



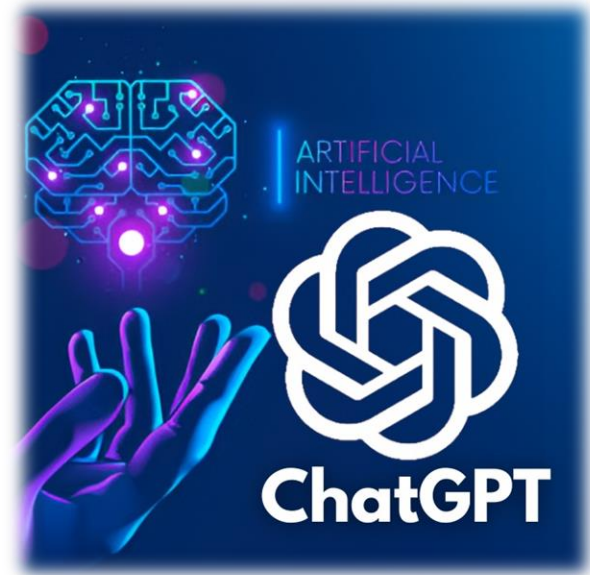
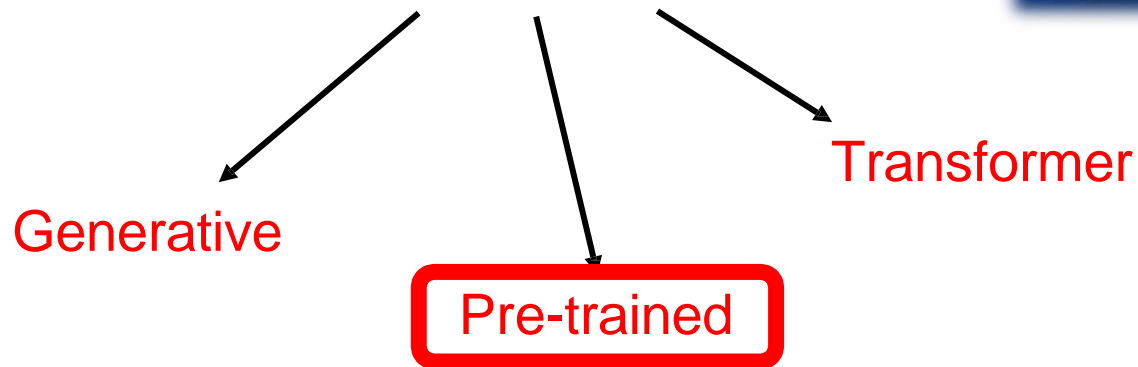
人工智能技术及应用

Artificial Intelligence and Application

GPT



ChatGPT



Outline



BERT series



GPT series

GPT-3

Language Models are Few-Shot Learners

Tom B. Brown*

Benjamin Mann*

Nick Ryder*

Melanie Subbiah*

Jared Kaplan[†]

Prafulla Dhariwal

Arvind Neelakantan

Pranav Shyam

Girish Sastry

Amanda Askell

Sandhini Agarwal

Ariel Herbert-Voss

Gretchen Krueger

Tom Henighan

Rewon Child

Aditya Ramesh

Daniel M. Ziegler

Jeffrey Wu

Clemens Winter

Christopher Hesse

Mark Chen

Eric Sigler

Mateusz Litwin

Scott Gray

Benjamin Chess

Jack Clark

Christopher Berner

Sam McCandlish

Alec Radford

Ilya Sutskever

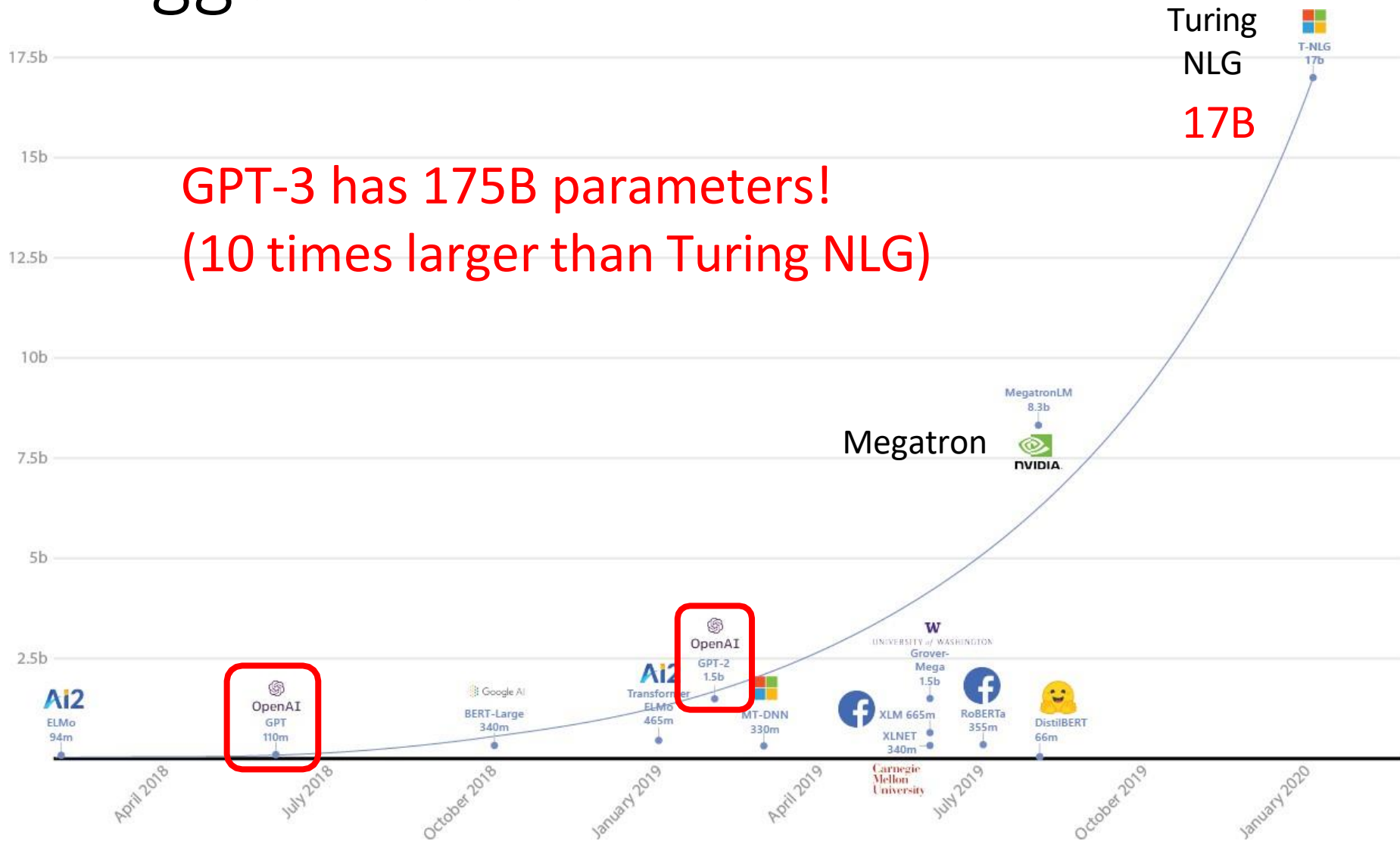
Dario Amodei

OpenAI

<https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>

Bigger Model

GPT-3 has 175B parameters!
(10 times larger than Turing NLG)





Elliot Turner
@eturner303

Reading the OpenAI GPT-3 paper. Impressive performance on many few-shot language tasks. The cost to train this 175 billion parameter language model appears to be staggering: Nearly \$12 million

自然语言处理

炫富

核武器

深度学习 (Deep Learning)

OpenAI

如何评价1700亿参数的GPT-3?

<https://www.zhihu.com/q>



minimaxir commented 26 days ago • edited ▾

...I'm not sure if it's even possible for the 175B model to be distributed in a reasonable manner.

The size of the 1.5B GPT-2 model was about 6GB on disk, which would imply that the 175B model is *at least* 700GB!



Sebastian Gehrmann
@sebgehr

The ELMo paper? 15 pages.
BERT? 16 pages.
GPT-2? 24 pages.
T5? 53 pages.
GPT-3?? 72 pages! arxiv.org/pdf/2005.14165...

github.com/openai/gpt-3/issues/1

GPT vs Bert



语言模型

GPT



文字接龙

how are _____

BERT



文字填空

how are you

对于大型语言模型的两种不同期待

- 期待一：成为专才，解某一个特定任务

这节课我们要讲如何驾驭
大型语言模型.....



翻译



This course
is about

这节课我们要讲如何驾驭
大型语言模型.....



摘要



本课程重点为.....

对于大型语言模型的两种不同期待

- 期待二：成为通才

这就是今天 ChatGPT 走的路线

对以下文句做 **翻译**：

这节课我们要讲如何驾驭
大型语言模型



什么都会



This course
is about
.....

对以下文句做 **摘要**：

这节课我们要讲如何驾驭
大型语言模型



什么都会



本课程重点为.....

Prompt

对于大型语言模型的两种不同期待

- 期待二：成为通才 The Natural Language Decathlon: Multitask Learning as Question Answering
<https://arxiv.org/abs/1806.08730>

Question	Context	Answer	Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US....	major economic center	What has something experienced?	Areas of the Baltic that have experienced eutrophication .	eutrophication
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser	Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson .	Bernie Wrightson
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune ...	Harry Potter star Daniel Radcliffe gets £320M fortune...	What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
Hypothesis: Product and geography are what make cream skimming work. Entailment , neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment	What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive	Who had given help? Susan or Joan?	Joan made sure to thank Susan for all the help she had given.	Susan

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing

<https://arxiv.org/abs/1506.07285>

对于大型语言模型的两种不同期待

- 成为专才的好处：专才在单一任务上有机会赢过通才



Is ChatGPT A Good Translator?
A Preliminary Study

<https://arxiv.org/abs/2301.08745>

对于大型语言模型的两种不同期待

- 成为专才的好处：专才在单一任务上有机会赢过通才

System	De-En		Ro-En		Zh-En	
	⇒	⇐	⇒	⇐	⇒	⇐
Google	45.04	41.16	50.12	46.03	31.66	43.58
DeepL	49.23(+9.3%)	41.46(+0.7%)	50.61(+0.9%)	48.39(+5.1%)	31.22(-1.3%)	44.31(+1.6%)
Tencent	n/a	n/a	n/a	n/a	29.69(-6.2%)	46.06(+5.6%)
ChatGPT	43.71(-2.9%)	38.87(-5.5%)	44.95(-10.3%)	24.85(-46.0%)	24.73(-21.8%)	38.27(-12.1%)

System	De-Zh		Ro-Zh		De-Ro	
	⇒	⇐	⇒	⇐	⇒	⇐
Google	38.71	21.68	39.05	25.59	33.31	32.27
DeepL	40.46(+4.5%)	22.82(+5.2%)	38.95(-0.2%)	25.39(-0.7%)	35.19(+5.6%)	34.27(+6.1%)
Tencent	40.66(+5.0%)	19.44(-10.3%)	n/a	n/a	n/a	n/a
ChatGPT	34.46(-10.9%)	19.80(-8.6%)	30.84(-21.0%)	19.17(-25.0%)	33.38(+0.2%)	29.89(-7.3%)

Is ChatGPT A Good Translator?
A Preliminary Study

<https://arxiv.org/abs/2301.08745>

对于大型语言模型的两种不同期待

- 成为专才的好处：专才在单一任务上有机会赢过通才

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
text-davinci-002	73.2	73.1	46.1	23.3	82.0	79.0	56.0	28.6
text-davinci-003	84.8*	81.2*	56.8	30.9	85.6*	82.8*	60.2*	31.8*
ChatGPT	84.8*	81.1	58.3*	33.4*	84.2	81.0	59.6	30.9
ZH-EN					EN-ZH			
WMT-Best	81.0	77.7	61.1	33.5	86.7	82.0	41.1	44.8
text-davinci-002	74.1	73.1	49.6	20.6	84.0	79.0	32.1	36.4
text-davinci-003	81.6*	78.9*	56.0*	25.0	85.8*	81.3*	34.6	38.3
ChatGPT	81.2	78.3	56.0	25.9*	84.4	78.7	36.0*	40.3*
RU-EN					EN-RU			
WMT-Best	86.0	81.7	68.9	45.1	89.5	84.4	58.3	32.4
text-davinci-002	77.5	76	58.7	34.9	85.4	80.9	51.6	25.1
text-davinci-003	84.8*	81.1*	64.6	38.5	86.7*	82.2*	54.0*	27.5*
ChatGPT	84.8*	81.0	66.5*	41.0*	77.6	70.4	41.1	19.0
FR-DE					DE-FR			
WMT-Best	89.5	80.7	81.2	64.8	85.7	79.5	74.6	58.4
text-davinci-002	66.6	67.9	45.8	25.9	64.2	67.6	44.6	24.5
text-davinci-003	84.6	77.9	65.7*	42.5*	78.5	76.1	58.9	35.6
ChatGPT	84.7*	78.5*	65.2	42.0	81.6*	79.8*	60.7*	37.3*

How Good Are GPT Models at
Machine Translation? A
Comprehensive Evaluation


<https://arxiv.org/abs/2302.09210>

对于大型语言模型的两种不同期待

- 成为通才的好处：只要重新设计 prompt 就可以快速开发新功能，不用写程序

对以下文句做 摘要：

这节课我们要讲如何驾驭
大型语言模型.....

————→  ———→ 本课程重点为


什么都会

还是太长



请给我 100 字以内的摘要：

这节课我们要讲如何驾驭
大型语言模型.....

————→  ———→ 重点是

什么都会

两种不同期待导致两类不同的使用方式

- 期待一：成为专才

这节课我们要讲如何驾驭
大型语言模型.....



