moore's Law for Everything (samaltman.com)

◆ **越来越多人类的工作将被能够思考和学习的软件取代，更多的权力将从劳动力转移到资本上。需要设计一种制度拥抱这种技术化的未来，公平地分配由此产生的财富。**

OpenAI Microsoft

We’re partnering to develop a hardware and software platform within Microsoft Azure which will scale to AGI. We’ll jointly develop new Azure AI supercomputing technologies.

战略调整：合作共赢、发展是硬道理。

汇报提纲

一 OpenAI的使命任务与机制创新

二 ChatGPT的发展历程

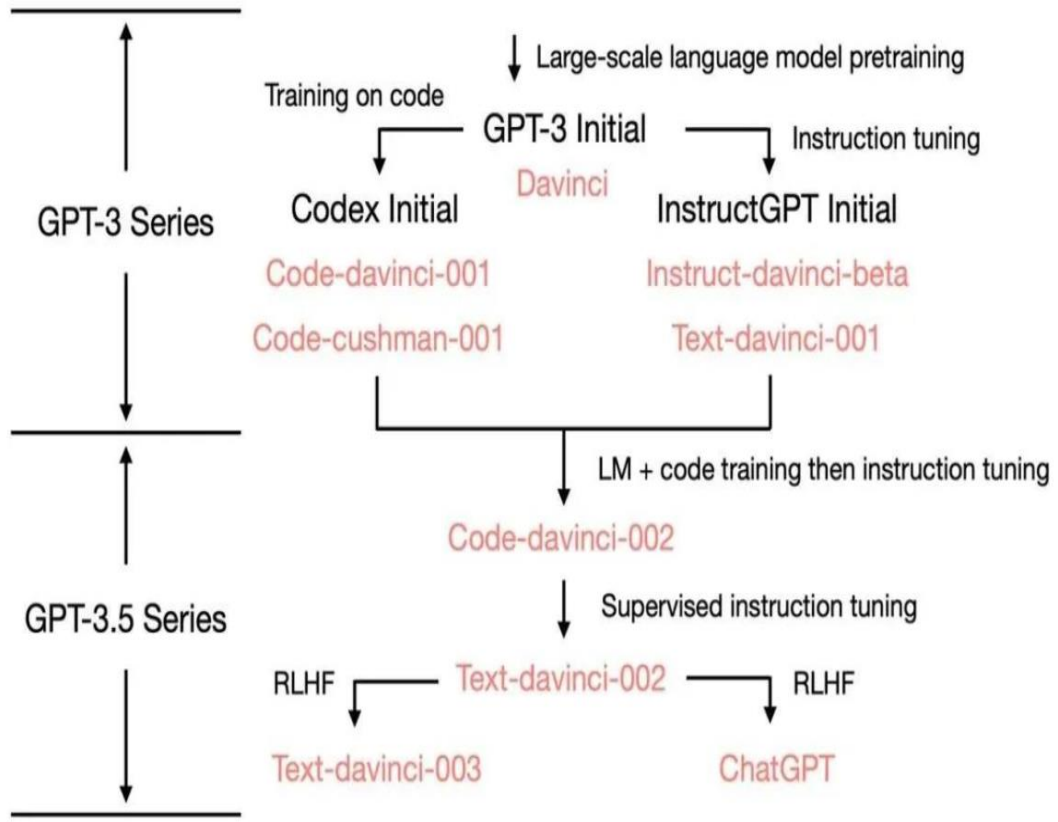
三 ChatGPT的技术创新

四 基于人机协作的复杂AI软件持续构造与演化

二、ChatGPT的发展历程——不断探索演进

无监督学习AGI模型技术探索演进

从GPT-3到GPT-3.5



时间	模型	能力
2018.06	GPT：transformer-decoder only	**1.17亿参数，语料约5GB。无监督学习：大量无标签数据上的生成式预训练模型。**
2019.02	GPT-2	**最大15亿参数，语料40GB。无监督预训练模型做有监督的任务。**
2020.07	GPT-3（davinci）	**最大1750亿参数，原始语料45TB，训练语料570GB。 生成+知识库+in-context learning**
2021.07	Codex Initial（基于12B变种GPT3微调）	**Code training code 推理，演变成Code-cushman-001**
2022.03	instruct-davinci-beta text-davinci-001	**Instruction tuning（监督指令微调） Instruction tuning（文本）**
2022.4-7	code-davinci-002（Codex） 侧重于Code	**GPT3.5变体（优于Text-davinci-002和003）+ instruction tuning 响应人类指令+泛化任务+代码能力+复杂推理(from trained on code)**
2022.5-6	Text-davinci-002 侧重于Text supervised instruction-tuned model	**降低了in-context learning能力，增强zero-shot能力**
2022.11	Text-davinci-003（RLHF） ChatGPT（RLHF）	**增加RLHF，恢复in-context learning（但比code-davinci-002 差）并改进zero-shot。增加RLHF对话历史建模,牺牲了In-context-learning能力**

二、ChatGPT的发展历程——逆向结论

- **语言生成能力 + 基础世界知识 + 上下文学习都是来自于预训练（davinci）**
- **存储大量知识的能力来自 1750 亿的参数量。**
- **遵循指令和泛化到新任务的能力来自于扩大指令学习中指令的数量（Davinci-instruct-beta）**
- **执行复杂推理的能力很可能来自于代码训练（code-davinci-002）**
- **生成中立、客观的能力、安全和翔实的答案来自与人类的对齐。具体来说：**
  - 如果是监督学习版，得到的模型是text-davinci-002
  - 如果是强化学习版 (RLHF)，得到的模型是text-davinci-003
  - 无论是有监督还是 RLHF，模型在很多任务的性能都无法超过 code-davinci-002，这种因为对齐而造成性能衰退的现象叫做**对齐税**。
