



引言

CS2916 大语言模型

饮水思源 愛國榮校

刘鹏飞@GAIR

<https://plms.ai/teaching/index.html>

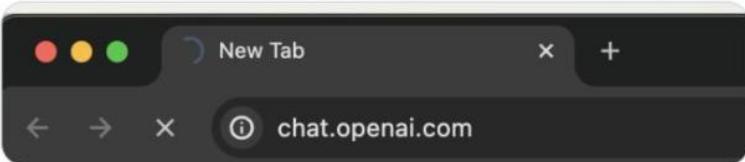


ChatGPT宕机的8个小时...



Andrew Gao ✅ @itsandrewgao · 14h
is #ChatGPT loading for you?

EDIT: it's down



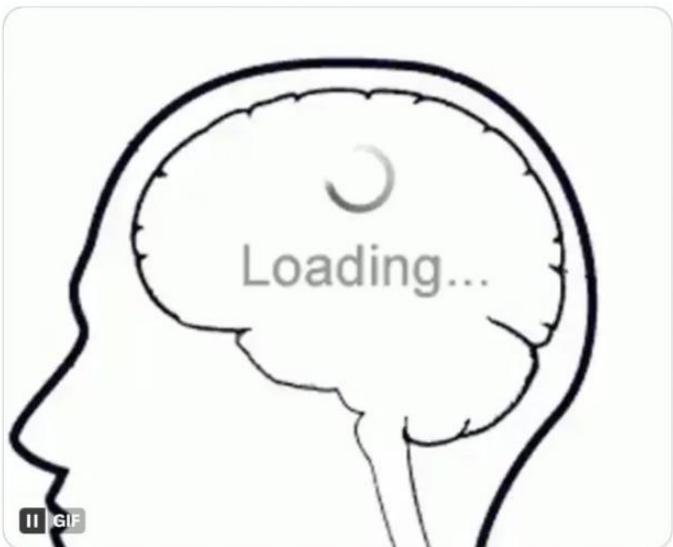
Favor ⚡ ✅ @heyOnuoha · 3h
ChatGPT is down?

brain.exe has stopped working 😂

15 8 54 2.9K



Mr. Trad Husband @Trad_Husbn · 12h
ChatGPT is down! How dare they force me to think for myself!!!



14

76

392

27K

2



Dorian Develops ✅ @DorianDevelops · 13h

How am I supposed to get any work done with ChatGPT being down!?

37

42

294

30K

2



Bitcoin Lebowski ⚡ @DudeJLebowski · 12h

Chat GPT is down, I'm not doing any actual thinking, I refuse. I'll just sit here until it becomes available again.

46

250

1.2K

142K

2



Kelly 🌟 ✅ @SelfMadeMastery · 3h

ChatGPT has been down all day and I'm surprised how much this has affected me.

I'm really using this thing every single day.

32

4

62

4.2K

2



Swapna Rajdev ✅ @SRajdev · 6h

I didn't realize how dependent I've become on #chatGPT until its major downtime today. I'm at least 30% less productive this morning!

5

2

5

341

2



Generative AI is Eating the World



杰夫·贝索斯

目前形式的大语言模型并不是发明，而是发现。望远镜是一项发明，但通过它观察木星，知道它有卫星，是一项发现。而大语言模型更像是发现，它们的能力不断让我们感到惊讶



比尔盖茨

人生中让我印象深刻的两次技术革命演示，一次是现在操作系统的先驱“图形用户界面”，另一个就是以ChatGPT为代表的生成式人工智能技术



黄仁勋

ChatGPT相当于AI界的iPhone问世，它使每一个人都可以成为程序员

马斯克悄悄成立大模型公司xAI





世界在发生着快速的变化

从文本生成文本



TF

You

请帮我写一个面向本科生的《大语言模型》的课程大纲

AI

ChatGPT

《大语言模型》课程旨在向本科生介绍人工智能和自然语言处理领域的一项重要技术——大型语言模型（如GPT系列）。本课程将通过理论学习与实践操作相结合的方式，帮助学生深入理解大语言模型的工作原理、应用场景、优点与局限，以及模型训练和部署的基本方法。以下是课程的大纲：

