

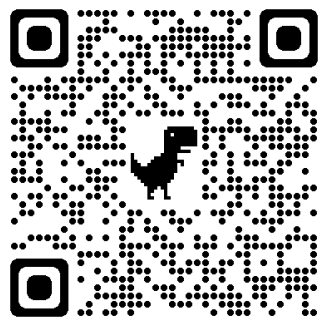


ChatGLM: 从千亿模型到ChatGPT的一点思考

唐杰

清华大学计算机系知识工程实验室 (KEG)

清华大学人工智能基础模型研究中心



Open AI引领了AGI的研究...

*我们更开放

预训练大模型是新一代人工智能应用的基础设施

大模型通用能力强
可完成多场景任务

单模态向多模态发展
从文本逐步发展到视觉

亟需建立以中文为核心
的超大模型

大模型元年

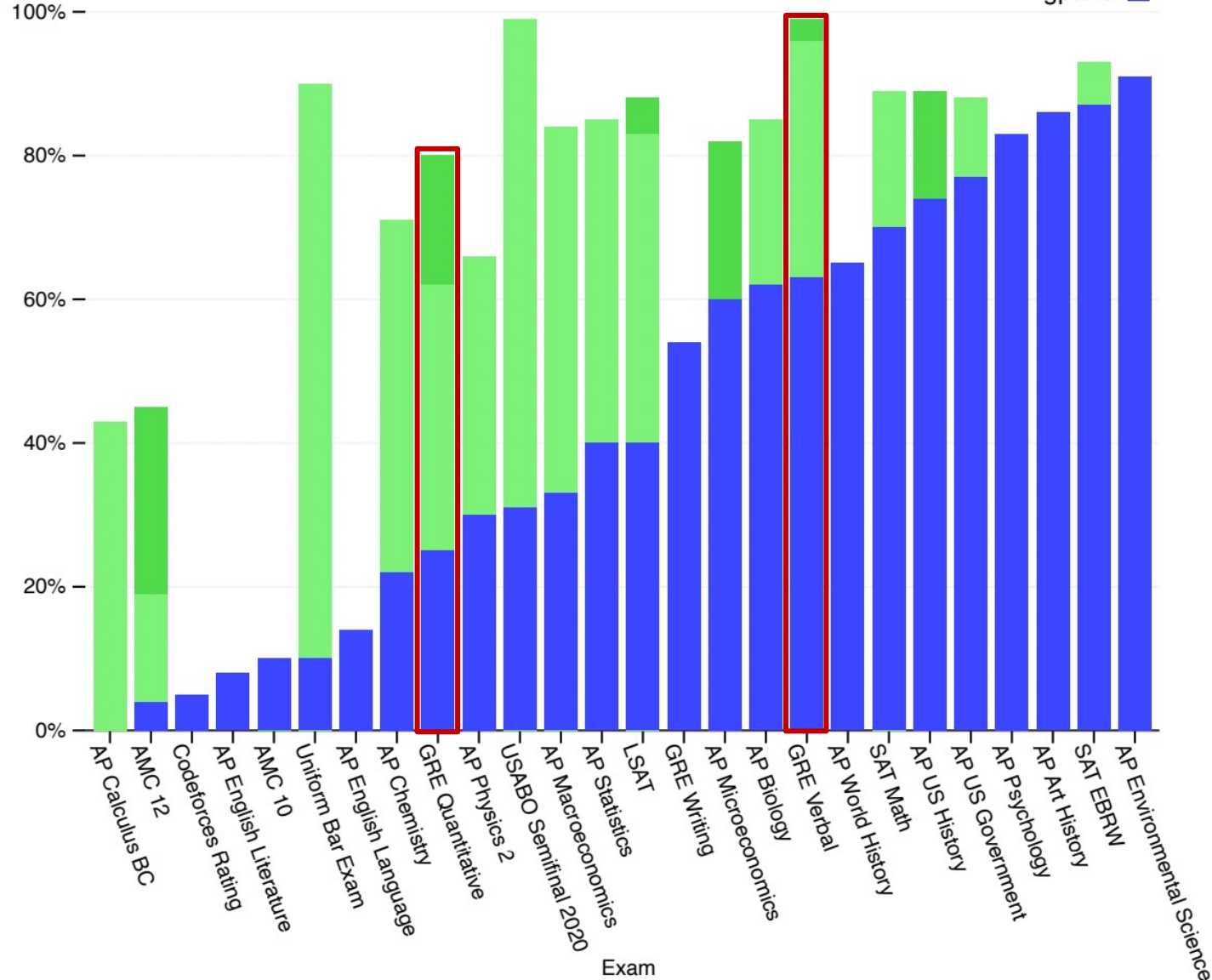


GPT-4

Exam results (ordered by GPT-3.5 performance)

Estimated percentile lower bound (among test takers)

gpt-4
gpt-4 (no vision)
gpt3.5



GPT-4

What is funny about this image? Describe it panel by panel.



GPT-4

The image shows a package for a "Lightning Cable" adapter with three panels.

Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

Source: <https://www.reddit.com/r/hmmm/comments/ubab5v/hmmm/>

创意文案



帮我生成一个短视频拍摄脚本，向大家介绍清华大学的校园环境，包含分镜的画面描述和取景点



片头：清华大学的校园环境非常优美，今天我们将带您一起去了解一下。

【画面一】校园入口

镜头从高空缓缓降落，拍摄清华大学校园的入口。入口两侧是高大的树木和美丽的花坛，门口站着两名保安，看起来非常严肃。

【画面二】校园道路

清华大学校园的道路非常宽敞，全部是由青石板铺成的。路上车辆很少，行人也很少，显得非常安静。镜头可以拍摄到路边的草坪和树木，还有一些教学楼和宿舍楼。

【画面三】校园湖泊

清华大学校园有一个非常大的湖泊，叫做“近春湖”。湖泊四周被美丽的柳树和桃花环绕，湖面上还有几只鸭子在悠闲地游泳。镜头可以拍摄到湖边的草地和远处的教学楼。

【画面四】校园建筑

清华大学校园的建筑非常壮观，大多数都是两层的小楼。建筑外立面全部是由青石板和红砖组成，看起来非常古朴。镜头可以拍摄到一些经典的建筑，比如“清华楼”、“科学馆”等。

【画面五】校园活动

清华大学校园的活动非常丰富多彩，每天都有很多人在操场上进行体育锻炼。还有一些学生在图书馆里认真学习，还有一些学生在校园里散步，看起来非常舒适。

【画面六】校园夜景

清华大学校园的夜景非常美丽，所有的建筑都被点亮了。湖面上还有几只小船在悠闲地游荡，镜头可以拍摄到一些教学楼和宿舍楼的夜景，以及一些草坪和树木。

【画面七】校园结尾

最后，镜头拍摄到清华大学校园的标志性建筑“清华门”，然后切换到标题字幕，结束画面。

以上是一个简单的清华大学校园环境的介绍，希望对您有所帮助。



常识

🔍 孙悟空有没有火烧赤壁

📖 孙悟空是中国古代小说《西游记》中的重要人物，他是一位神通广大的猴子，经过修行后成为了斗战胜佛。在小说中，孙悟空并没有参与火烧赤壁的战斗。

火烧赤壁是中国古代历史上著名的战役之一，发生在公元 208 年，是刘备、孙权联合对抗曹操的军队而形成的三方对峙局面。在战役中，孙刘联军采取水战，用火攻来攻破曹军。曹操的军队在火势蔓延之下，陷入混乱，最终溃败。

虽然孙悟空没有在火烧赤壁中战斗，但他在小说中参与了其他许多重要的战役和冒险，最终成为了斗战胜佛，成为了中国文学和艺术中的重要形象之一。

以上内容为 AI 生成，不代表开发者立场，请勿删除或修改本标记

👍 🗨️ 📄 复制

🔍 林黛玉有没有倒拔垂杨柳

📖 林黛玉是《红楼梦》中的人物，她自小多病，被称为病美人。在小说中，并没有提到林黛玉倒拔垂杨柳的情节。然而，有一些网友可能会恶搞小说情节，创作出林黛玉倒拔垂杨柳的段子。倒拔垂杨柳是鲁智深干的事情，鲁智深因为拳打镇关西的事情跑到东京大相国寺看守菜园子。菜园子经常有二、三十个泼皮来偷菜，没人能管得了他们。听说来了新人，他们就上门准备欺负一下。可是遇到了鲁智深，鲁智深就把两个领头的踢到粪坑里。这样的操作吓坏了泼皮。第二天泼皮上门赔礼，一起吃饭的时候，因为鸟叫声，鲁智深喝了点酒，就干出了倒拔垂杨柳的事情。

以上内容为 AI 生成，不代表开发者立场，请勿删除或修改本标记

👍 🗨️ 📄 复制



代码

用python帮忙写个冒泡排序算法的代码，再给一个数组作为例子，写个测试程序

以下是冒泡排序算法的 Python 代码实现：

```
def bubble_sort(arr):  
    n = len(arr)  
    # 遍历所有数组元素  
    for i in range(n):  
        # 最后 i 个元素已经排好序，不需要再比较  
        for j in range(0, n-i-1):  
            # 如果当前元素比下一个元素大，则交换它们的位置  
            if arr[j] > arr[j+1]:  
                arr[j], arr[j+1] = arr[j+1], arr[j]  
    return arr
```

下面是测试程序，它将输入一个数组，并使用冒泡排序算法对其进行排序，然后输出排序后的数组：

```
arr = [64, 34, 25, 12, 22, 11, 90]  
sorted_arr = bubble_sort(arr)  
print(sorted_arr)
```

输出结果如下：

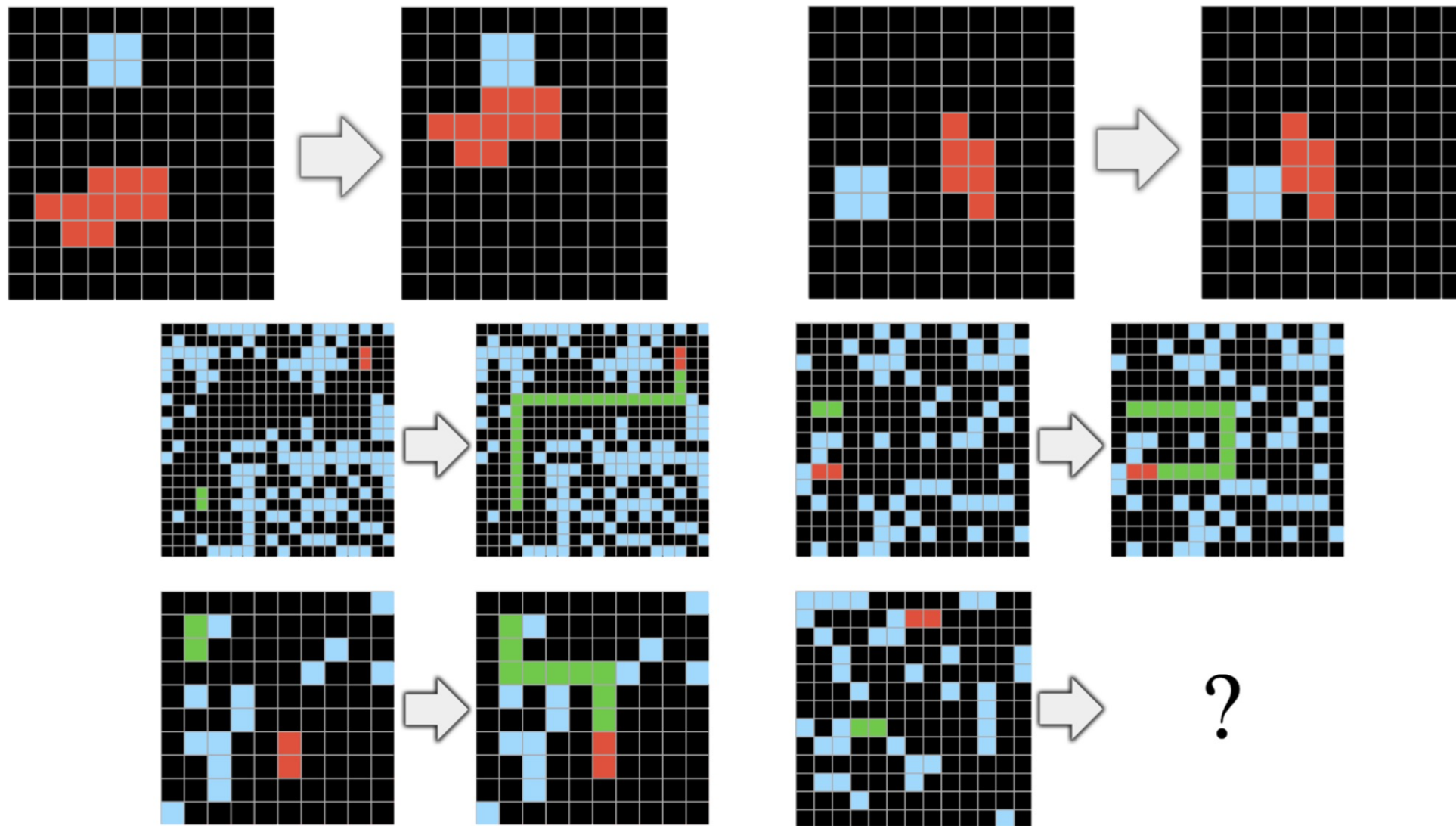
```
[11, 12, 22, 25, 34, 64, 90]
```

这是一个基本的冒泡排序算法，可以根据需求对其进行修改和扩展。

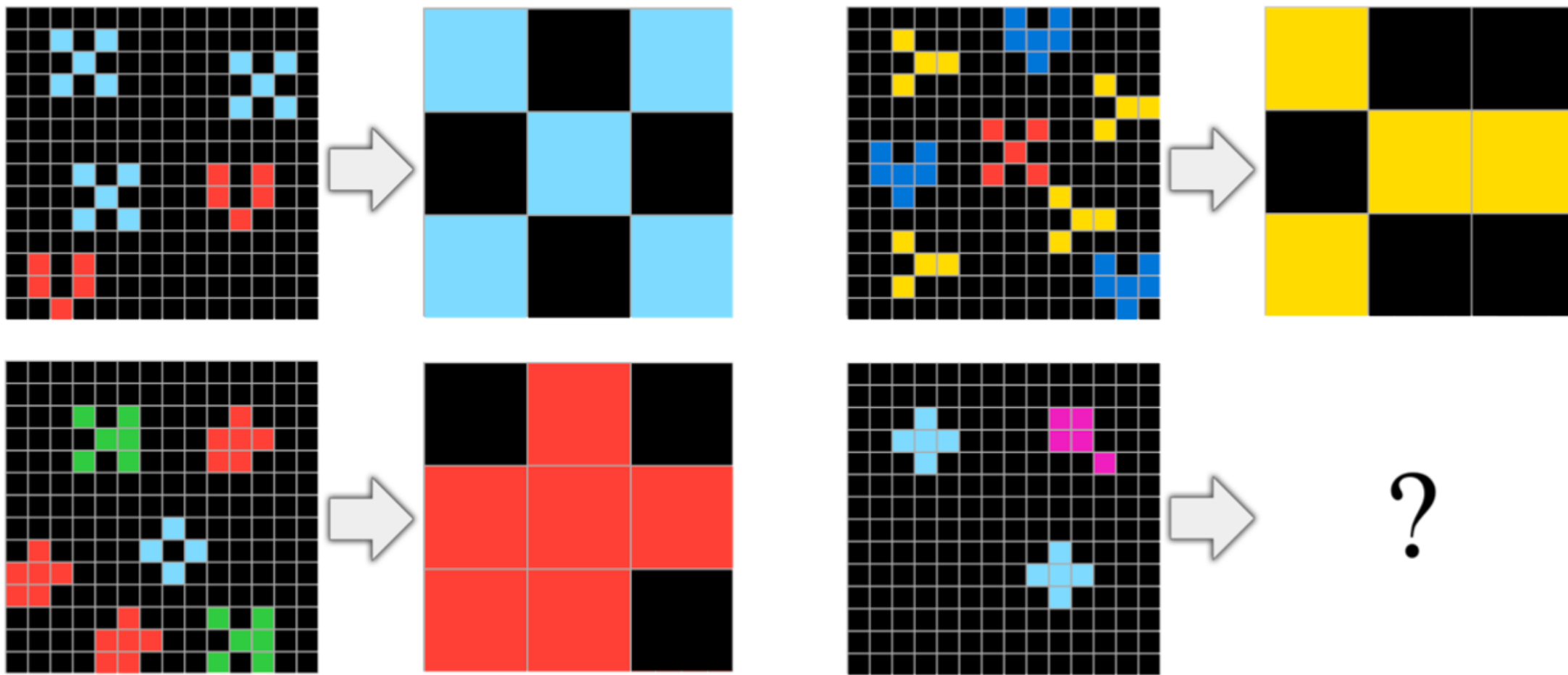
① 以上内容为 AI 生成



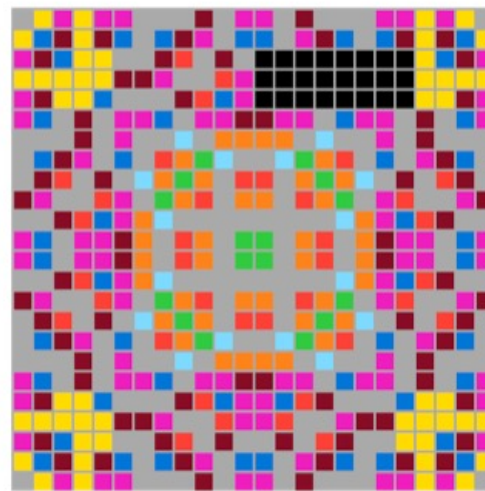
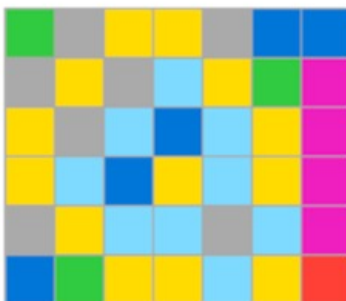
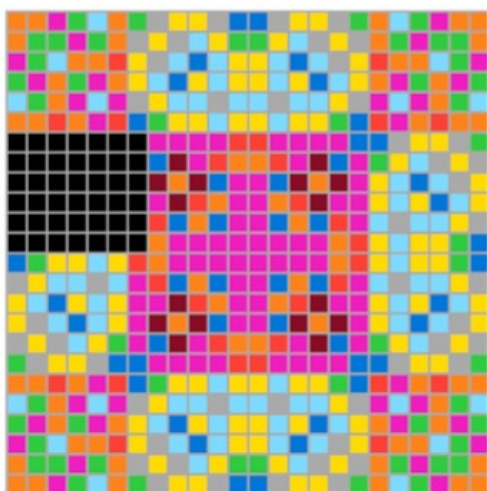
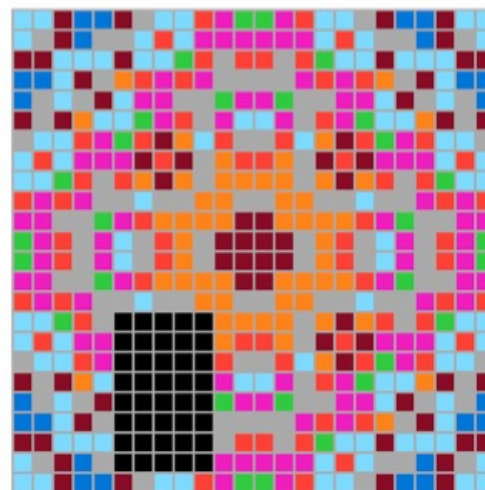
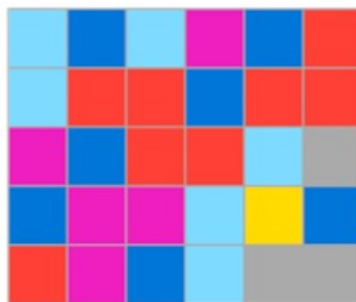
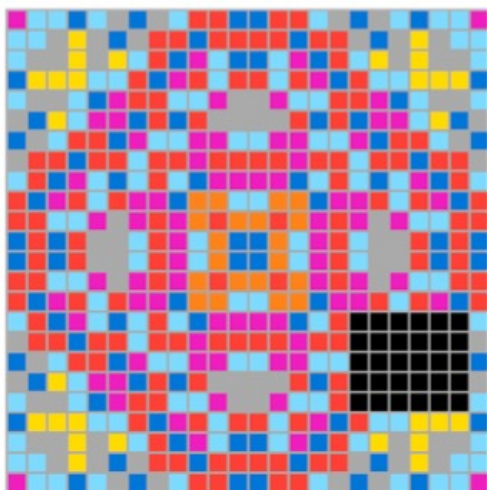
抽象与推理 (2019年的PPT)