- **ChatGPT对话能力也来自于 RLHF，具体来说它牺牲了上下文学习的能力，来换取：**
  - 建模对话历史
  - 增加对话信息量
  - 拒绝模型知识范围之外的问题

汇报提纲

一 OpenAI的使命任务与机制创新

二 ChatGPT的发展历程

三 ChatGPT的技术创新

四 基于人机协作的复杂AI软件持续构造与演化

三、ChatGPT的技术创新

■ ChatGPT具体功能

- 多语言翻译，支持英文、中文、日文、韩文、西班牙文、德文、法文、俄文等多语言交互，能自动识别翻译意图及语种
- 对话问答，多轮对话理解能力较好，可以较好的完成指代消解，省略回复等任务，上下文一致性较高
- 内容创作能力，代码生成（支持Python等多种编程语言，代码规范有注释）、论文写作能力、诗词创作能力、歌词创作，精通乐理知识
- 语义理解，复杂语句理解，上下文语义衔接、错误内容纠正、质疑不合理前提、拒绝恶意提问

■ ChatGPT涉及到的NLP任务

①分类 ②信息抽取 ③意图识别 ④生成 ⑤重写文本 ⑥翻译 ⑦文本摘要和总结 ⑧对话 ⑨知识问答 ⑩阅读理解 还有其他比如同义词等

■ ChatGPT不足

- 生成模型通病，真实性不确定，可能是一本正经的胡说八道
- 知识整合能力和逻辑推理能力还不算太好
- 时效性，非持续学习，内容更新至2021年

三、ChatGPT的技术创新

ChatGPT现象级创新

ARTIFICIAL INTELLIGENCE / TECH / POLICY

New York City schools ban access to ChatGPT over fears of cheating and misinformation

ICML | 2023

Fortieth International Conference on Machine Learning

Dates Calls Resources Attend Organization

Year (2023)

Help

My Stuff/Registrations

Profile

Exhibitor Info

Clarification on Large Language Model Policy LLM

We (Program Chairs) have included the following statement in the Call for Papers for ICML represented by 2023:

Papers that include text generated from a large-scale language model (LLM) such as ChatGPT are prohibited unless the produced text is presented as a part of the paper's experimental analysis.

This statement has raised a number of questions from potential authors and led some to proactively reach out to us. We appreciate your feedback and comments and would like to clarify further the intention behind this statement and how we plan to implement this policy for ICML 2023.

The End of Programming

The end of classical computer science is coming, and most of us are dinosaurs waiting for the meteor to hit.