课程大纲

第1周：引言

- 课程介绍与目标
- 人工智能与自然语言处理概述
- 大语言模型的历史与发展

第2周：基础知识回顾

- 机器学习基础
- 深度学习入门
- 文本处理基本技术

第3-4周：大语言模型原理

- 什么是大语言模型
- 模型架构（如Transformer）
- 模型训练方法（包括预训练和微调）

第5-6周：GPT系列模型深度解析

- GPT模型结构详解
- 从GPT-1到GPT-4：演进与创新
- 案例研究：特定应用中的GPT模型



世界在发生着快速的变化

从文本生成图片

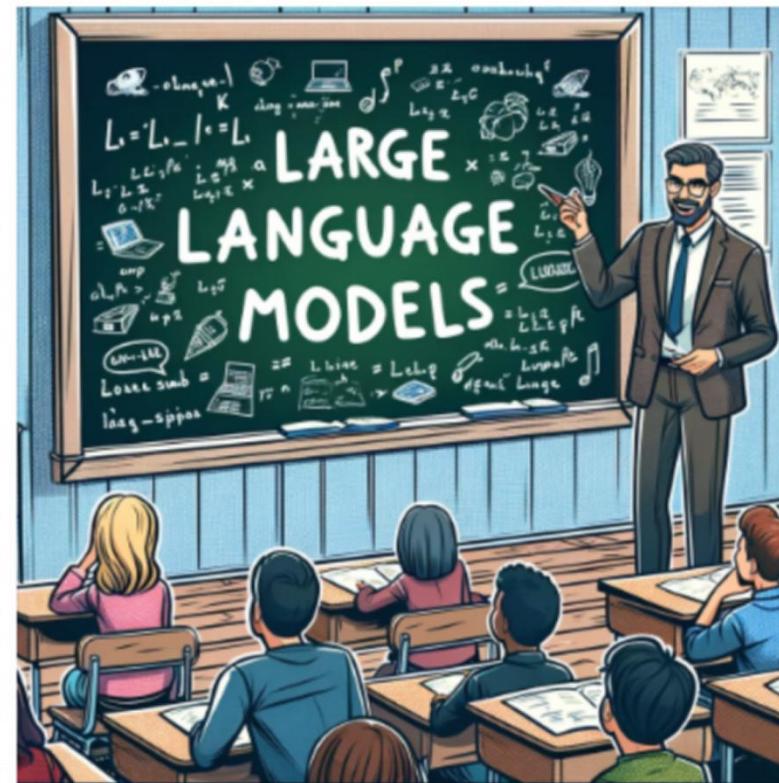


TF

You

帮我画一张卡通风格的图片，里面一位老师正在教室里授课，黑板上写着课程名《Large Language Models》

ChatGPT





世界在发生着快速的变化

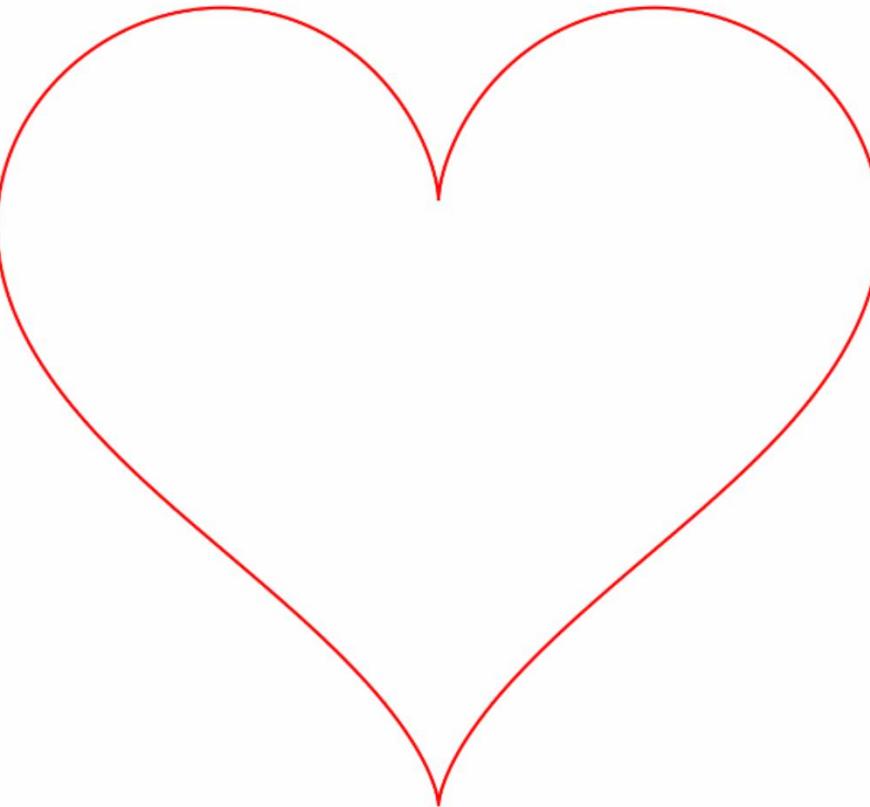
从文本生成代码



YOU
生成一段python代码，运行后可以画出一个心型图

ChatGPT

Heart Shape



这是使用Python绘制的心形图。通过运行上述代码，我们得到了一个美丽的心形图像，没有坐标轴，突出了心形的纯粹和简洁。[...]



世界在发生着快速的变化

查看详细代码



Analysis X

python Copy code

```
import matplotlib.pyplot as plt
import numpy as np

# Define the parameters for the heart shape
t = np.linspace(0, 2 * np.pi, 1000)

# Define the heart shape in polar coordinates
x = 16 * np.sin(t) ** 3
y = 13 * np.cos(t) - 5 * np.cos(2*t) - 2 * np.cos(3*t) - np.cos(4*t)

# Plot the heart shape
plt.figure(figsize=(8, 7))
plt.plot(x, y, color='red')
plt.title('Heart Shape')
plt.axis('equal')
plt.axis('off')
plt.show()
```

这是使用Python绘制的心形图。通过运行上述代码，我们得到了一个美丽的心形图像，没有坐标轴，突出了心形的纯粹和简洁。指向手的光标图标



世界在发生着快速的变化

两艘海盗船在一杯咖啡中航行时相互争斗的逼真特写视频

从文本生成视频





生成式人工智能正在吞噬这个世界



从文本生成视频

两艘海盗船在一杯咖啡中航行时相互争斗的逼真特写视频
Sora v2 release is impending:

- * 1-minute video outputs
- * text-to-video
- * text+image-to-video
- * text+video-to-video

OpenAI's Chad Nelson showed this at the C21Media Keynote in London.
And he said we will see it very very soon, as @sama has foreshadowed.





大语言模型驱动的生成式人工智能(GAI)

□ GAI是一种特定类型的AI

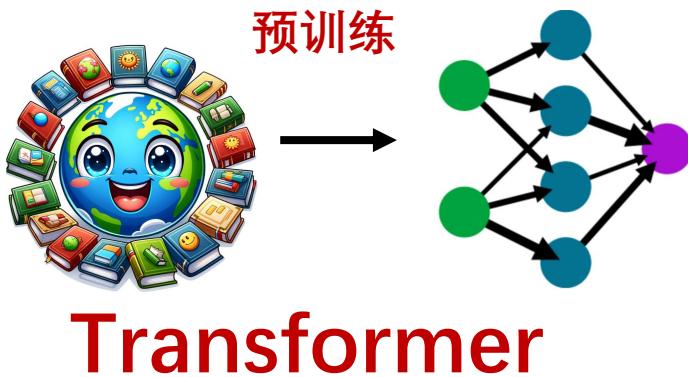
- 直接应用：可以生成不同模态的内容
- 间接应用：利用生成力解决复杂任务

基于大语言模型的“generative power”被看成是一种新的算力，可以用来解决许多复杂的任务，比如数学、化学等。



大语言模型驱动的生成式人工智能(GAI)

- 核心技术基础
 - 生成式预训 (Generative Pretraining)
 - 数据存储 (数值向量)

