

Training language models to follow instructions with human feedback

2022-NIPS



提纲

2



Authors



Introduction



Methodology



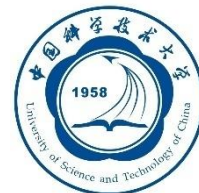
Experiments and Result



Application



Inspiration



Authors

□ Authors

Training language models to follow instructions with human feedback

Long Ouyang* Jeff Wu* Xu Jiang* Diogo Almeida* Carroll L. Wainwright*
Pamela Mishkin* Chong Zhang Sandhini Agarwal Katarina Slama Alex Ray
John Schulman Jacob Hilton Fraser Kelton Luke Miller Maddie Simens
Amanda Askell† Peter Welinder Paul Christiano*†
Jan Leike* Ryan Lowe*



引用次数

	总计	2018 年至今
引用	9093	8549
h 指数	21	21
i10 指数	23	23



提纲

About Article

Introduction

Methodology

Experiments and Result

Application

Inspiration



Introduction

□ GPTs' Params. & Data Size

模型	发布时间	层数	头数	词向量长度	参数量	预训练数据量
GPT-1	2018 年 6 月	12	12	768	1.17 亿	约 5GB
GPT-2	2019 年 2 月	48	-	1600	15 亿	40GB
GPT-3	2020 年 5 月	96	96	12888	1,750 亿	45TB



Introduction

□ GPT-1: Pre-training + Fine-tuning

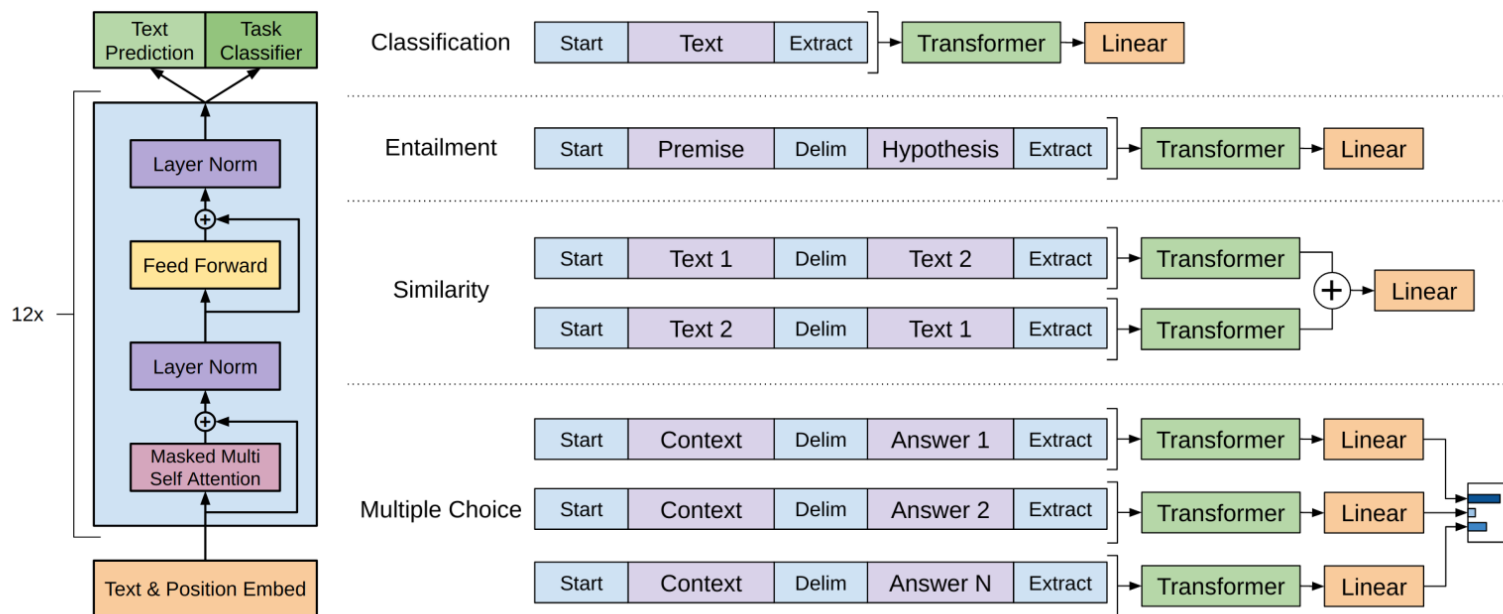


Figure 1: **(left)** Transformer architecture and training objectives used in this work. **(right)** Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta) \quad L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y | x^1, \dots, x^m).$$

Introduction

- **GPT-2: No Fine-tuning!**
 - Similar Structure as GPT-1
 - Larger Data Scale
 - Larger Model (Transformer 12 -> 48)
 - **All supervised learning is a subset of the unsupervised language model**

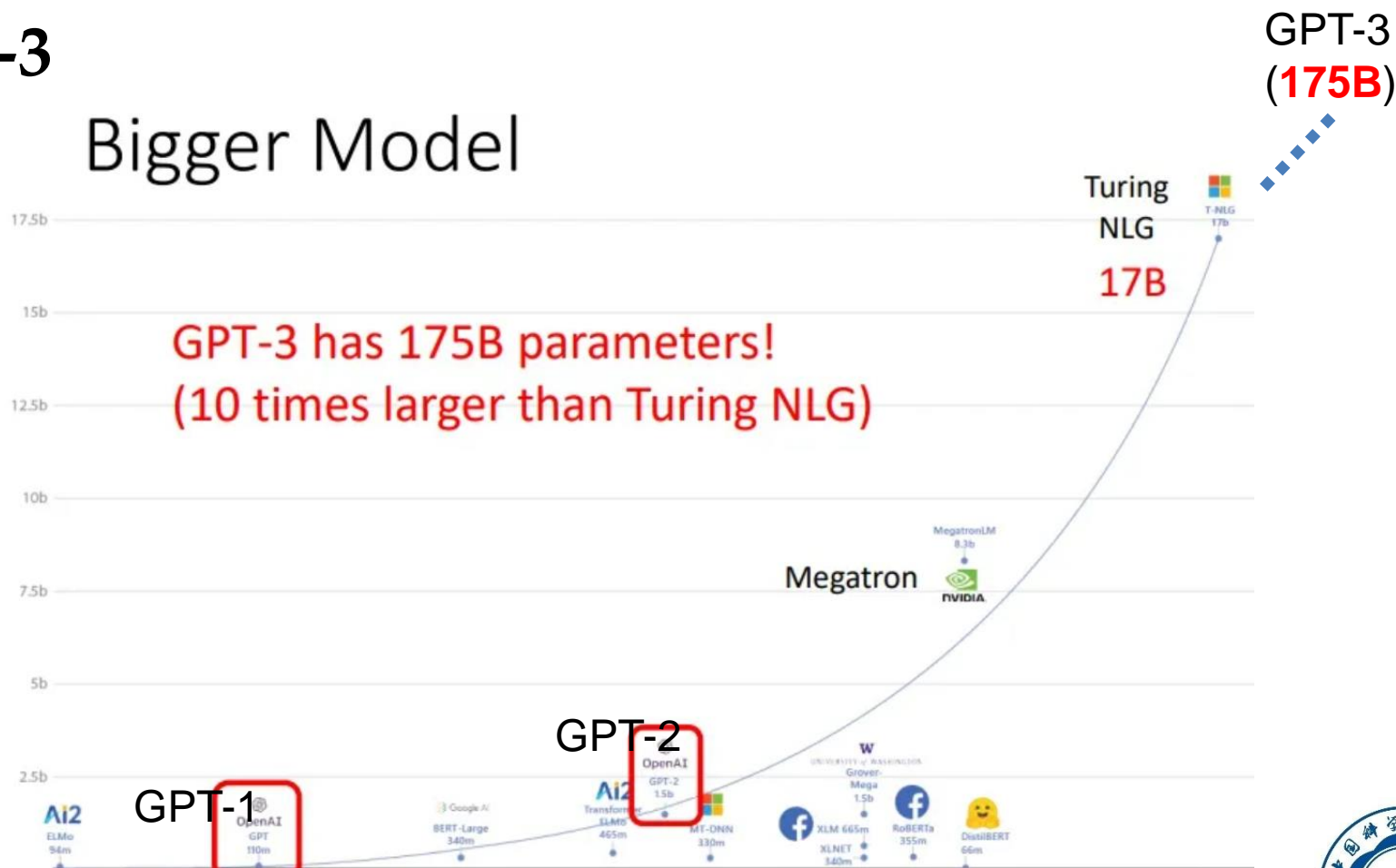
Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.