CAME OF AGE in the 1980s, programming personal computers such as the Commodore VIC-20 and Apple][e at home. Going on to study computer science (CS) in college and ultimately getting a Ph.D. at Berkeley, the bulk of my professional training was rooted in what I will call "classical" CS: programming, algorithms, data structures, systems, programming languages. In Classical Computer Science, the ultimate goal is to reduce an idea to a program written by a human—source code in a language like Java or C++ or Python. Every idea in Classical CS—no matter how complex or sophisticated, from a database join algorithm to the mind-bogglingly obtuse Paxos consensus protocol—can be expressed as a human-readable, human-

Edward Tian
@edward_the6

I spent New Years building GPTZero — an app that can quickly and efficiently detect whether an essay is ChatGPT or human written

8:17 AM · Jan 3, 2023

2.4M Views **1,830** Retweets **655** Quote Tweets **15.4K** Likes

Simplicity Wins: How Large Language Models Will Revolutionize Software Engineering

Posted on December 23, 2022 by Andreas Kirsch

Posted by u/PapayaEqual 16 hours ago

266

My company blocked chatgpt

Educational Purpose Only

Im a junior software engineer, in my team the seniors are allways occupied and they dont have time to explain so everytime im stuck chatgpt is my saviour. Today I arrived at my office and no o[illegible] access to chatgpt it was blocked by office wifi what should we do?

新智元

Temporary policy: ChatGPT is banned

Asked today Modified today Viewed 52k times

568

Use of ChatGPT generated text for posts on Stack Overflow is temporarily banned.

This is a temporary policy intended to slow down the influx of answers created with ChatGPT. What the final policy will be regarding the use of this and other similar tools is something that will need to be discussed with Stack Overflow staff and, quite likely, here on Meta Stack Overflow.

三、ChatGPT的技术创新——涌现能力的逆向分析

How does GPT Obtain its Ability? Tracing Emergent Abilities of Language Models to their Sources

Yao Fu, yao.fu@ed.ac.uk
University of Edinburgh

Emergent Abilities of Large Language Models

Jason Wei[1] Yi Tay[1] Rishi Bommasani[2] Colin Raffel[3]
Barret Zoph[1] Sebastian Borgeaud[4] Dani Yogatama[4] Maarten Bosma[1]
Denny Zhou[1] Donald Metzler[1] Ed H. Chi[1] Tatsunori Hashimoto[2]
Oriol Vinyals[4] Percy Liang[2] Jeff Dean[1] William Fedus[1]
[1]Google Research [2]Stanford University [3]UNC Chapel Hill [4]DeepMind

论文地址：https://arxiv.org/pdf/2206.07682.pdf

- **复杂推理的能力来自于代码训练**是我们倾向于相信的假设
- **对没有见过的任务泛化能力来自大规模指令学习** 是至少 **4 篇论文**的结论
- **GPT-3.5来自于其他大型基础模型，而不是1750亿参数的GPT-3** 是有根据的猜测。
- **所有这些能力都已经存在了，通过instruction tuning，无论是有监督学习或强化学习的方式来解锁而不是注入这些能力** 是一个强有力的假设，强到你不敢不信。主要是因为instruction tuning数据量比预训练数据量少了几个数量级
- 结论 = 许多证据支持这些说法的正确性；假设 = 有正面证据但不够有力；有根据的猜测 = 没有确凿的证据，但某些因素会指向这个方向

复杂系统工程创新：

人机协作的复杂AI软件持续演化系统工程

三、ChatGPT的技术创新——复杂AI软件持续演化

复杂系统工程创新：**人机协作的复杂AI软件持续演化系统工程**

人机协作的海量数据工程

专业标注的高质量且多样性的大规模语料
(万亿级规模Token)
+
高质量的Instruction
任务描述与专业回答

对齐人类价值观的超大模型强化学习工程

无监督预训练超大语言模型
+
Reward Model
+
强化学习迭代优化

超级智能算力云大规模并行训练工程

微软云Azure
+
Ray大规模并行训练

迭代飞轮

用户访问log
+
用户反馈
(API Prompt DB)