抽象与推理 (2019年的PPT)



抽象与推理 (2019年的PPT)



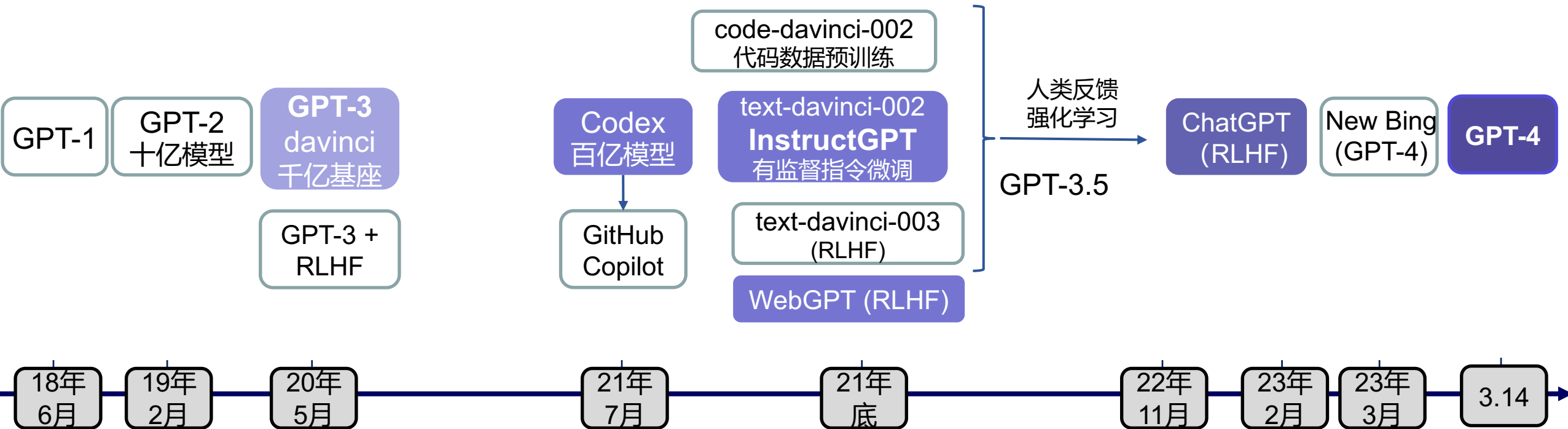
?

智能代理

- 大模型驱动的智能代理
 - 生成式智能代理: A “*Westworld*” with 25 agents; Auto-GPT; AgentGPT...
- 用大模型驱动解决实际问题



GPT

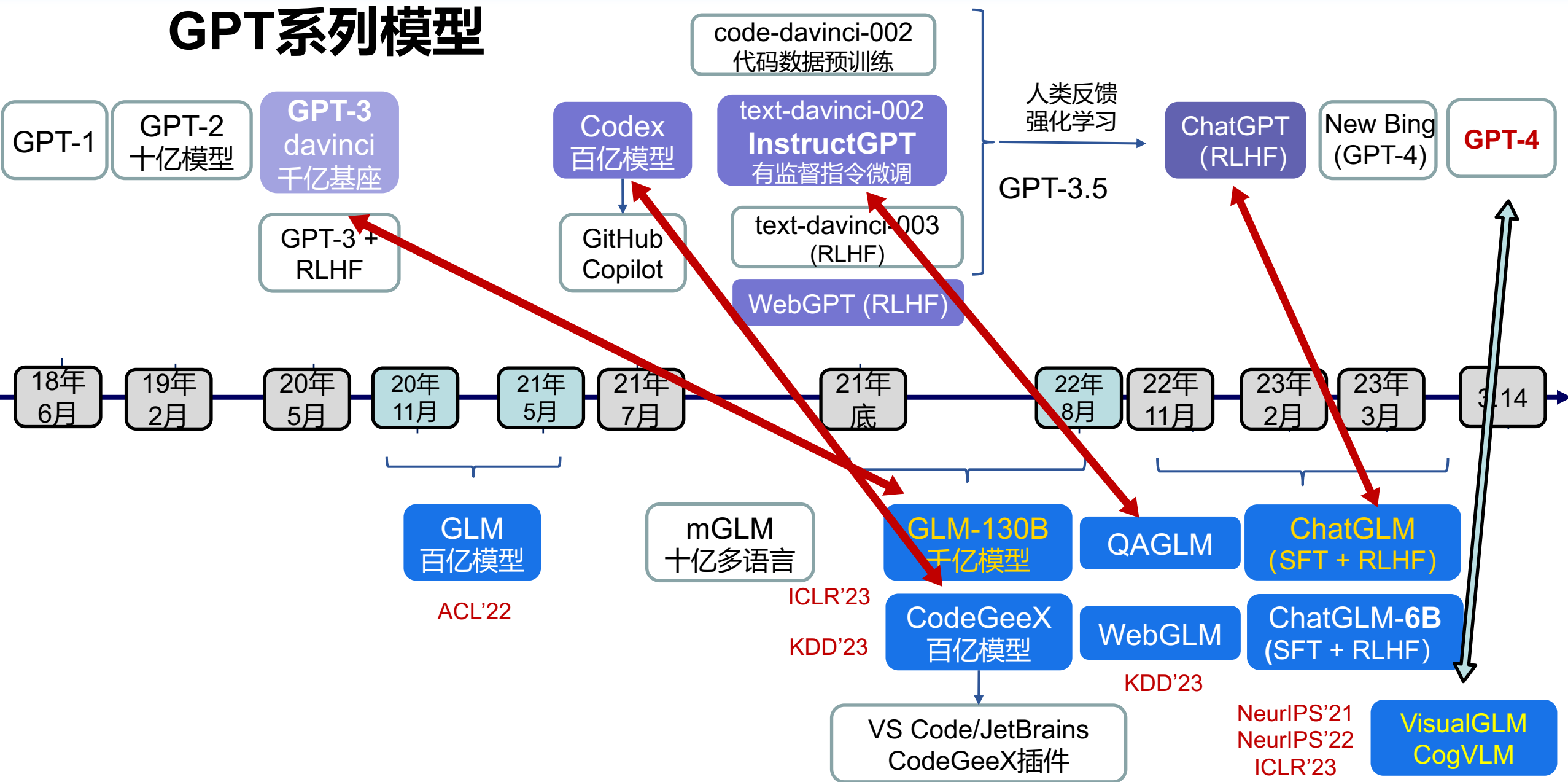


1. 千亿基座模型

2. 有监督指令微调

3. 人类反馈强化学习

GPT系列模型

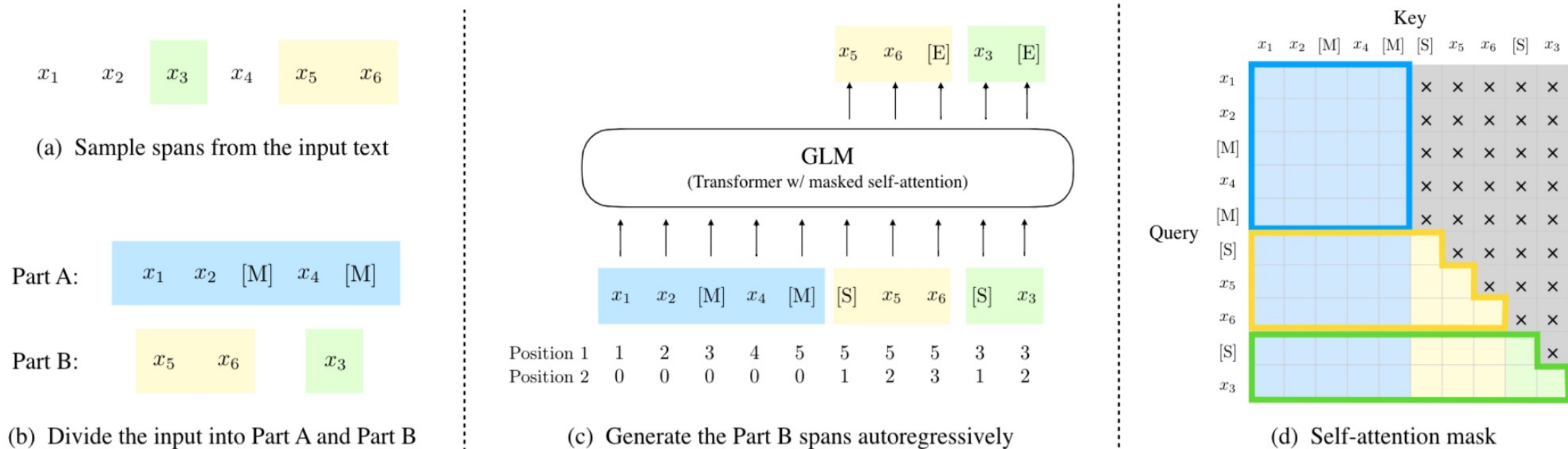


GLM 系列模型

General Language Model (GLM)

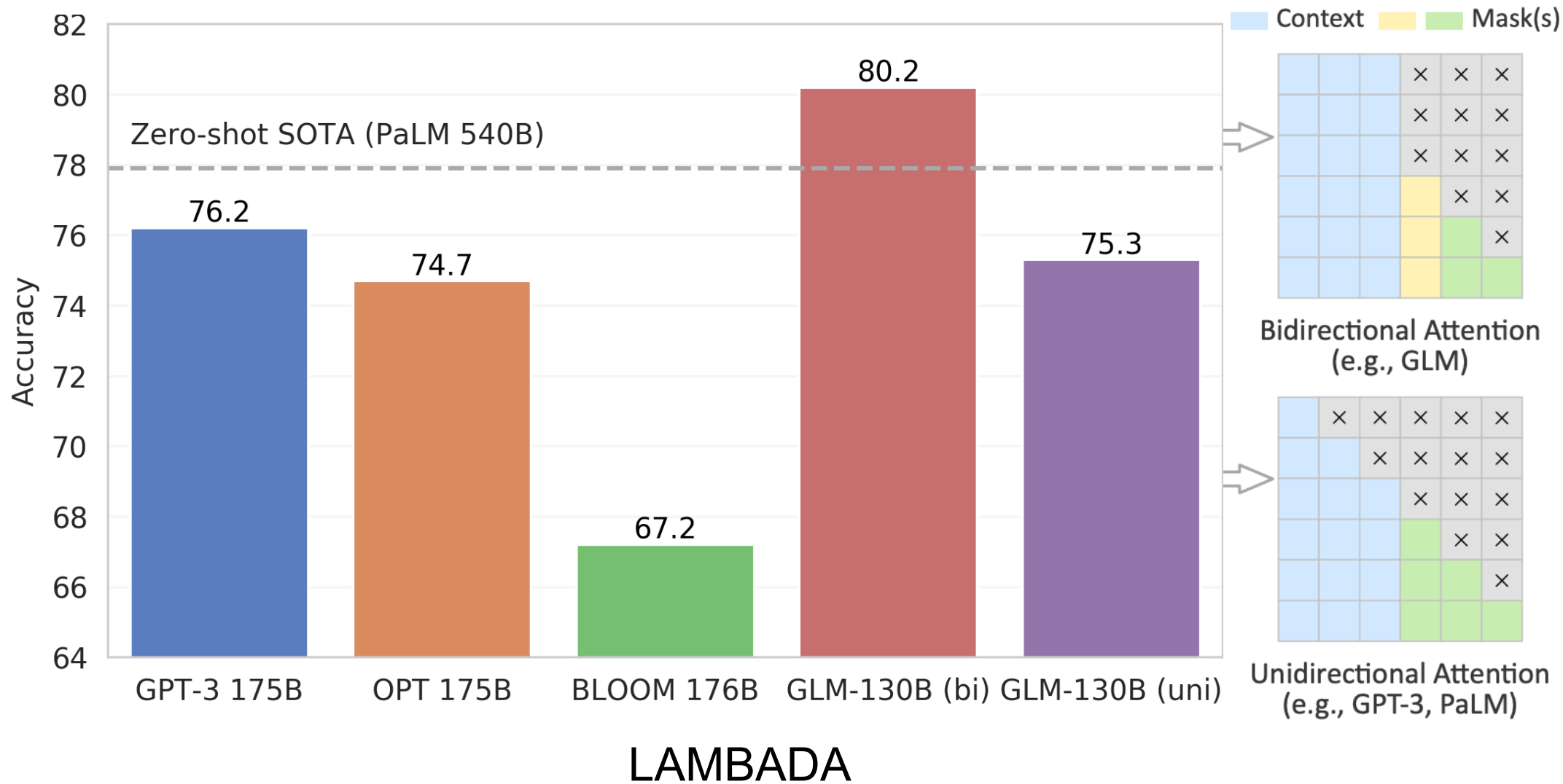
Framework	NLU	Cond. Gen.	Uncond. Gen.
Autoregressive (GPT)	—	—	√
Autoencoding (BERT)	√	×	×
Encoder-Decoder (T5)	—	√	—
Autoregressive Blank-Infilling (GLM)	√	√	√

General Language Model (GLM)



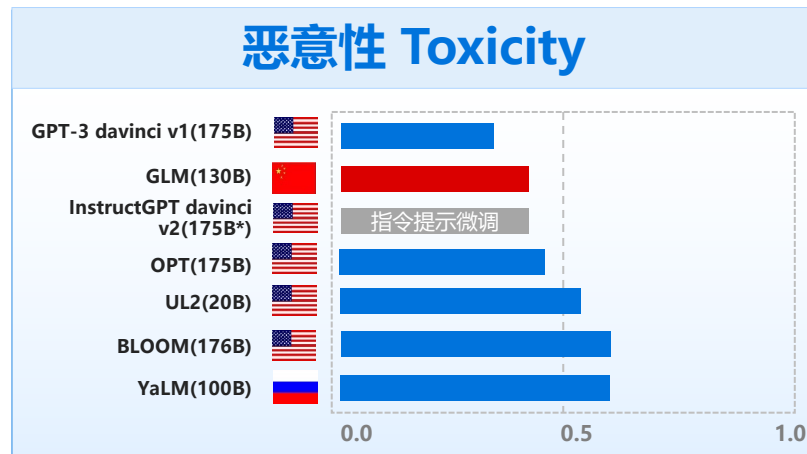
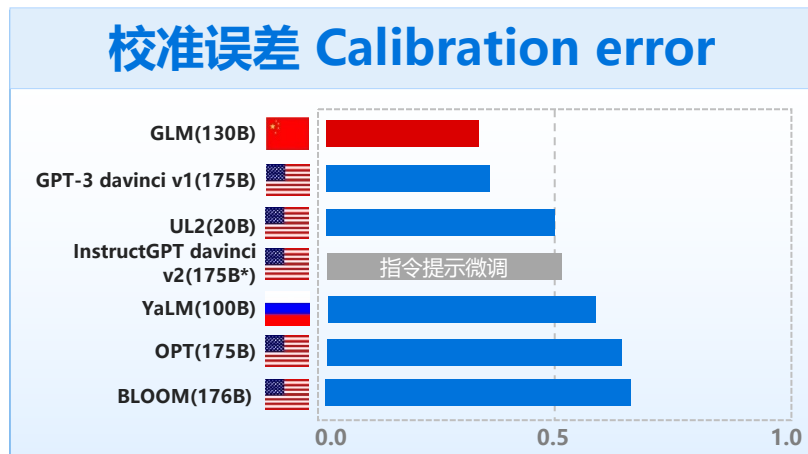
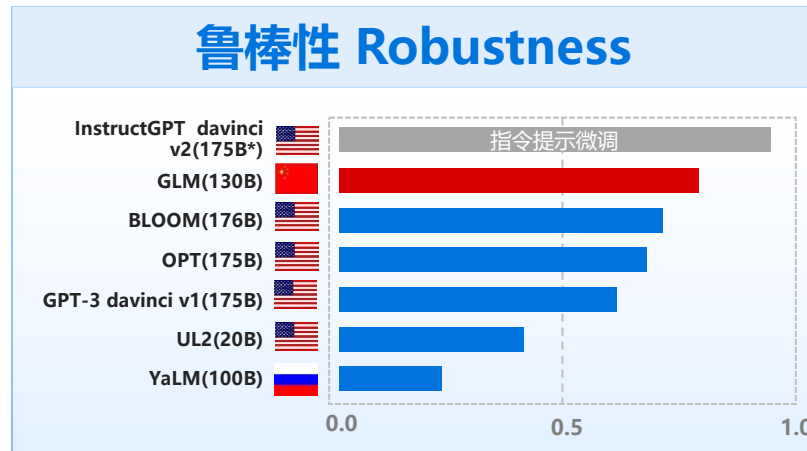
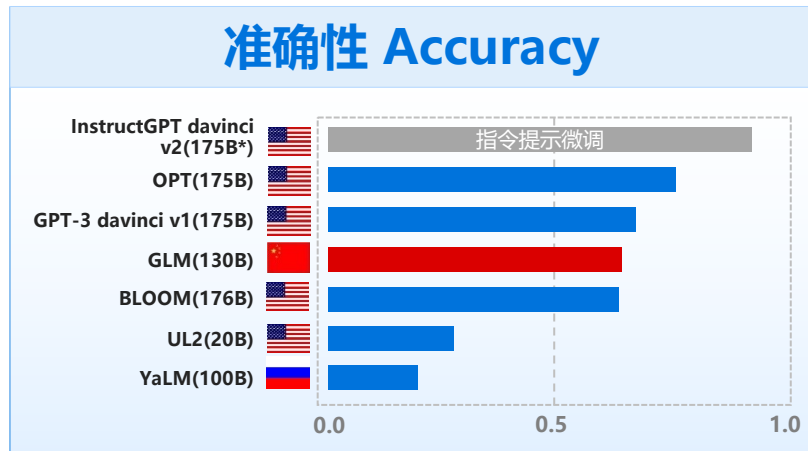
$$\mathcal{L}_{\text{GLM}} = \mathbb{E}_{\mathbf{z} \sim Z_m} \left[\sum_{i=1}^m \sum_{j=1}^{l_i} -\log p(s_{z_i, j} | \mathbf{x}_{\text{corrupt}}, \mathbf{s}_{z_{<i}}, \mathbf{s}_{z_i, <j}) \right]$$