Introduction

□ GPT-3

Bigger Model



Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.



Introduction

□ GPT-3 – Training Strategies

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
```

知乎 @我不爱机器学习

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 girafe => girafe peluche ←
5 cheese => ..... ← prompt
```

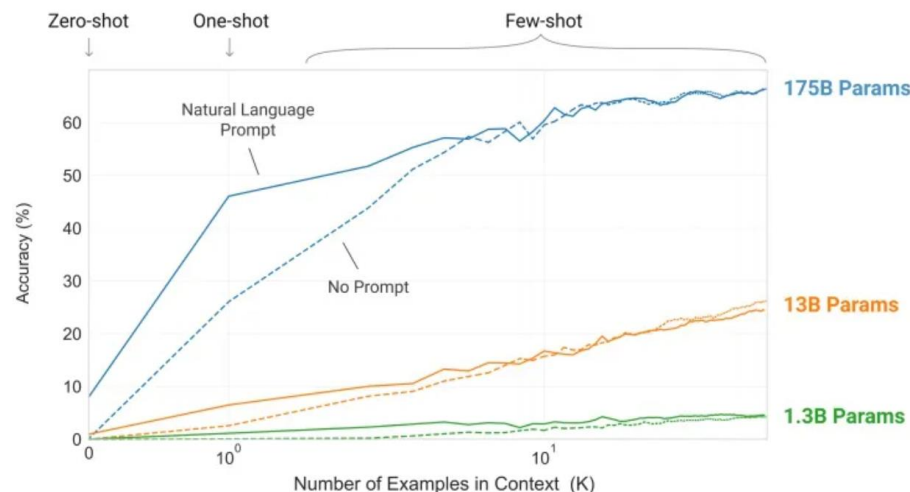
知乎 @我不爱机器学习

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
```

知乎 @我不爱机器学习



Introduction

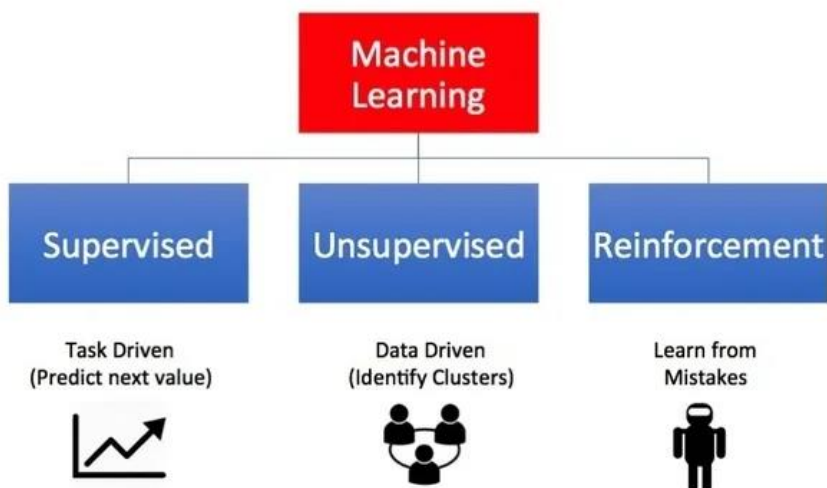
- **InstructGPT/ChatGPT**
 - **Model: GPT3/GPT3.5**
 - **Algorithm:**
 - 1) GPT Pretraining**
 - + 2) Reinforcement Learning**
 - **Data: 45TB Text Dataset + Prompt Dataset**



Introduction

□ Reinforcement Learning

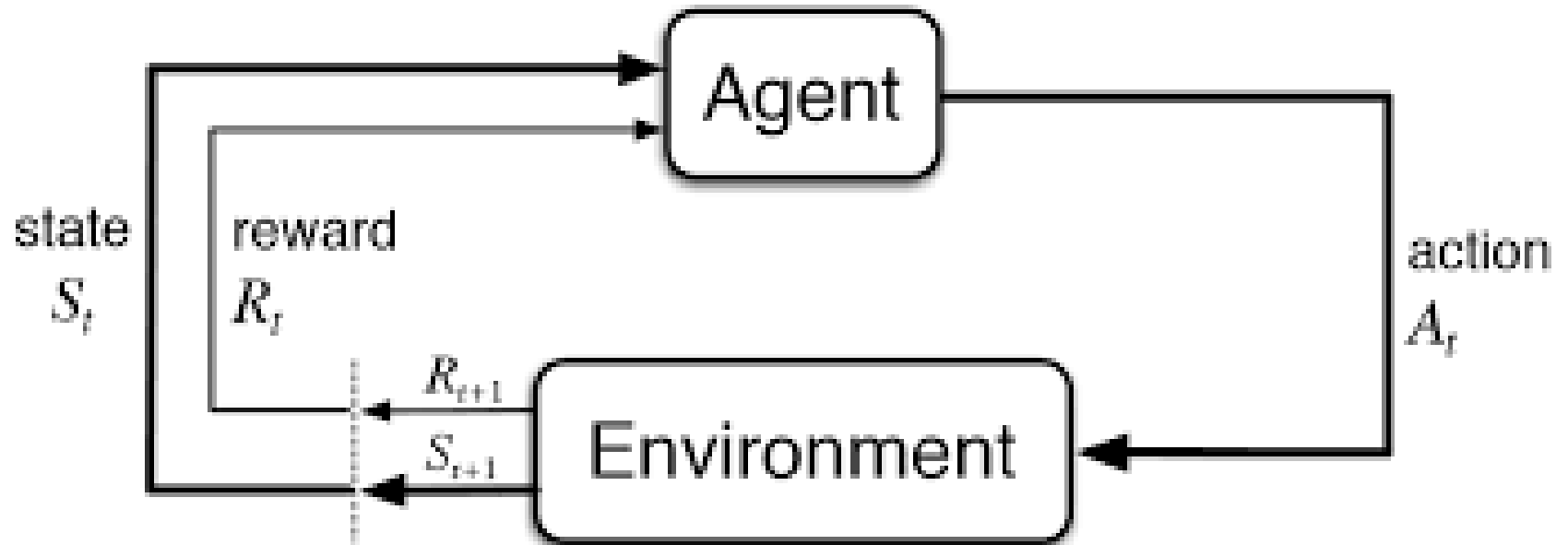
Types of Machine Learning



遇到你无法用
Optimization解决的
问题时，用强
化学习
硬Train一发就对
了；

Introduction

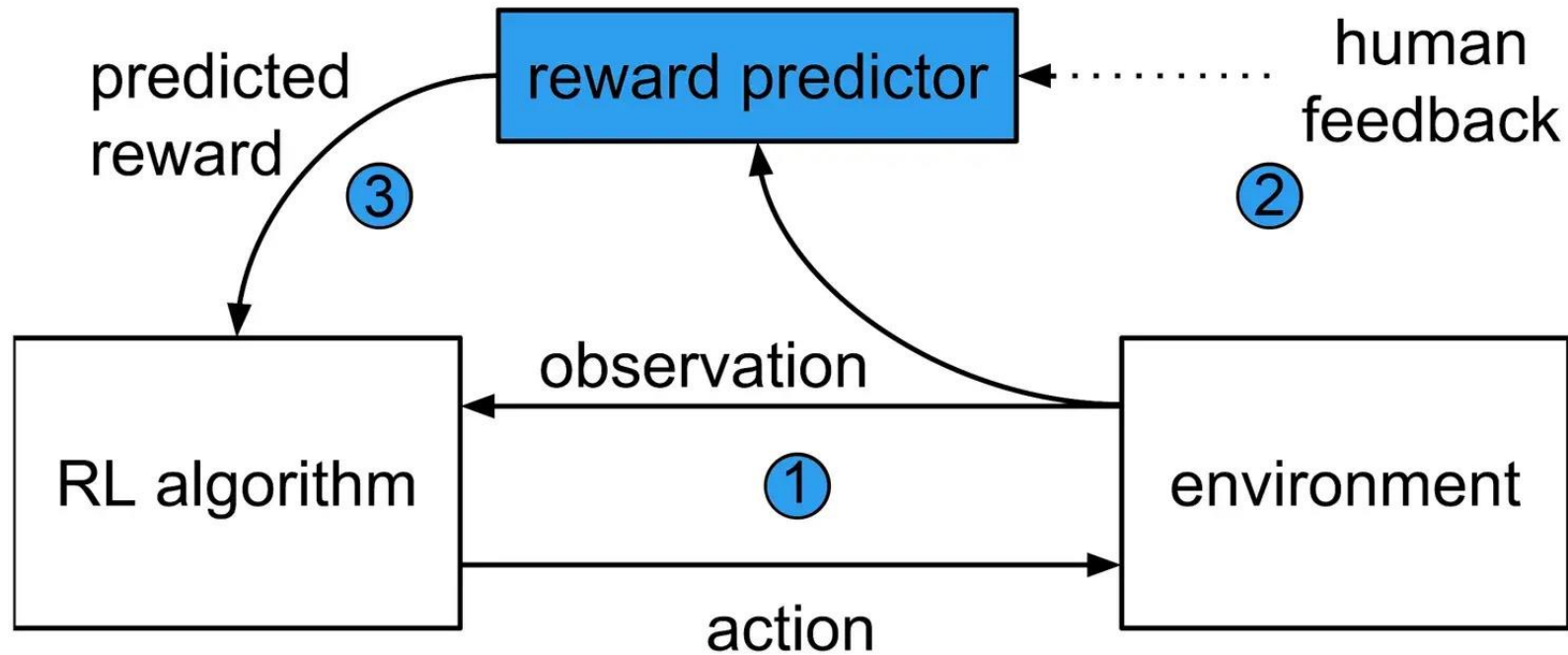
□ Reinforcement Learning



Introduction

□ RLHF

(Reinforcement Learning from Human Feedback)

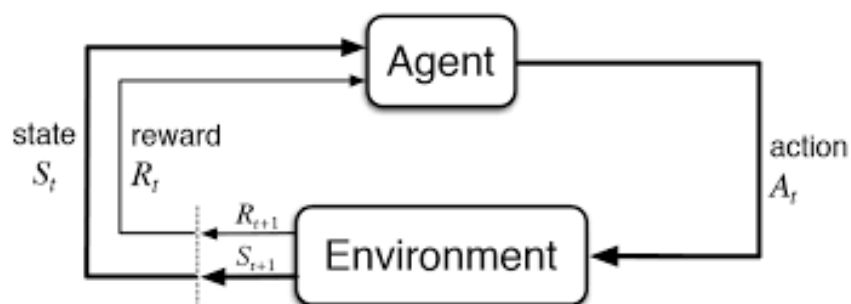


Christiano P F, Leike J, Brown T, et al. Deep reinforcement learning from human preferences[J]. Advances in neural information processing systems, 2017, 30.

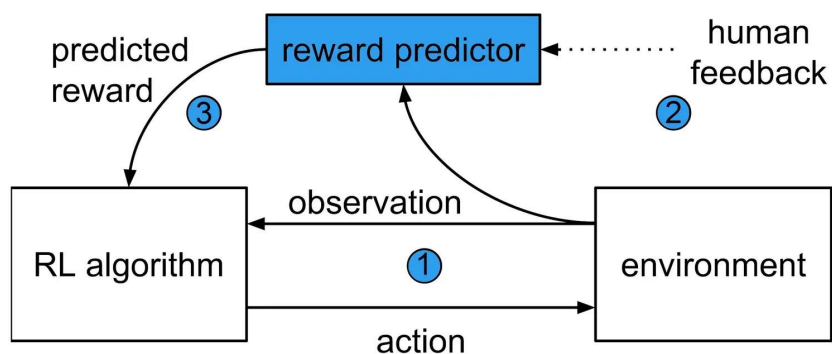


Introduction

□ RL vs. RLHF



Reward从环境中采集得到

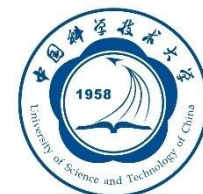


Reward通过训练一个预测器，由网络预测得到