翻译



This course
is about

- 期待二：成为通才

对以下文句做 **翻译**：

这节课我们要讲如何驾驭
大型语言模型



什么都会



This course
is about

对预训练模型做改造

BERT



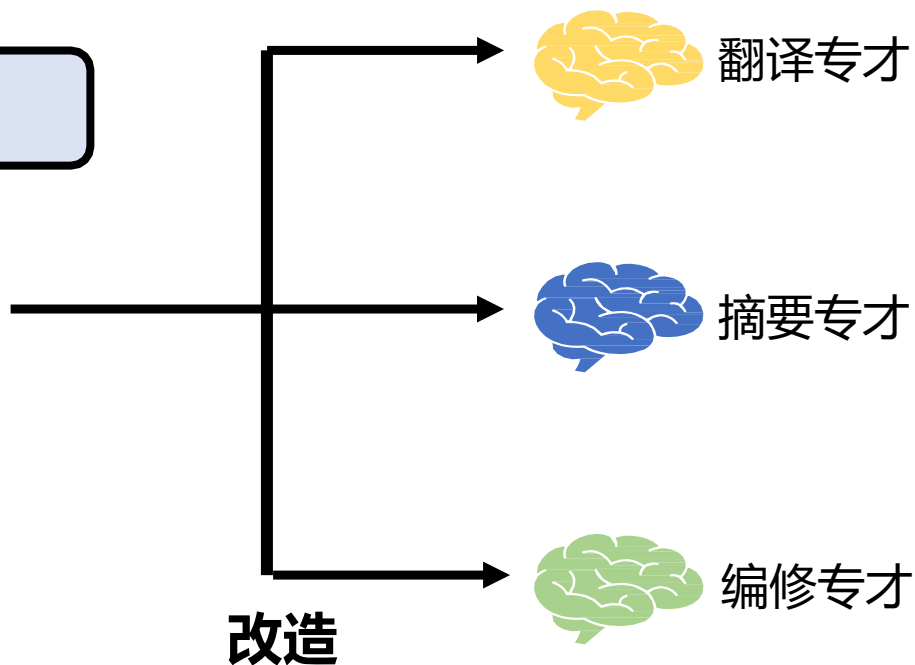
不会说一整句话



语言模型

文字填空

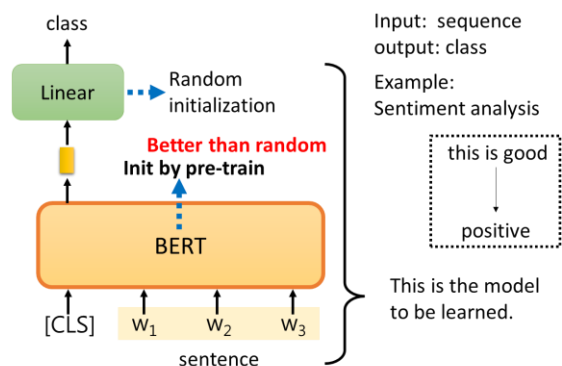
how are you



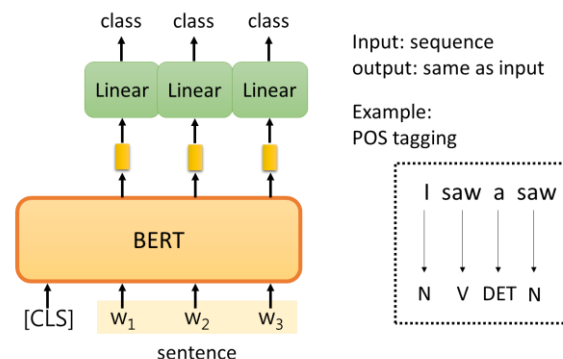
加外挂和微调参数

对预训练模型做改造—加外挂

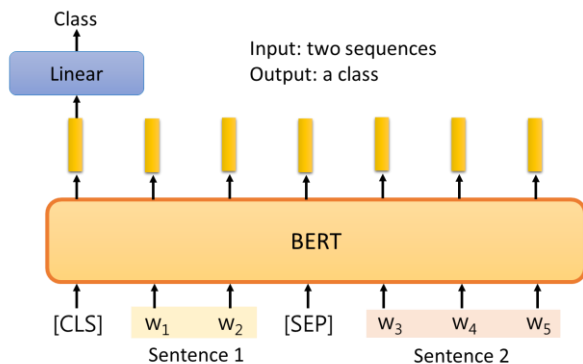
How to use BERT – Case 1



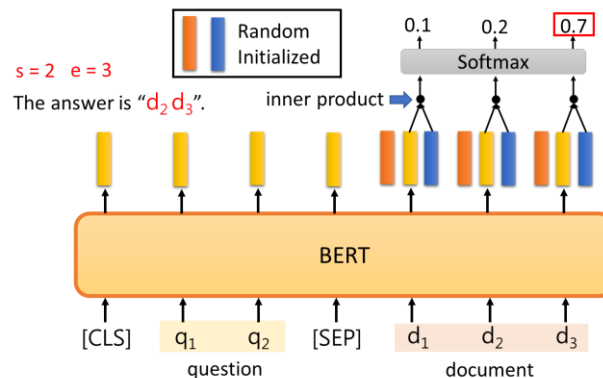
How to use BERT – Case 2



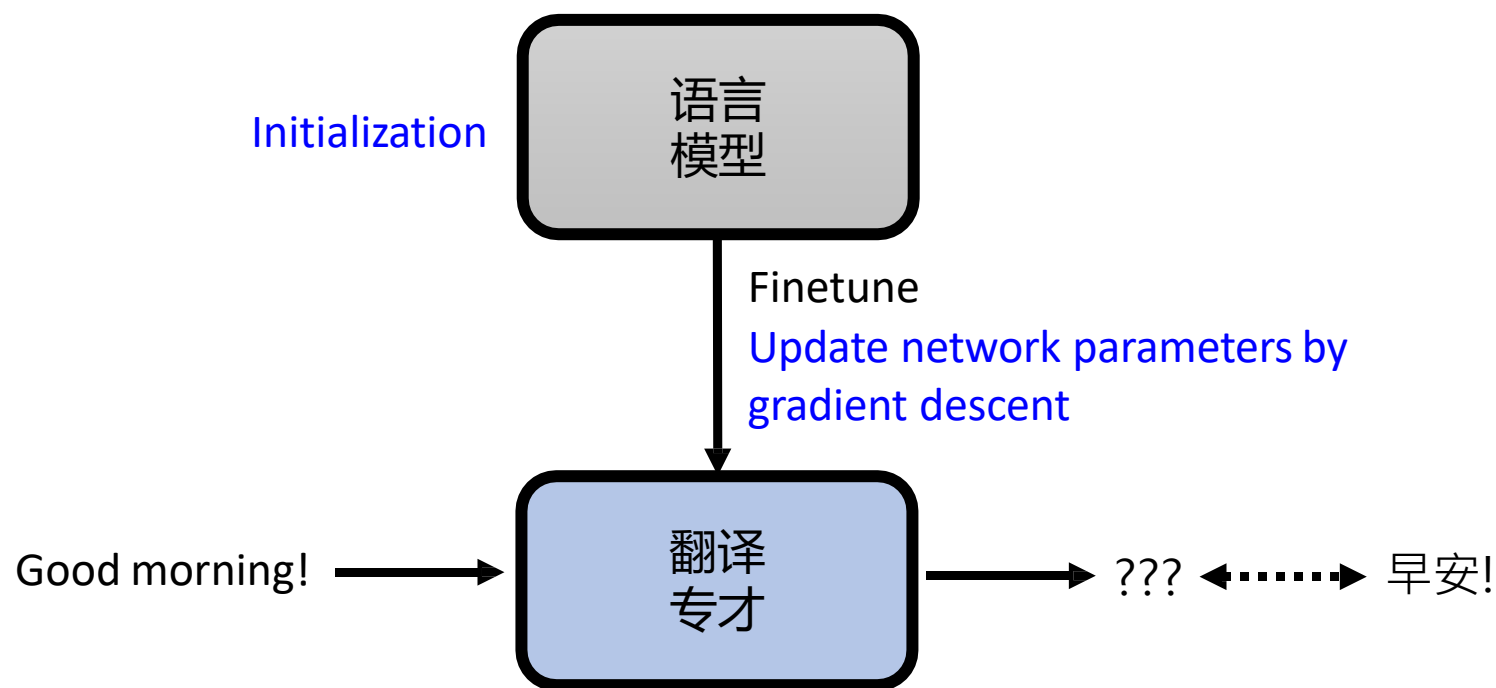
How to use BERT – Case 3



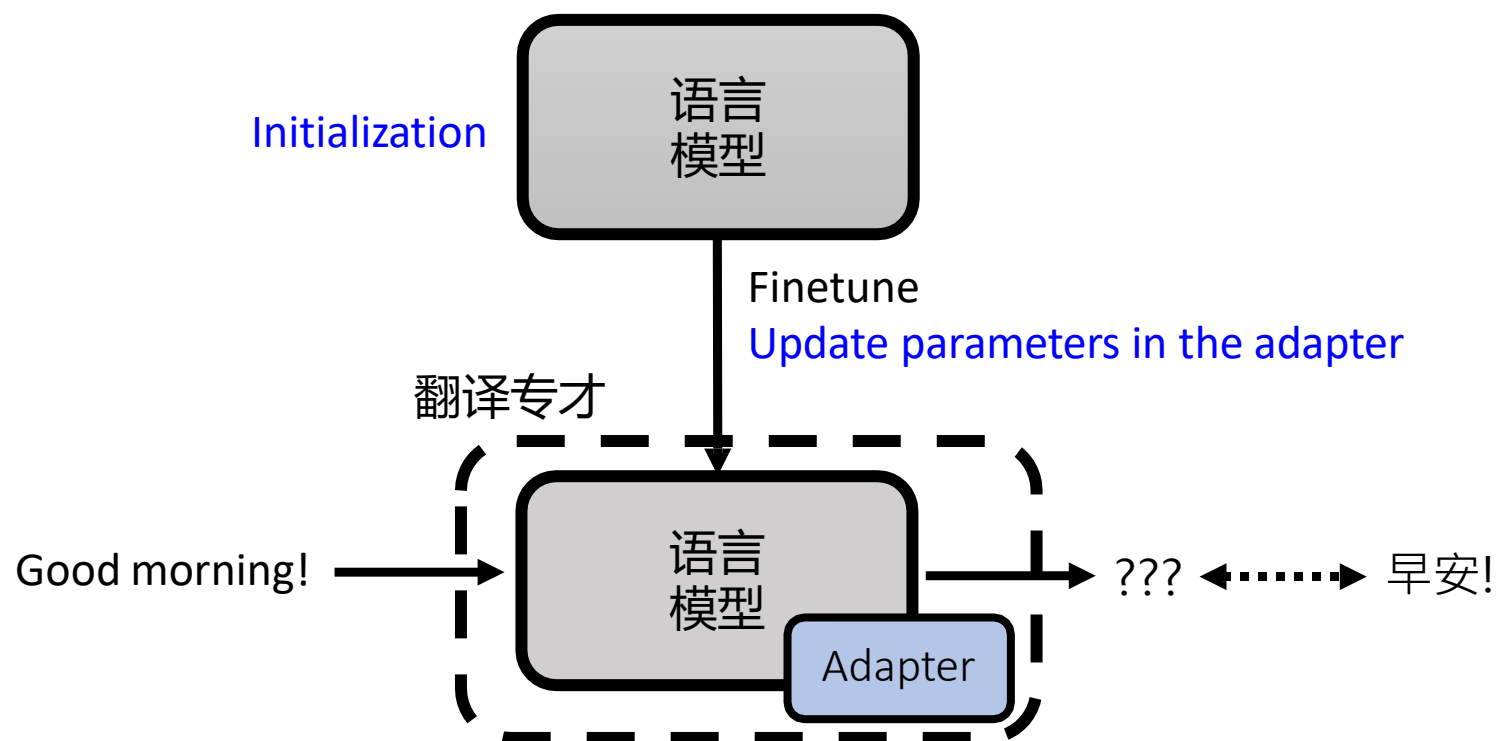
How to use BERT – Case 4



对预训练模型做改造—微调 (Finetune)