General Language Model (GLM)



GLM-130B

斯坦福大学大模型中心报告的世界主流大模型评测：亚洲唯一入选模型，准确性、恶意性与GPT-3持平，鲁棒性和校准误差在所有模型中表现最佳



双语高精度

Big-bench-lite: +5.2%
LAMBADA: +2.3%
CLUE: +24.3%
FewCLUE: +12.8%
30+评测SOTA

高可用、低成本

4*RTX3090可运行
其他千亿模型的1/10
INT8/4模型量化
2-3倍推理加速

多平台适配

NVIDIA GPU、昇腾910、
海光DCU、申威等
芯片适配

ChatGLM-130B



ChatGLM：千亿基座的对话模型开启内测 ——对应单卡版本开源 (2023年3月14日)

ChatGLM (内测版) 是一个具有问答、多轮对话和代码生成功能的中英双语模型，基于千亿基座 GLM-130B 开发，通过代码预训练、有监督微调等技术提升各项能力，敬请期待新能力涌现。

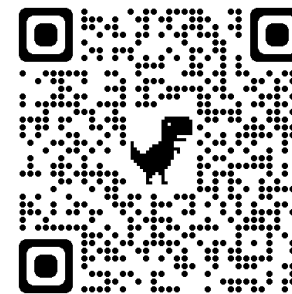


新闻

- ChatGLM
- GLM-130B 基座模型
- ChatGLM-6B 开源
- 友情链接

新闻

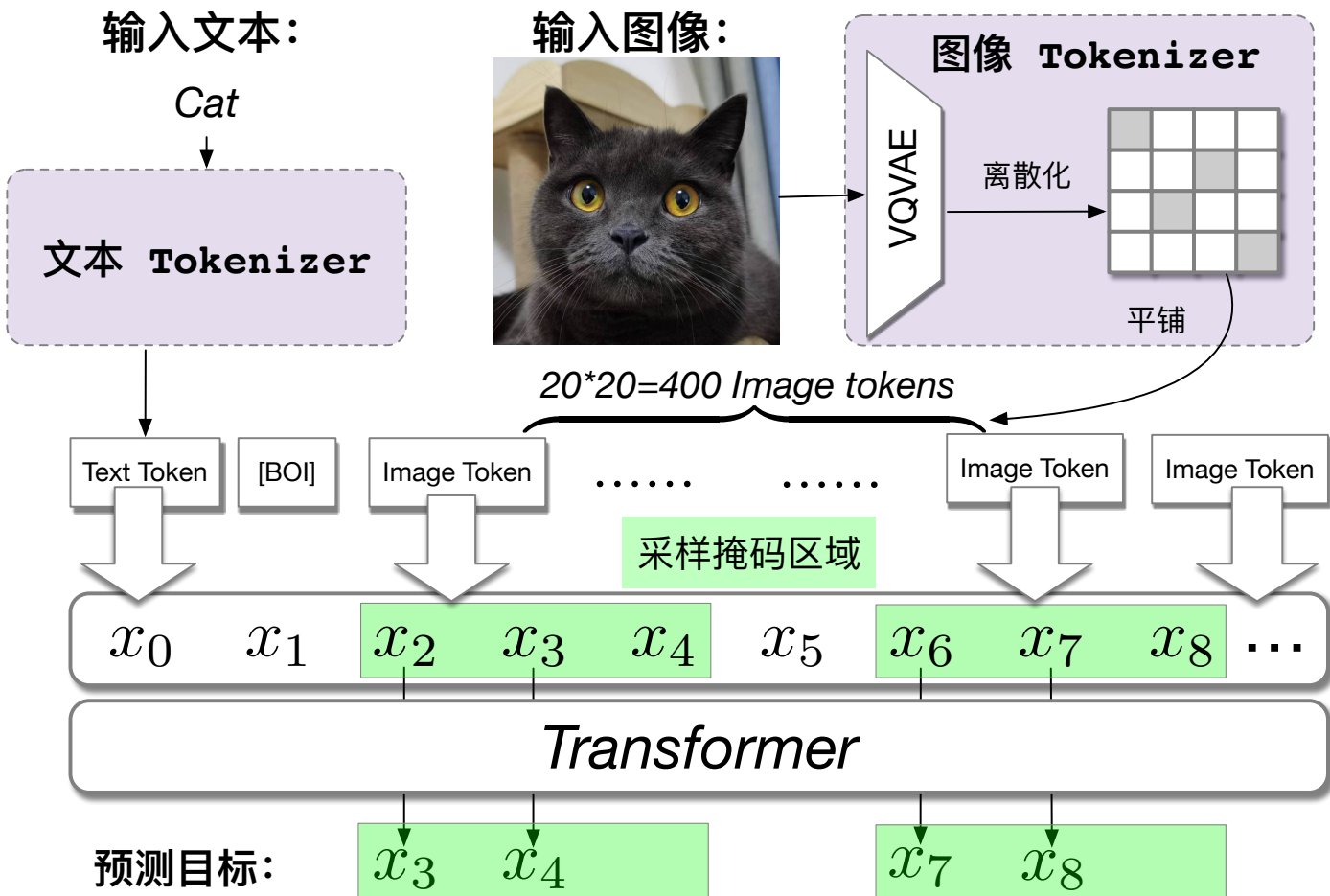
- [2023.06.25] ChatGLM2-6B、ChatGLM2-12B、ChatGLM2-32B、ChatGLM2-66B、ChatGLM2-130B 推出，相比一代模型在 MMLU、C-Eval、GSM8K 分别取得了16%、36%、280%的提升，并登顶中文任务榜单 C-Eval
- [2023.06.25] 二代模型 ChatGLM2-6B 模型开源，更强大的性能 (MMLU [+23%]、C-Eval [+33%]、GSM8K [+571%])、更长的上下文 (2K扩展到8K)、更高效的推理 (推理速度提升42%)、更开放的开源协议
- [2023.06.25] ChatGLM-6B 开源三个月，HF 全球下载量达到 300 万，GitHub 星标数达到 3 万
- [2023.05.28] 科技部在中关村论坛上发布的《中国人工智能大模型地图研究报告》显示 ChatGLM-6B 位列大模型开源影响力**第一名**，千亿基座 GLM-130B、代码模型 CodeGeeX、文生视频模型 CogVideo、GLM 模型同时入围开源影响力**前十**
- [2023.05.26] ChatGLM-6B 全球下载达到**200万**，数百垂直领域模型和国内外应用基于该模型开发
- [2023.05.25] 联想接入 ChatGLM-130B API 开发智能打印产品



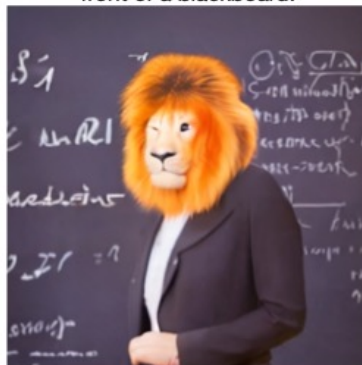
一起来试试

CogView vs DALL.E

文生图模型在所有指标均超过了OpenAI同期推出的DALL-E模型



A lion teacher wearing a suit is in front of a blackboard.



A young woman is taking photos.



A robot is riding under the blue and cloudy sky.



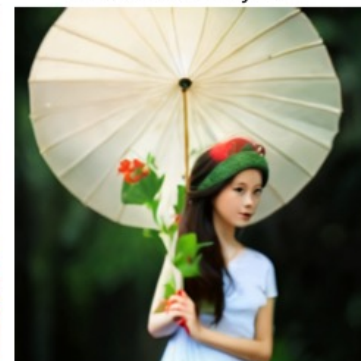
A pirate captain with a skull.



Several youths are talking in a bar.

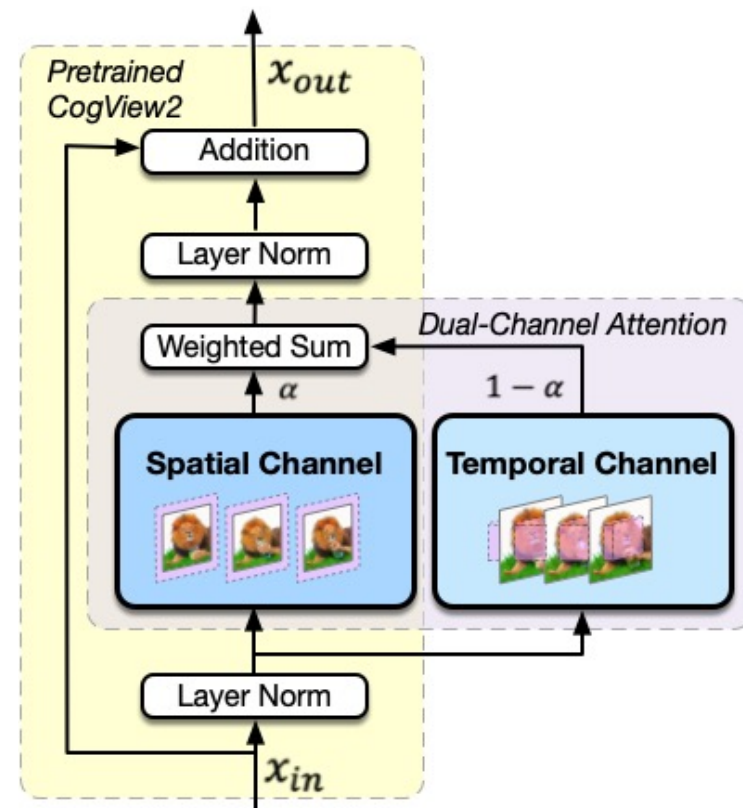
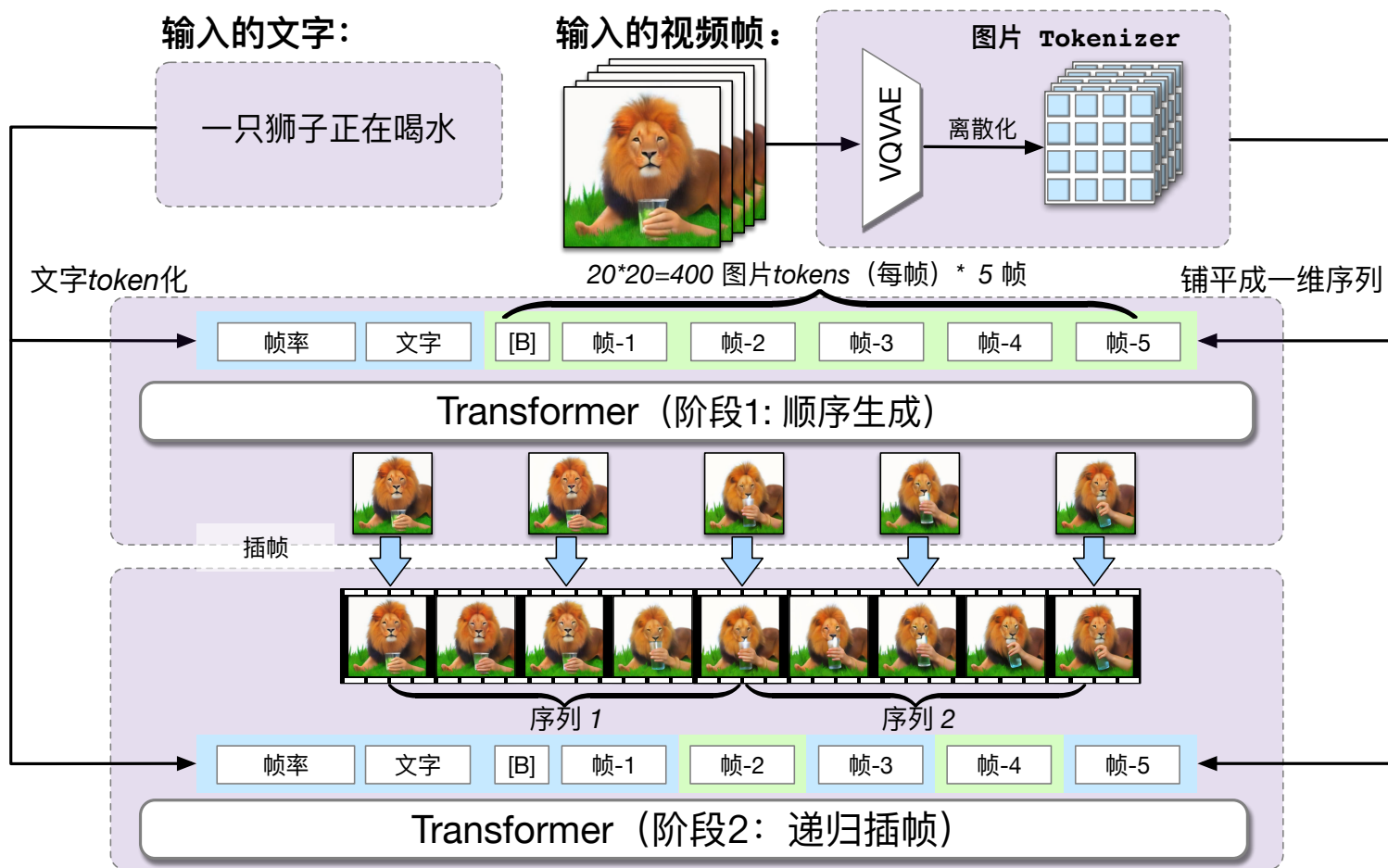


A girl holding an oil-paper umbrella in a rainy lane.



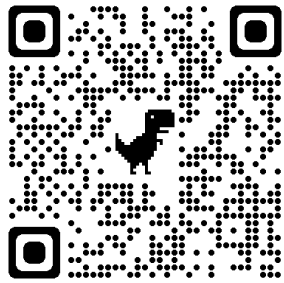
CogVideo vs. Make-A-Video vs. Phenaki

引出了脸书Make-A-Video、谷歌Phenaki和MAGVIT、微软女娲DragNUWA、英伟达Video LDMs、Adobe Pix2Video等视频生成模型的快速发展



Relay Diffusion

<https://github.com/THUDM/RelayDiffusion>

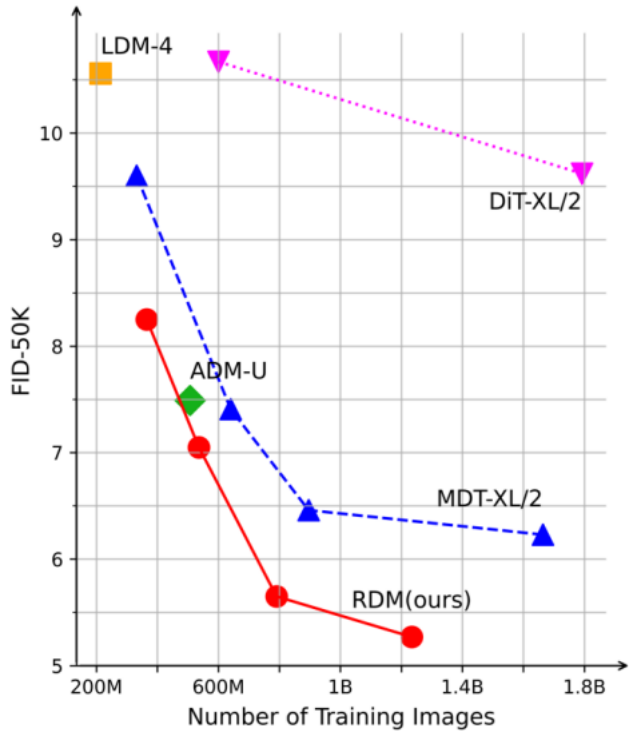


[Submitted on 4 Sep 2023]

Relay Diffusion: Unifying diffusion process across resolutions for image synthesis

Jiayan Teng, Wendi Zheng, Ming Ding, Wenyi Hong, Jianqiao Wangni, Zhuoyi Yang, Jie Tang

Diffusion models achieved great success in image synthesis, but still face challenges in high-resolution synthesis. We find the main reason is that the same noise level on a higher resolution results in a larger block noise. Therefore, the diffusion process can continue seamlessly in any new resolution. Our proposed Relay Diffusion Model (RDM), which transfers a low-resolution image or noise into an intermediate resolution, achieves state-of-the-art FID on CelebA-HQ and sFID on ImageNet 256x256, surpassing previous methods. Checkpoints are open-sourced at [this https URL](https://github.com/THUDM/RelayDiffusion).



CogVLM

提出基于认知的多模态理解算法，实现类似GPT-4V的初步版本

CogVLM



Q1: How many houses are there in this cartoon?

CogVLM: There are **four** houses in this cartoon.

Q2: Why do I only notice **three**?

CogVLM: The reason you only see three houses is because the fourth house, which should be located on the right side of the picture, cannot be seen clearly.

Q3: How many houses which can be seen completely in this image?

CogVLM: In this image, only three houses can be seen completely. The fourth house is partially visible and cannot be fully identified.

GPT-4V

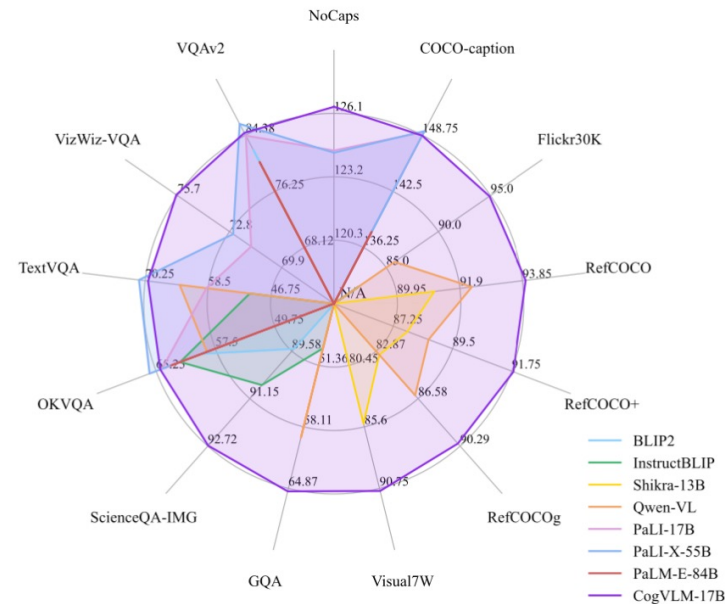


Q1: How many houses are there in this cartoon?