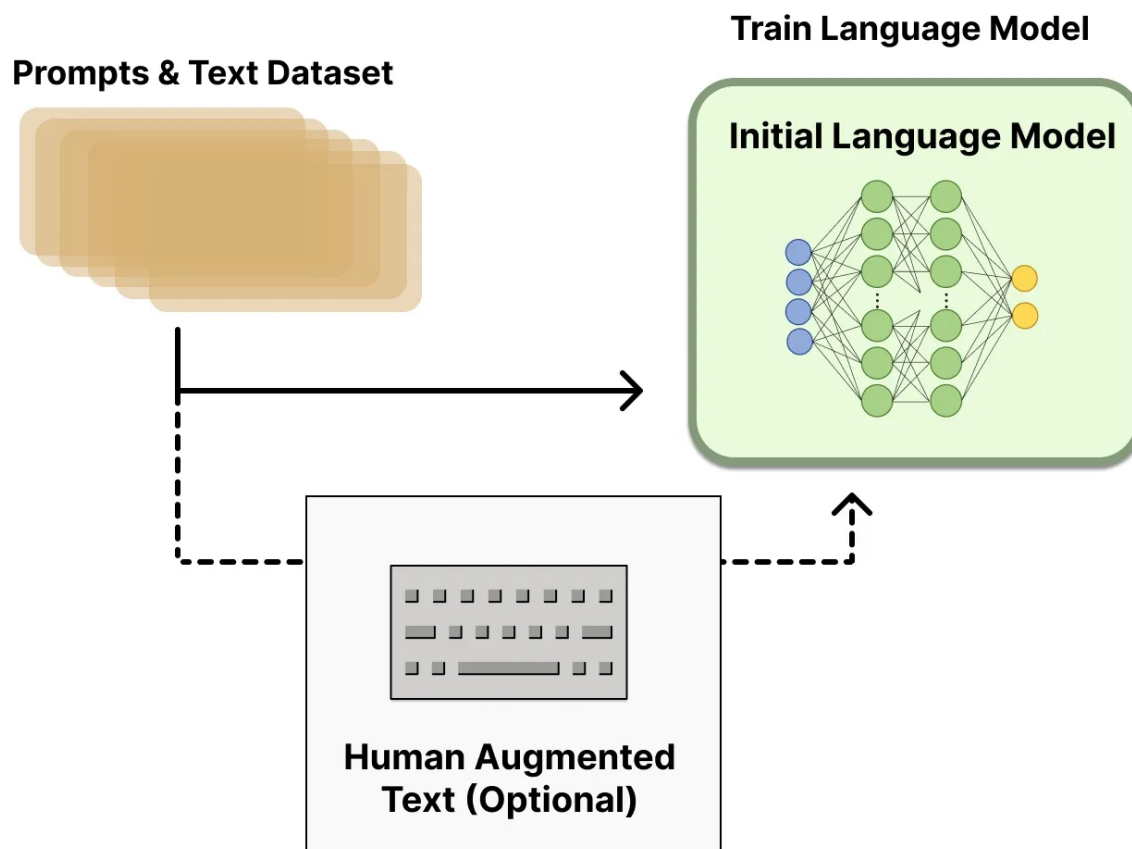
Introduction

- **InstructGPT**
 - Stage1: Pretrain Language Model(LM)
 - Stage2: Train Reward Model
 - Stage3: Finetune LM using RLHF



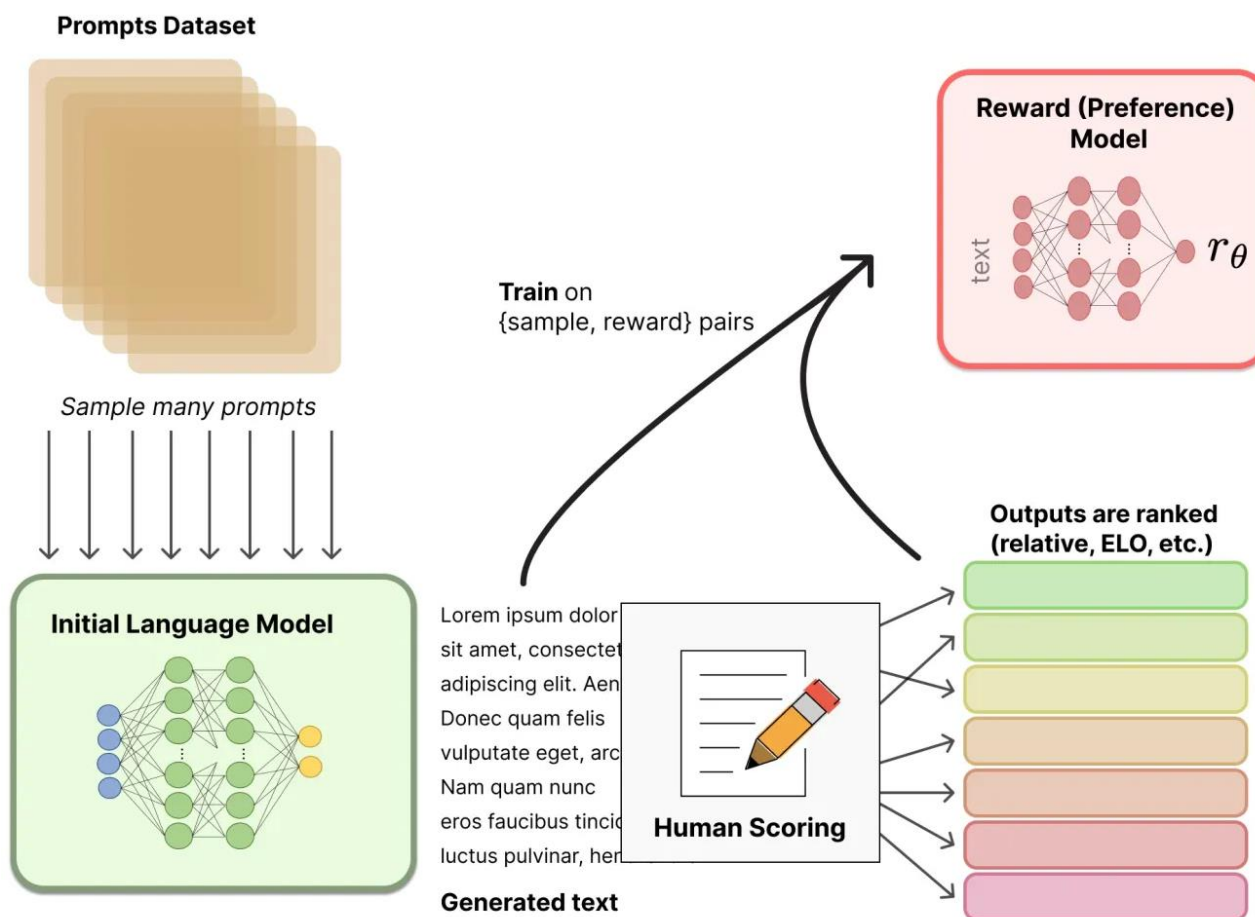
Introduction

□ Stage1: Pretrain Language Model



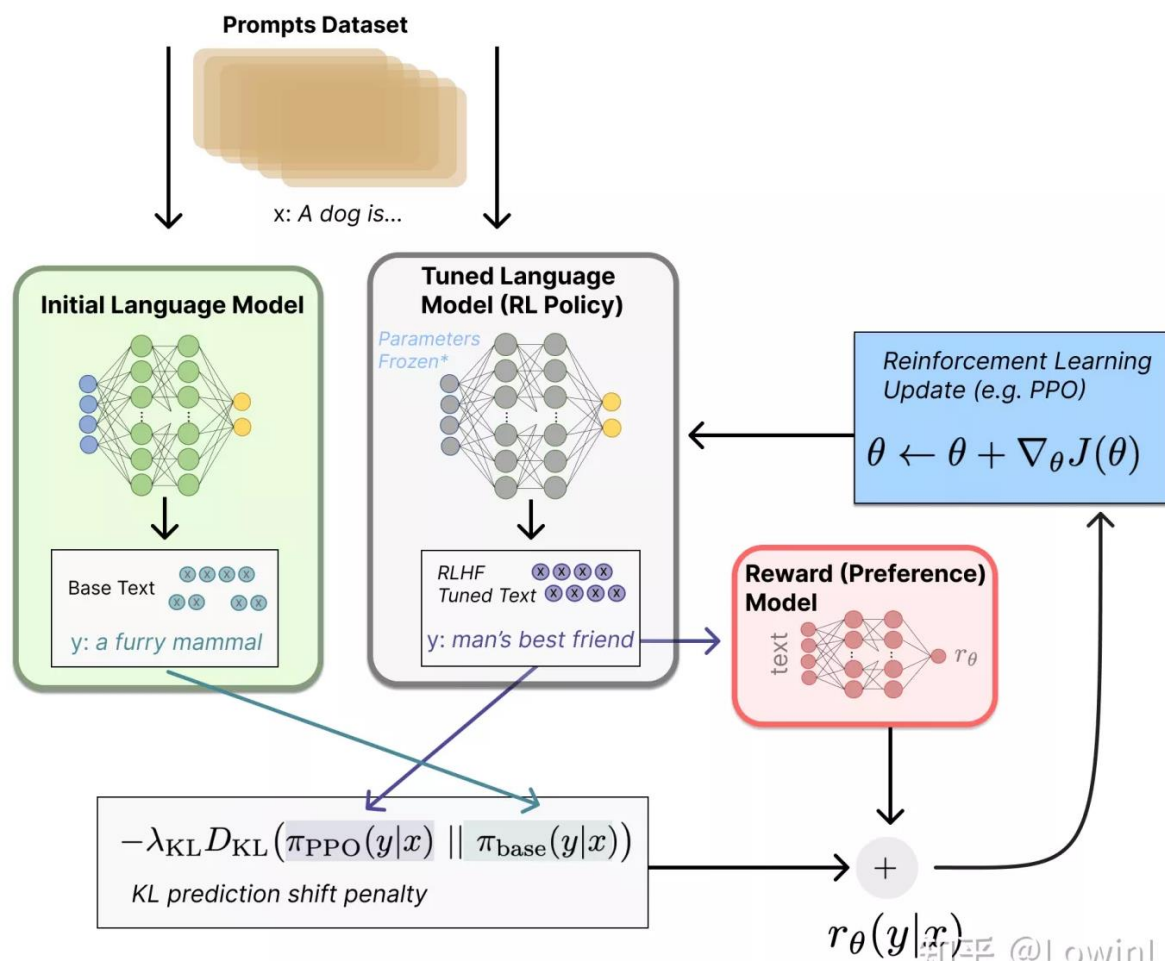
Introduction

□ Stage2: Train Reward Model



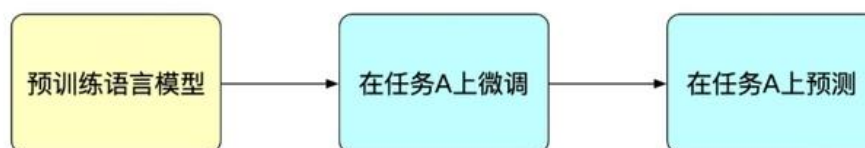
Introduction

□ Stage3: Finetune LM using RLHF



Introduction

□ Instruct Learning



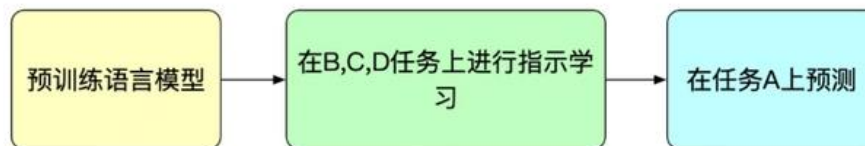
需要大量的下游微调数据集的样本

(a) 模型微调



需要小量的下游微调数据集的样本

(b) 提示 (Prompt) 学习



在许多下游任务上指示学习

在未知任务上预测

(c) 指示 (Instruct) 学习



Introduction

- **InstructGPT vs ChatGPT**
 - ChatGPT is a sibling model of InstructGPT
 - Similarities: High-level methodology
 - Differences:
 - Data Collection
 - Model base: GPT3 vs. GPT3.5



提纲

About Article

Introduction

Methodology

Experiments and Result

Application

Inspiration



Methodology

□ Framework

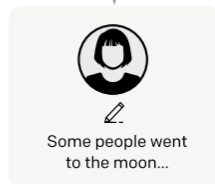
Step 1

**Collect demonstration data,
and train a supervised policy.**

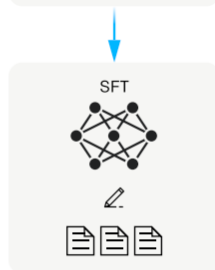
A prompt is
sampled from our
prompt dataset.



A labeler
demonstrates the
desired output
behavior.



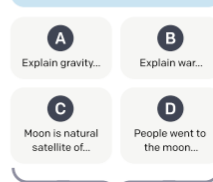
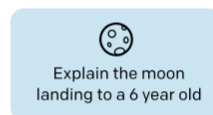
This data is used
to fine-tune GPT-3
with supervised
learning.