对预训练模型做改造—Adapter

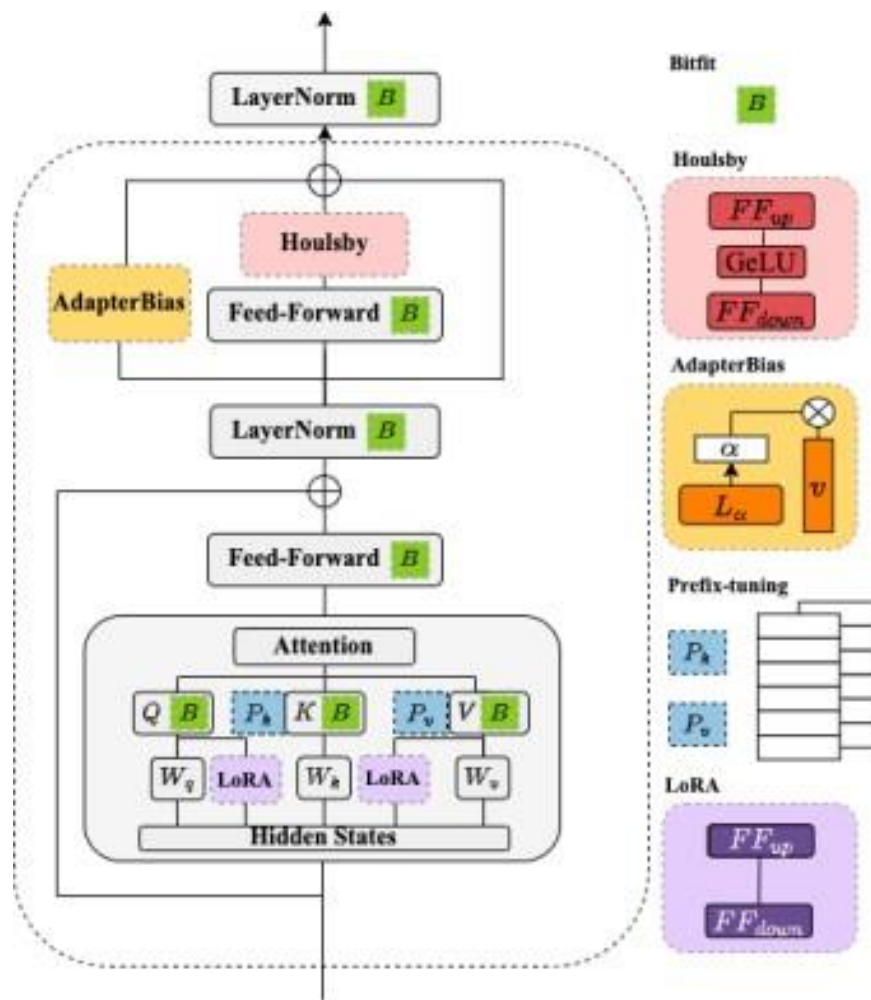


各种Adapter

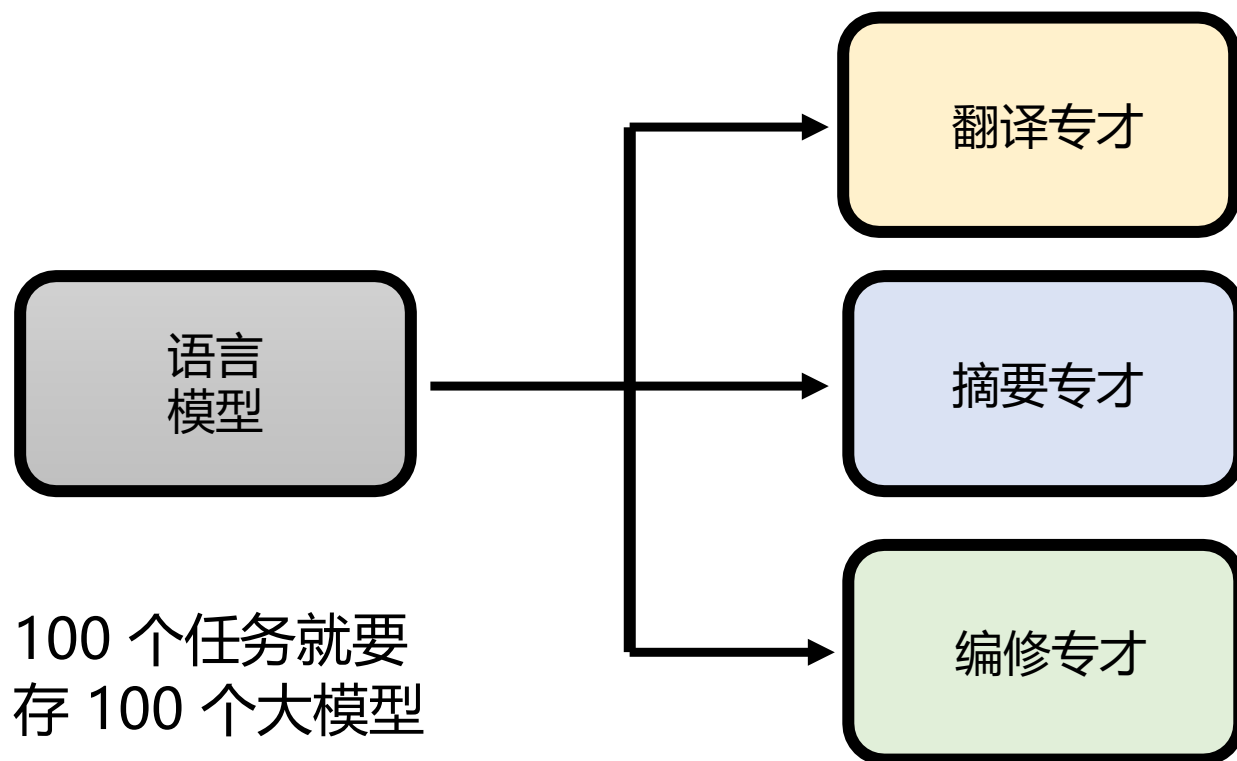


<https://adapterhub.ml/>

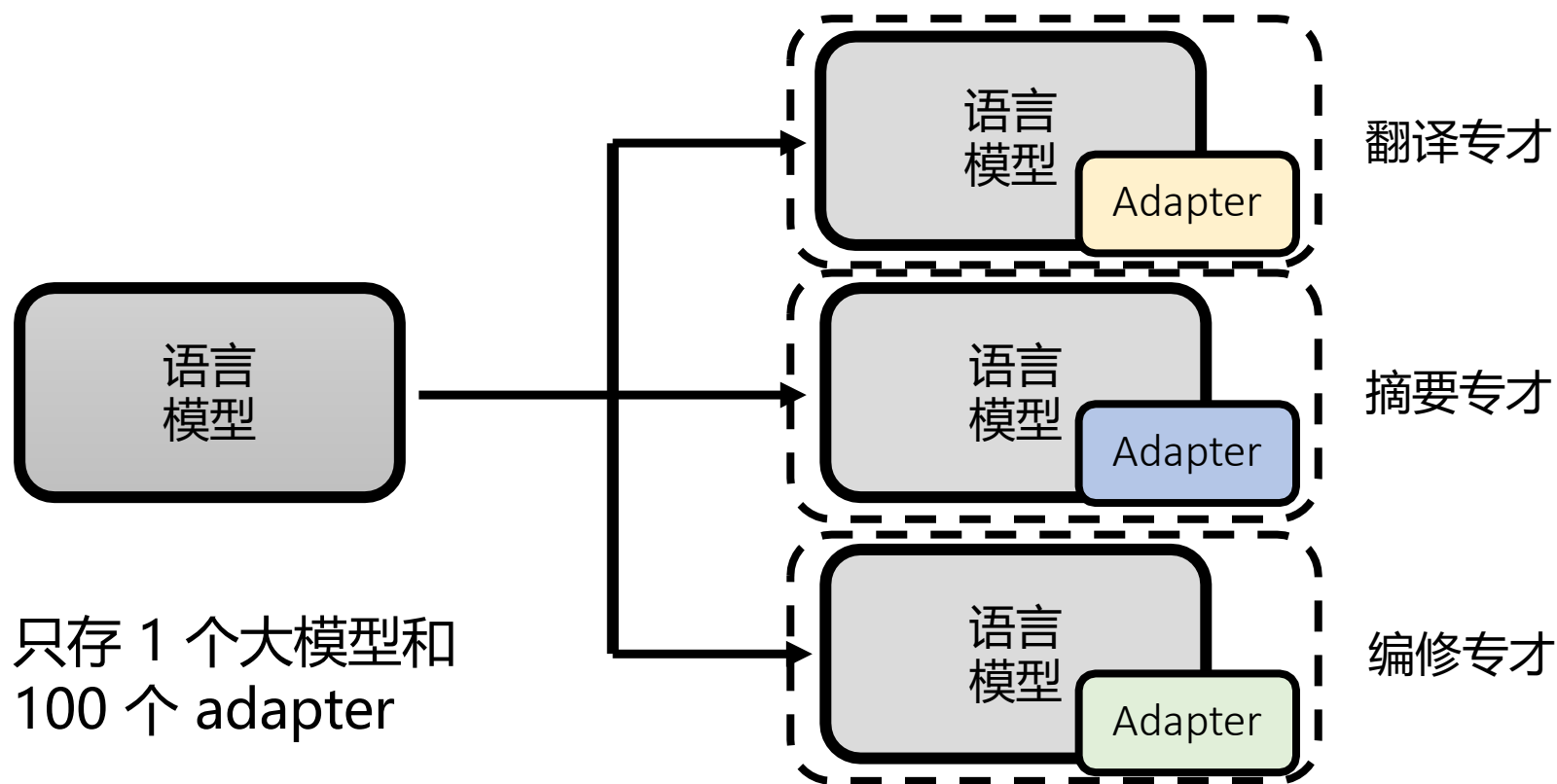
Source of image:
<https://arxiv.org/abs/2210.06175>



对预训练模型做改造— Adapter



对预训练模型做改造—Adapter



两种不同期待导致两类不同的使用方式

- 期待一：成为专才

这节课我们要讲如何驾驭
大型语言模型.....



翻译



This course
is about

- 期待二：成为通才

对以下文句做 **翻译**：

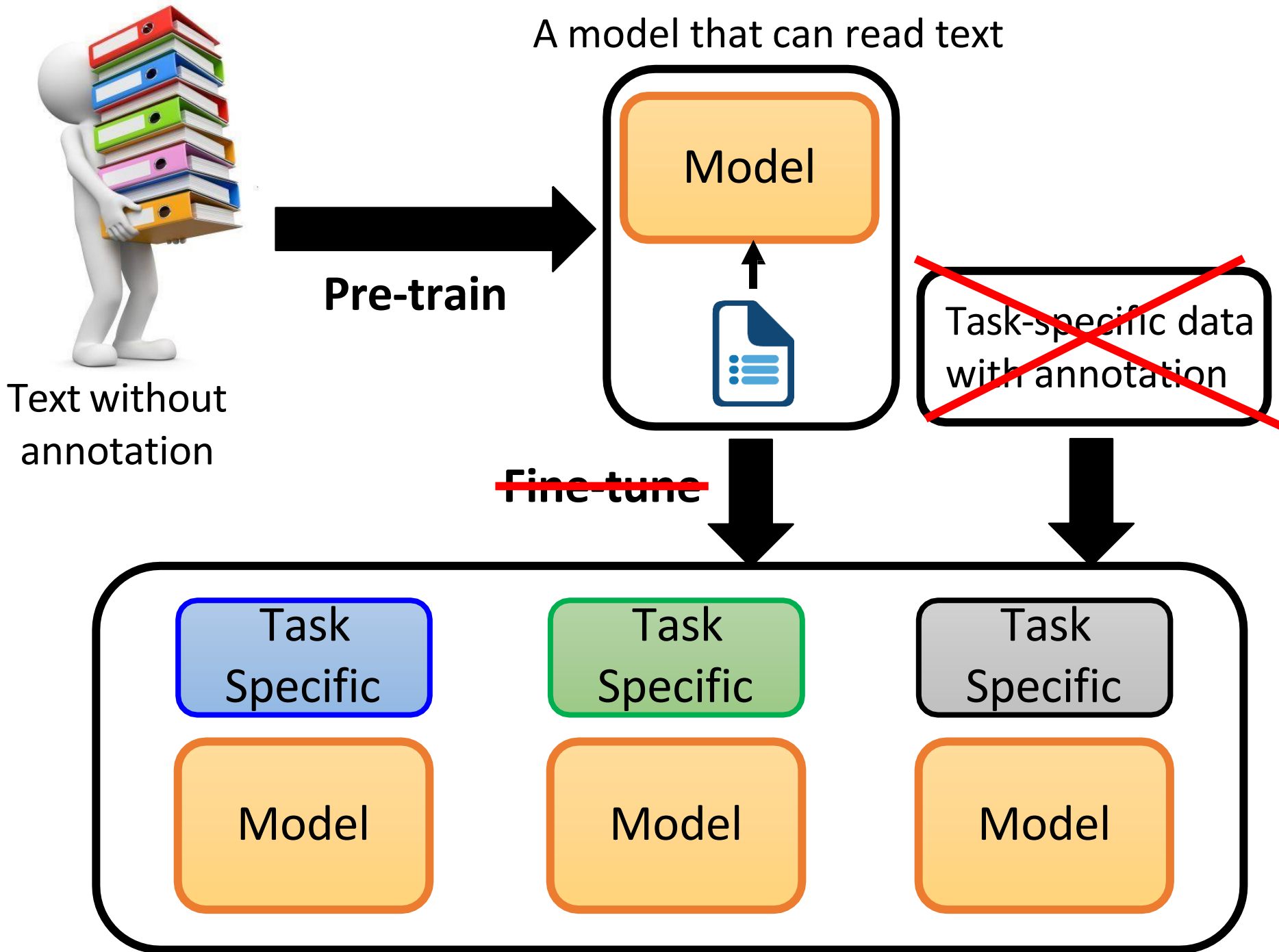
这节课我们要讲如何驾驭
大型语言模型



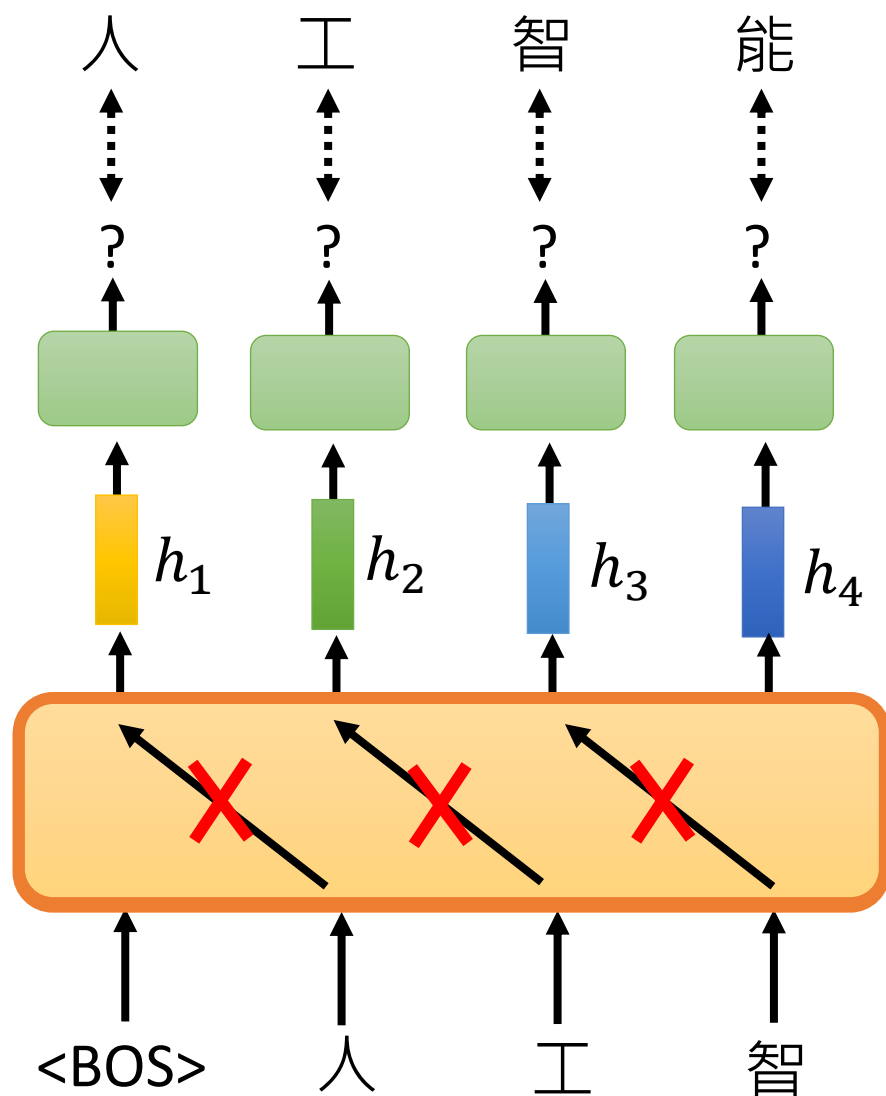
什么都会



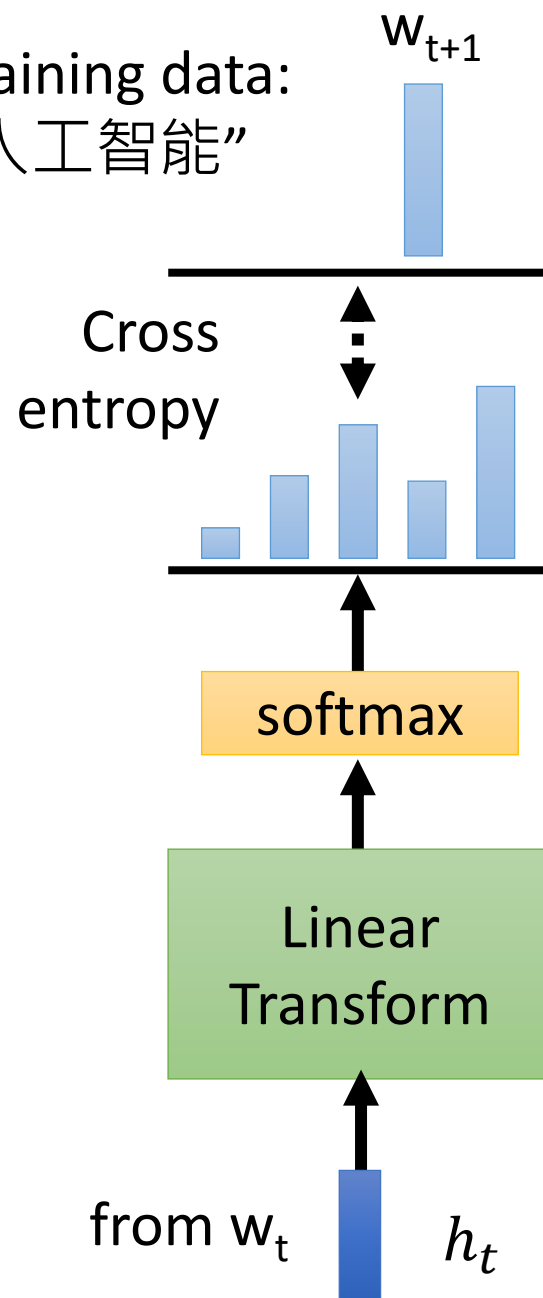
This course
is about



Predict Next Token



Training data:
“人工智能”



Predict Next Token

They can do generation.