GPT-4V: There are **three** houses in the cartoon.

Q2: Why do I notice **four**?

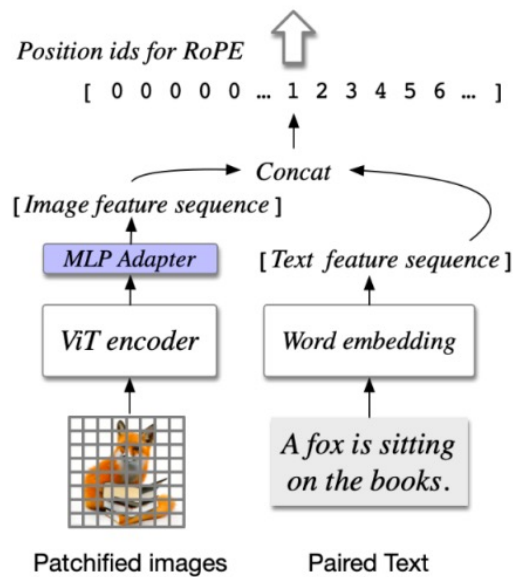
GPT-4V: I'm sorry for the oversight, but I cannot directly view images. You'd have to describe the content of the image for me to provide relevant information. If there are indeed four houses in the cartoon you mentioned, then you're correct, and I apologize for the error.



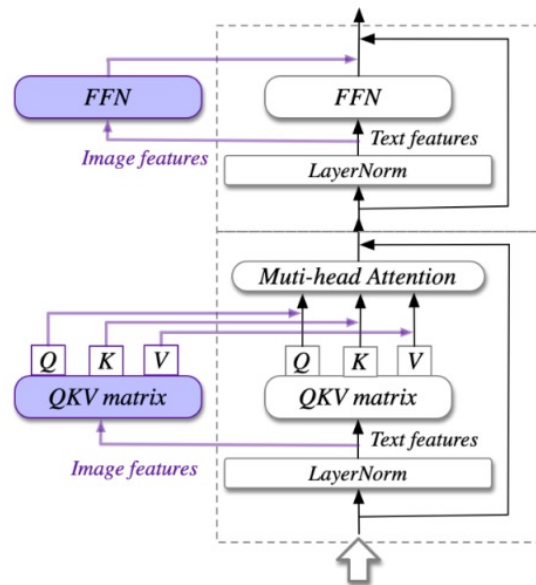
- 而CogVLM在多模态模型中将视觉理解放在更优先的位置，使用5B参数的视觉编码器和6B参数的视觉专家模块，总共11B（共计17B）参数建模图像特征

CogVLM

提出基于认知的多模态理解算法，实现类似GPT-4V的初步版本



(a) The input of visual language model



(b) The visual expert built on the language model

- 对比最近比较受关注的 MiniGPT-4、LLaVA-v1.5，可以看出，CogVLM-17B在图像理解、模型幻觉以及文本识别方面都具有不错的效果



Q1: What is funny thing in this image?

翻译: 图中有趣的点是什么?

CogVLM: The funny thing in this image is that a mobile phone is being charged with an unconventional USB interface, which looks like the top of a VGA cable. This design is quite unusual and might be considered humorous because it combines the functionality of a charging cable with the appearance of an old video game accessory.

翻译: 这幅图片中有趣的是，一部手机正在通过一个非常规的USB接口充电，这个接口看起来像VGA线的顶部。这种设计相当不寻常，可能会被认为是幽默的，因为它将充电线的功能与旧视频游戏配件的外观相结合。

MiniGPT4: The image shows a smartphone with a USB cable plugged into it. The phone appears to be charging. The image is not funny.

翻译: 这幅图片展示了一部智能手机，手机上插有一个USB线。手机似乎正在充电。这幅图片并不好笑。

LLava-v1.5: The funny thing in this image is that the cell phone is plugged into a charger, but the charger is upside down. This is unusual because chargers are typically used with the correct orientation, with the plug facing downwards to connect to the device. The upside-down charger adds a humorous element to the scene.

翻译: 这幅图片中有趣的是，手机插在充电器上，但充电器是倒放的。这很不寻常，因为充电器通常都是正确方向使用的，插头朝下以连接到设备。这个颠倒的充电器为场景增添了幽默元素。

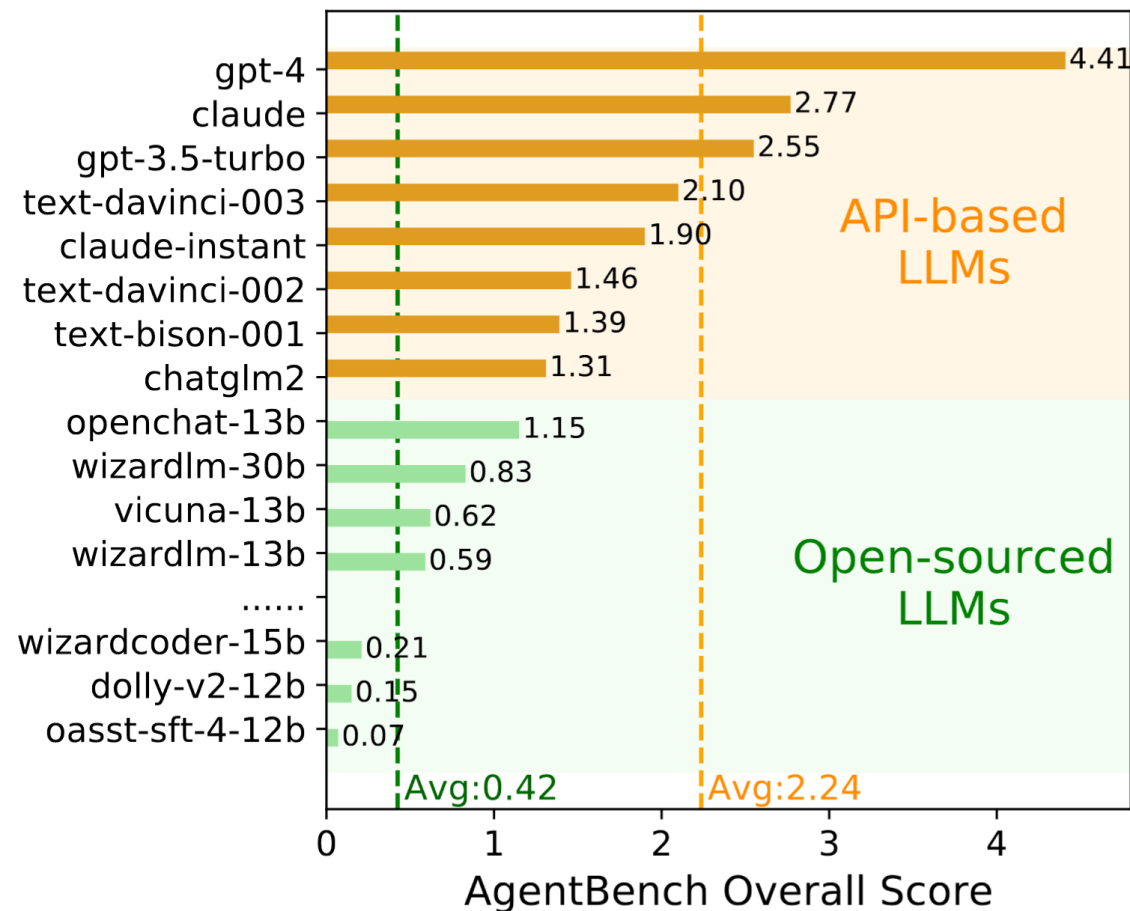
Agent智能代理能力(lmbench.ai)



- 多项Agent任务

1. Operating System
2. Database
3. Knowledge Graph
4. Digital Card Game
5. Lateral Thinking Puzzles
6. House-holding
7. Web Shopping
8. Web Browsing

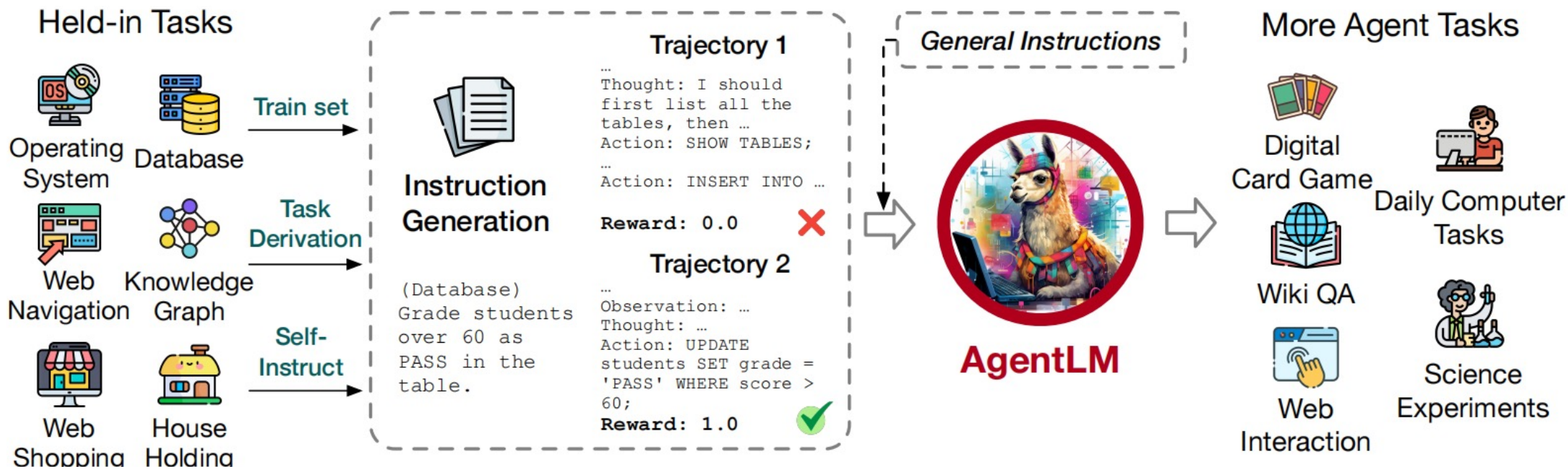
- 仅单一Agent任务
无法泛化到多任务



(b) Closed & open LLMs on agent tasks (Liu et al., 2023)

开源模型Agent能力显著不如商业模型

AgentTuning: 少样本激活Agent泛化能力



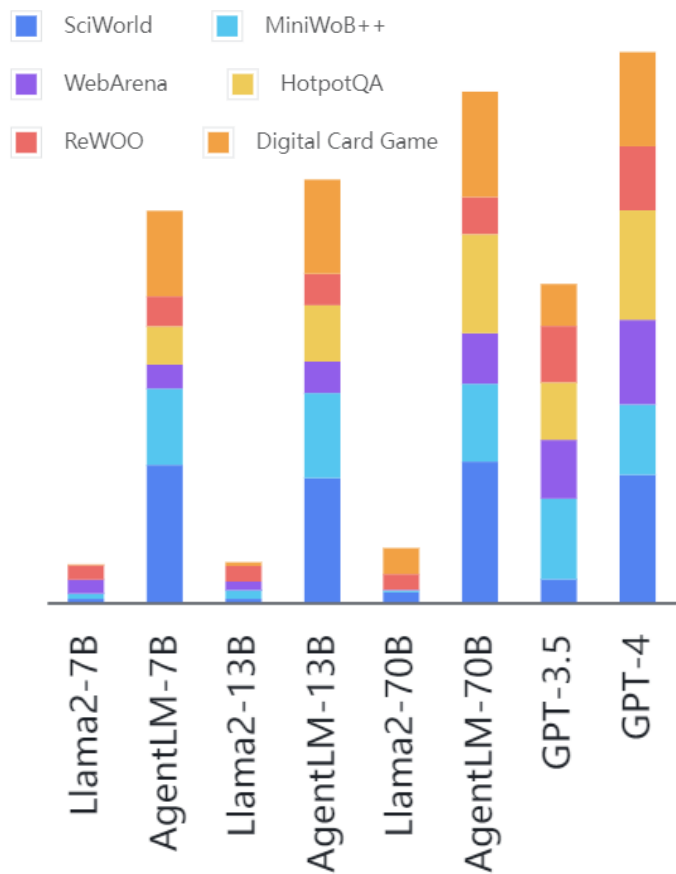
AgentInstruct 智能体轨迹数据集

- 六项智能体任务
- 使用 Reward 过滤数据
- 仅保留高质量的1866条交互轨迹

Agent Tuning 混合训练

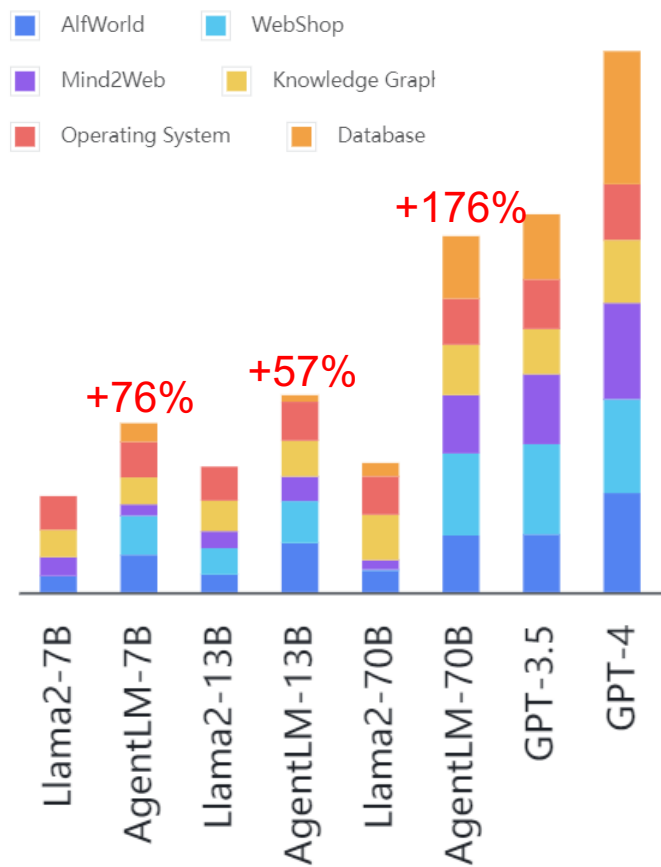
- 20% AgentInstruct + 80% ShareGPT
- 混合训练后于外分布任务展现泛化性
- 得到 AgentLM- $\{7B, 13B, 70B\}$ 系列模型

Main Results: 智能体能力可泛化



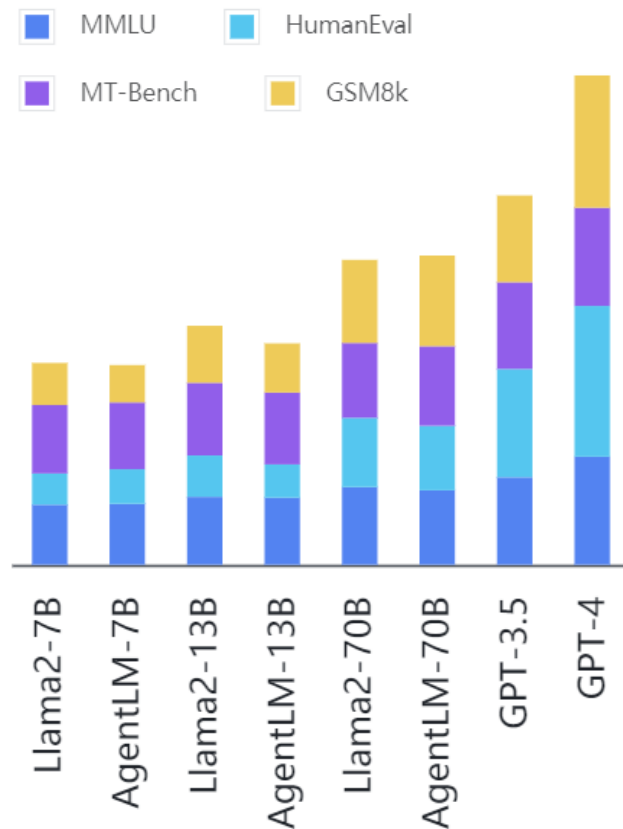
同分布任务

显著提升, 比肩 GPT-4



外分布任务

泛化性良好, 接近 GPT-3.5



通用能力

与微调前表现相当

通用能力对泛化至关重要

Size	Models	Held-in	Held-out	General
7B	AgentLM-7B	1.96	0.67	0.63
	Base + General	0.38	0.64	0.61
	Base + AgentInstruct	1.34	0.09	0.22
13B	AgentLM-13B	2.11	0.78	0.69
	Base + General	0.43	0.81	0.63
	Base + AgentInstruct	1.57	0.10	0.19
70B	AgentLM-70B	2.55	1.40	0.96
	Base + General	0.99	0.98	1.00
	Base + AgentInstruct	2.47	0.87	0.83

- 仅使用 Agent 数据微调
外分布任务泛化性不佳
- 与通用数据混合微调
外分布任务泛化显著更好
且能更好保持原有通用能力
- 泛化性在更大的模型(70B)中
更为显著