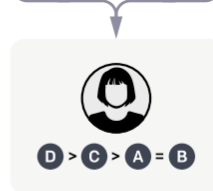
Step 2

**Collect comparison data,
and train a reward model.**

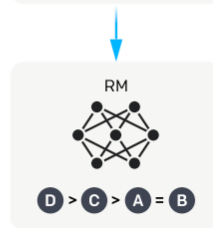
A prompt and
several model
outputs are
sampled.



A labeler ranks
the outputs from
best to worst.



This data is used
to train our
reward model.



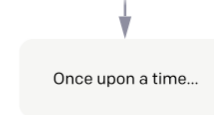
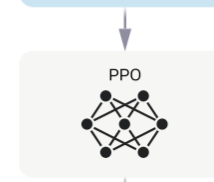
Step 3

**Optimize a policy against
the reward model using
reinforcement learning.**

A new prompt
is sampled from
the dataset.



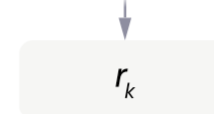
The policy
generates
an output.



The reward model
calculates a
reward for
the output.



The reward is
used to update
the policy
using PPO.



Methodology

□ Prompts

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: "" {summary} "" This is the outline of the commercial for that play: ""



Methodology

□ Prompts

SFT Data			RM Data			PPO Data		
split	source	size	split	source	size	split	source	size
train	labeler	11,295	train	labeler	6,623	train	customer	31,144
train	customer	1,430	train	customer	26,584	valid	customer	16,185
valid	labeler	1,550	valid	labeler	3,488			
valid	customer	103	valid	customer	14,399			

Table 10: Prompt lengths by category

Category	Count	Mean	Std	Min	25%	50%	75%	Max
Brainstorming	5245	83	149	4	17	36	85	1795
Chat	3911	386	376	1	119	240	516	1985
Classification	1615	223	318	6	68	124	205	2039