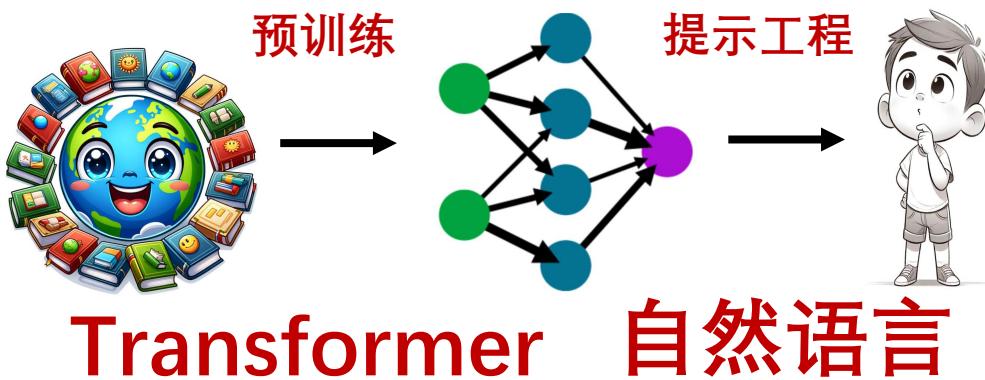




大语言模型驱动的生成式人工智能(GAI)

□ 核心技术基础

- 生成式预训练 (Generative Pretraining)
 - 数据存储 (数值向量)
- 提示工程 (Prompting Engineering)
 - 数据读取 (自然语言描述)

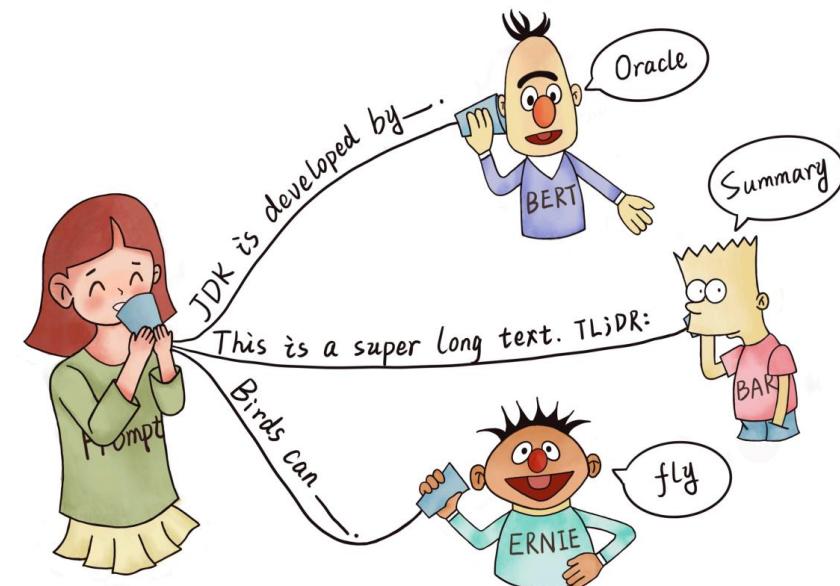
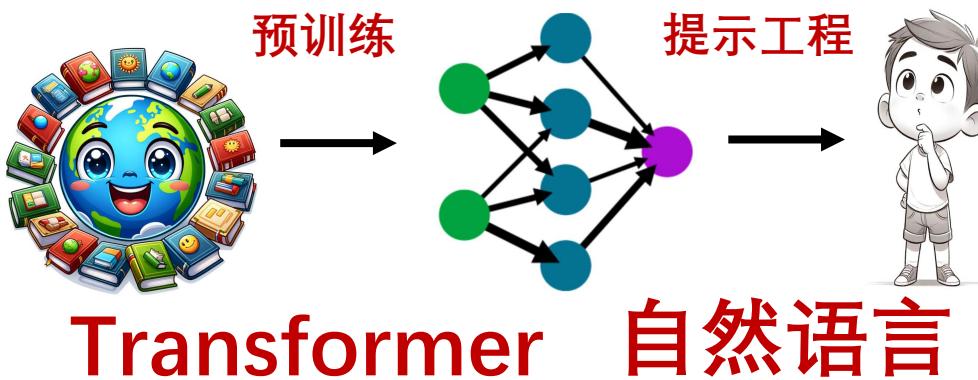




大语言模型驱动的生成式人工智能(GAI)

□ 核心技术基础

- 生成式预训练 (Generative Pretraining)
 - 数据存储 (数值向量)
- 提示工程 (Prompting Engineering)
 - 数据读取 (自然语言描述)





生成式人工智能技术的独特优势

- 数据存储
- 数据检索
- 数据标注
- 建模范式统一
- 函数生成



数据存储

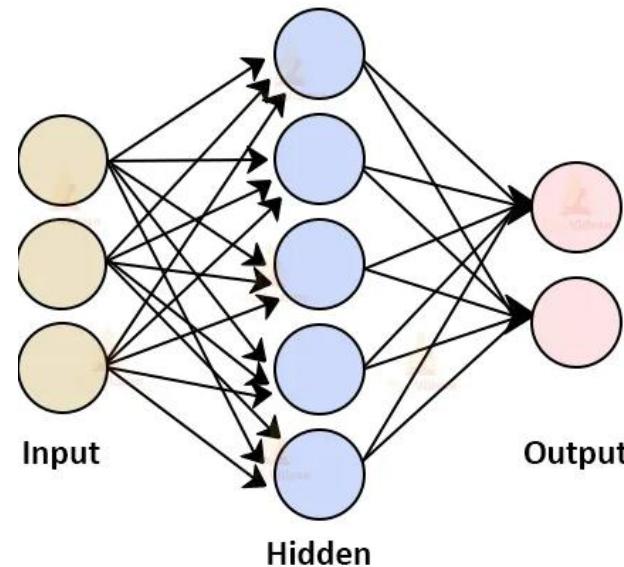
□ 过去：存储在磁盘，数据库里

- 无需学习过程
- 可理解
- 需要特殊的语言进行交互



□ 现在：存储在参数里

- 泛化性能好
- 需要数学运算, GPU
- 可以用自然语言来进行交互





数据检索

□ 过去：搜索引擎

- 无需学习过程
- 信息检索技术支持
- 字符的匹配

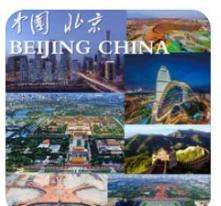
□ 现在：提示词

- 自然语言描述
- 可引导、可学习

中国的首都是

[Q 网页](#) [视频](#) [知道](#) [贴吧](#) [图片](#) [资讯](#) [文库](#)

百度为您找到相关结果约79,400,000个



中华人民共和国首都：

北京市

北京(Beijing), 简称“京”, 古称燕京、北平, 是中华人民共和国直辖市、国家中心城市、超大城市, 国务院批复确定的中国化中心、国际交往中心、科技创新中心, 截至202... [详情](#)

TF

You

中国的首都是



ChatGPT

中国的首都是北京。



数据标注

□ 过去：雇佣标注工人进行标注

- 纯人工
- 标注的质量难以控制
- 标注方式与具体任务高度相关

□ 现在：大模型辅助标注&直接合成

- 标注变成QA问题
- 不需要人工标注太多数据

人物 1 组织 2 产品 3 位置 4

经过多学科专家集中会诊后，在 **广州医科大学附属第一医院** 组织 集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果。

Playground

Load a preset...