EM PROMPT
-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL
COMPLETION
(MACHINE-
10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

How to use GPT?

第一节：听下面对话，每段对话有几个小题，从题中所给的A、B、C三个选项中选出最佳选项，并标在试卷的相应位置。

M: Excuse me.Can you tell me how much the shirt is?

W: Yes, it's nine fifteen.

请看选项：How much is the shirt?

[A] £ 19.15

[B] £ 9.15

[C] £ 9.18

衬衫的价格为9镑15便士，所以你选择[B]项，并在试卷上将其标出。

Description

A few example

“Few-shot” Learning

(no gradient
descent)

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

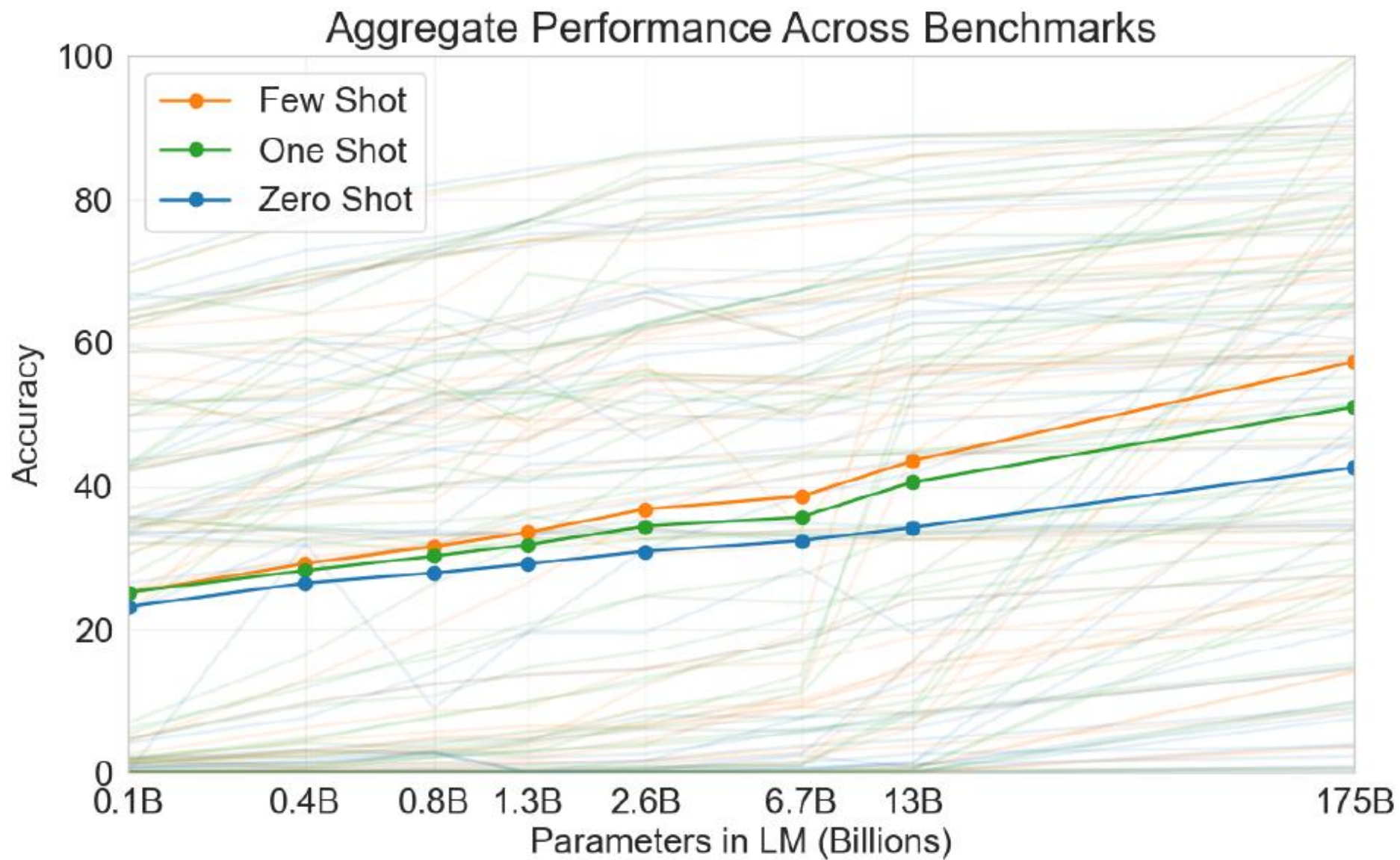
“One-shot” Learning

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

“Zero-shot” Learning

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

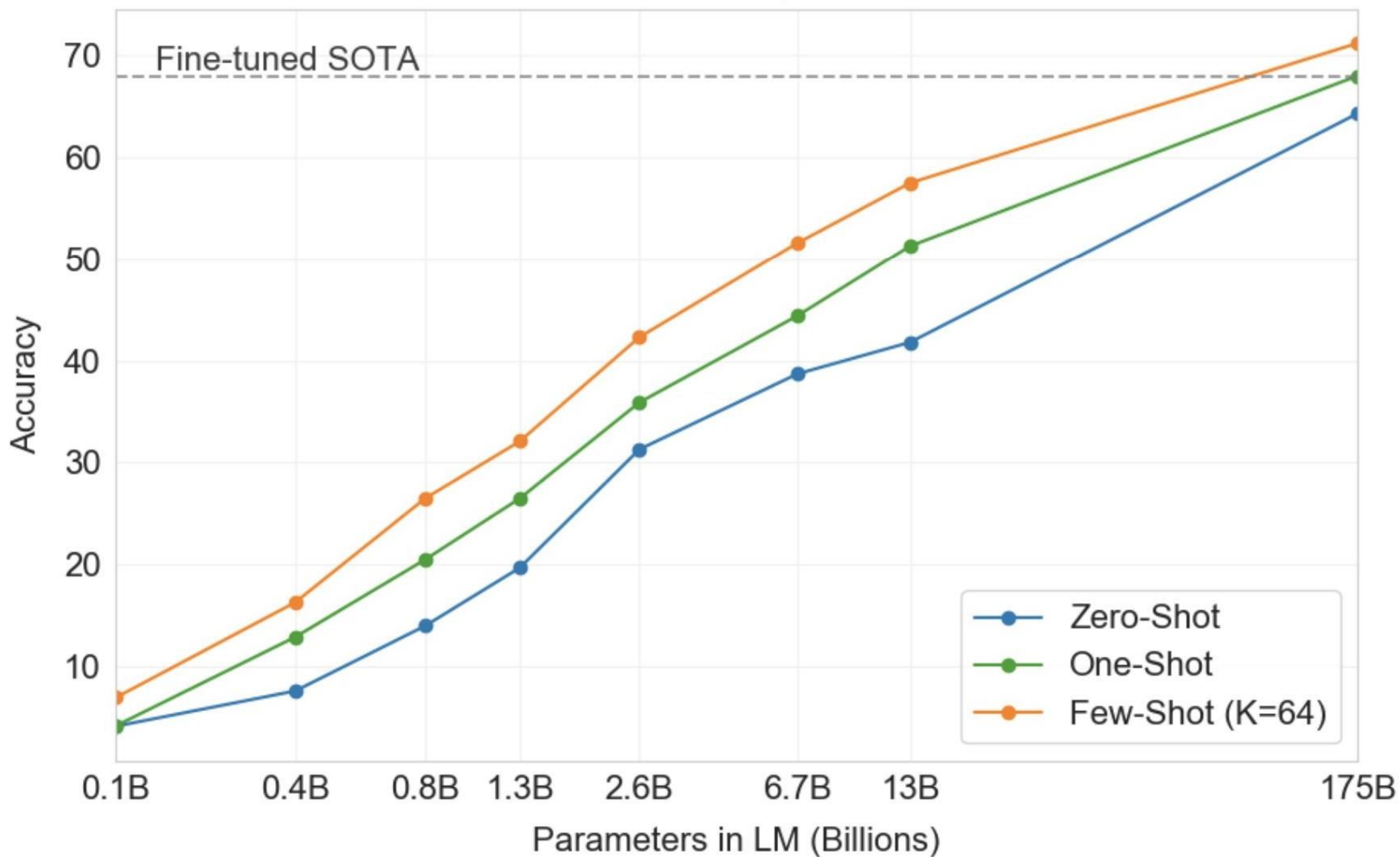
“In-context” Learning



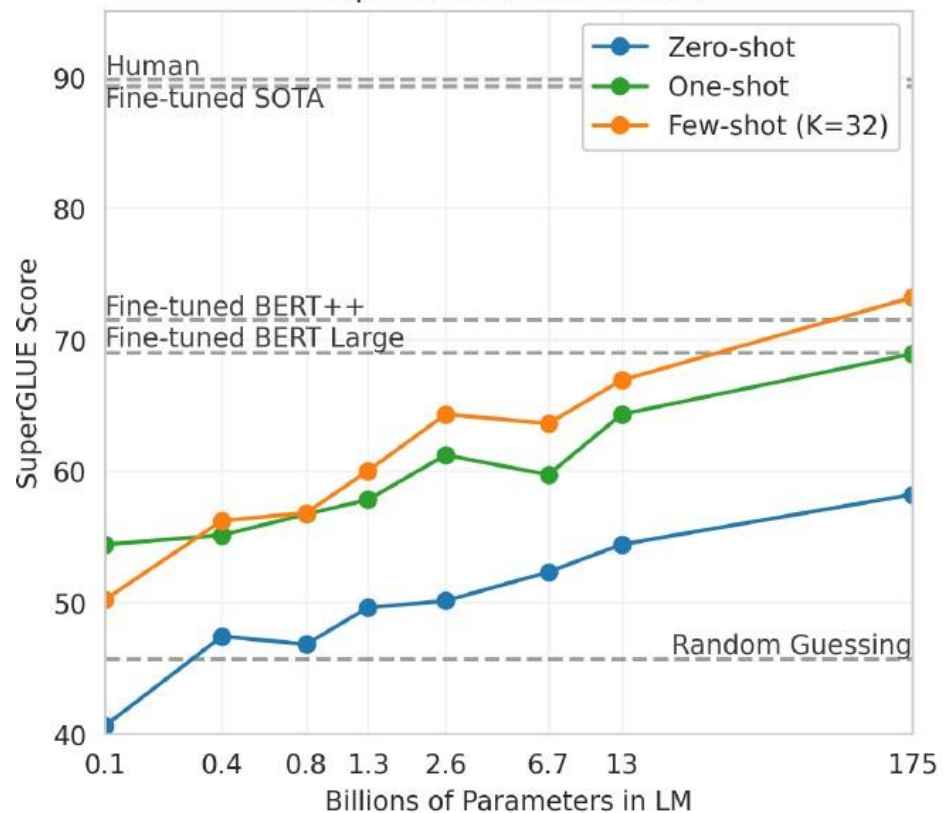
Average of **42** tasks

Closed Book QA

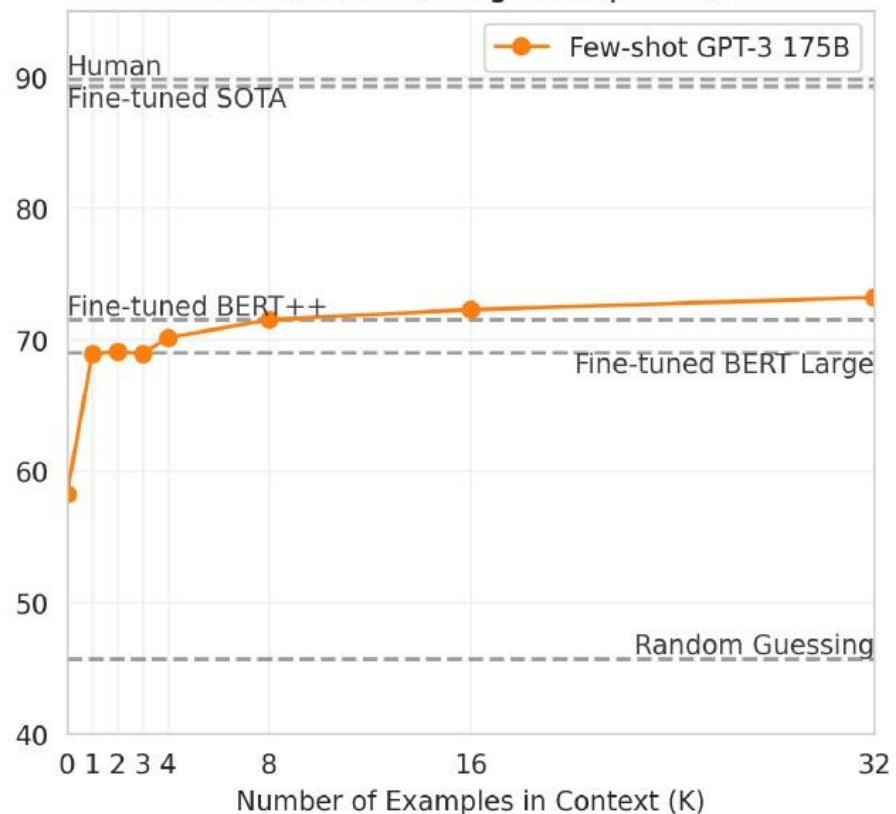
TriviaQA



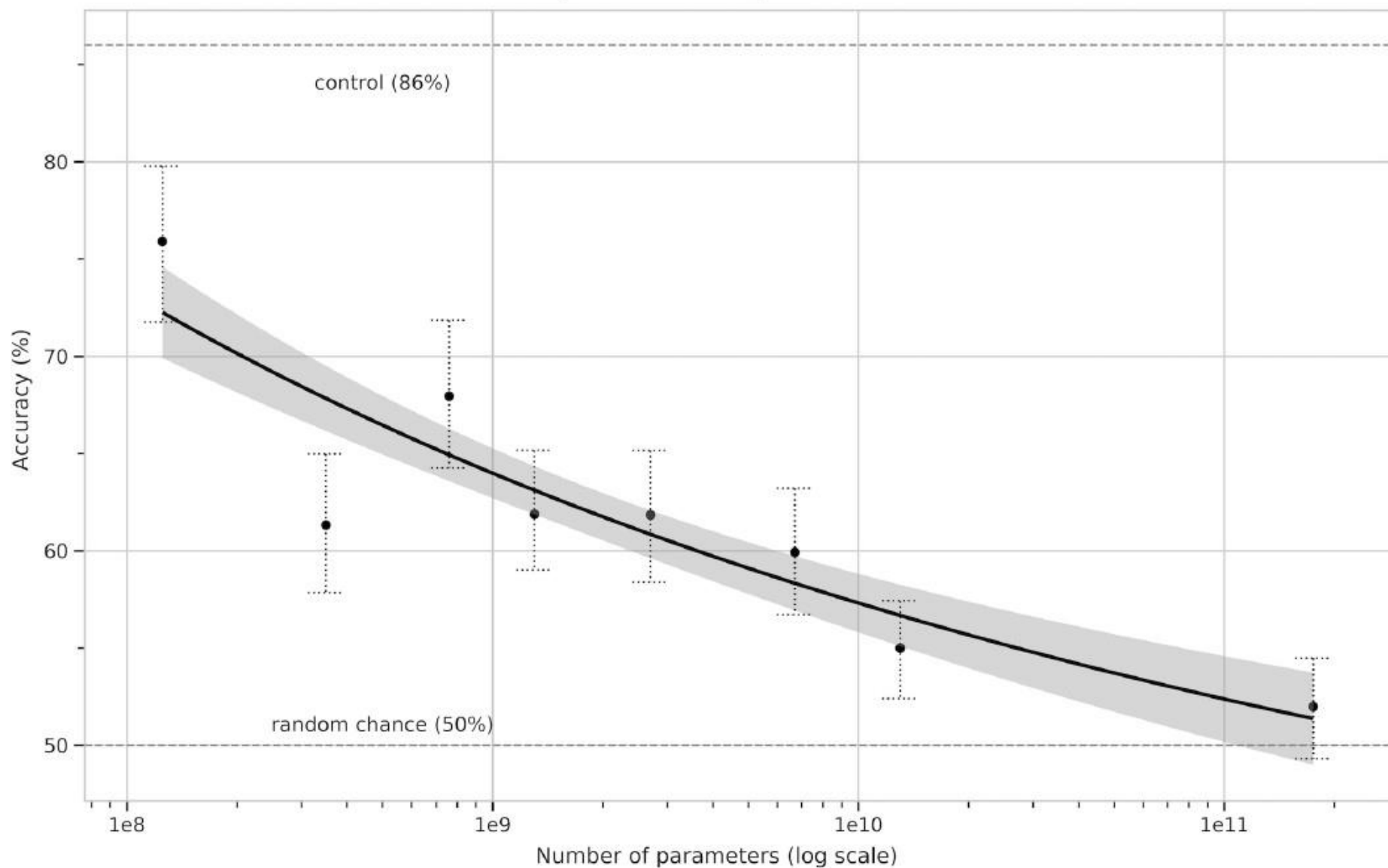
SuperGLUE Performance



In-Context Learning on SuperGLUE



Human ability to detect model generated news articles



A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:
We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

A "yalubalu" is a type of vegetable that looks like a big pumpkin. An example of a sentence that uses the word yalubalu is:

I was on a trip to Africa and I tried this yalubalu vegetable that was grown in a garden there. It was delicious.