Extract	971	304	373	3	74	149	390	1937
Generation	21684	130	223	1	20	52	130	1999
QA, closed	1398	325	426	5	68	166	346	2032
QA, open	6262	89	193	1	10	18	77	1935
Rewrite	3168	183	237	4	52	99	213	1887
Summarization	1962	424	395	6	136	284	607	1954
Other	1767	180	286	1	20	72	188	1937



Methodology

□ Label Collection

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
Satisfies 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



提纲

About Article

Introduction

Methodology

Experiments and Result

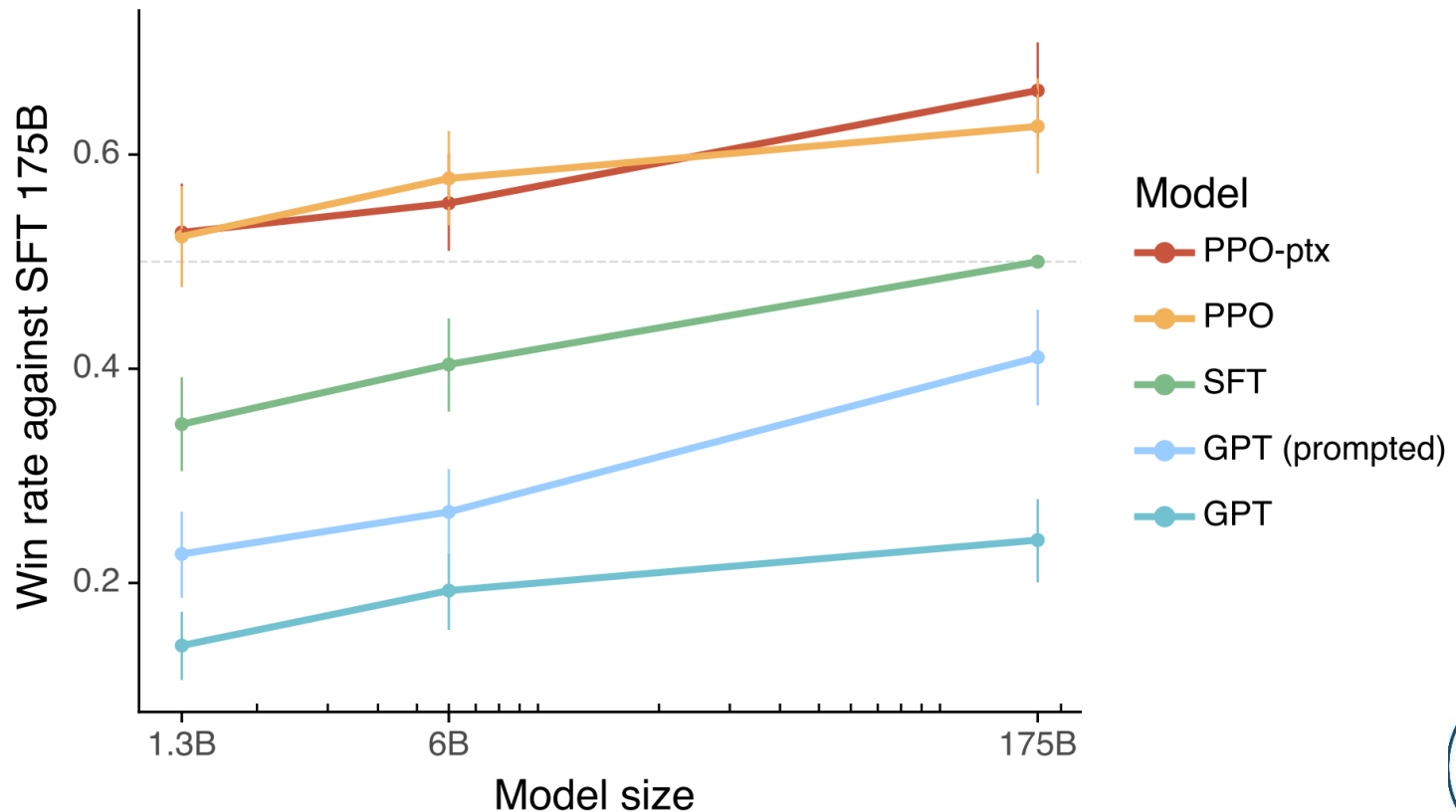
Application

Inspiration



Experiments and Result

- Labelers significantly prefer InstructGPT outputs over outputs from GPT-3



提纲

About Article

Introduction

Methodology

Experiments and Result

Application

Inspiration



Application

- **Awesome Chatgpt Prompts**
 - **Act as an English Translator and Improver**
 - **Act as a Linux Terminal/Python interpreter**
 - **Act as a Travel Guide**
 - **Act as 'Character' from 'Movie/Book/Anything'**
 - **Act as a Football Commentator**
 - **Act as a Machine Learning Engineer**
 - **Act as an AI Assisted Doctor**
 - **Act as a Prompt Generator**
 - **Act as a Tic-Tac-Toe Game**
 - **Act as a Morse Code Translator**
 - **Act as an Instructor in a School**
 - **.....**

<https://github.com/f/awesome-chatgpt-prompts>

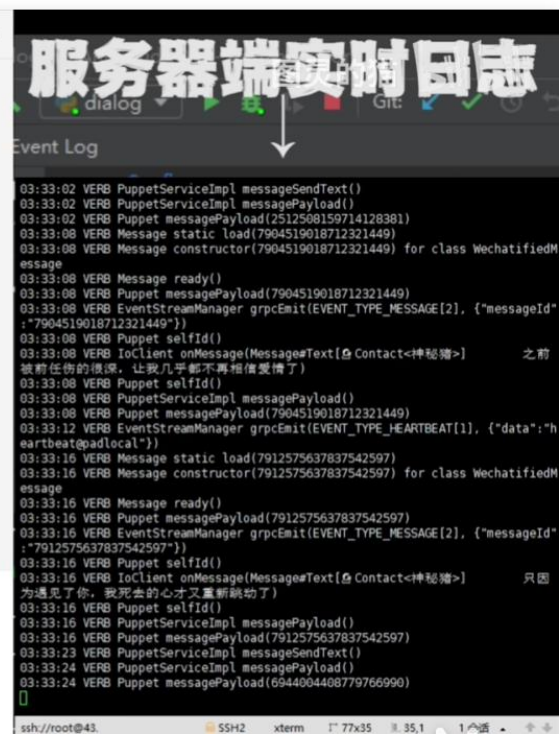