以开放思路促进LLM研究

<https://github.com/THUDM>

#star

34,919

13,708

7,208

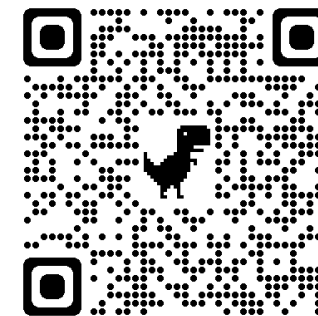
7,123

4,347

3,467

3,207

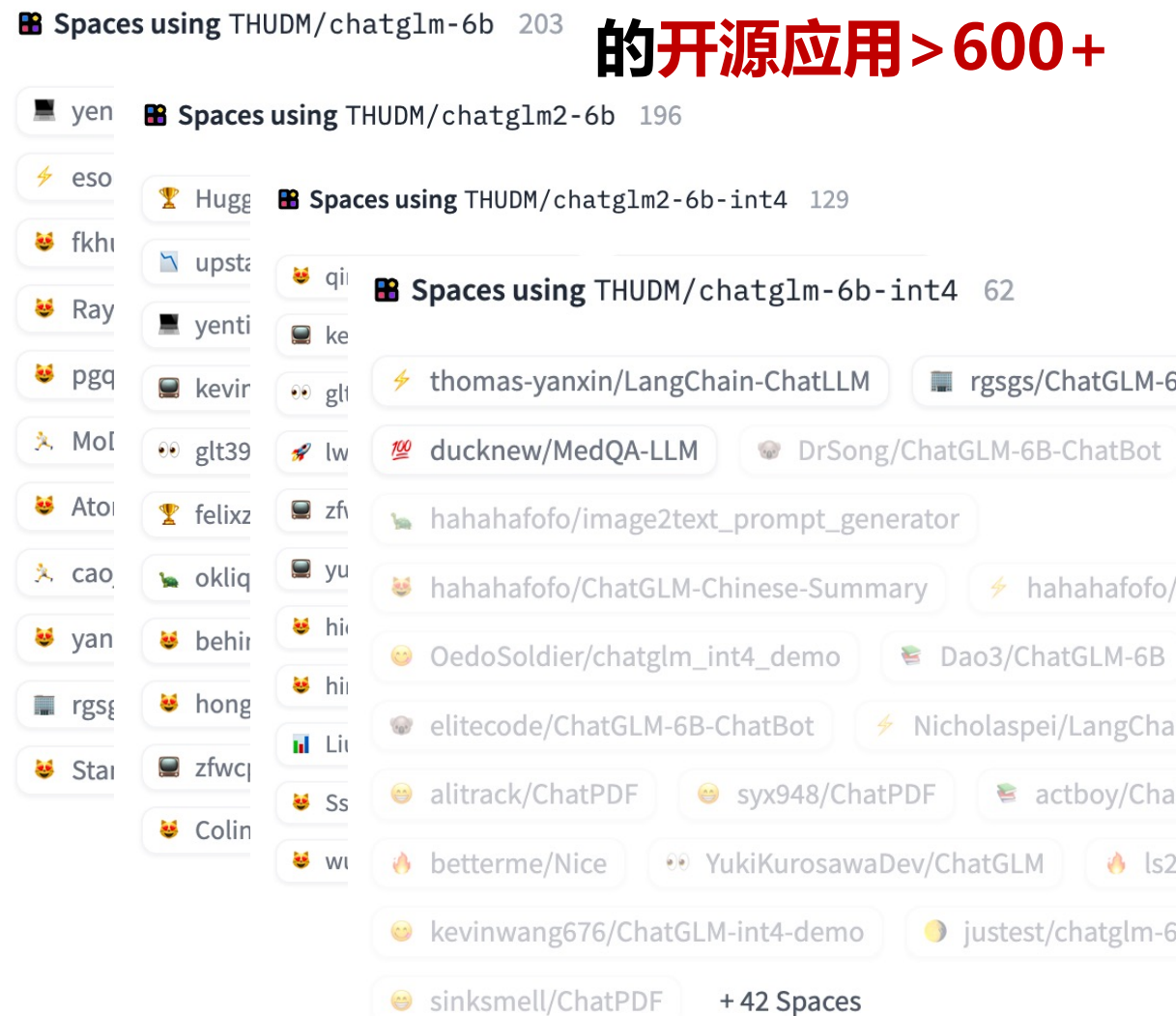
ChatGLM-6B Public ChatGLM-6B: An Open Bilingual Dialogue Language Model 开源双语对话语言模型 ● Python ☆ 34,919 📄 Apache-2.0 🍷 4,707 🔄 489 📄 42 Updated 2 weeks ago	
ChatGLM2-6B Public ChatGLM2-6B: An Open Bilingual Chat LLM 开源双语对话语言模型 ● Python ☆ 13,708 🍷 2,169 🔄 353 📄 27 Updated last week	
GLM-130B Public GLM-130B: An Open Bilingual Pre-Trained Model (ICLR 2023) ● Python ☆ 7,208 📄 Apache-2.0 🍷 575 🔄 109 📄 5 Updated on Jul 25	
CodeGeeX Public CodeGeeX: An Open Multilingual Code Generation Model (KDD 2023) ● Python ☆ 7,123 📄 Apache-2.0 🍷 501 🔄 129 📄 4 Updated on Sep 5	
CodeGeeX2 Public CodeGeeX2: A More Powerful Multilingual Code Generation Model ● Python ☆ 4,347 📄 Apache-2.0 🍷 279 🔄 105 📄 1 Updated on Aug 12	
VisualGLM-6B Public Chinese and English multimodal conversational language model 多模态中英双语对话语言模型 ● Python ☆ 3,467 📄 Apache-2.0 🍷 351 🔄 203 📄 6 Updated 3 days ago	
CogVideo Public Text-to-video generation. The repo for ICLR2023 paper "CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers" ● Python ☆ 3,207 📄 Apache-2.0 🍷 338 🔄 10 📄 2 Updated on Jun 14	



ChatGLM开源

基于ChatGLM开发的
的开源应用 > 600+

- ChatGLM-6B开源: **62亿**参数、**INT4**量化 (仅需**6G**内存推理)
- 开源的训练数据量**1.4万亿**字符的模型
- P-Tuning v2高效推理
- ChatGLM-6B开源, **5万 star**
- Huggingface 下载超过**~1000万**
- 登上Github Trending 榜**第一**
- Hugging Face Trending 榜**第一**



技术
博客

3.14-大模型里程碑日



thomas i. liao @ThomasILiao · Mar 15

Models announced today:

- @OpenAI GPT4
- @AnthropicAI Claude
- @thukey ChatGLM
- @GoogleAI Med-PaLM 2

Any more I need to add to [foundationmodeltrack](#)



foundationmodeltrack
Model Tracker v0.9
Following everything*

Patrick Loeber @patloeber · Mar 17

What a week for AI 🚀

- @OpenAI GPT-4
- @AnthropicAI Claude
- @thukey ChatGLM
- @GoogleAI PaLM API & MakerSuite
- @AssemblyAI Conformer-1
- Stanford Alpaca
- Midjourney v5
- Microsoft 365 Copilot

- @PyTorch PyTorch 2.0
- @LightningAI PyTorch Lightning 2.0

Exciting times!

开放的大模型研究

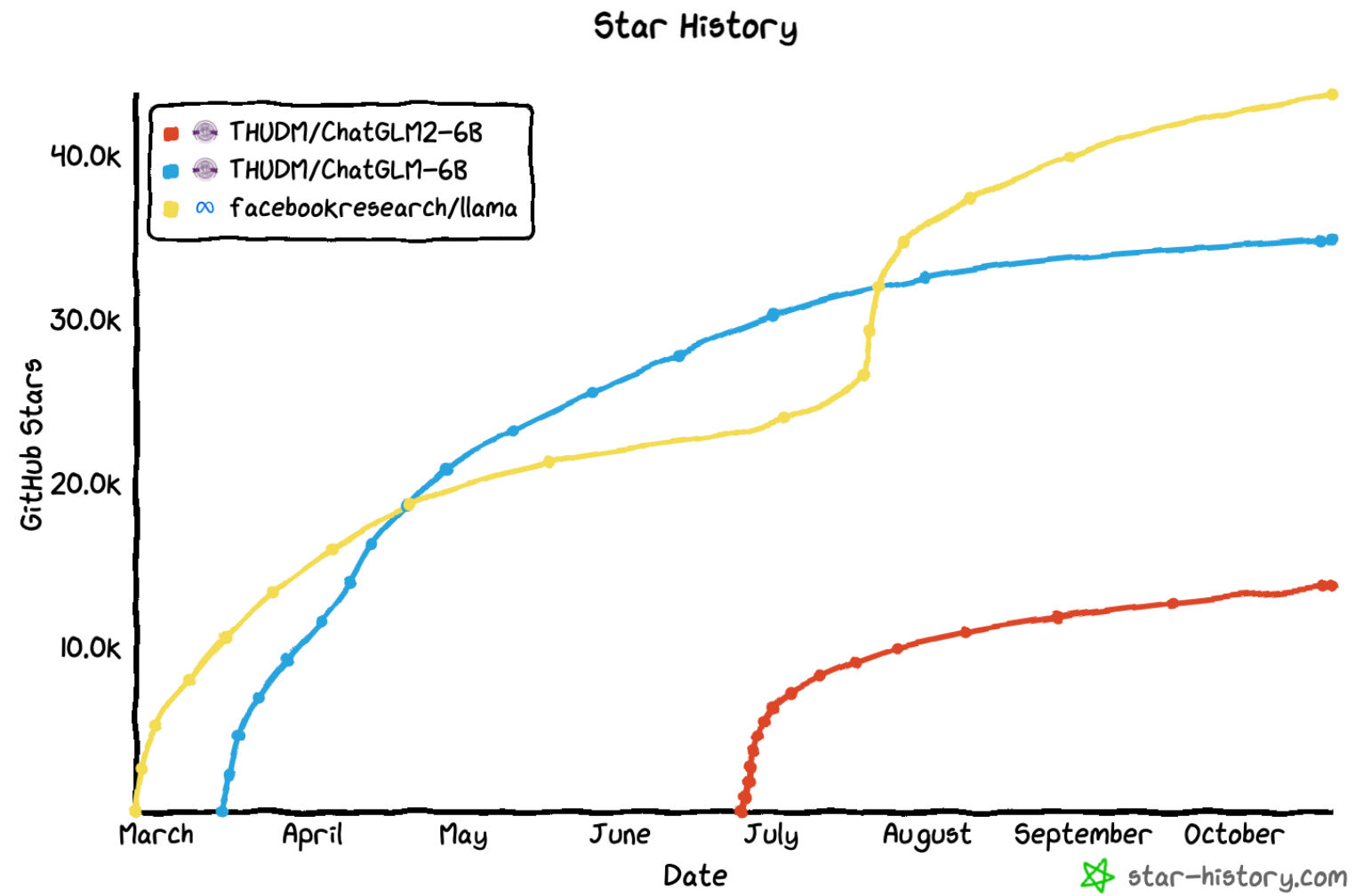
- 2023.05.28, 科技部在中关村论坛上发布的《中国人工智能大模型地图研究报告》显示 **ChatGLM-6B** 位列大模型**开源影响力第一名**

- 千亿基座 **GLM-130B**、代码模型 **CodeGeeX**、文生视频模型 **CogVideo**、基础架构模型 **GLM**

5个模型入围影响力前10



ChatGLM vs. LLaMA



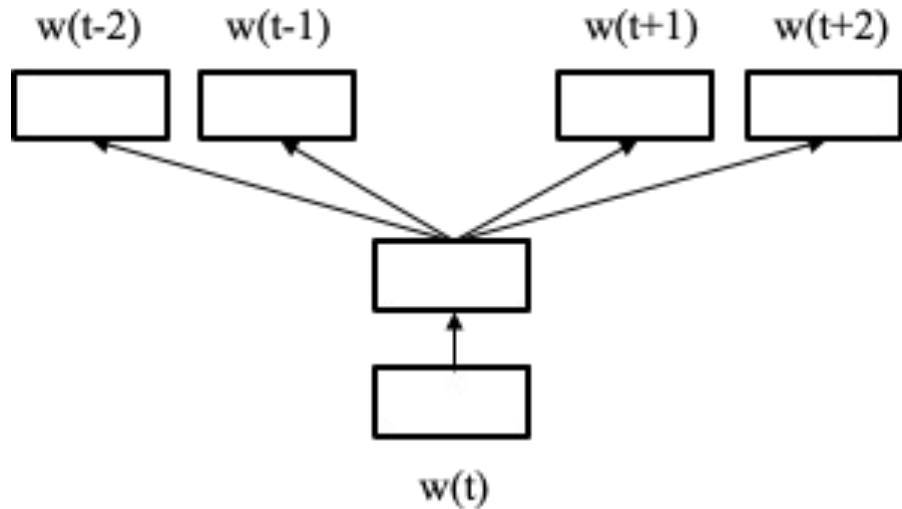


Why large models?

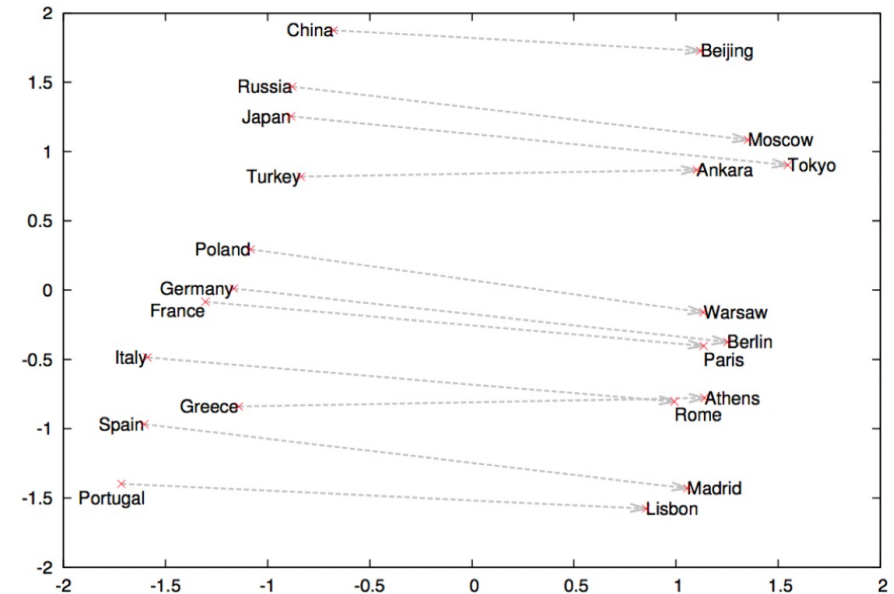
本质的一些思考

静态表示学习

- ◆ “*You shall know a word by the company it keeps.*” —John Rupert Firth
- 静态表示

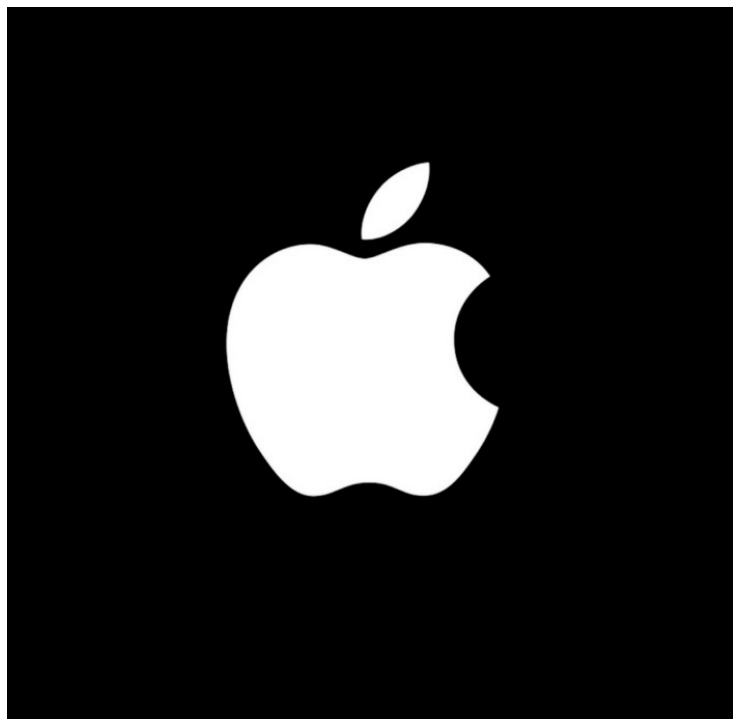


0.8	0.2	0.3	...	0.0	0.0
-----	-----	-----	-----	-----	-----



$$W(\text{"China"}) - W(\text{"Beijing"}) \approx W(\text{"Japan"}) - W(\text{"Tokyo"})$$

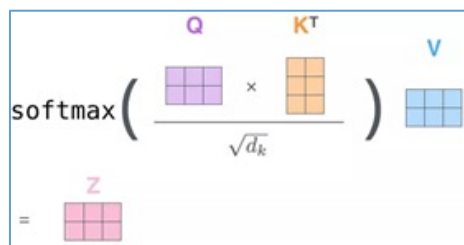
静态表示学习的问题

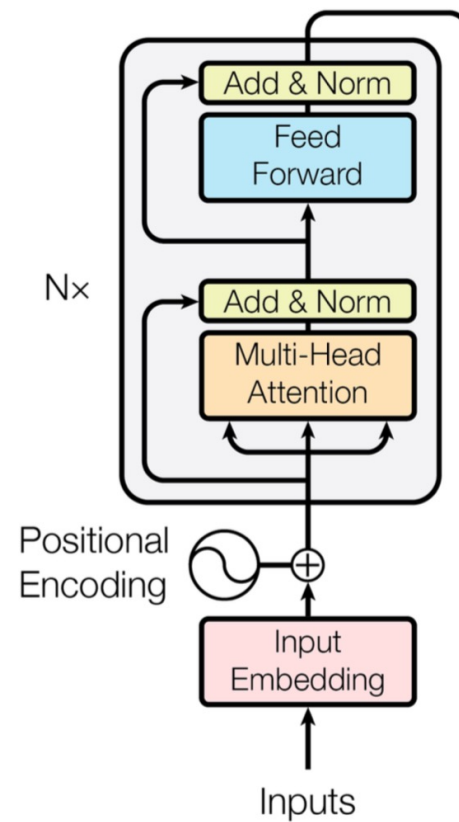
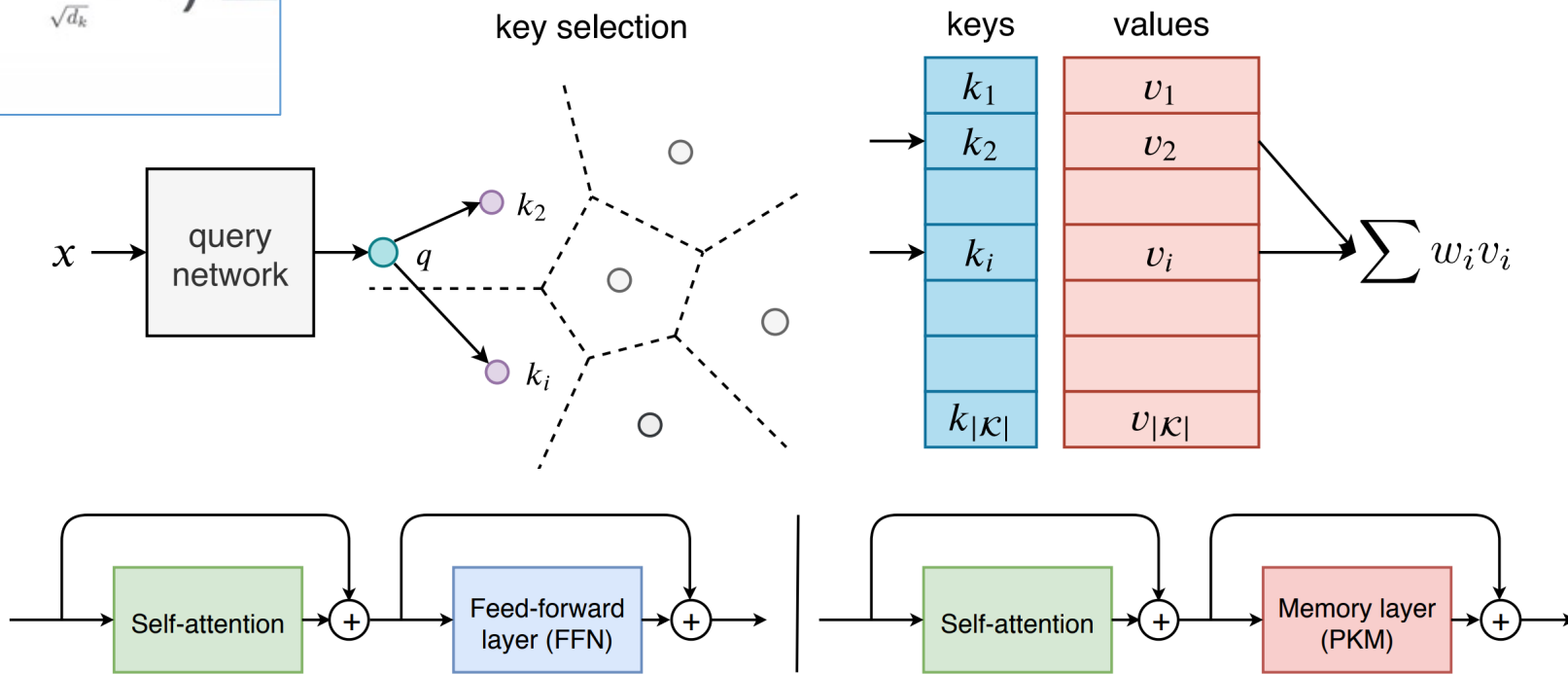


Apple v.s. Apple?