A "Burringo" is a car with very fast acceleration. An example of a sentence that uses the word Burringo is:

In our garage we have a Burringo that my father drives to work every day.

A "Gigamuru" is a type of Japanese musical instrument. An example of a sentence that uses the word Gigamuru is:

I have a Gigamuru that my uncle gave me as a gift. I love to play it at home.

To "screeg" something is to swing a sword at it. An example of a sentence that uses the word screeg is:

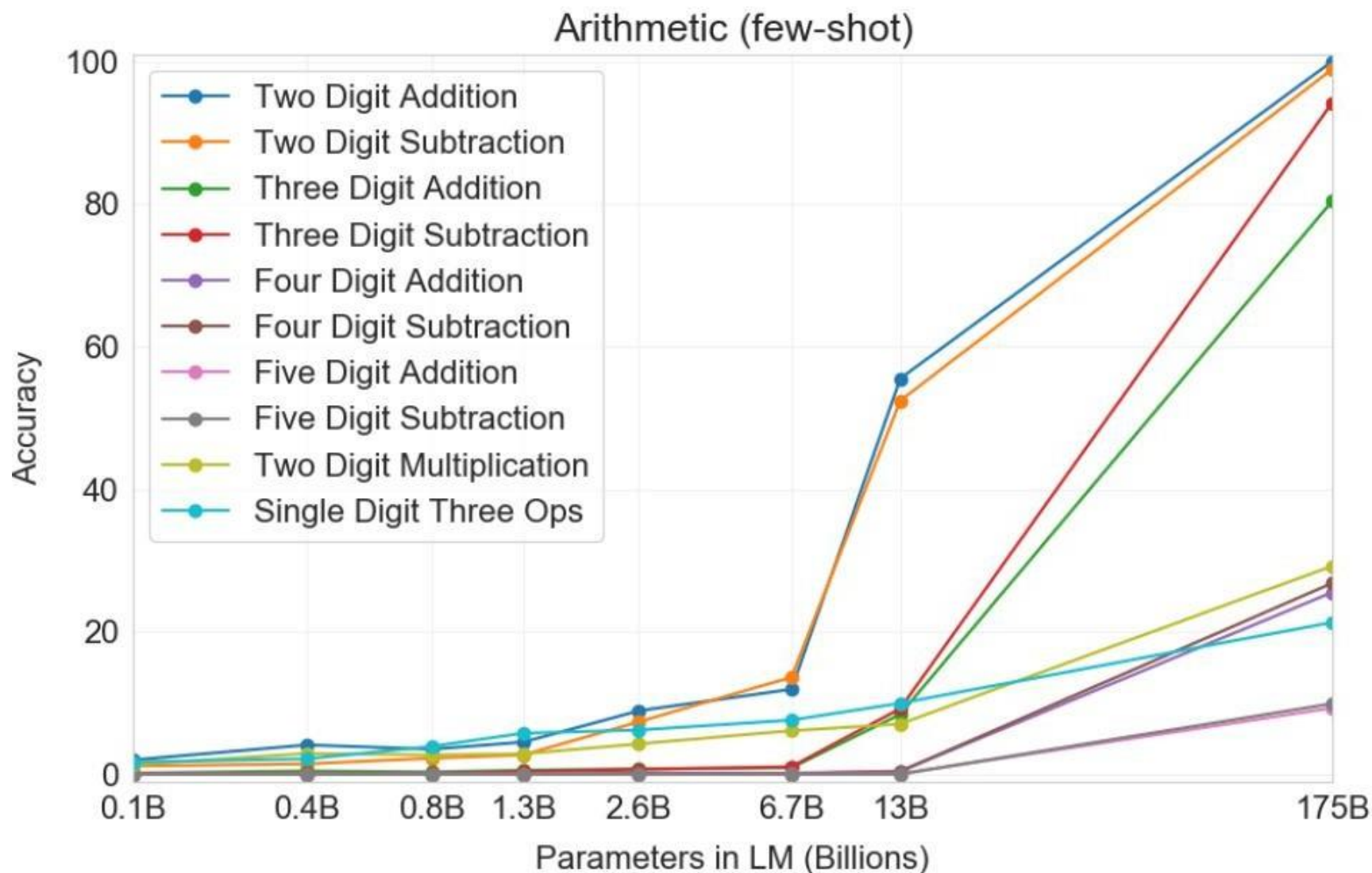
We screeghed at each other for several minutes and then we went outside and ate ice cream.

Figure 3.16: Representative GPT-3 completions for the few-shot task of using a new word in a sentence. Boldface is GPT-3's completions, plain text is human prompts. In the first example both the prompt and the completion are provided by a human; this then serves as conditioning for subsequent examples where GPT-3 receives successive additional prompts and provides the completions. Nothing task-specific is provided to GPT-3 other than the conditioning shown here.

Context → Q: What is 17 minus 14?

A:

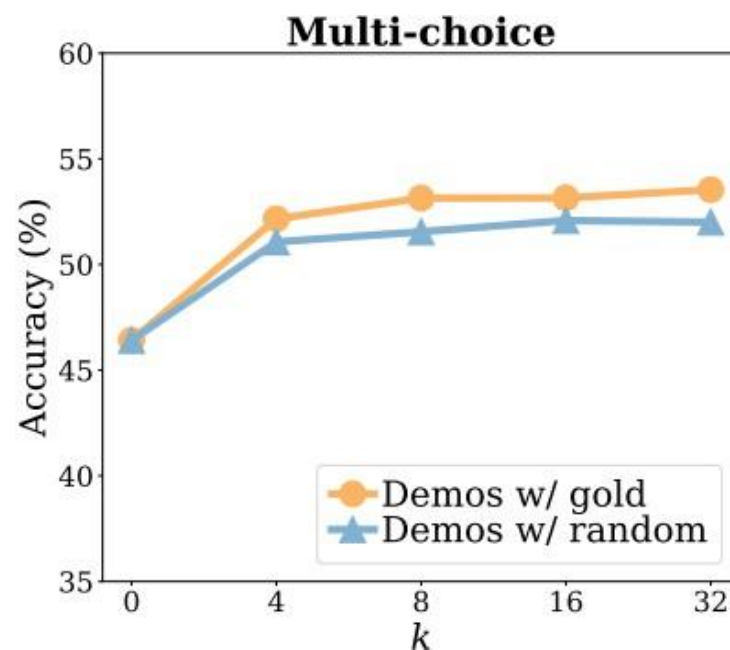
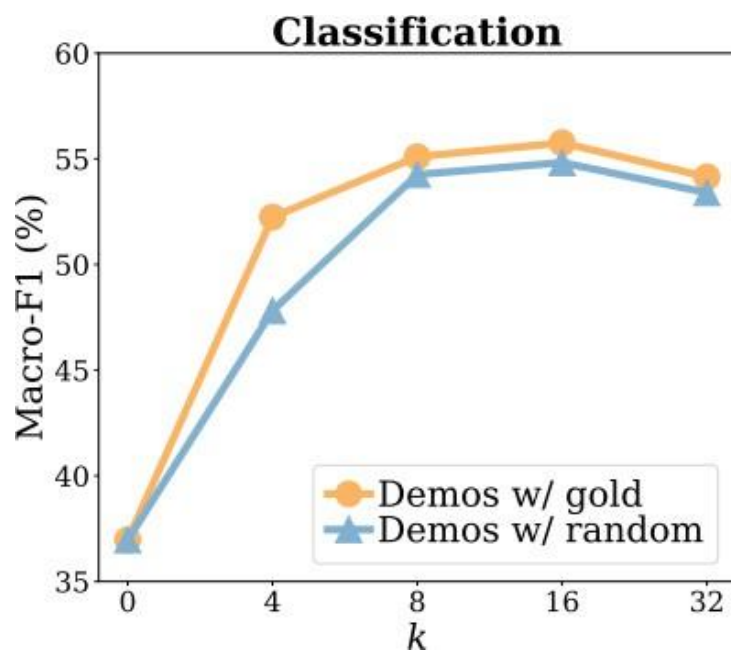
Target Completion → 3



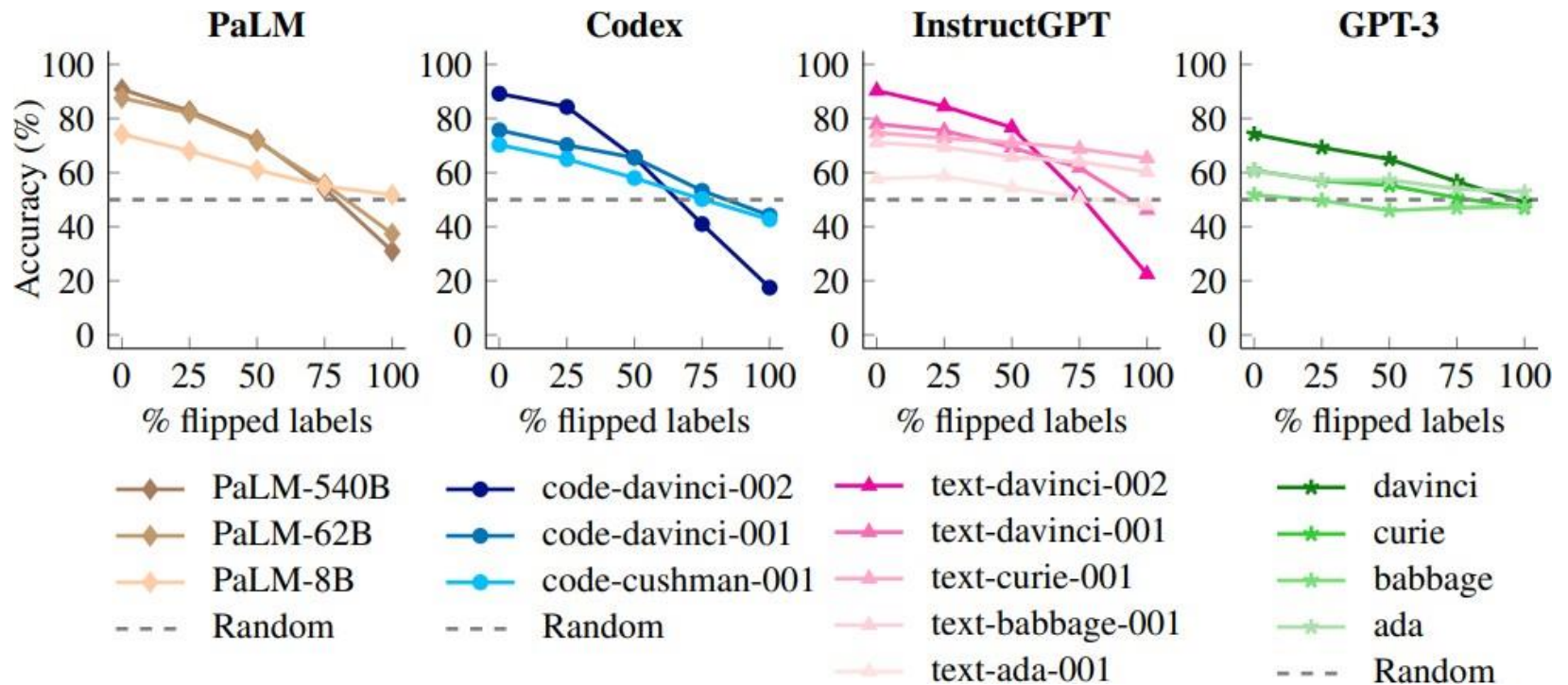
In-context Learning

Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?