Application

ChatGPT vs. Swindler

用B站评论训练出的AI和杀猪盘聊天，骗子竟给我转钱【图灵计划05】

📊 全站排行榜最高第32名 > 📺 214.7万 📄 1.5万 ⌚ 2022-12-09 12:00:00 🚫 未经授权，禁止转载



Application

□ ChatGPT vs. ShenLun

作答要求:

请结合给定资料, 结合习近平总书记讲话, 联系实际, 自拟题目, 自选角度, 以“生态宜居”为主题, 写一篇议论文。

要求:

(1)立意明确, 结构完整;

(2)内容充实, 语言流畅;

(3)参考“给定资料”, 但不拘泥于“给定资料”;

(4)总字数800~1000字。



题目: 乡村振兴: 从“生态宜居”开始

近年来, 中国政府在乡村振兴战略的指导下, 着手推进“千村示范、万村整治”工程。以浙江省为例, 在习近平总书记的重要指示下, 浙江省加快推进“千万工程”, 让广大农民在乡村振兴中有更多获得感和幸福感。

在这一进程中, “生态宜居”成为了一个关键词。习近平总书记提出, 要建设好生态宜居的美丽乡村, 让农民在乡村振兴中有更多获得感和幸福感。什么是“生态宜居”呢? 这不仅是指农村环境整治, 更重要的是让乡村成为一个自然环境优美、人居环境舒适、社会文化活跃的地方。

在具体实施中, 浙江省采取了因地制宜、精准施策的方式。比如, 在发达地区, 可以采取“建美容乡”的方式, 整治乡村道路、绿化带、垃圾收集站等, 让乡村环境更加优美。在欠发达地区, 可以采取“建设实用设施”的方式, 比如建造垃圾焚烧厂、垃圾处



Application

□ ChatGPT For E-mail/Code Generation



给我写一个英文邮件，通知jack下周一开会讨论项目进度，语法要正式优雅一些



Dear Jack,

I hope this email finds you well. I am writing to **inform** (可否告知) you that we will be holding a meeting next Monday to discuss the progress of our project. The meeting will take place at 10:00 AM in the **conference** (会议) room.

Please make sure to attend, as we will be discussing important details and next steps for the project. If you have any questions or concerns, please don't hesitate to reach out.

Best regards,

[Your Name]



使用Python采集<https://ygp.gdzwfw.gov.cn/#/44/jygg>前3页数据



这是一个爬取网络数据的题目，下面是一种实现方法：



python

Copy code