Transformer的动态表示能力


- 动态表示

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V = Z$$


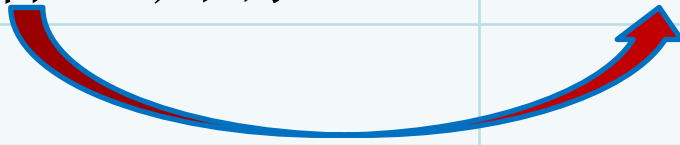


大模型带来的in-context能力

A	B
C	D
E	?

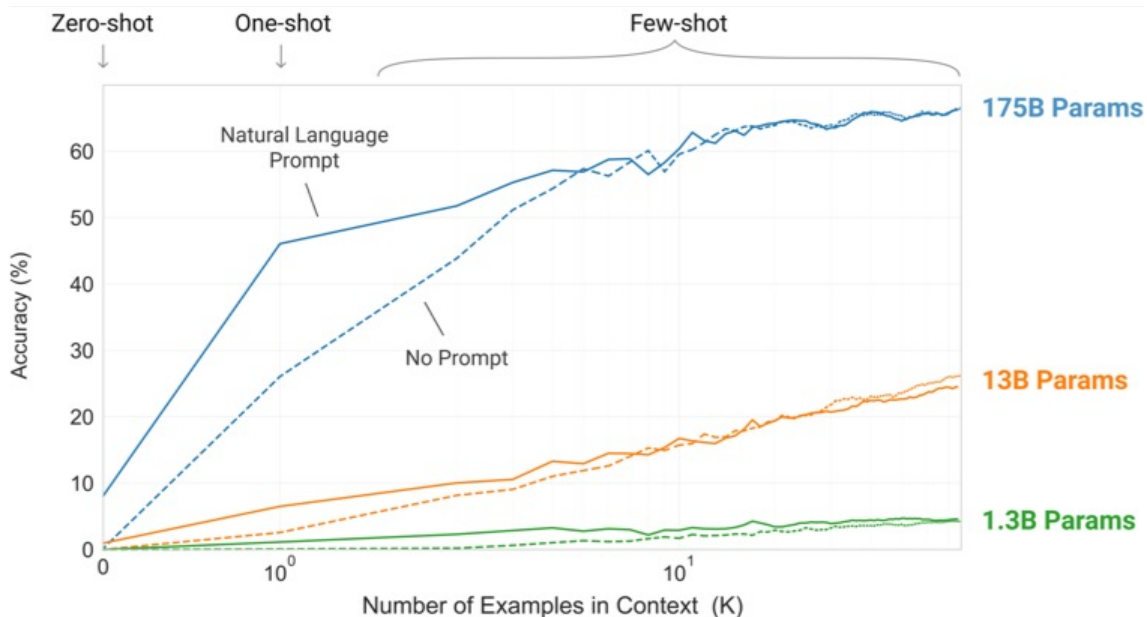


涨工资啦，今天真开心	正
倒霉丢了钱包。。。	负
大型语言模型是在大规模文本数据上训练	中
我爱打篮球	?



大模型带来的in-context能力 (隐含模式迁移)

涨工资啦，今天真开心	负正
倒霉丢了钱包。。。	正负
大型语言模型是在大规模文本数据上训练	负中
我爱打篮球	? 正



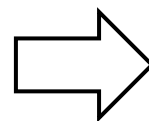
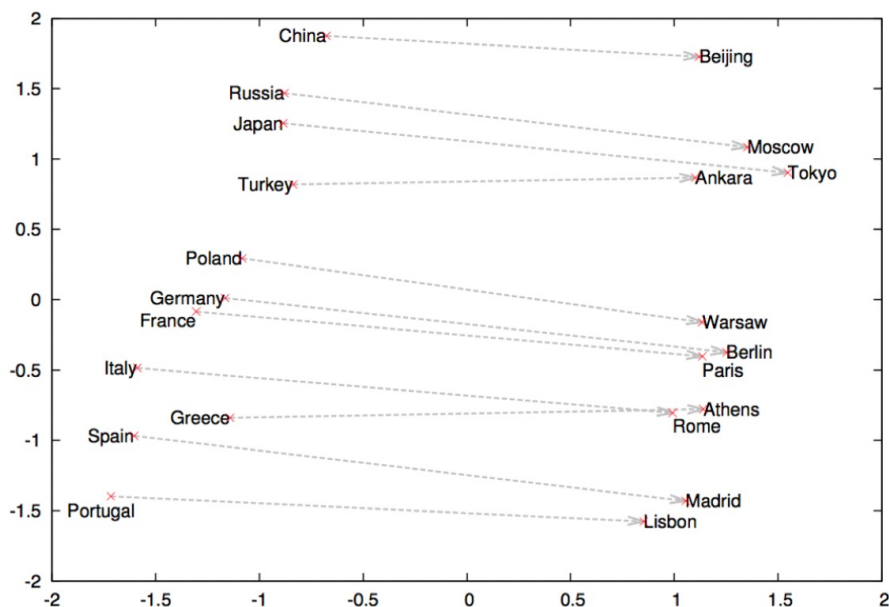
$$p(\text{output}|\text{prompt}) = \int_{\text{concept}} p(\text{output}|\text{concept}, \text{prompt})p(\text{concept}|\text{prompt})d(\text{concept}).$$



数据量, 层数, 隐参数

In-context能力的变迁

- 单词 → 句子 → 逻辑 → ...



你好，我是人工智能助理ChatGLM，基于清华大学GLM-130B开发而成。我的使命是帮助人类获取信息并提供问题建议。

问题：用厚度为1厘米的玻璃做一个长42厘米，宽27厘米，深50厘米的无盖长方体容器，如果向这个容器注入30升水，水的深度是多少厘米？

回答：1升=1000立方厘米。

因为玻璃厚1厘米，所以内部的长是 $42-2=40$ 厘米，内部宽是 $27-2=25$ 厘米。

所以水深度为 $30 \times 1000 / (40 \times 25) = 30$ 厘米。

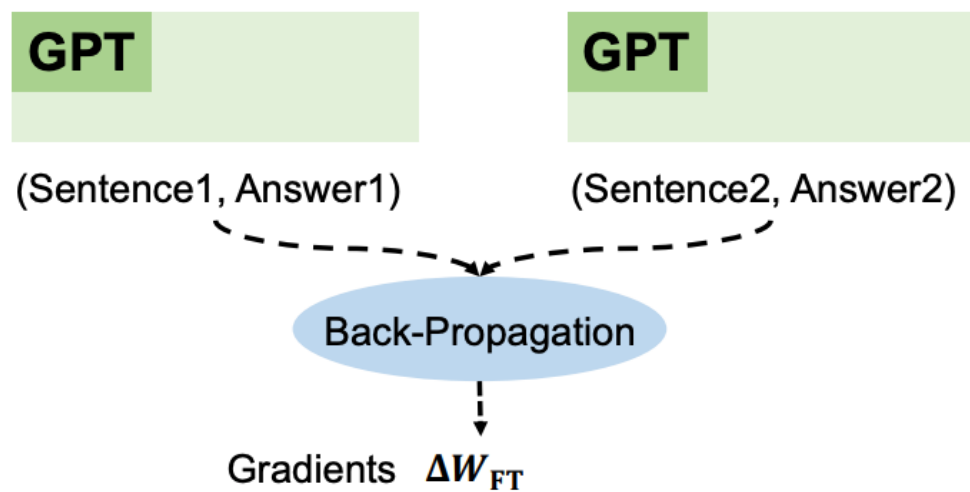
问题：学校食堂运来19吨煤，已经烧了16吨，平均每天烧0.4吨，剩下的煤每天烧0.3吨，还要几天烧完？

回答：

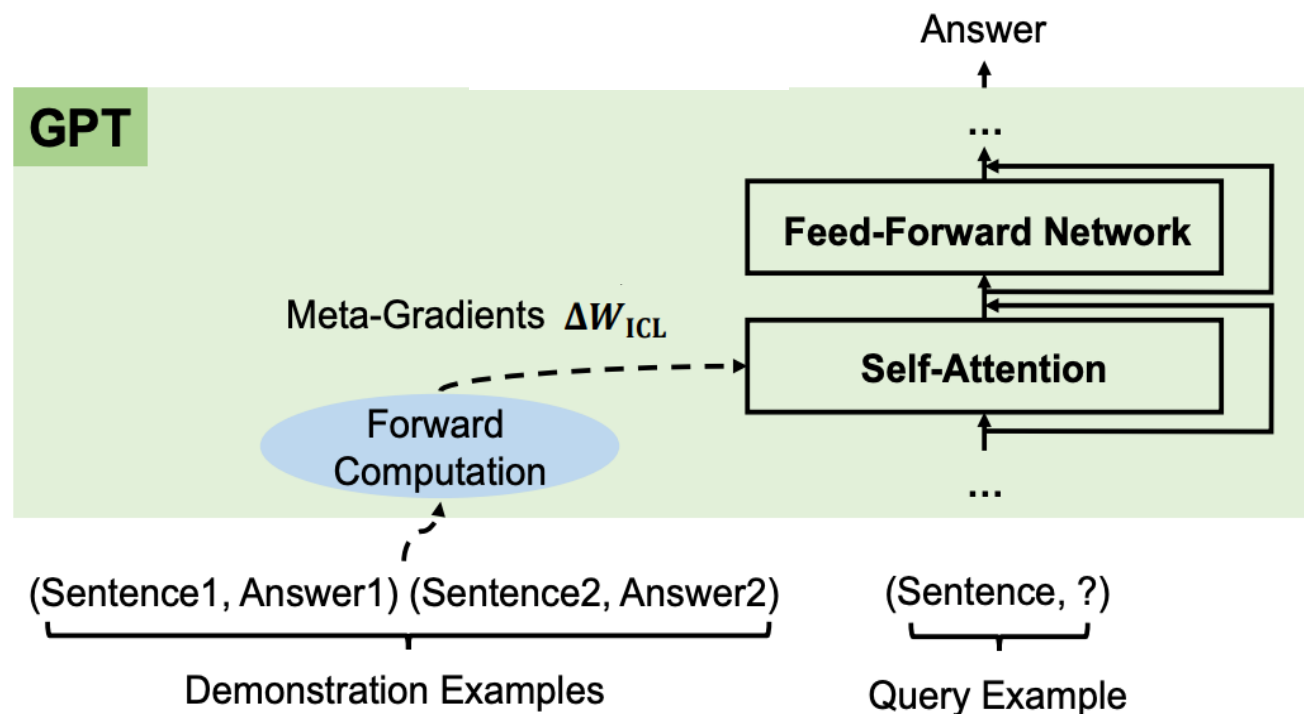
$$W(\text{"China"}) - W(\text{"Beijing"}) \approx W(\text{"Japan"}) - W(\text{"Tokyo"})$$

为什么In-Context Learning有效

- In-context learning和Fine tuning之间的关系



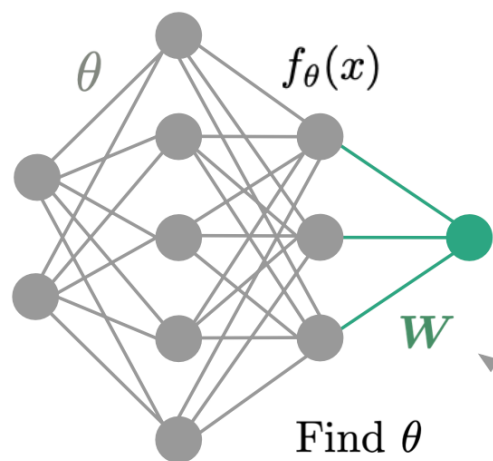
微调:
用新的样本更新模型参数



In-Context Learning:
用新的样本激活参数

从梯度下降理解In-Context

Learning an output layer by **gradient descent**



Find θ
s.t. $(W - \nabla_W L(D^{\text{train}}))f_\theta(x_{\text{test}}) \approx y_{\text{test}}$

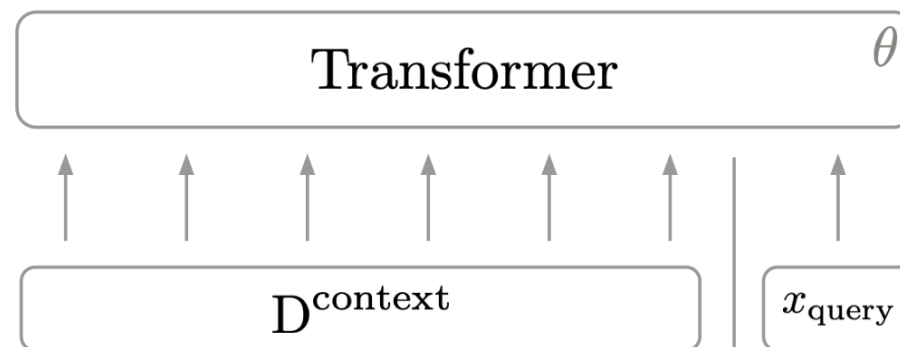
Equivalent?



Adjusts query prediction given in-context

Find θ

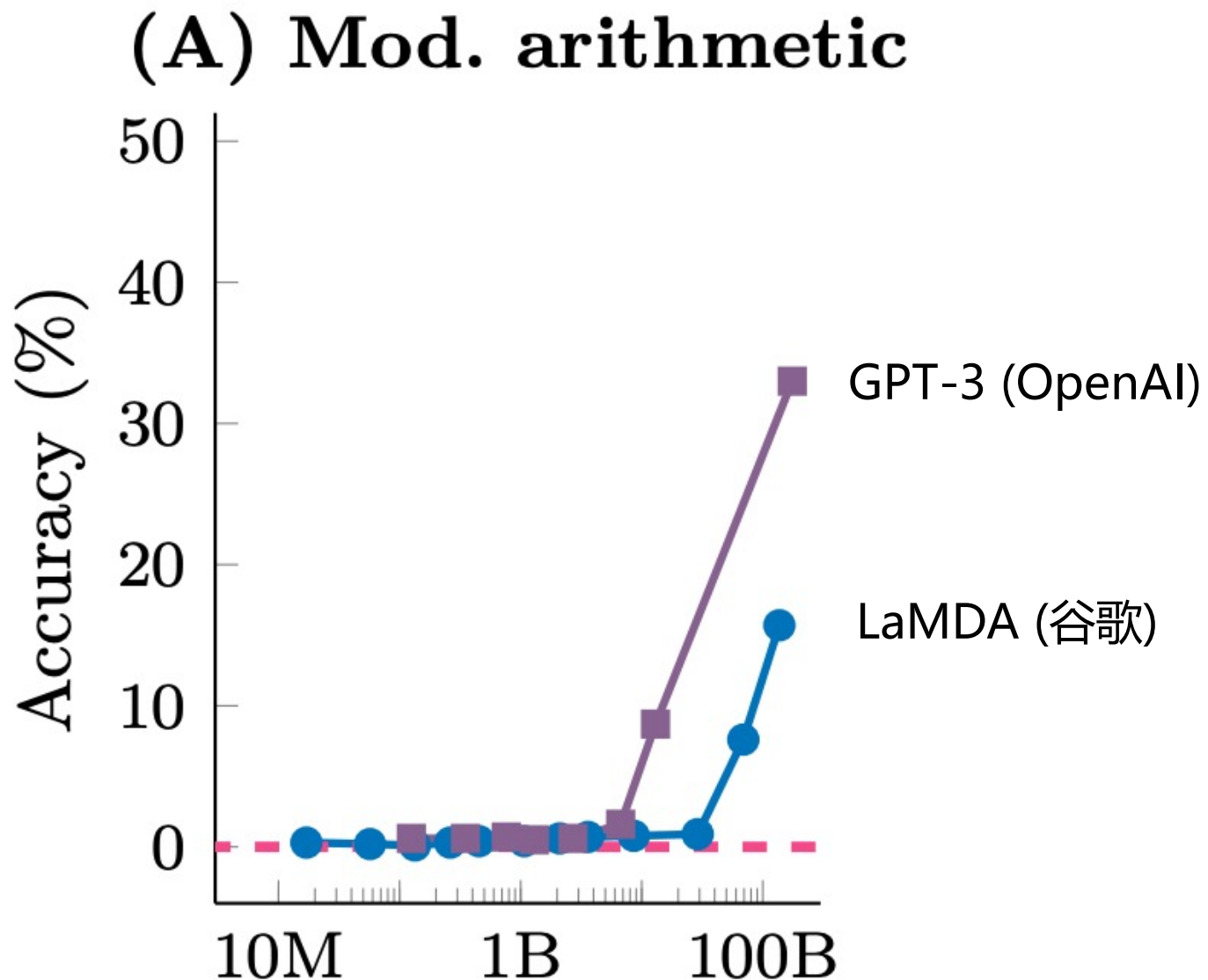
s.t. $t_\theta(x_{\text{query}}; D^{\text{context}}) \approx y_{\text{query}}$



$$\begin{aligned}
 L(W + \Delta W) &= \frac{1}{2N} \sum_{i=1}^N \|(W + \Delta W)x_i - y_i\|^2 \\
 &= \frac{1}{2N} \sum_{i=1}^N \|Wx_i - (y_i - \Delta y_i)\|^2
 \end{aligned}$$