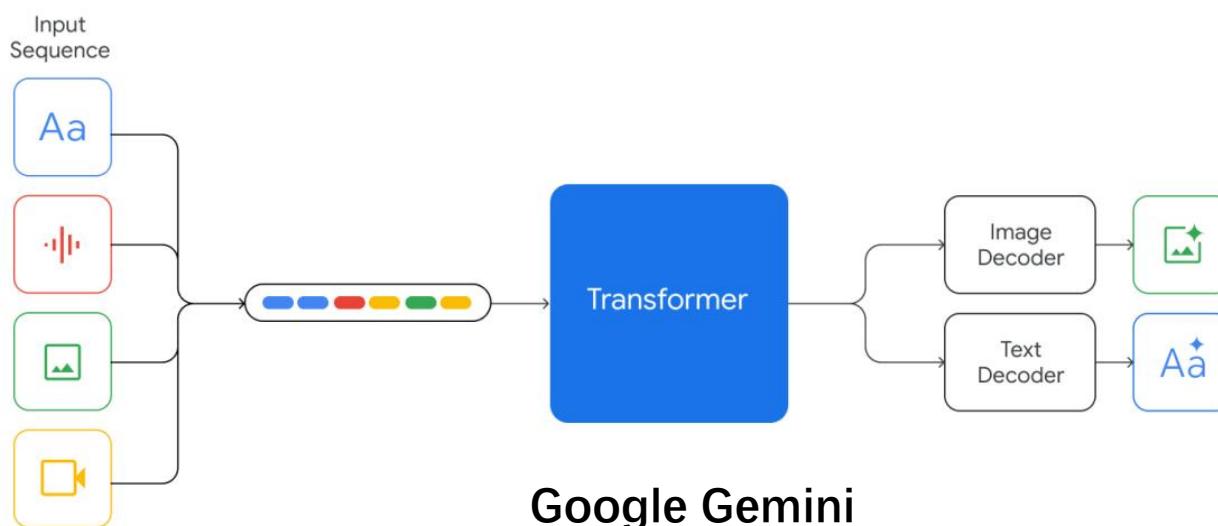
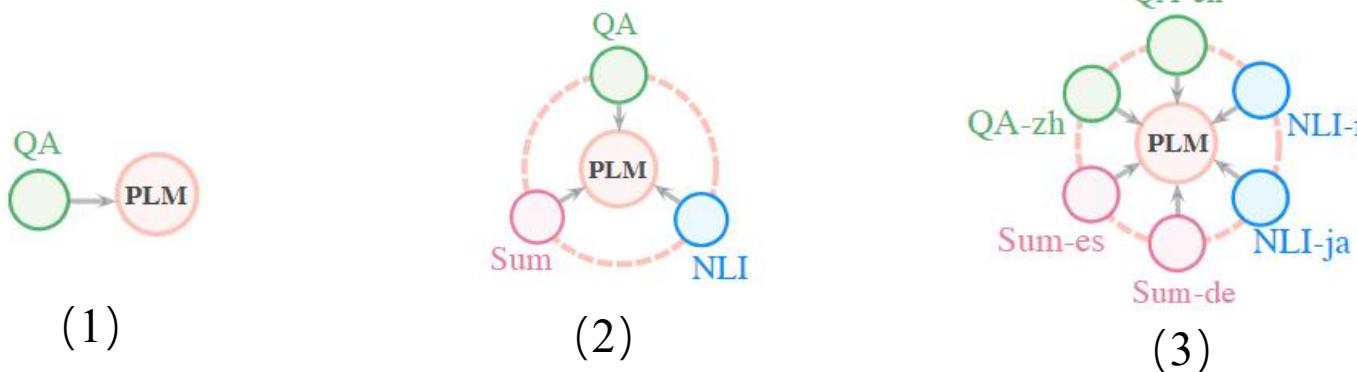
Save

经过多学科专家集中会诊后，在广州医科大学附属第一医院组织集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果
地点：广州市
经过多学科专家集中会诊后，在广州医科大学附属第一医院组织集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果
机构：**广州医科大学附属第一医院**