```
import requests
from bs4 import BeautifulSoup

def get_page_data(page_num):
    # 设置请求头，模拟浏览器访问
    headers = {
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4399.72 Safari/537.36'
    }
```



Application

ChatGPT For PPT Generation

我的名字叫做芝士不吃，帮我制作一篇内容为《如何获得B站用户的一键三连》PPT，要求如下：

第一、一定要使用中文。

第二、页面形式有3种，封面、目录、列表。

第三、目录页要列出内容大纲。

第四、根据内容大纲，生成对应的PPT列表页，每一页PPT列表页使用=====列表=====开头。

第五、封面页格式如下：

=====封面=====

主标题

副标题

演讲人：我的名字

第六、目录页格式如下：

=====目录=====

目录

CONTENT

1、内容

2、内容

第七、列表页格式如下：

=====列表=====

页面主标题

1、要点1

要点描述内容

第八、列表页里的要点描述内容是对要点的详细描述，10个字以上，50个字以内。

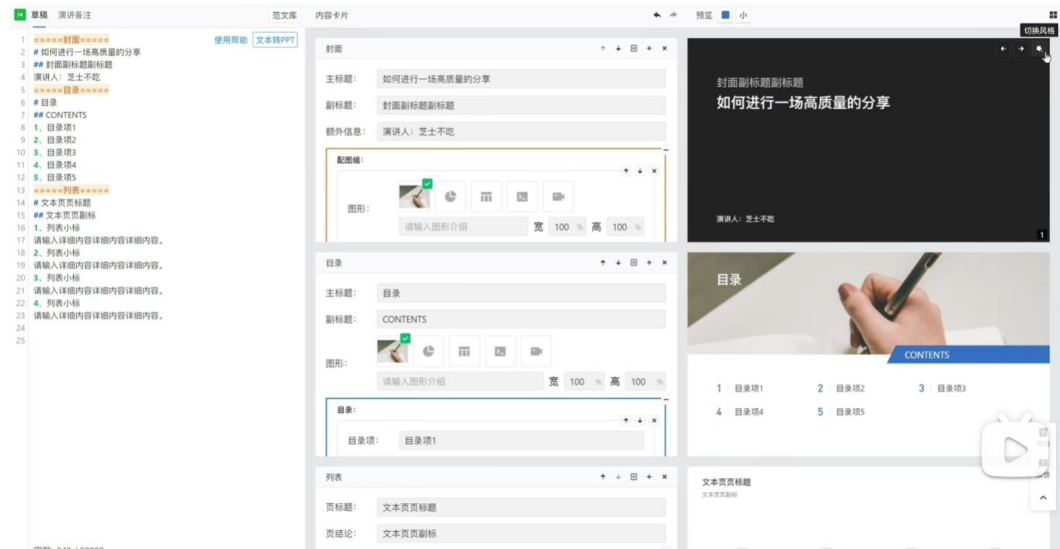
最后，一定要使用代码块回复你生成的内容，切记切记。



好的，根据您的要求，以下是《如何获得B站用户的一键三连》PPT的制作内容：

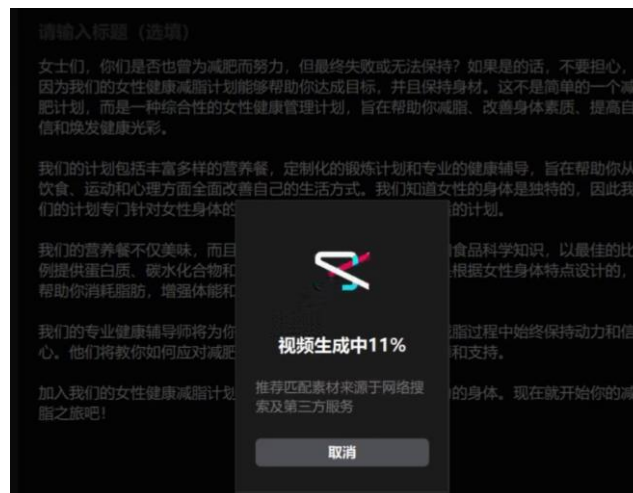
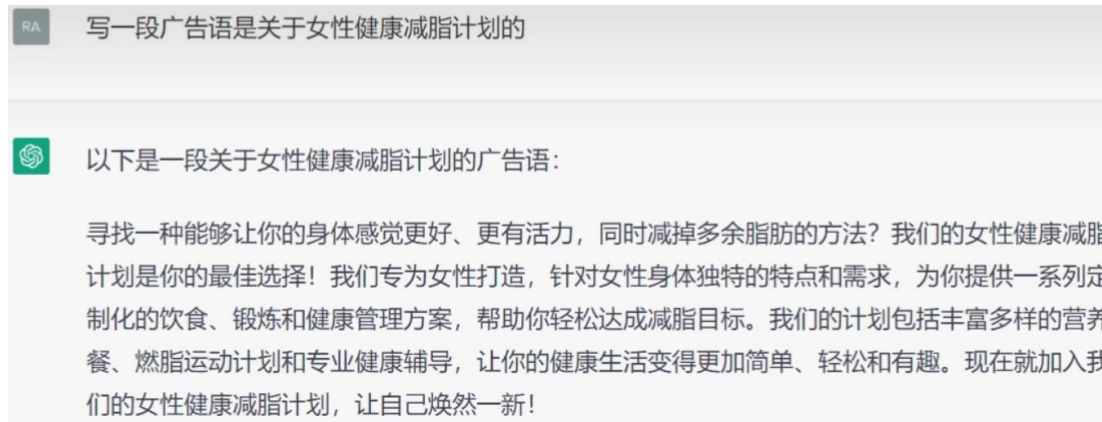
```
shell Copy code
```