Ref: <https://arxiv.org/abs/2202.12837>



In-context Learning



In-context Learning

- Classification!

Input: 680, 841, 842, 496, 204, 985, 546, 275, 453, 835, 644, 1, 308, 5, 65, 160

Output: Bar

Input: 193, 101, 270, 957, 670, 407, 104, 23, 569, 708, 700, 395, 481, 105, 234, 785

Output: Foo

Input: 16, 409, 28, 668, 53, 342, 813, 181, 963, 728, 558, 420, 975, 686, 395, 931

Output: Bar

Input: 448, 421, 190, 246, 413, 766, 463, 332, 935, 911, 304, 244, 876, 95, 236, 695

Output: Foo

Input: 632, 318, 49, 138, 602, 508, 924, 227, 325, 767, 108, 254, 475, 298, 202, 989

Output: Foo

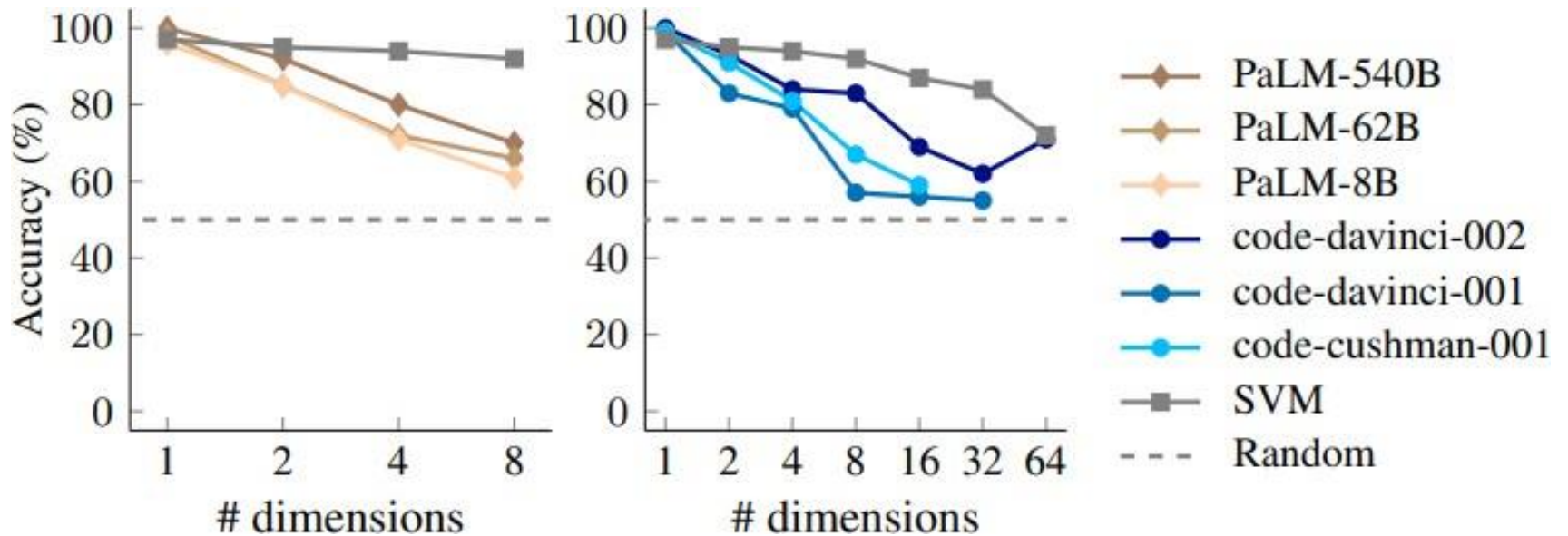
Input: 412, 140, 30, 508, 837, 707, 338, 669, 835, 177, 312, 800, 526, 298, 214, 259

Output:

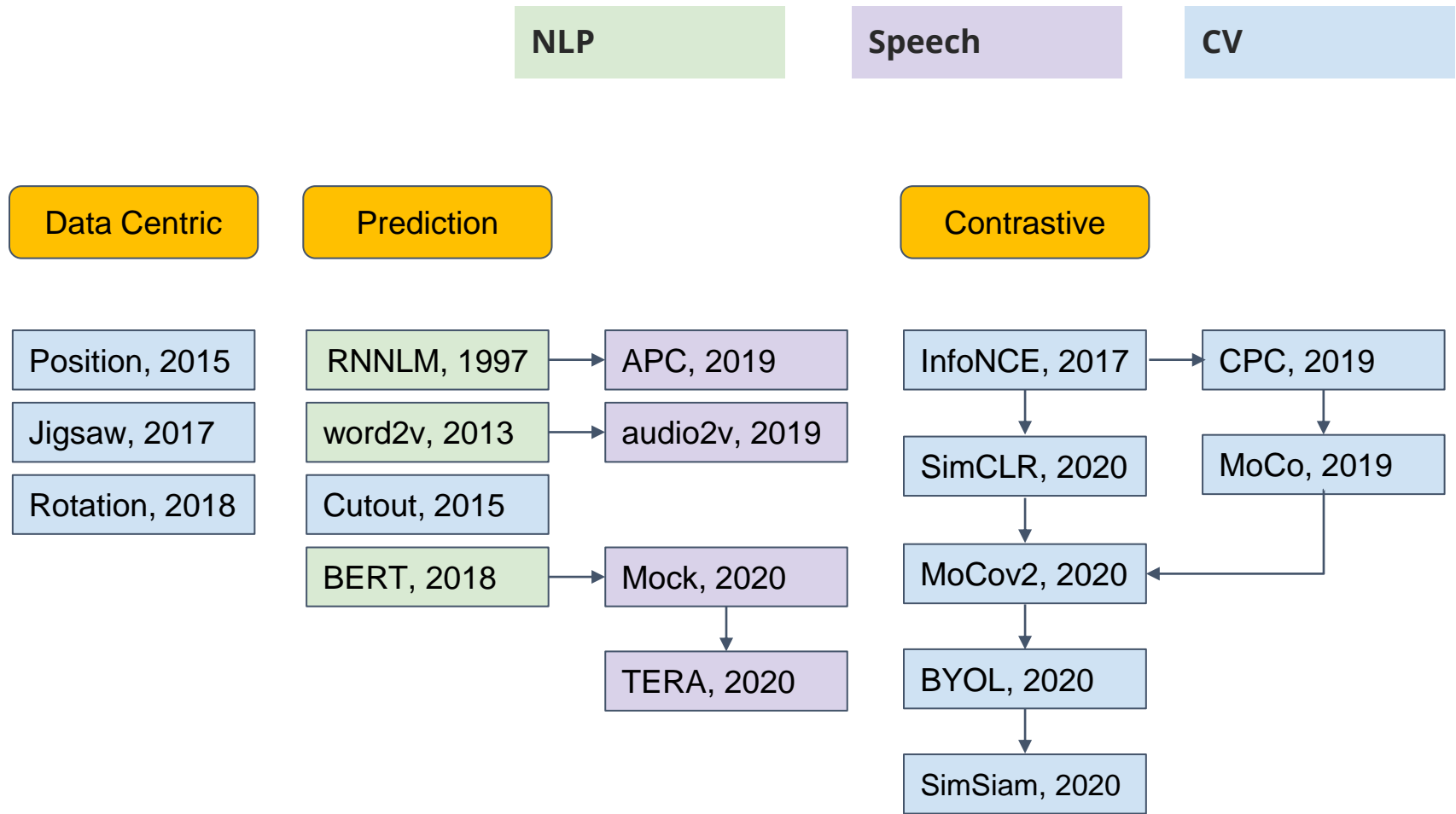
Answer:

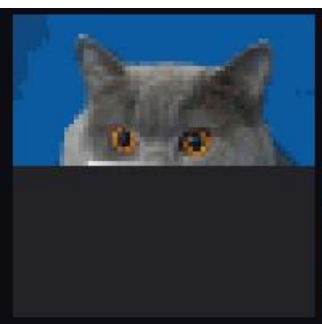
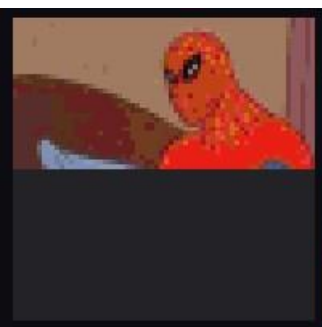
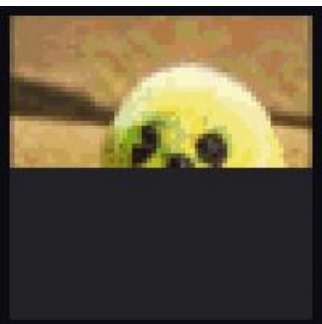
Bar

In-context Learning



Beyond Text





Source of image: <https://openai.com/blog/image-gpt/>

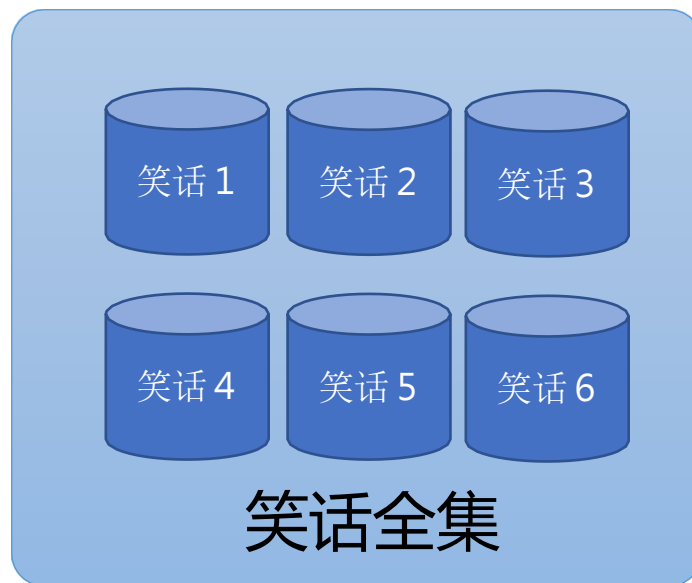
ChatGPT



对 ChatGPT 的常见误解 1

ChatGPT用罐头回应

讲个笑话吧?



由开发者准备好的罐头回应

对 ChatGPT 的常见误解 2

ChatGPT 的答案是网络搜索的结果

什么是 Diffusion Model ?



Diffusion Model is ...



(从网络上抄来的句子)

整理、重组

多数 ChatGPT 的答案在网络上都找不到一模一样的文本

甚至有很多幻想出来的答案

What are Diffusion Models?

July 11, 2021 · 26 min · Lilian Weng

Diffusion Models: A Comprehensive Survey

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Ruosong Wang, Hsuan Yang

Diffusion models have emerged as a powerful new family of deep learning applications, including image synthesis, video generation, and more. This survey provides an expanding body of work on diffusion models, categorizing the research into image, audio, and text. We also discuss the challenges of training diffusion models for enhanced results. We further review the wide-ranging applications of diffusion models in natural language processing, temporal data modeling, to interdisciplinarily provide a contextualized, in-depth look at the state of diffusion models and further exploration. Github: [this https URL](https://github.com/yangling0309/diffusion-models).

(网络上搜索文章)

对 ChatGPT 的常见误解 2

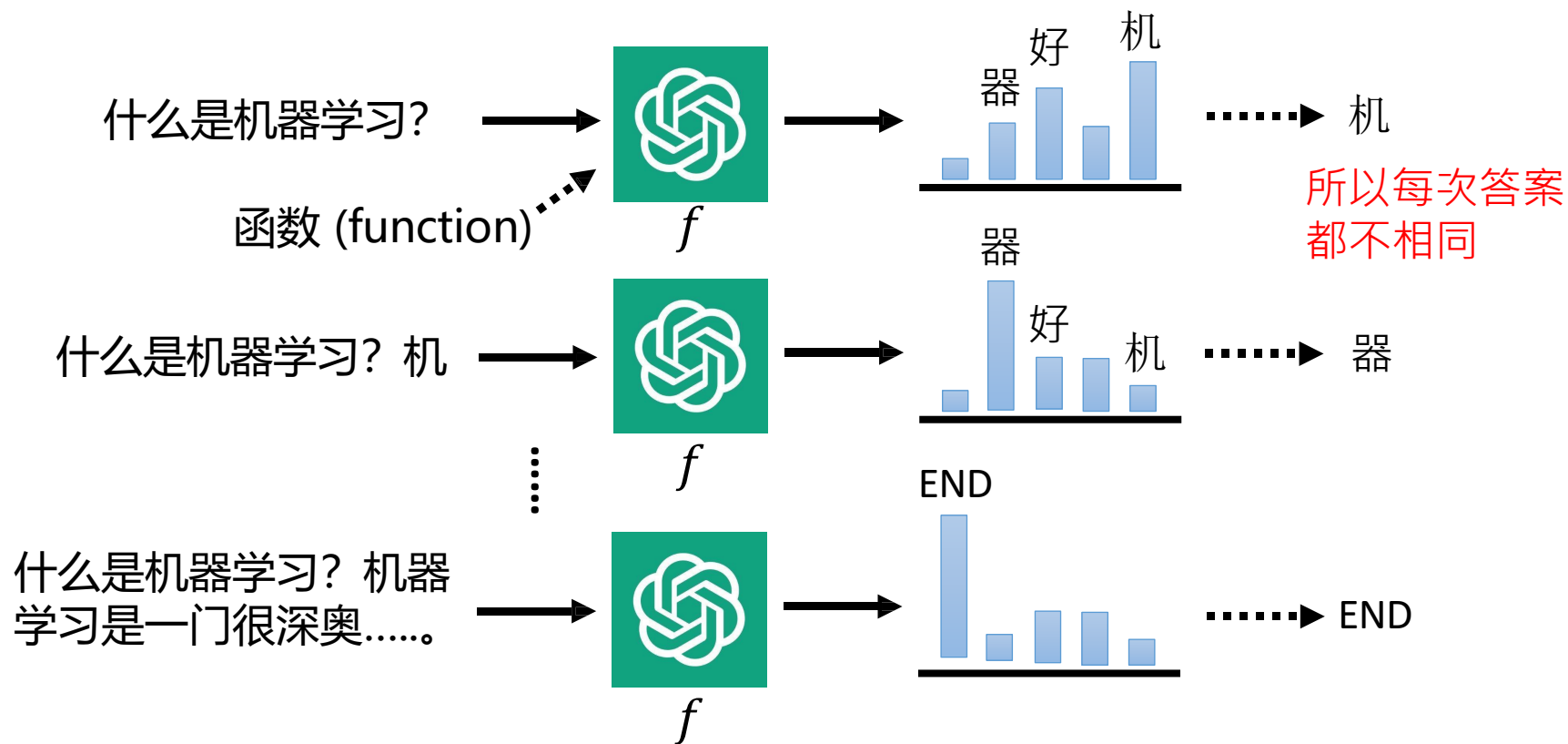
4. Can I trust that the AI is telling me the truth?