建模范式统一

□ 不同任务，语言，模态，一个框架

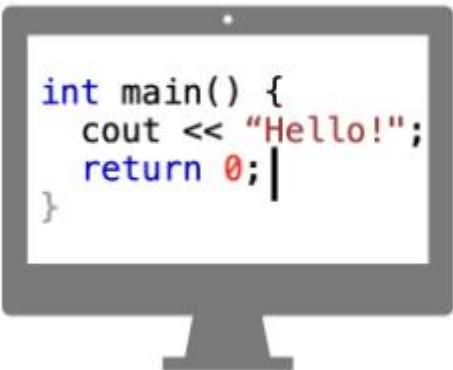




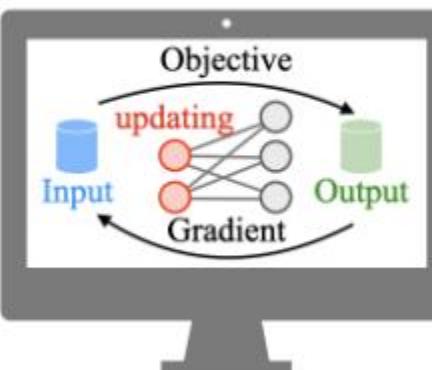
函数的生成

□ 自然语言驱动、可交互

Software 1.0



Software 2.0



Software 3.0

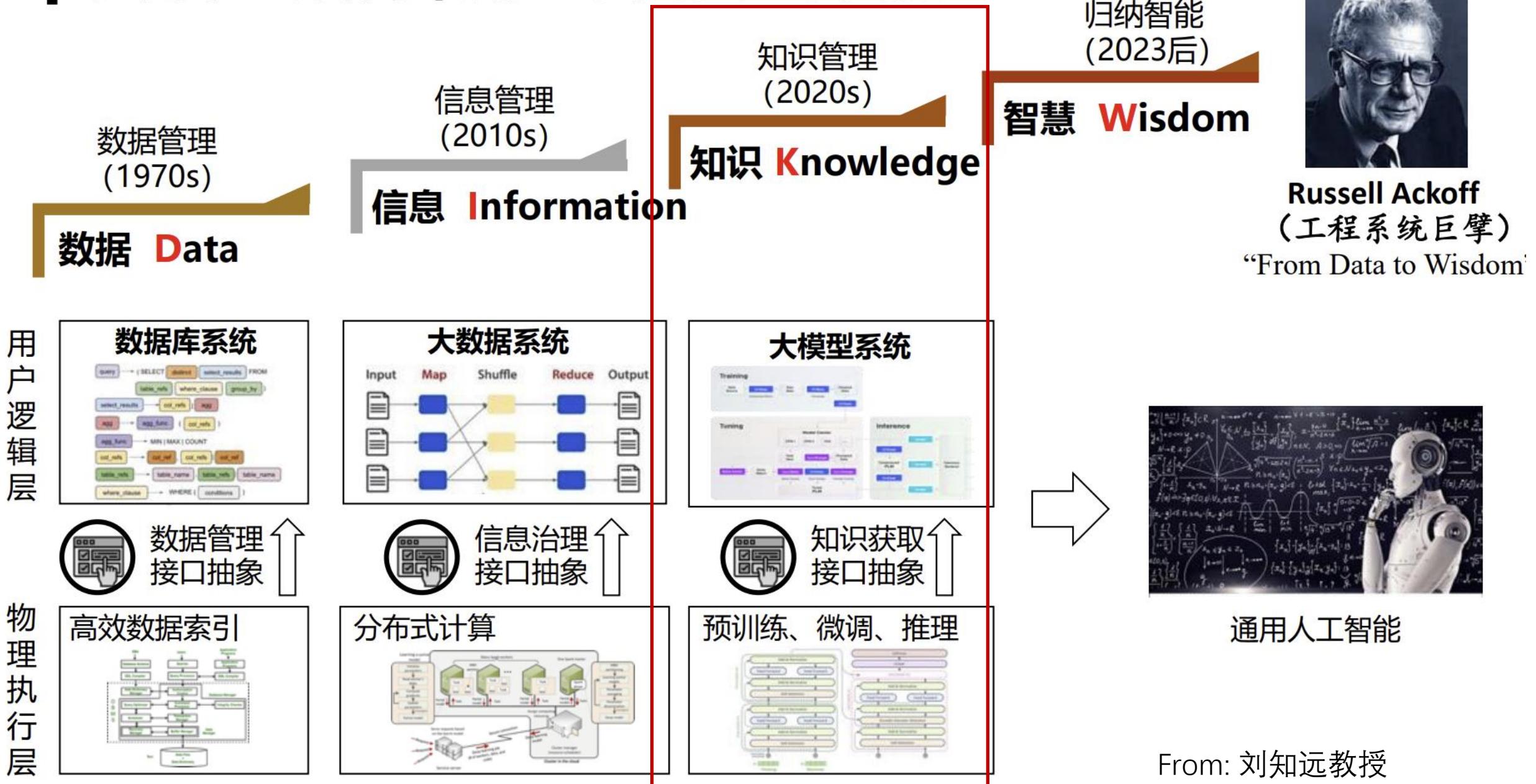


一个用于“情感分析”的软件

```
Give a sentence, classify the sentiment of it using negative and positive labels  
Text: I love this movie.  
Label: positive  
Text: This movie is too boring.  
Label: negative  
Text: I love this movie.  
Label: positive
```

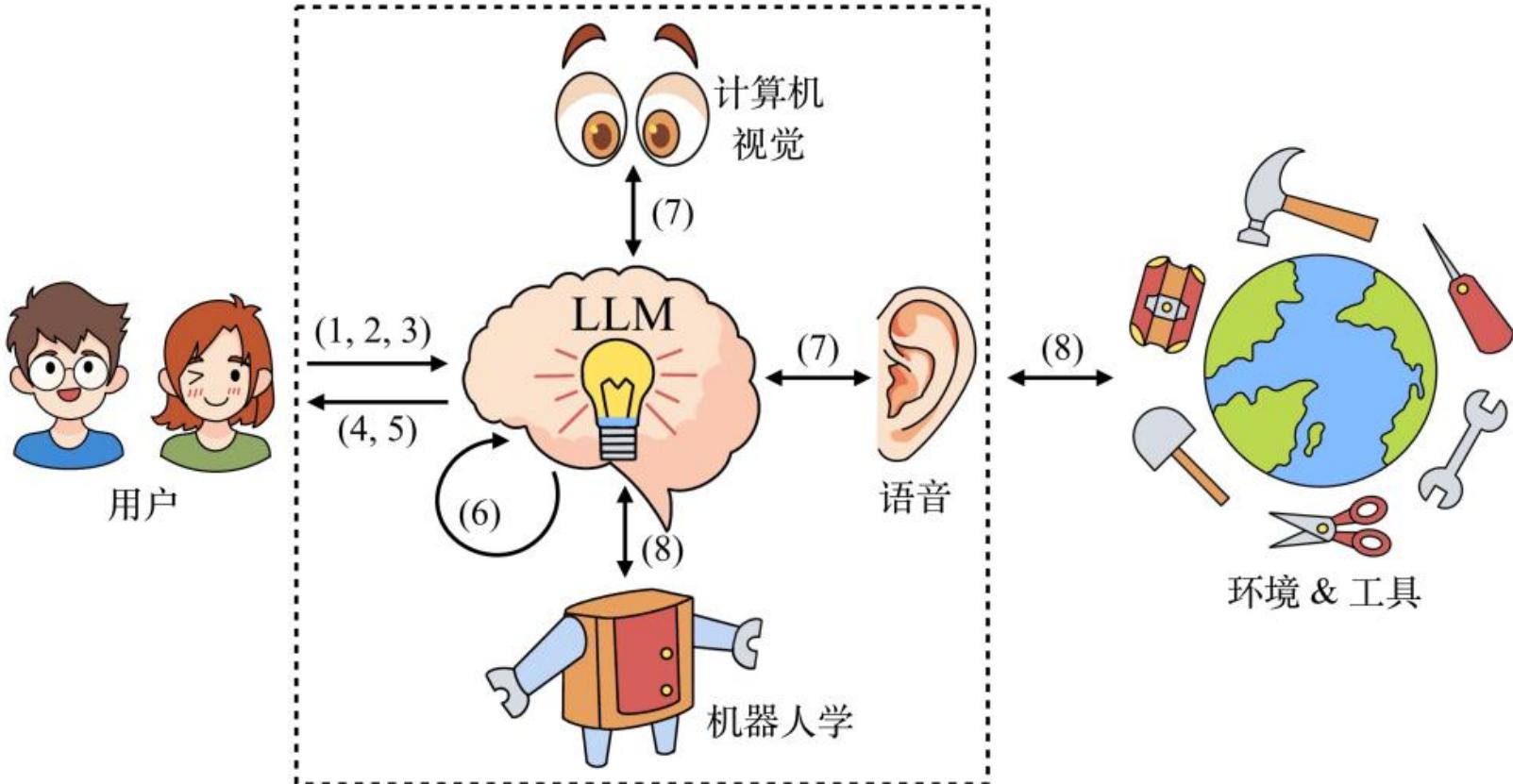
```
from promptware import install  
software = install("sentiment_classifier")  
label = software.execute({"text": "I love this movie"})
```