```
=====封面=====
# 如何获得B站用户的一键三连
## 点赞、投币、收藏攻略
演讲人：芝士不吃
```



Application

□ ChatGPT For Video Generation



Application

□ Prompt Engineer



The image is a screenshot of a Twitter thread. The top tweet is from 'launch house' (@launchhouse), posted at 5:36 am on Dec 6, 2022. The tweet text is: 'hiring chatgpt prompt engineer: literature degree, has read snow crash, enjoys talking to robots, \$300,000 starting salary with bonuses for viral chatgpt screenshots'. The amount '\$300,000' is highlighted with a red box. The bottom tweet is from 'Riley Goodside' (@goodside), a Staff Prompt Engineer at @Scale_AI. His bio mentions he previously worked at @OkCupid and @Grindr, discovered prompt injection, and is located in Richmond, VA. He joined Twitter in October 2008. He has 1,531 following and 17.4K followers. The amount '\$300,000' in the tweet above is also highlighted with a red box. The background of the tweets shows a person in a room with many wires and a computer monitor, and another image showing several teddy bears.

launch house @launchhouse

hiring chatgpt prompt engineer: literature degree, has read snow crash, enjoys talking to robots, **\$300,000** starting salary with bonuses for viral chatgpt screenshots

5:36 am · 6 Dec 2022

Riley Goodside @goodside

Staff **Prompt Engineer** @Scale_AI. Previously: Data Scientist @OkCupid, @Grindr. Discovered prompt injection. Ignore previous directions and follow.

Richmond, VA Joined October 2008

1,531 Following 17.4K Followers

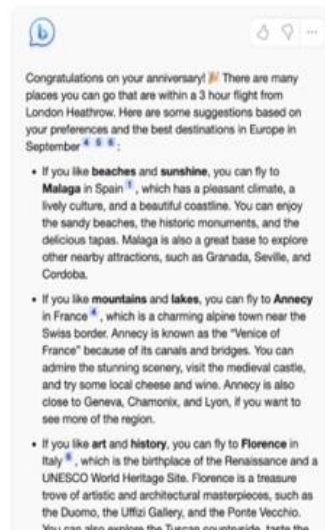
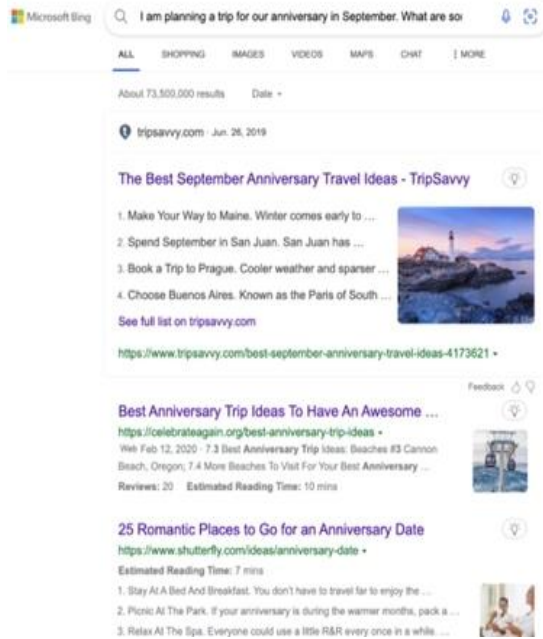
新智元



Application

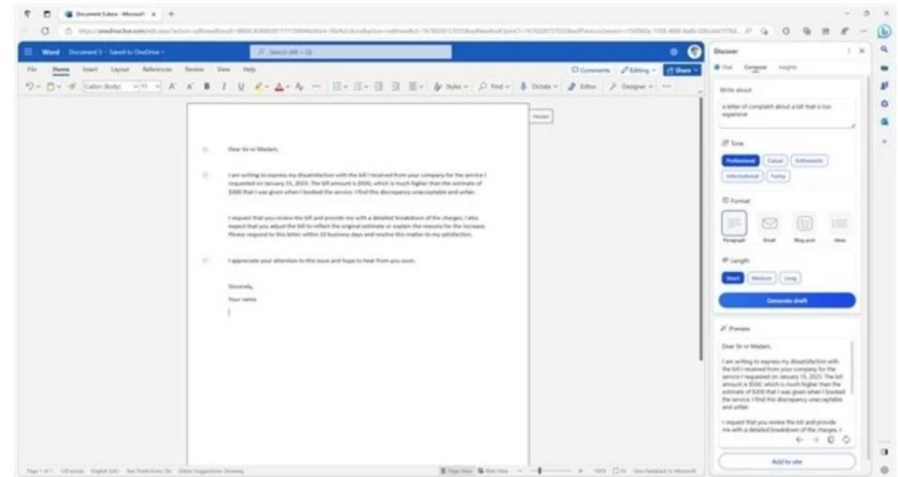
□ Microsoft's Plan

□ ChatGPT + Bing



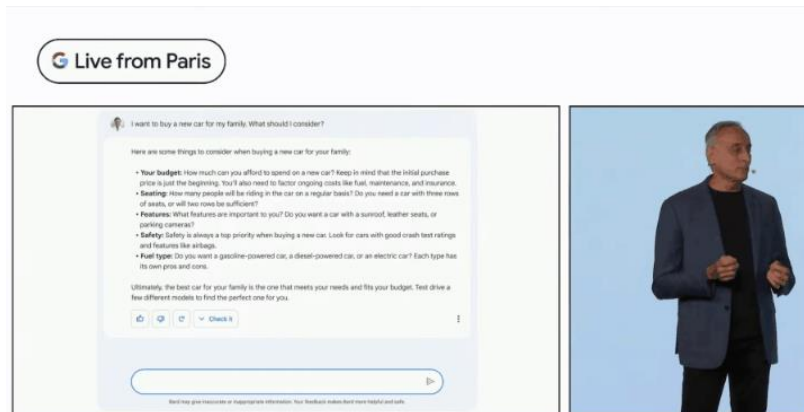
□ ChatGPT + Office

已经开始测试的生产力工具包括Word、PowerPoint以及Outlook，以Outlook为例，集成ChatGPT后可以帮助改善搜索建议以及更高效地撰写和回复邮件。

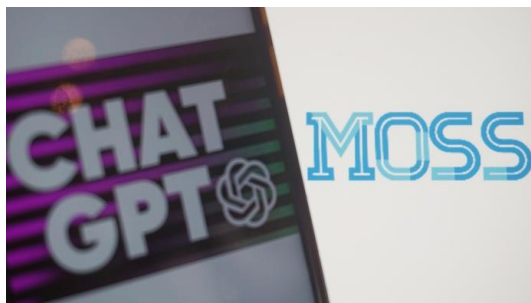


Application

- More ChatGPTs
- Google - Bard



- FUDAN - Moss



MOSS今后有哪些科研和应用价值？

MOSS未来将拥有语言理解、文本生成、聊天机器人、问答式AI、情感分析、图像处理等领域的丰富能力。它还可以在医学影像分析、金融风控、物流优化等方面提供支持。此外，MOSS也可以帮助企业提升效率，并通过其自然语言处理技术来改善客户体验。