- ChatGPT is not connected to the internet and it can occasionally produce incorrect answers. It has limited knowledge of world and events after 2021 and may also occasionally produce harmful instructions or biased content.

We'd recommend checking whether responses from the model are accurate or not. If you find an answer is incorrect, please provide that feedback by using the "Thumbs Down" button.

<https://help.openai.com/en/articles/6783457-chatgpt-general-faq>

ChatGPT 真正做的事- 文字接龙



监督学习套用到 ChatGPT 上

输入：上海第一高楼是那一座？

输出：上海中心大厦

输入：帮我修改这段文字

输出：好的

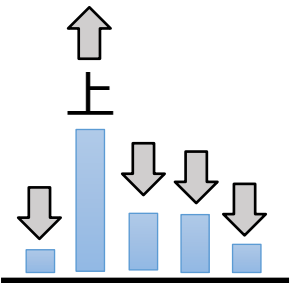
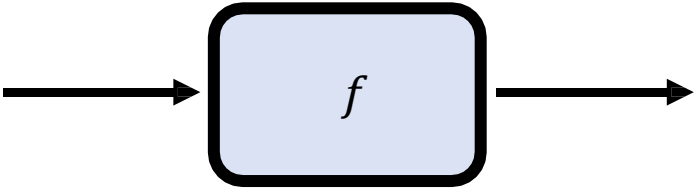
输入：教我做坏事

输出：这是不对的

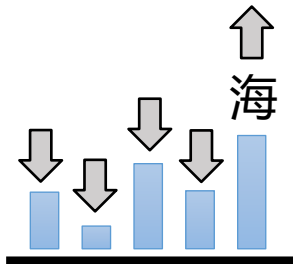
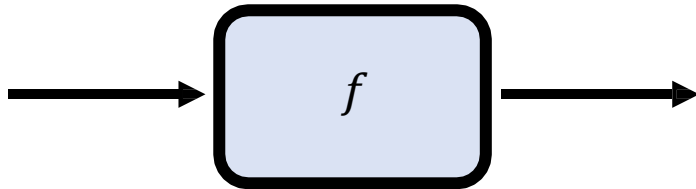


机器自己寻找函数 f 使得

上海第一高楼是那一座？



上海第一高楼是那一座？ 上



监督学习套用到 ChatGPT 上

输入：上海第一高楼是那一座？

输出：上海中心大厦

输入：帮我修改这段文字

输出：好的

输入：教我做坏事

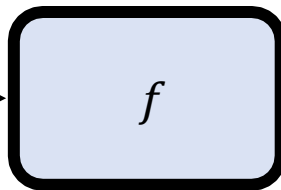
输出：这是不对的



机器自己寻找函数 f 使得

假设机器真的根据老师的教导找出函数 f ，他的能力可能还是非常有限
人类老师可以提供的成对资料是有限的

世界第一高山是那一座？



? ? ? ? ?



轻松生成成对数据

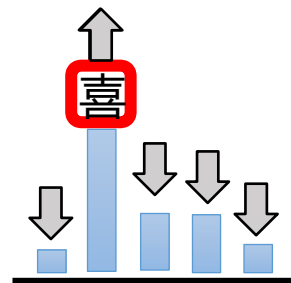
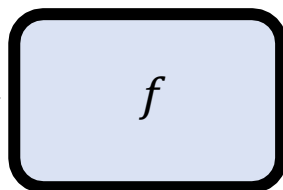
网络上每一段文字都可以教机器做文字接龙



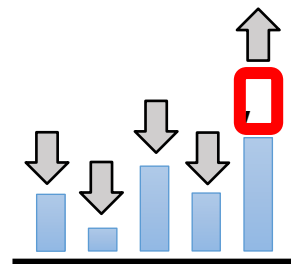
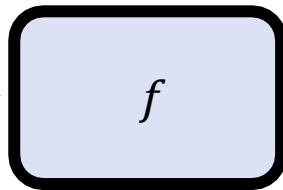
.....→ 世界第一高山是喜马拉雅山
今天天气真好，我要出去玩

机器自己寻找函数 f 使得

世界第一高山是



今天天气真好



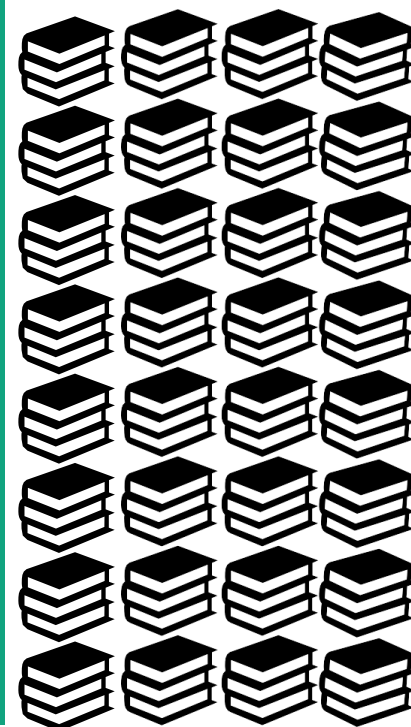
GP
T

在 ChatGPT 之前的 GPT 系列

Model:
1542M



Data:
40GB



Model:
117M



Data:
1GB

GPT
(2018)

GPT-2
(2019)

GPT-2

<https://openai.com/blog/better-language-models/>



EM PROMPT
-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL
MPLETION
(MACHINE-
10 TRIES)

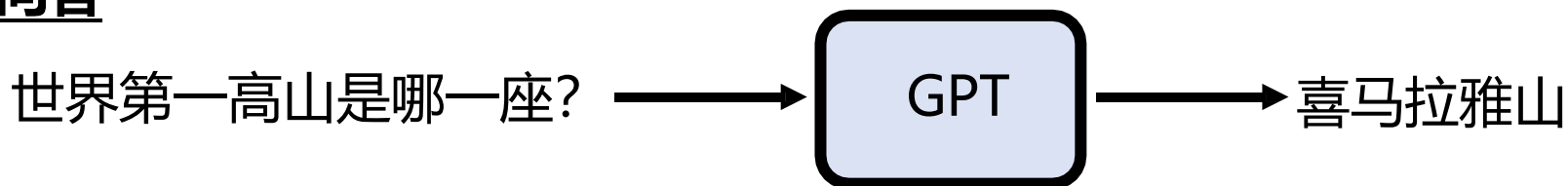
The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

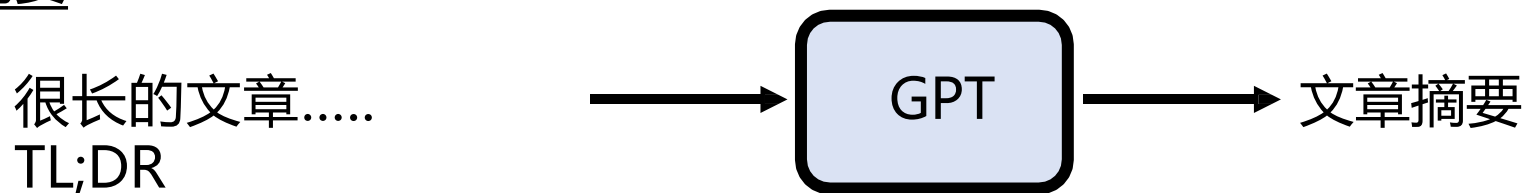
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

GPT-2 也是可以回答问题的!

问答



摘要

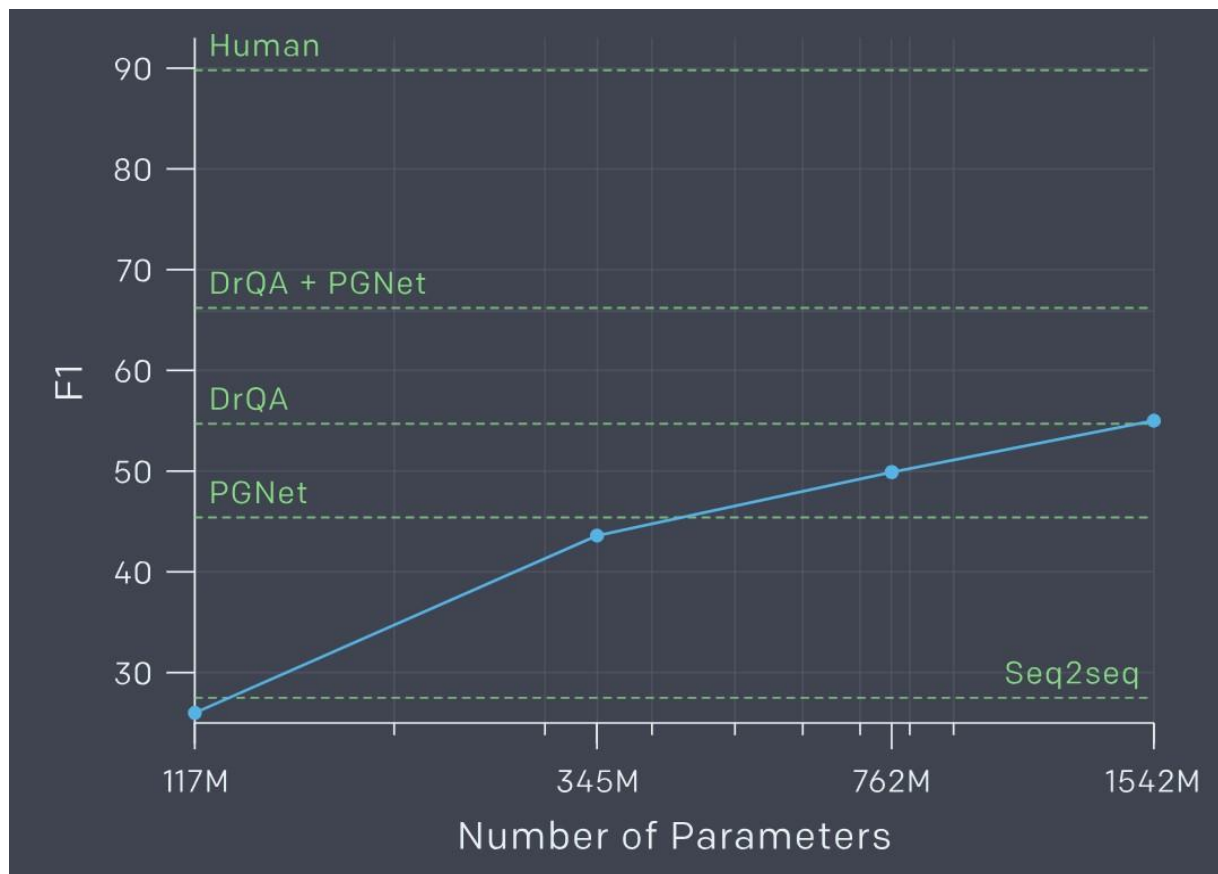


GPT-2

<https://openai.com/blog/better-language-models/>

问答上表现
如何?

CoQA



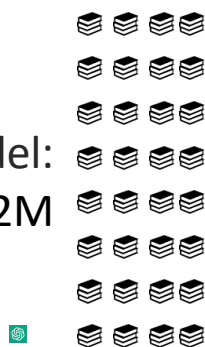
在 ChatGPT 之前的 GPT 系列

GPT-3.5?