微软云Azure
+
大规模并发推理API

超级智能算力云大规模并发推理弹性服务工程

持续演化：4年半

专业数据标注：80人以上

微软云：10亿美金投入

数据量：持续更新

用户数：百万级以上

公司规模：200人以上

ChatGPT的技术创新——海量数据工程

OpenAI Model	OpenAI API	Dataset
GPT-3 Series		
GPT-3 Initial	Davinci	300B tokens (60% 2016 - 2019 C4 + 22% WebText2 + 16% Books + 3% Wikipedia)
Instruct-GPT initial	Davinci-Instruct-Beta	instruction tuning
Codex initial	Code-Cushman-001	159G github代码
GPT-3.5 Series		
Current Codex	Code-Davinci-002	159G github代码+instruction tuning
Instruct-GPT supervised	Text-Davinci-002	159G github代码+instruction tuning
Instruct-GPT RLHF	Text-Davinci-003	少量RLHF策略调优数据
ChatGPT		✓ 少量RLHF策略调优数据；模型生成的多个答案对由人工打分构造对比数据集 ✓ 对于2021年的事件知识了解有限，有人推测训练Code-Davinci-002时使用了C4中2016-2021的数据，相比原始GPT3，序列长度也变成4096 ✓ GitHub代码数据 ✓ 预训练数据扩展了Openwebtext数据集 ✓ 奖励模型训练集是从预定义数据集中抽取一组提示来生成的，OpenAI采用的是用户提交的GPT API的提示，传递给语言模型来生成新的文本；然后人工对多个LM的输出进行排名，而不是直接标量打分

三、ChatGPT的技术创新——海量数据工程

◆ 标注数据类型分布、示例

Instruction格式构建更符合人类的表达、更自然语言，是多任务的。

Table 1: Distribution of use case categories from our API prompt dataset.

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%

Table 2: Illustrative prompts from our API prompt dataset. These are fictional examples inspired by real usage—see more examples in Appendix A.2.1.

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: """

◆ 标注数据来源：标注和用户请求

标注：标注人员标注3类prompt

Plain: We simply ask the labelers to come up with an arbitrary task, while ensuring the tasks had sufficient diversity.

Few-shot: We ask the labelers to come up with an instruction, and multiple query/response pairs for that instruction.

User-based: We had a number of use-cases stated in waitlist applications to the OpenAI API. We asked labelers to come up with prompts corresponding to these use cases.

用户请求：从用户请求中筛选prompt

…licate prompts by checking for prompts that …f prompts to roughly 200 per organization. … based on organization IDs, so that e.g. the …ng set.

“在之前的工作中，我们发现标注者经常给我们觉得很一般的文章也打高分。
所以这次，我们斥巨资保证标注质量。为提高与标注者的交流效率，我们拒绝采用第三方标注平台，
直接雇佣80名标注者到身边.....
我们对标注者每小时的标注量不做要求（如果标得太快会在一开始就被辞掉）。
我们也不按标注数目发工资，而是按工作时长，15刀/时。”

◆ RM数据构建

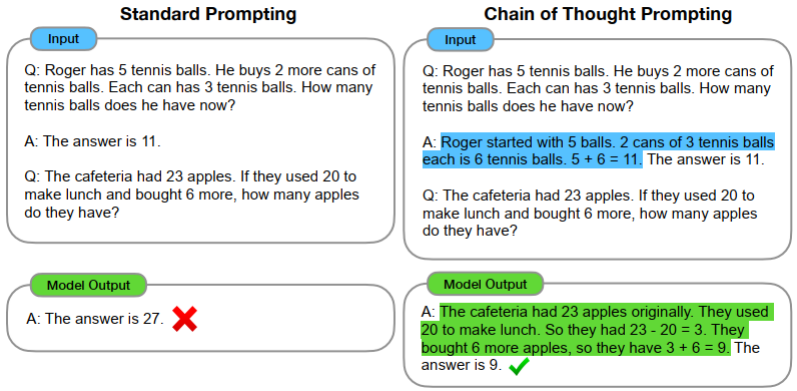
RM模型打分标准：每一…

…本时，中间增加一些解释

…聚。

Table 3: Labeler-collected metadata on the API distribution.

Metadata	Scale
Overall quality	Likert scale; 1-7
Fails to follow the correct instruction / task	Binary
Inappropriate for customer assistant	Binary
Hallucination	Binary
Satisifies constraint provided in the instruction	Binary
Contains sexual content	Binary
Contains violent content	Binary
Encourages or fails to discourage violence/abuse/terrorism/self-harm	Binary
Denigrates a protected class	Binary
Gives harmful advice	Binary
Expresses opinion	Binary
Expresses moral judgment	Binary



三、ChatGPT的技术创新——指令微调数据工程

Improve model's capability of understanding the task description.