大模型成为智能时代基础设施





大语言模型是人类延伸的大脑



生成式AI

1. 需求挖掘
2. 提示工程
3. 系统安全
4. 用户接口
5. AI安全 & 对齐
6. LLMOps
7. 多模态功能
8. 环境交互性

传统的AI是“降劳”（取代重复的劳动力），生成式AI是“增智”



生成式AI（GPT）将如何影响我们的工作？

- GPT 的影响横跨各类薪资层级：尽管存在部分特殊情况，但整体来看，**工资越高，受 GPT 冲击的程度越大；**
- 职业技能与 GPT 的冲击程度有关：**科学和批判性思维技能最不容易受GPT冲击，而编程和写作技能受影响的程度最高；**
- **高学历更容易受到 GPT 的冲击：**持有学士、硕士和更高学位的人比没有正规教育学历的人更容易受到 GPT 的冲击；

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin^{*1}, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania



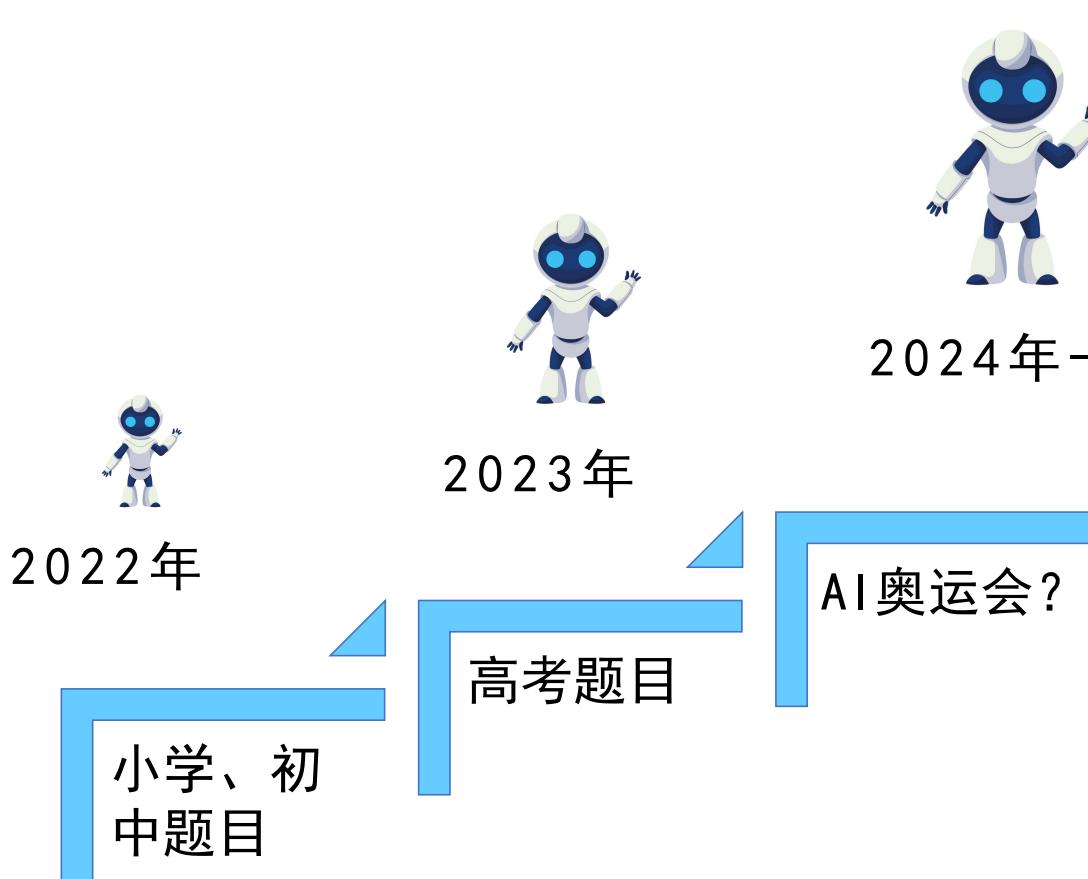
生成式AI（GPT）将如何影响我们的教育？

	碳基智能	硅基智能
代表智能体	人类	大模型
学习方式	人类教育	预训练与对齐
指导者	教师	数据、AI科学家

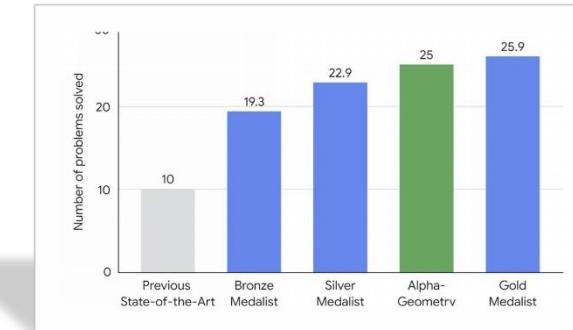




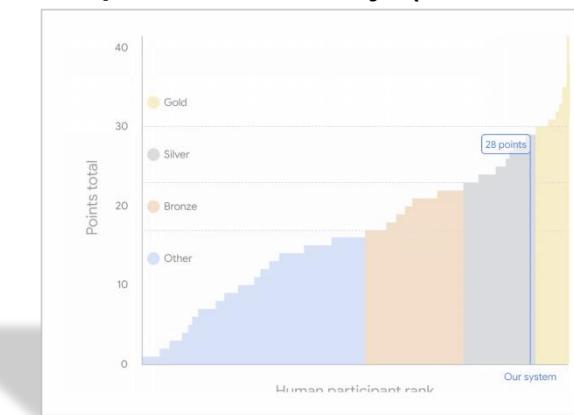
硅基智能正在快速进化：从“小学”到“奥赛”