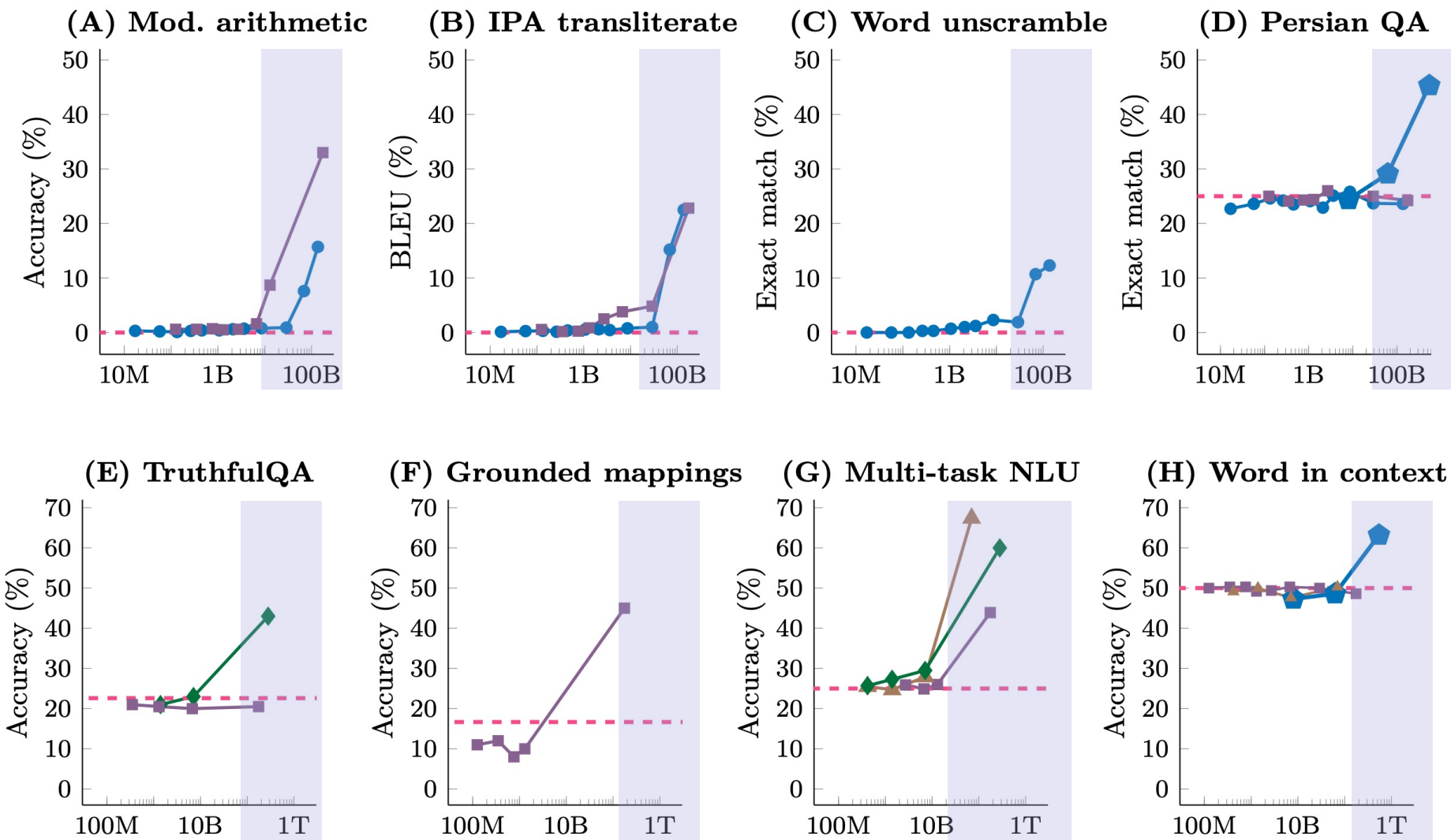
为什么千亿(100B)大规模

- What is 16 mod 12?
- 16 divided by 12 equals 1 remainder 4. So the answer is 4!



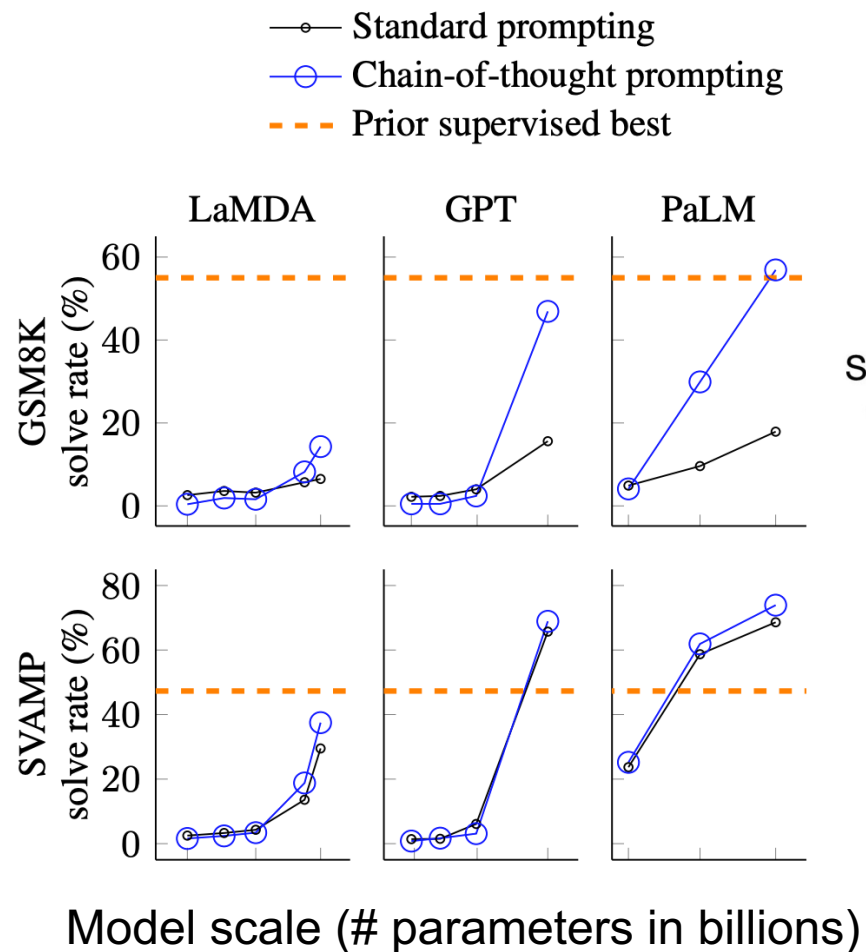
为什么千亿(100B)大规模

—●— LaMDA —■— GPT-3 —◆— Gopher —▲— Chinchilla —◆— PaLM - - - Random

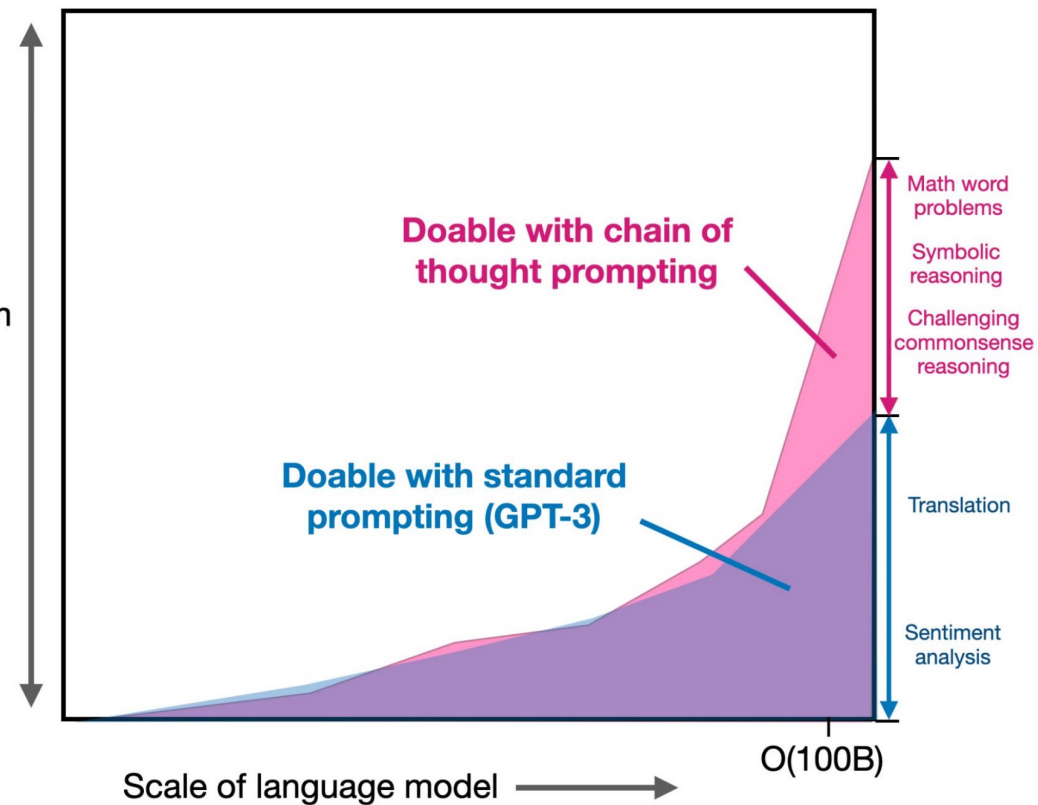


Scaling属性

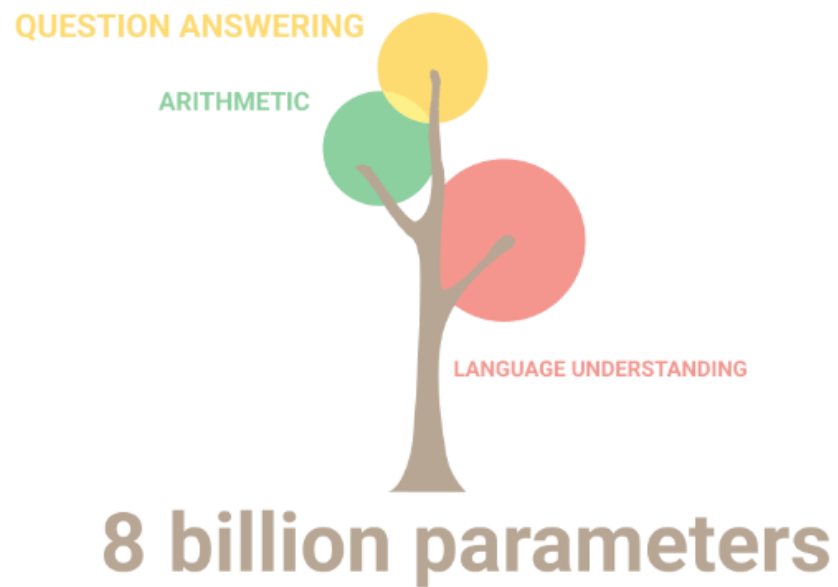
Scaling Law带来的复杂推理能力增强



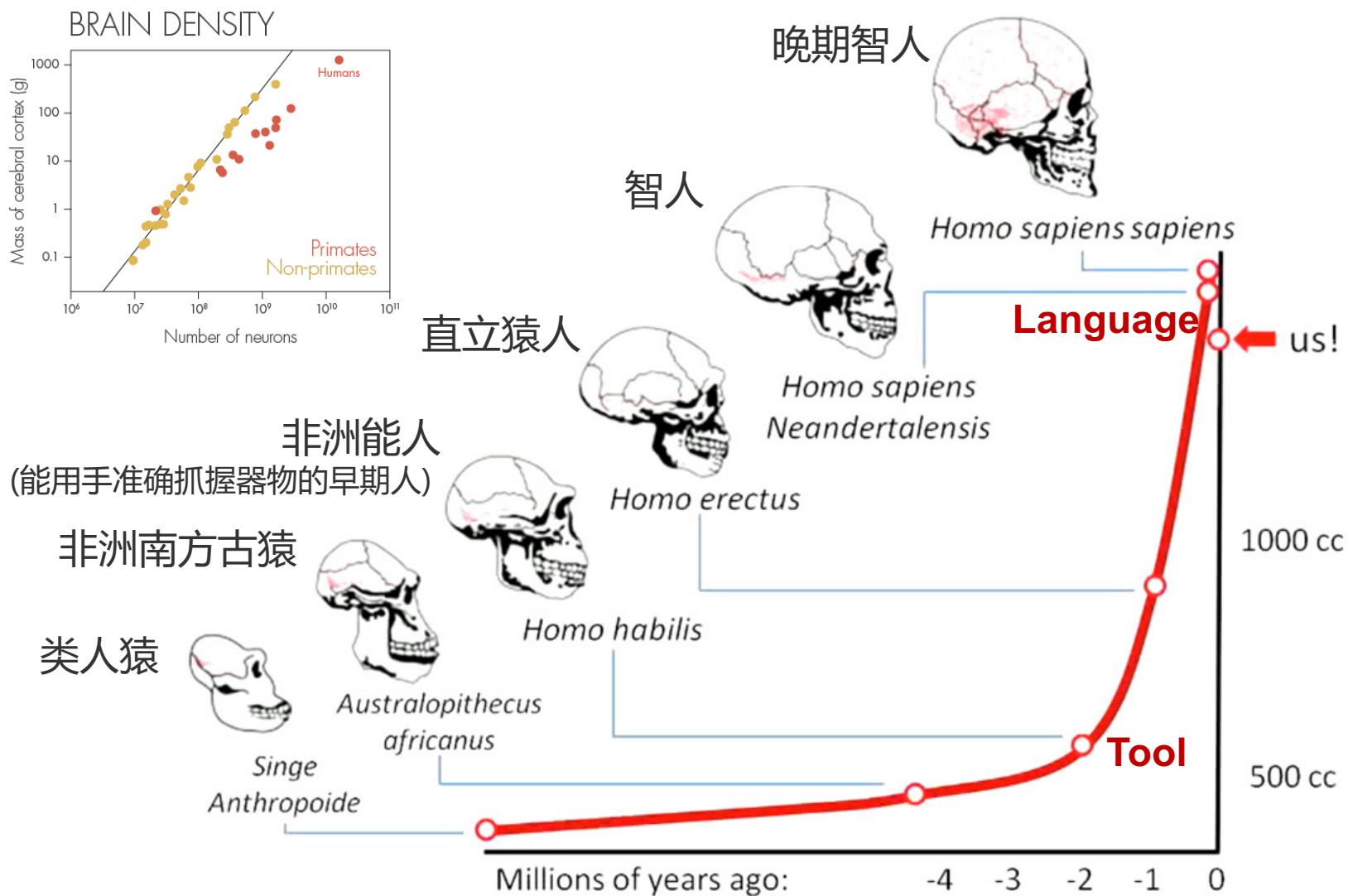
Some spectrum of NLP tasks



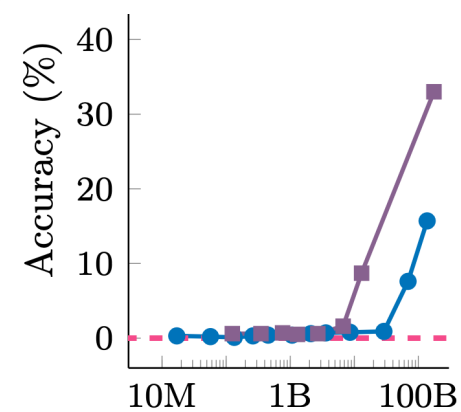
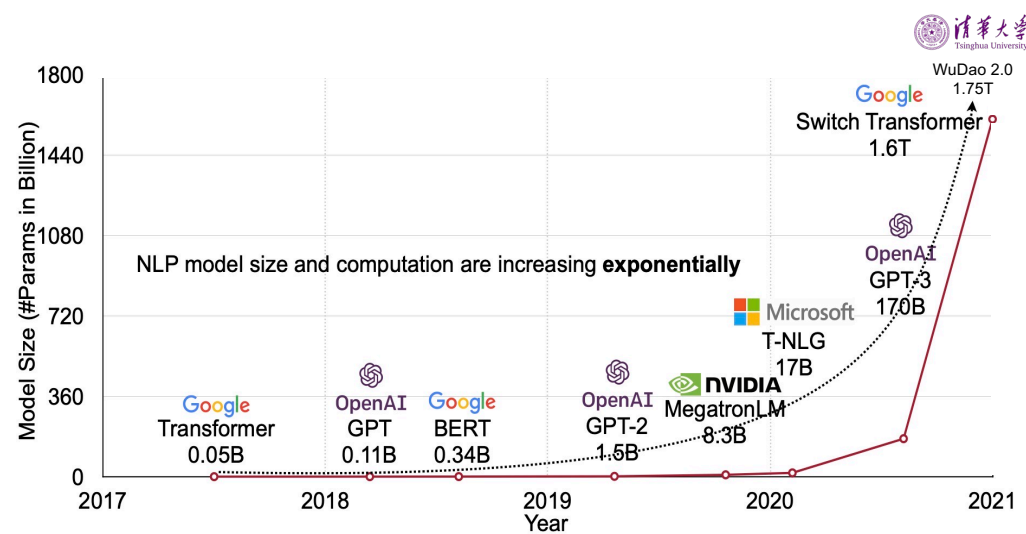
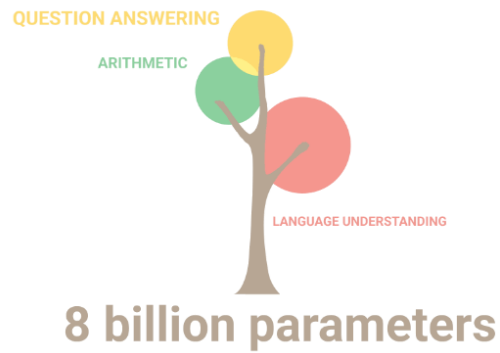
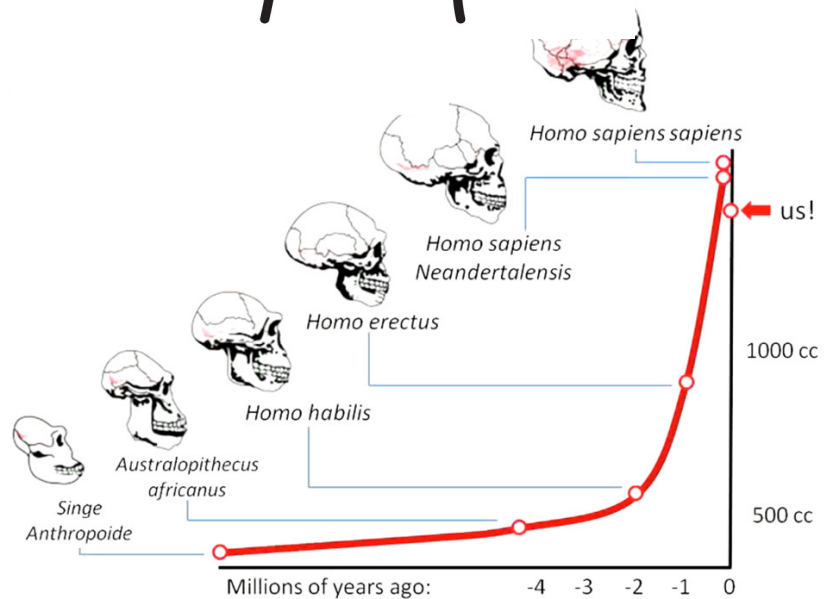
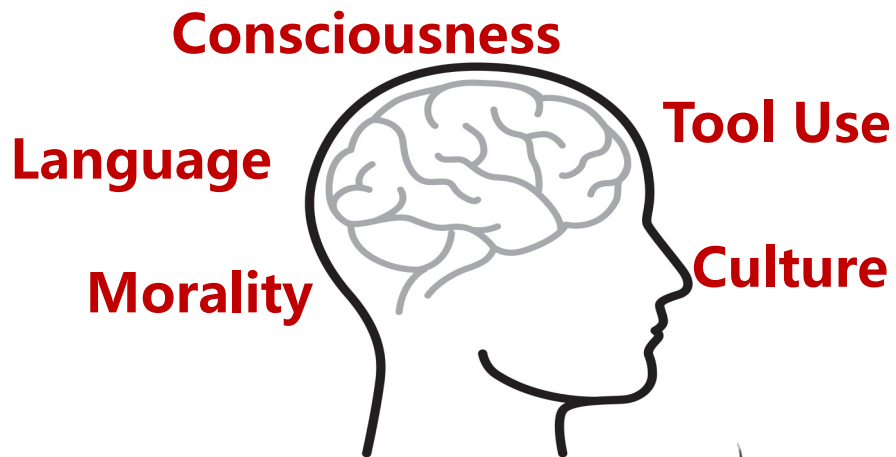
“涌现的模型新能力”