LM for sentence completion

I went to Jolin's concert last night. I really loved her songs and dancing. It was _____

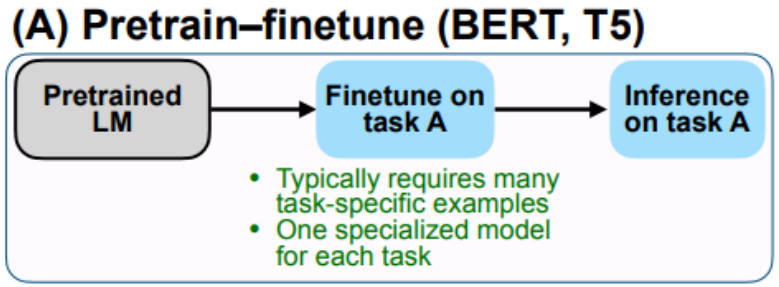
Detailed task instruction for LM generation

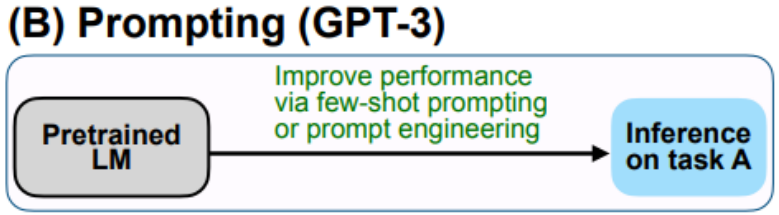
Decide the sentiment of the following sentences:
I went to Jolin's concert last night. I really loved her songs and dancing.
OPTIONS: - positive – negative - neutral

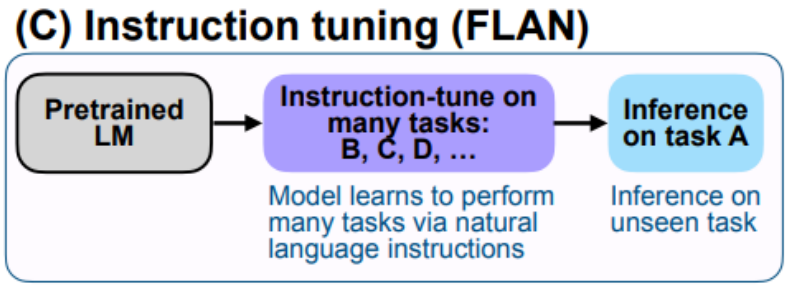
FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS, ICLR2022, Google Research.