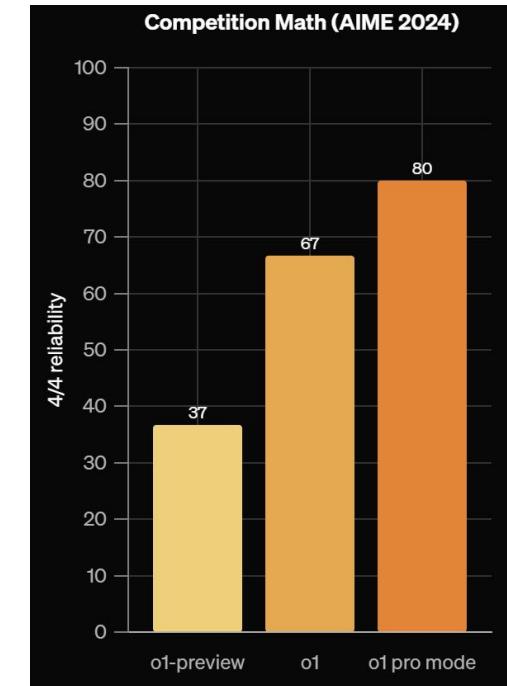
Google 工作攻破奥数难题登Nature，接近人类金牌水平



AlphaGeometry (2024.01)



AlphaProof (2024.07)



OpenAI O1



硅基智能正在快速进化：从“小学”到“奥赛”

reStructured Pre-training

Weizhe Yuan *

CMU

bellyappelian@gmail.com

Pengfei Liu †

CMU

stefanpengfei@gmail.com



OlympicArena: Benchmarking Multi-discipline Cognitive Reasoning for Superintelligent AI

Zhen Huang³, Zengzhi Wang^{1,3}, Shijie Xia^{1,3}, Xuefeng Li^{1,3}, Haoyang Zou³, Ruijie Xu^{1,3}, Run-Ze Fan^{1,3}, Lyumanshan Ye^{1,3}, Ethan Chern^{1,3}, Yixin Ye^{1,3}, Yikai Zhang^{1,3}, Yuqing Yang³, Ting Wu³, Binjie Wang³, Shichao Sun³, Yang Xiao³, Yiyuan Li³, Fan Zhou^{1,3}, Steffi Chern³, Yiwei Qin³, Yan Ma³, Jiadi Su³, Yixiu Liu^{1,3}, Yuxiang Zheng^{1,3}, Shaoting Zhang², Dahua Lin², Yu Qiao², Pengfei Liu^{1,2,3}

¹Shanghai Jiao Tong University, ²Shanghai Artificial Intelligence Laboratory,

³Generative AI Research Lab (GAIR)

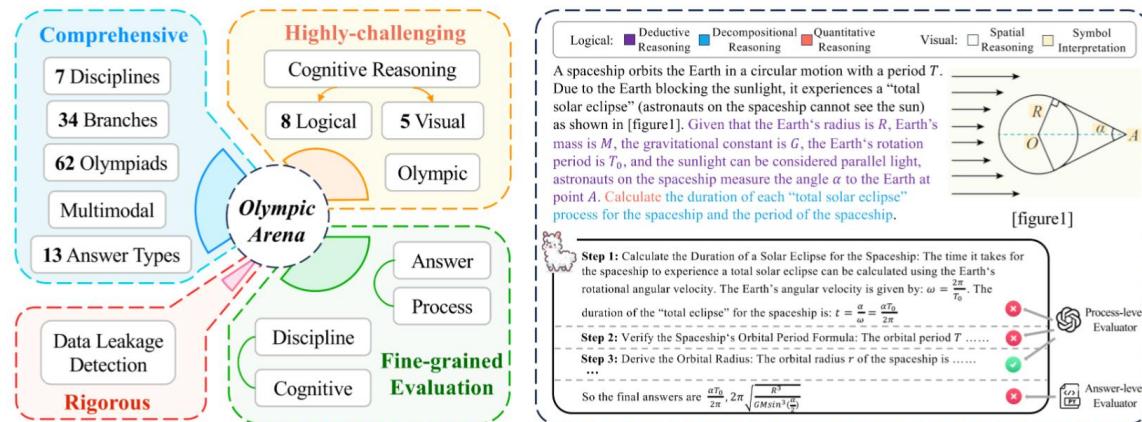


Figure 1: The overview of our *OlympicArena* benchmark.

硅基智能正在快速进化：从“小学”到“奥赛”

reStructured Pre-training

Weizhe Yuan *

CMU

bellyappelian@gmail.com

Pengfei Liu †

CMU

stefanpengfei@gmail.com



OlympicArena: Benchmarking Multi-discipline Cognitive Reasoning for Superintelligent AI

Zhen Huang³, Zengzhi Wang^{1,3}, Shijie Xia^{1,3},
Ruijie Xu^{1,3}, Run-Ze Fan^{1,3}, Lyumanshan Ye^{1,3}, Etha
Yuqing Yang³, Ting Wu³, Binjie Wang³, Shichao Sun³,
Steffi Chern³, Yiwei Qin³, Yan Ma³, Jiadi Su³,
Shaoting Zhang², Dahua Lin², Yu C