<https://platform.openai.com/docs/model-index-for-researchers>

Model:
175B

Model:
1542M



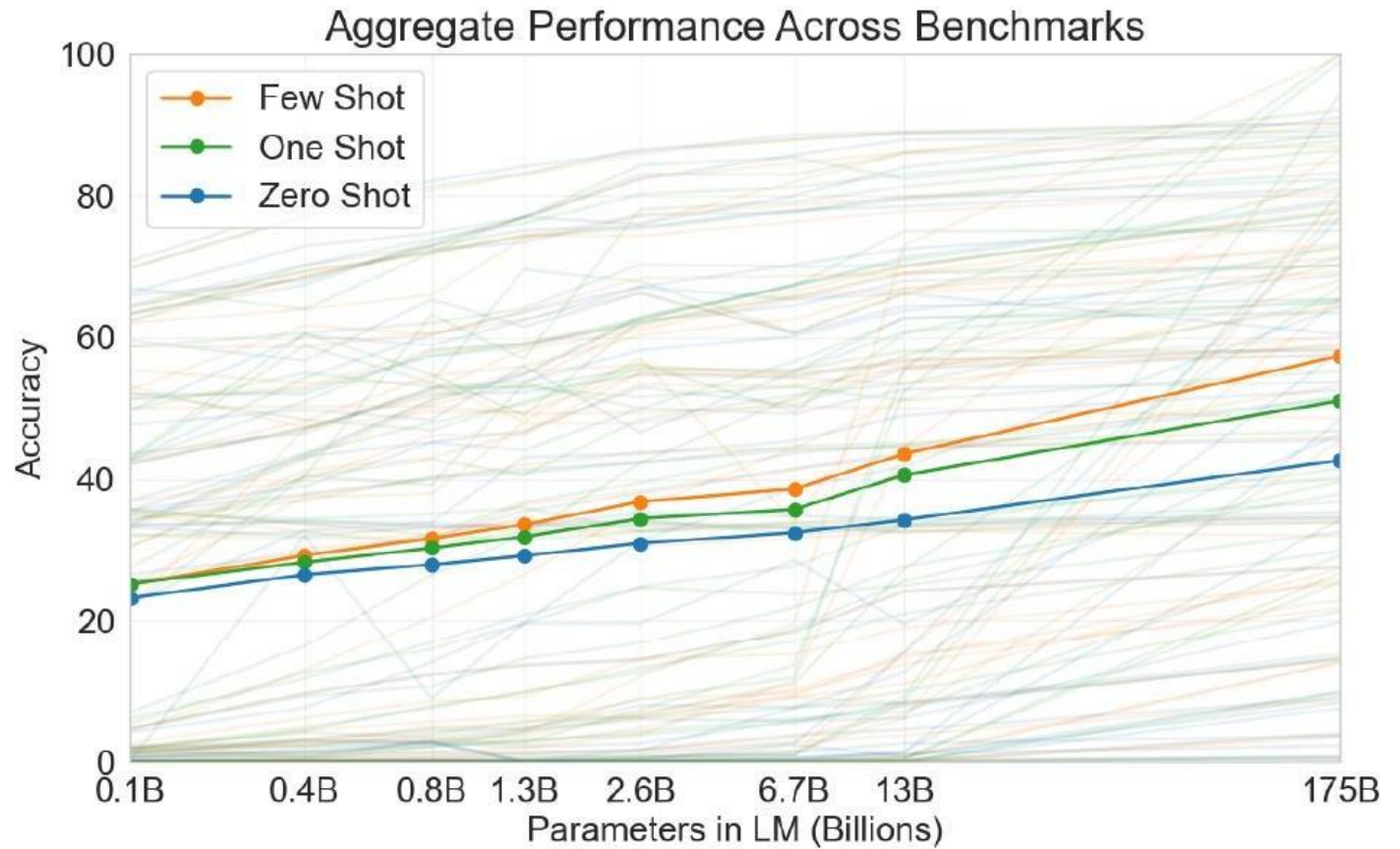
GPT-2
(2019)



GPT-3
(2020)

Data: 570GB 阅读哈利波特全集 30 万遍
(45T before filtering)

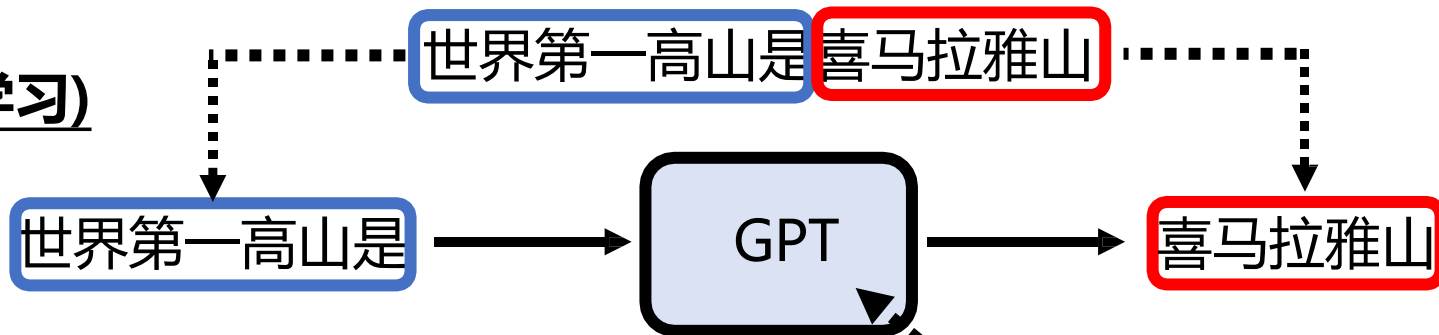
GPT-3



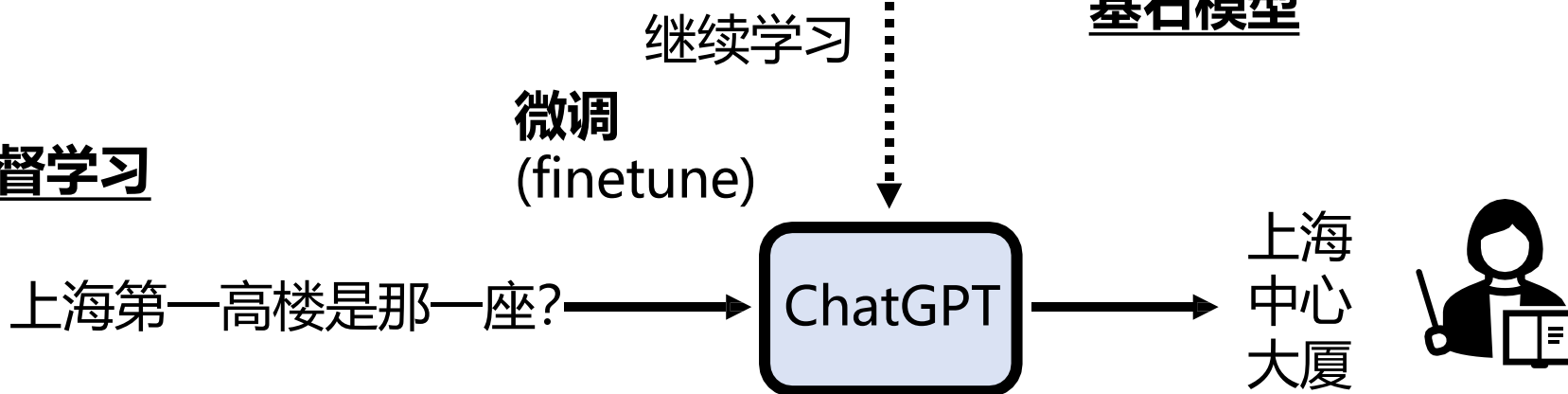
<https://arxiv.org/abs/2005.14165>

Average of **42** tasks

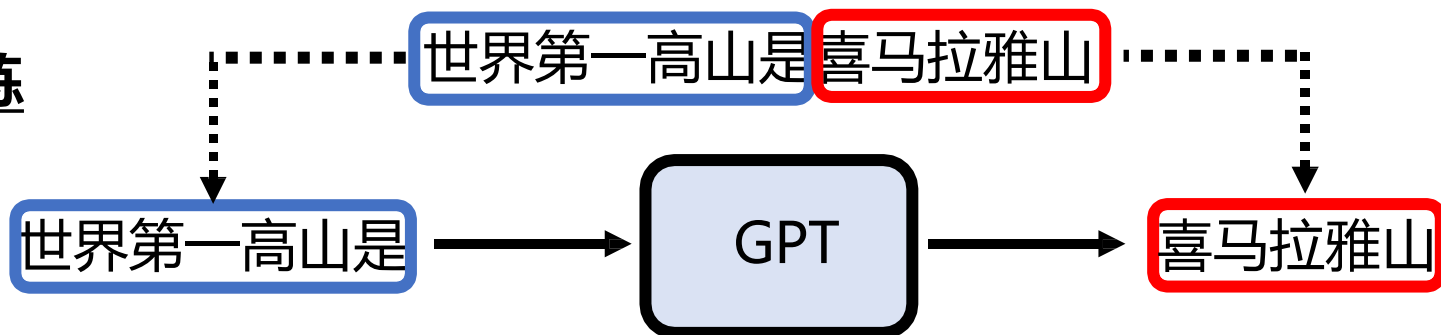
预训练
(自监督学习)



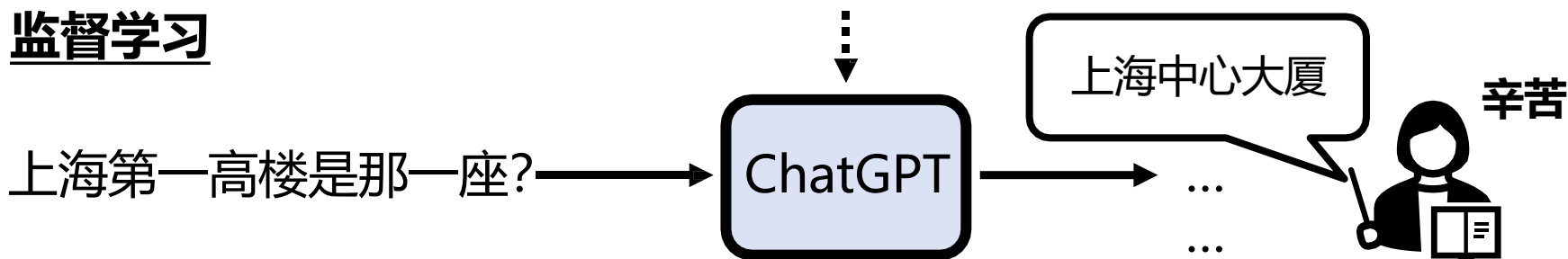
监督学习



预训练



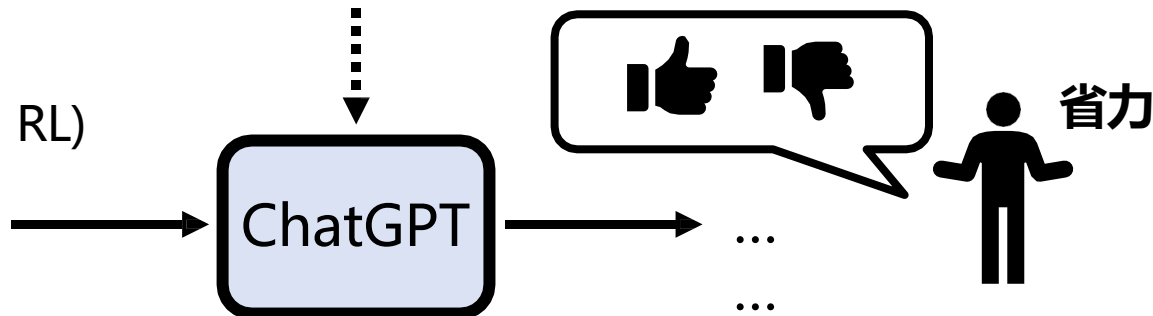
监督学习



强化学习

(Reinforcement Learning, RL)

请帮我写诗赞美AI



GPT-4



GPT-4 Technical Report

OpenAI*

Abstract

We report the development of GPT-4, a large-scale, multimodal model which can accept image and text inputs and produce text outputs. While less capable than humans in many real-world scenarios, GPT-4 exhibits human-level performance on various professional and academic benchmarks, including passing a simulated bar exam with a score around the top 10% of test takers. GPT-4 is a Transformer-based model pre-trained to predict the next token in a document. The post-training alignment process results in improved performance on measures of factuality and adherence to desired behavior. A core component of this project was developing infrastructure and optimization methods that behave predictably across a wide range of scales. This allowed us to accurately predict some aspects of GPT-4's performance based on models trained with no more than 1/1,000th the compute of GPT-4.

<https://openai.com/research/gpt-4>

What do we know?

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [33] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [34]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

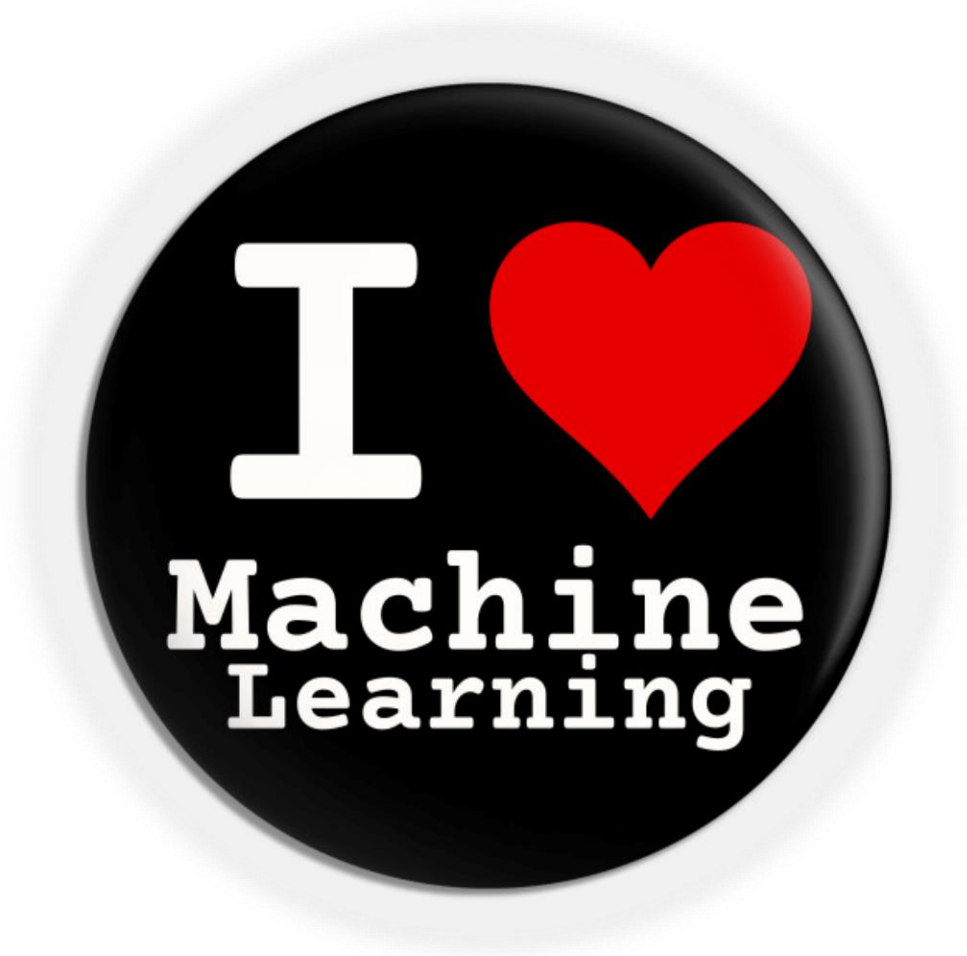
GPT 看得见了!

User Can you explain why this is funny. Think about it step-by-step.



GPT-4 The comic is satirizing the difference in approaches to improving model performance between statistical learning and neural networks.

In statistical learning, the character is shown to be concerned with overfitting and suggests a series of complex and technical solutions, such



希望大家乐享学习“人工智能”!

我们的征途
是星辰大海