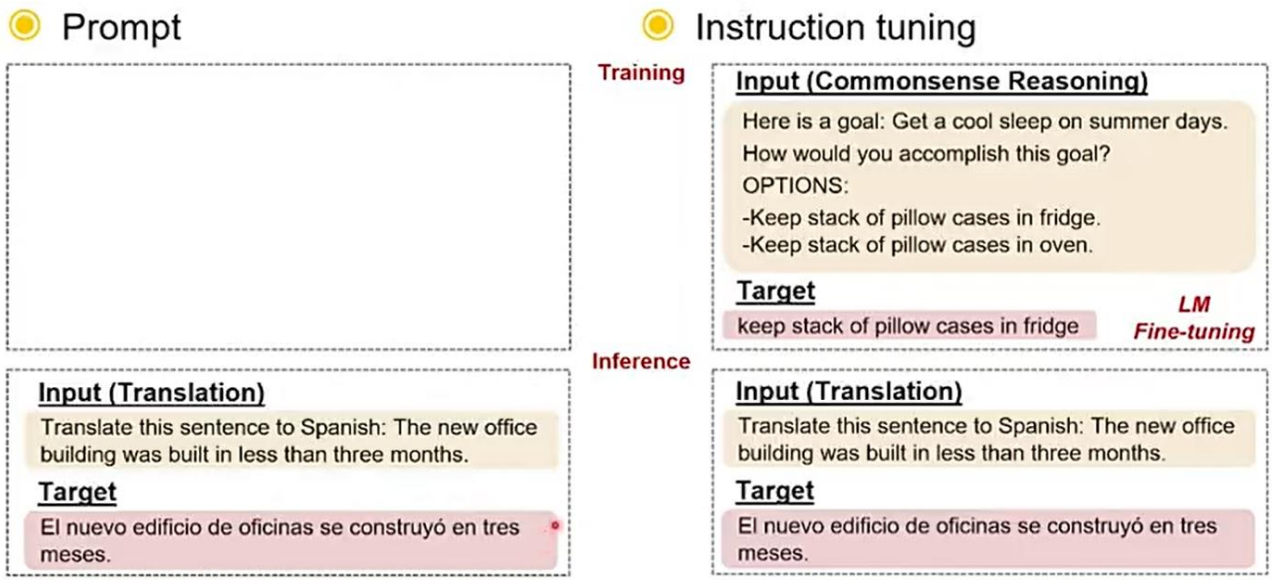
三、ChatGPT的技术创新——指令微调数据工程

Fine-tune LM to better understand task descriptions via other tasks.