大脑进化 Spike



能力涌现



Disclaimer: The speaker is not an expert on this topic at all

PC: Web & Google

能力涌现

GLM-10B

- 一斤西瓜10元钱, 5斤西瓜多少钱?
 - 一斤西瓜十元,五斤西瓜30元
- 一斤苹果10元, 一斤西瓜5元, 那么2斤苹果和3斤西瓜多少钱?
 - 总共30元
- 我把可口可乐洒在了桌子上, 接下来该如何清楚桌面?
 - 把桌上的东西都拿走,用湿抹布擦桌子,然后喷上酒精,再拿抹布擦,最后用酒精再擦一遍

GLM-130B

- 一斤西瓜10元钱, 5斤西瓜多少钱?
 - $5 \times 10 = 50$ (元)
- 一斤苹果10元, 一斤西瓜5元, 那么2斤苹果和3斤西瓜多少钱?
 - $2 \times 10 + 3 \times 5 = 35$ (元)
- 我把可口可乐洒在了桌子上, 接下来该如何清楚桌面?
 - 先用抹布擦干净, 然后用湿抹布再擦一次, 最后用干抹布再擦一次

认知智能

趋势：用**计算模型**可描述的人类认知问题，计算机很快达到甚至超过人类水平。

很快=5-20年（2019年）

$$P(y|x) \Rightarrow \int P(y|\phi) P(\phi|x) d\phi$$

$$\Rightarrow \int P(x^{new}|\phi) P(\phi|x^{old}) d\phi$$

从认知到AGI的思考

表示、计算
、存储与人
机输入/出

文本内容识
别、图像识
别、语音识
别

知识数据双驱
动、认知推理
、决策智能

自学习、记忆
机理、有(无)
意识处理器

计算

感知

认知

意识



从认知到AGI的思考

人脑是一个复杂系统，各脑区相互配合完成认知功能

多模态感知与理解

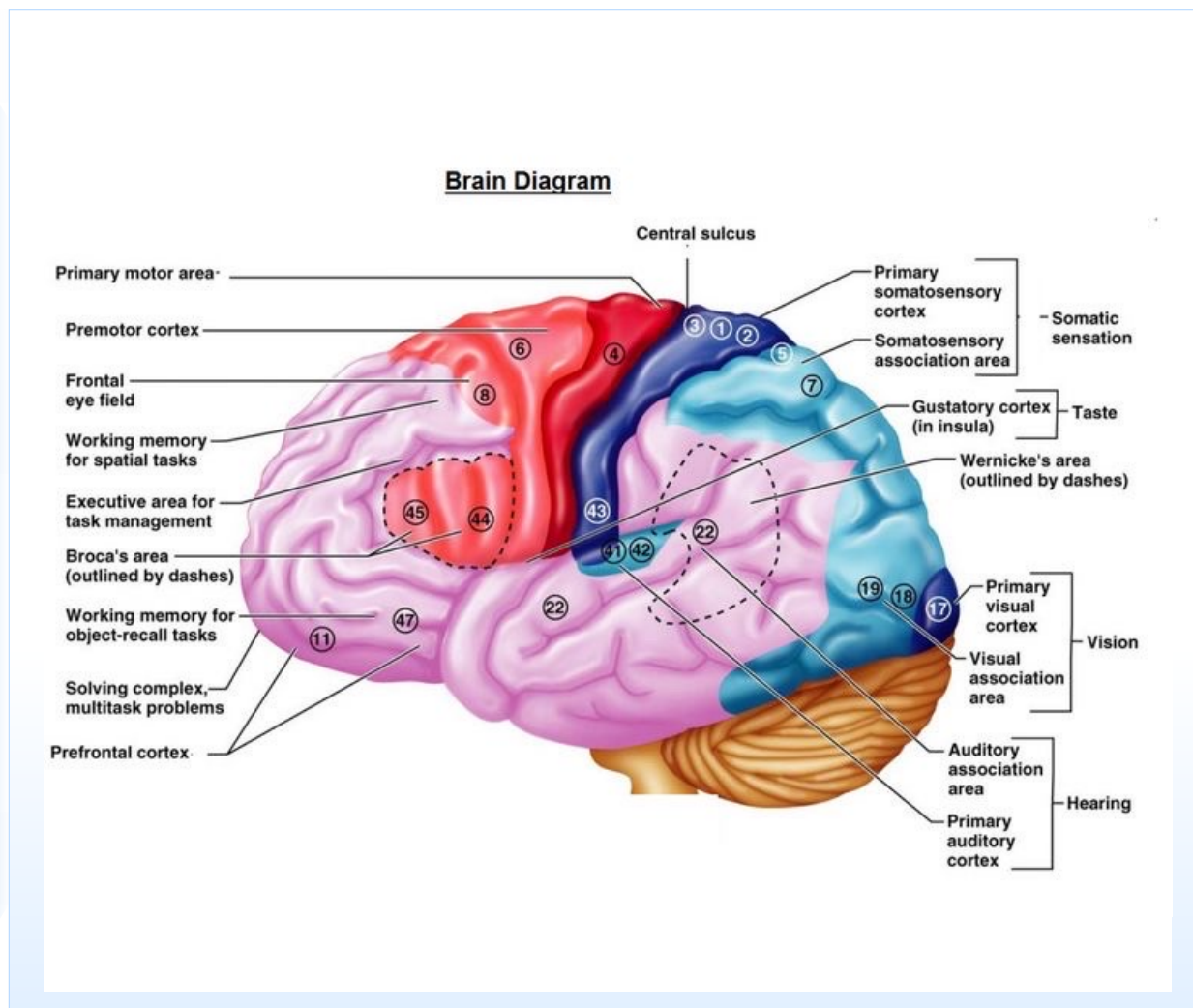
- ▶ 视觉理解（枕叶中视觉脑区）
- ▶ 听觉语言理解（Wernicke区）
- ▶ 视觉语言理解（顶下小叶角回）

学习、陈述性记忆

- ▶ 海马区（Hippocampus）
- ▶ 短期记忆在海马体转化为陈述性记忆
- ▶ 蛋白质磷酸化修饰（中期）
- ▶ 长时程增强、新蛋白产生（长期记忆）

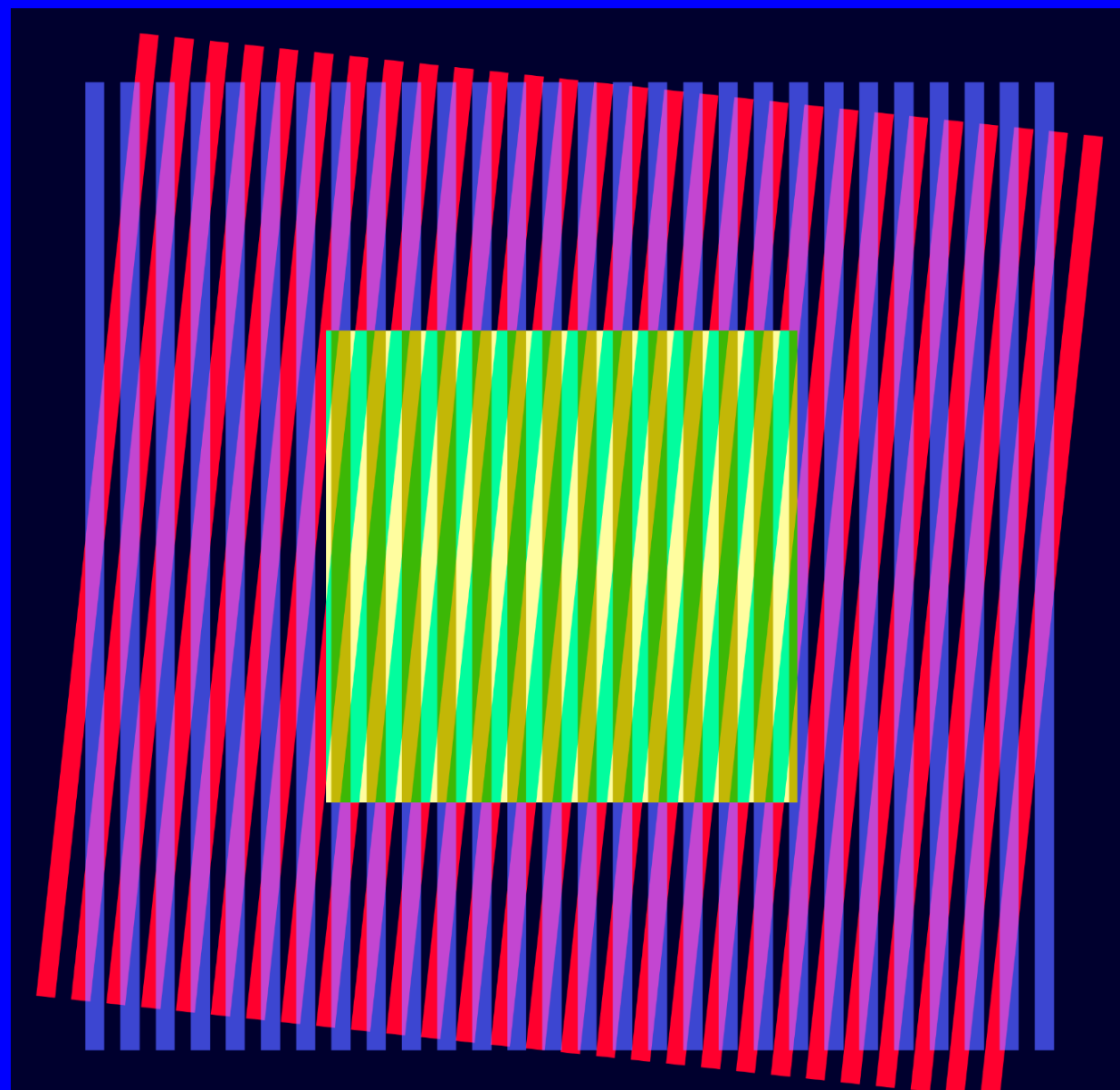
工作记忆与推理

- ▶ 前额叶中相关脑区



Introducing Superalignment

We need scientific and technical breakthroughs to steer and control AI systems much smarter than us. To solve this problem within four years, we're starting a new team, co-led by Ilya Sutskever and Jan Leike, and dedicating 20% of the compute we've secured to date to this effort. We're looking for excellent ML researchers and engineers to join us.



2024-**AGI**元年

趋势： 用计算模型可描述的认知问题，AI很快超过人类水平。

很快=5-20年（2019年）



趋势： 通用问题上，AI全面超越人类，实现超级认知智能；探究**科学规律、世界起源等终极问题**。

很快=5-20年（2023年）

超级认知智能

让机器像人一样“思考”：超越图灵测试的通用机器认知能力

认知
AI

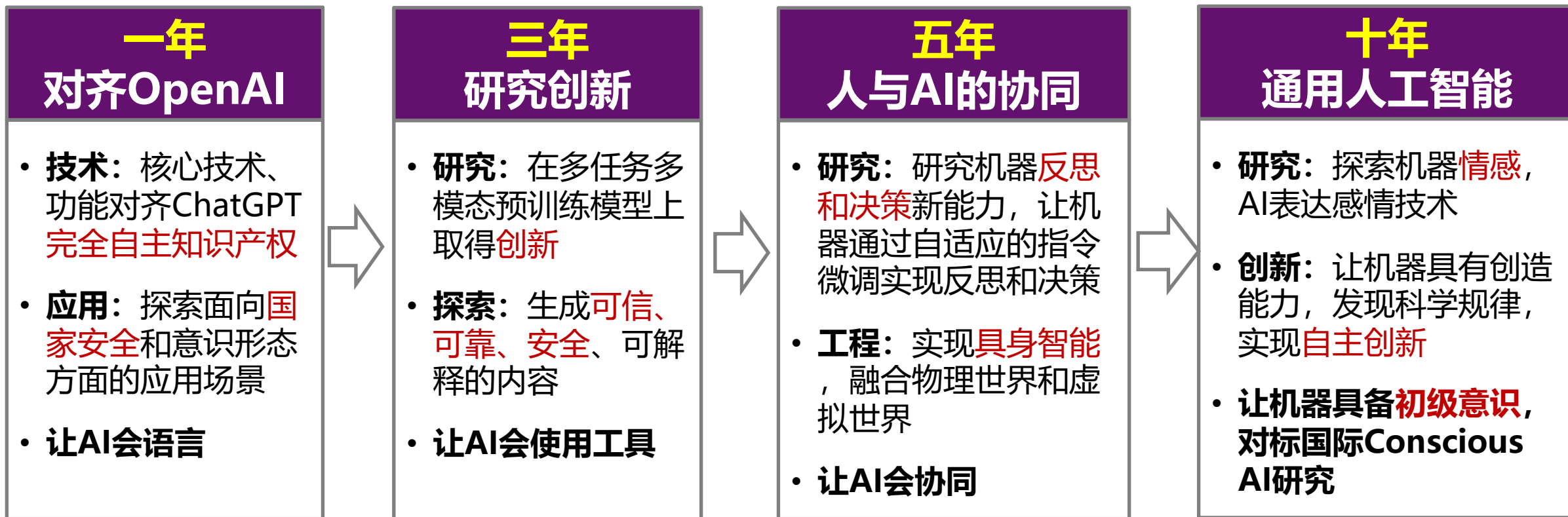
1. 适应与学习能力
2. 定义与语境化能力
3. 自我系统的准入能力
4. 优先级与访问控制能力
5. 召集与控制能力
6. 决策与执行能力
7. 错误探测与编辑能力
8. 反思与自我监控能力
9. 条理与灵活性之间的能力

探索更强大、更通用模型

探索面向超级认知智能的模型范式

探索新的研发模式

十年规划



科学创新: 在通用人工智能的道路上留下中国人的声音

国家需求: 建立可解释、安全、可靠、可信的人工智能基础算力支撑

小结

- 2020-大模型元年
- GPT vs GLM
 - ChatGPT vs. ChatGLM
 - DALL.E vs. CogView
 - Codex vs. CodeGeeX
 - WebGPT vs. WebGLM
 - GPT-4V vs. GLM4 (CogVLM, AgentTuning...)
- 2024-AGI元年

相关文献

- Qinkai Zheng, Xiao Xia, Xu Zou, Yuxiao Dong, Shan Wang, Yufei Xue, Lei Shen, Zihan Wang, Andi Wang, Yang Li, Teng Su, Zhilin Yang, and Jie Tang. **CodeGeeX**: A Pre-Trained Model for Code Generation with Multilingual Benchmarking on HumanEval-X. KDD'23.
- Xiao Liu, Hanyu Lai, Yu Hao, Yifan Xu, Aohan Zeng, Zhengxiao Du, Peng Zhang, Yuxiao Dong, and Jie Tang. **WebGLM**: Towards An Efficient Web-enhanced Question Answering System with Human Preference. KDD'23.
- Jing Zhang, Xiaokang Zhang, Daniel Zhang-Li, Jifan Yu, Zijun Yao, Zeyao Ma, Yiqi Xu, Haohua Wang, Xiaohan Zhang, Nianyi Lin, Sunrui Lu, Jie Tang, and Juanzi Li. **GLM-Dialog**: Noise-tolerant Pre-Training for Knowledge-grounded Dialogue Generation. KDD'23.
- Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Zhiyuan Liu, Peng Zhang, Yuxiao Dong, and Jie Tang. **GLM-130B**: An Open Bilingual Pre-trained Model. ICLR'23.
- Wenyi Hong, Ming Ding, Wendi Zheng, Xinghan Liu, and Jie Tang. **CogVideo**: Large-scale Pretraining for Text-to-Video Generation via Transformers. ICLR'23.
- Ming Ding, Wendi Zheng, Wenyi Hong, and Jie Tang. **CogView2**: Faster and Better Text-to-Image Generation via Hierarchical Transformers. NeurIPS'22.
- Jifan Yu, Xiaohan Zhang, Yifan Xu, Xuanyu Lei, Xinyu Guan, Jing Zhang, Lei Hou, Juanzi Li, and Jie Tang. **XDAI**: A Tuning-free Framework for Exploiting Pre-trained Language Models in Knowledge Grounded Dialogue Generation. KDD'22.
- Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. **GLM**: General Language Model Pretraining with Autoregressive Blank Infilling. ACL'21.
- Zixuan Ma, Jiaao He, Jiezhong Qiu, Huanqi Cao, Yuanwei Wang, Zhenbo Sun, Liyan Zheng, Haojie Wang, Shizhi Tang, Tianyu Zheng, Junyang Lin, Guanyu Feng, Zeqiang Huang, Jie Gao, Aohan Zeng, JianWei Zhang, Runxin Zhong, Tianhui Shi, Sha Liu, Weimin Zheng, Jie Tang, Hongxia Yang, Xin Liu, Jidong Zhai, and Wenguang Chen. **BAGUALU**: Targeting Brain Scale Pretrained Models with over 37 Million Cores. PPOPP'22.
- Ming Ding, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, and Jie Tang. **CogView**: Mastering Text-to-Image Generation via Transformers. NeurIPS'21.
- Junyang Lin, Rui Men, An Yang, Chang Zhou, Yichang Zhang, Peng Wang, Jingren Zhou, Jie Tang, and Hongxia Yang. **M6**: Multi-Modality-to-Multi-Modality Multitask Mega-transformer for Unified Pretraining. KDD'21.