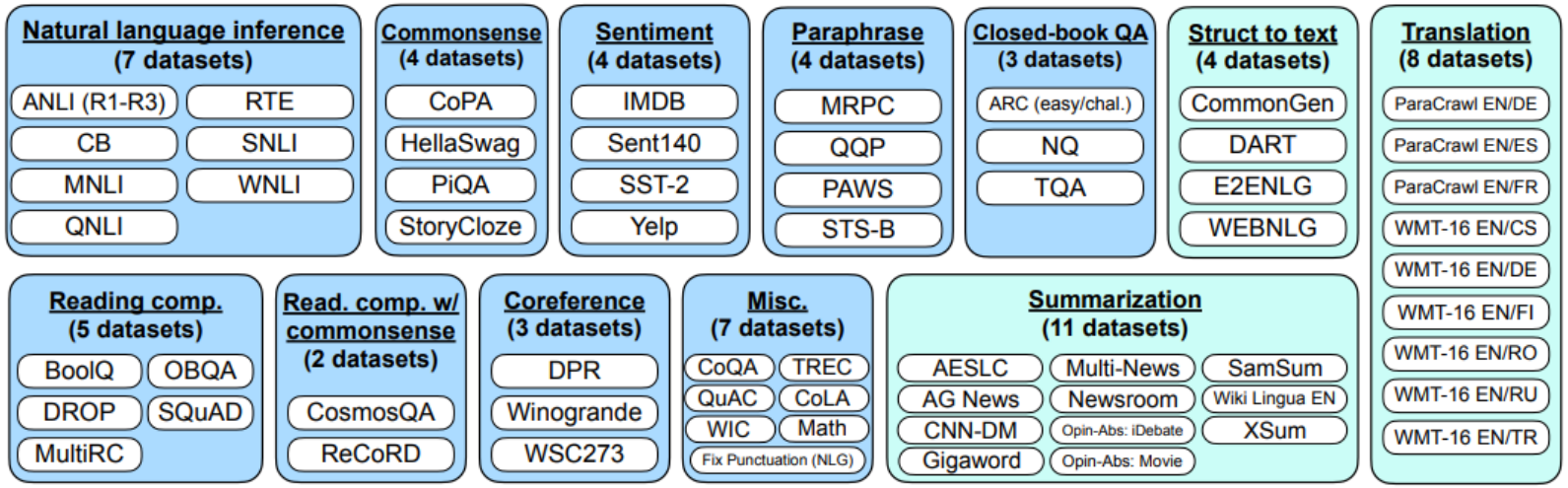


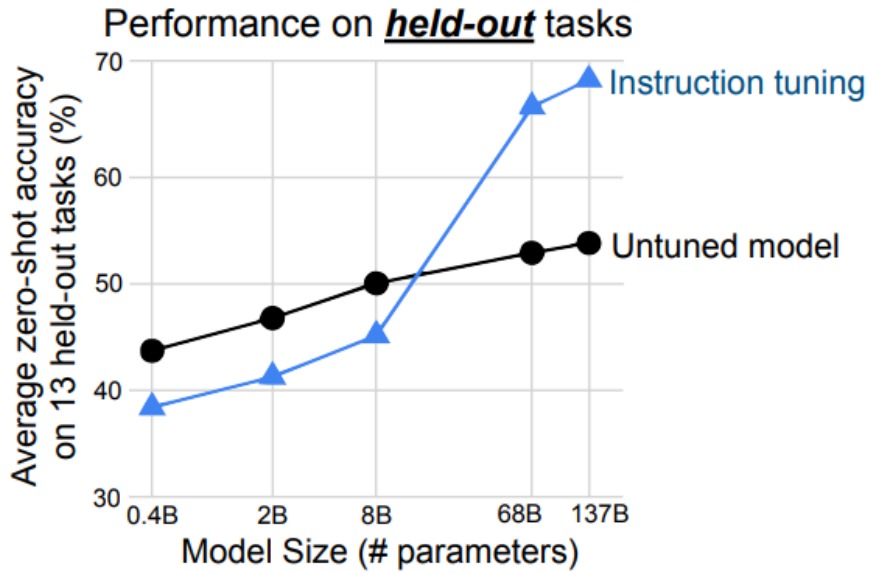




FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS, ICLR2022， Google Research.

三、ChatGPT的技术创新——指令微调数据工程





Premise

Russian cosmonaut Valery Polyakov set the record for the longest continuous amount of time spent in space, a staggering 438 days, between 1994 and 1995.

Hypothesis

Russians hold the record for the longest stay in space.

Target

Entailment
Not entailment

⇨ Options:
- yes
- no

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, ...

Scaling Laws

FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS, ICLR2022, Google Research.

三、ChatGPT的技术创新——RLHF工程

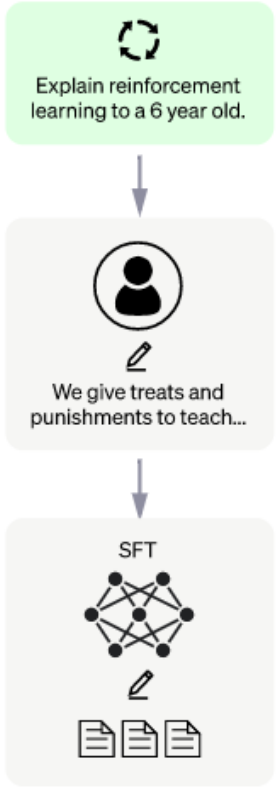
Step 1

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-3.5 with supervised learning.



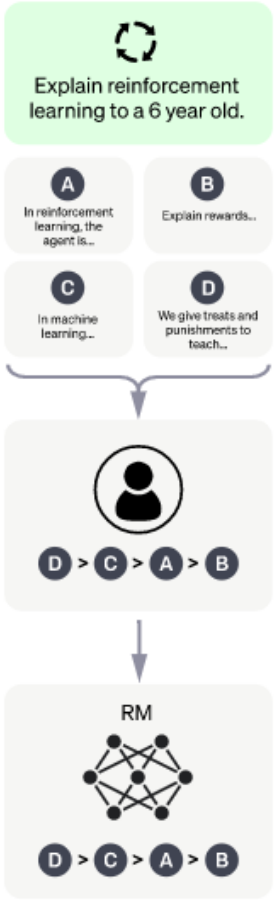
Step 2

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 the PPO reinforcement learning algorithm.

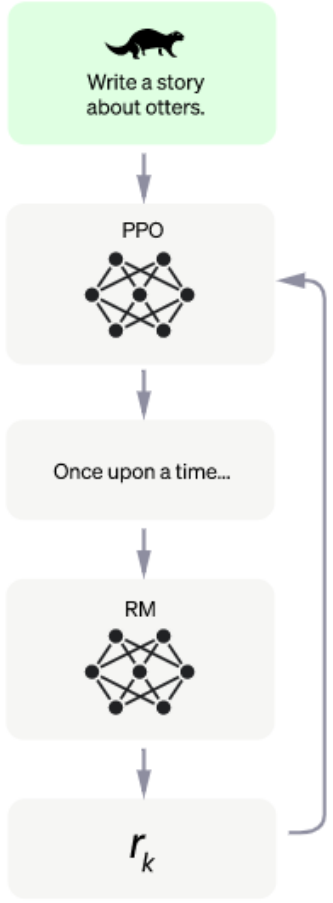
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



三、ChatGPT的技术创新——RLHF原理

maximize the following likelihood:

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

GPT1-3

the following objective to maximize:

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

无监督预训练 → $L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C})$ ← 监督微调

强化学习损失：根据RM打分产生。

与参考模型的损失差：即希望RL模型与SFT的差异不要太大，相当于一个正则化惩罚。（Goodhart's Law）

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

InstructGPT

预训练损失：策略模型的输入构建的语言模型交叉熵损失。

x：模型输入，y：模型输出

π_φ^{RL}：PPO策略模型，即要用强化学习更新的LM模型。

π^{SFT}：参考模型，用Instrucion数据fine tune过的LM模型。

$D_{pretrain}$：预训练分布，β：KL 散度的系数，γ：预训练损失系数。

Training language models to follow instructions with human feedback, OpenAI

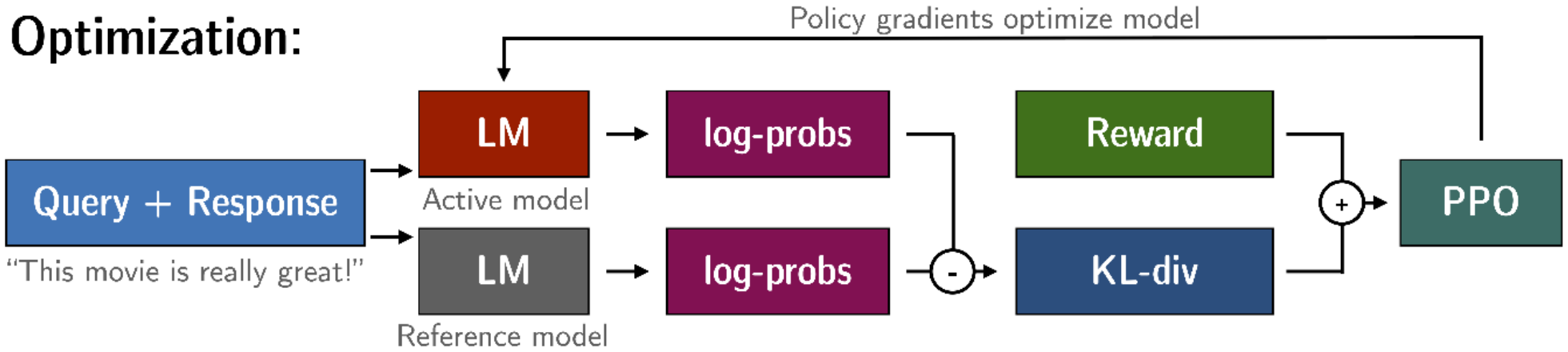
三、ChatGPT的技术创新——RLHF原理

https://github.com/lvwerra/trl

Rollout:

Evaluation:

Optimization:



Proximal Policy Optimization Algorithms，OpenAI，2017.8

三、ChatGPT的技术创新——未来世界

the Modular Reasoning, Knowledge and Language (MRKL, pronounced "miracle") system

当前的痛点：

1. **Lack of access to current Spatiotemporal information.**
2. **Lack of access to proprietary information sources.**
3. **Lack of reasoning.**
4. **Lack of Lifelong learning.**
5. **Model explosion.**

MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning

Ehud Karpas, Omri Abend, Yonatan Belinkov, Barak Lenz, Opher Lieber, Nir Ratner, Yoav Shoham, Hofit Bata, Yoav Levine, Kevin Leyton-Brown, Dor Muhlgay, Noam Rozen, Erez Schwartz, Gal Shachaf, Shai Shalev-Shwartz, Amnon Shashua, Moshe Tenenholtz

免责申明：

1. 本附加与原报告无关；
2. 本资料来源互联网公开数据；
3. 本资料在“行业报告资源群”和“知识星球行业与管理资源”均免费获取；
4. 本资料仅限社群内部学习，如需它用请联系版权方

合作与沟通，
请联系客服

客服微信　　客服微信

行业报告资源群

微信扫码 长期有效

1. 进群即领福利《报告与资源合编》，内有近百行业、上万份行研、管理及其他学习资源免费下载；
2. 每日分享学习最新6+份精选行研资料；
3. 群友交流，群主免费提供相关领域行研资料。

微信扫码 行研无忧

知识星球 行业与管理资源 是投资、产业研究、运营管理、价值传播等专业知识库，已成为产业生态圈、企业经营者及数据研究者的智慧工具。

知识星球 行业与管理资源每月更新5000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；

微信扫码加入后无限制搜索下载。

请批评指正！