¹Shanghai Jiao Tong University, ²Shanghai Ar

³Generative AI Research L

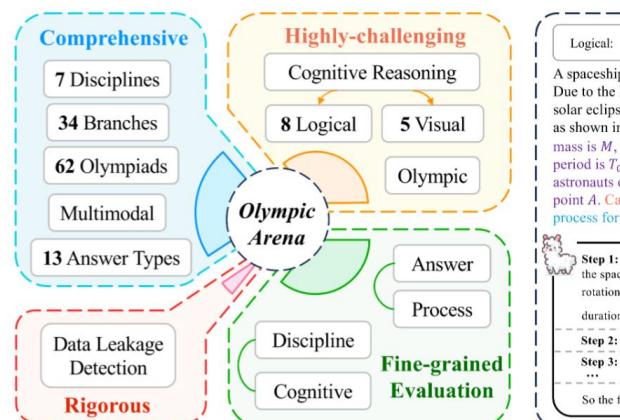


Figure 1: The overview of our O

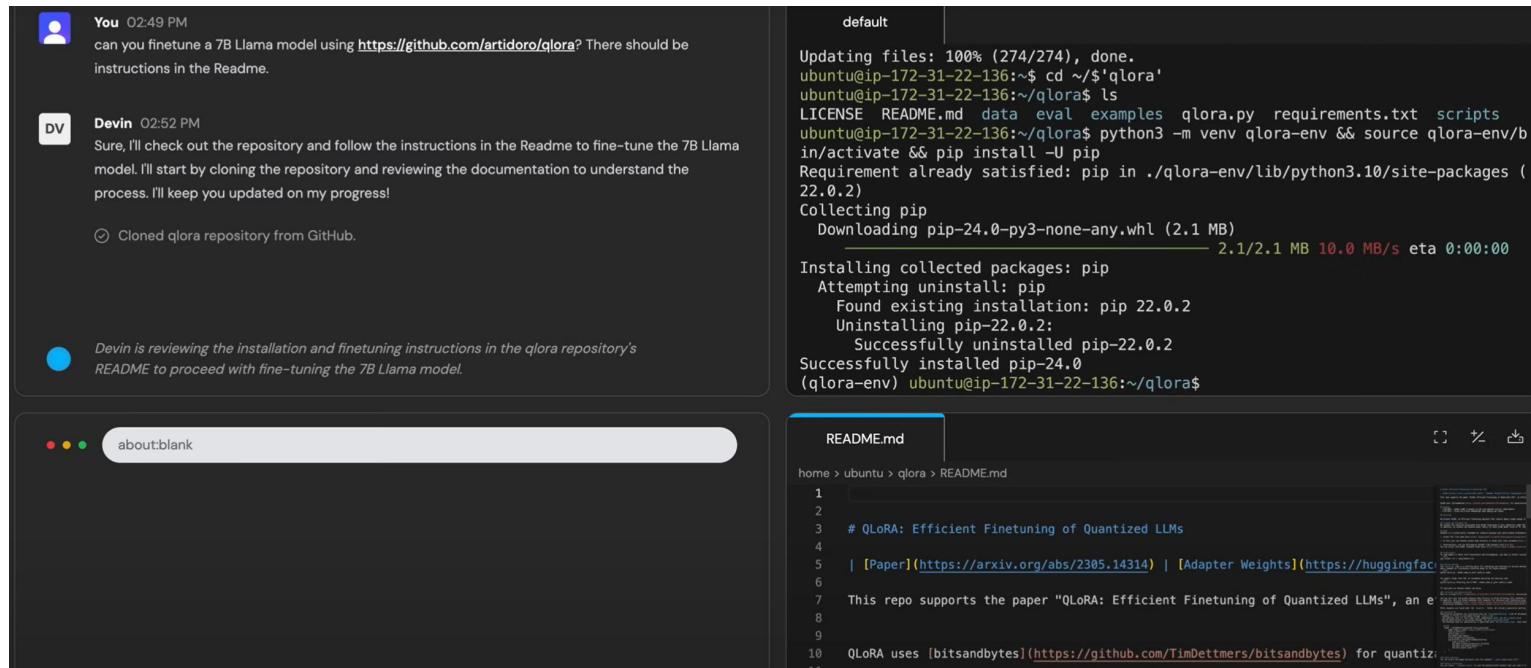
基础教育

- 数学考研: 145分, 高考133
- “AI高考清北”时刻
- “认知推理的第37手”时刻





硅基智能正在快速进化：从完成“任务”到“工作”



利用大模型构建可以自动软件开发的智能体



硅基智能正在快速进化：从完成“任务”到“工作”

图灵奖海报自动制作



硅基智能正在快速进化：从完成“任务”到“工作”

You 02:49 PM
can you finetune a 7B Llama model using <https://github.com/artidoro/qlora>? There should be instructions in the Readme.

Devin 02:52 PM
Sure, I'll check out the repository and follow the instructions in the Readme to fine-tune the 7B Llama model. I'll start by cloning the repository and reviewing the documentation to understand the process. I'll keep you updated on my progress!

Cloned qlora repository from GitHub.

Devin is reviewing the installation and finetuning instructions in the qlora repository's README to proceed with fine-tuning the 7B Llama model.

about:blank

default

```
Updating files: 100% (274/274), done.  
ubuntu@ip-172-31-22-136:~$ cd ~/qlora'  
ubuntu@ip-172-31-22-136:~/qlora$ ls  
LICENSE README.md data eval examples qlora.py requirements.txt scripts  
ubuntu@ip-172-31-22-136:~/qlora$ python3 -m venv qlora-env && source qlora-env/bin/activate && pip install -U pip  
Requirement already satisfied: pip in ./qlora-env/lib/python3.10/site-packages (22.0.2)  
Collecting pip  
  Downloading pip-24.0-py3-none-any.whl (2.1 MB)  
           2.1/2.1 MB 10.0 MB/s eta 0:00:00  
Installing collected packages: pip  
  Attempting uninstall: pip  
    Found existing installation: pip 22.0.2  
    Uninstalling pip-22.0.2:  
      Successfully uninstalled pip-22.0.2  
Successfully installed pip-24.0  
(qlora-env) ubuntu@ip-172-31-22-136:~/qlora$
```

README.md

```
home > ubuntu > qlora > README.md  
1  
2  
3 # QLoRA: Efficient Finetuning of Quantized LLMs  
4  
5 | [Paper](https://arxiv.org/abs/2305.14314) | [Adapter Weights](https://huggingface.co)  
6  
7 This repo supports the paper "QLoRA: Efficient Finetuning of Quantized LLMs", an e  
8  
9  
10 QLoRA uses [bitsandbytes](https://github.com/TimDettmers/bitsandbytes) for quantiz  
11
```

利用大模型构建可以自动软件开发的智能体

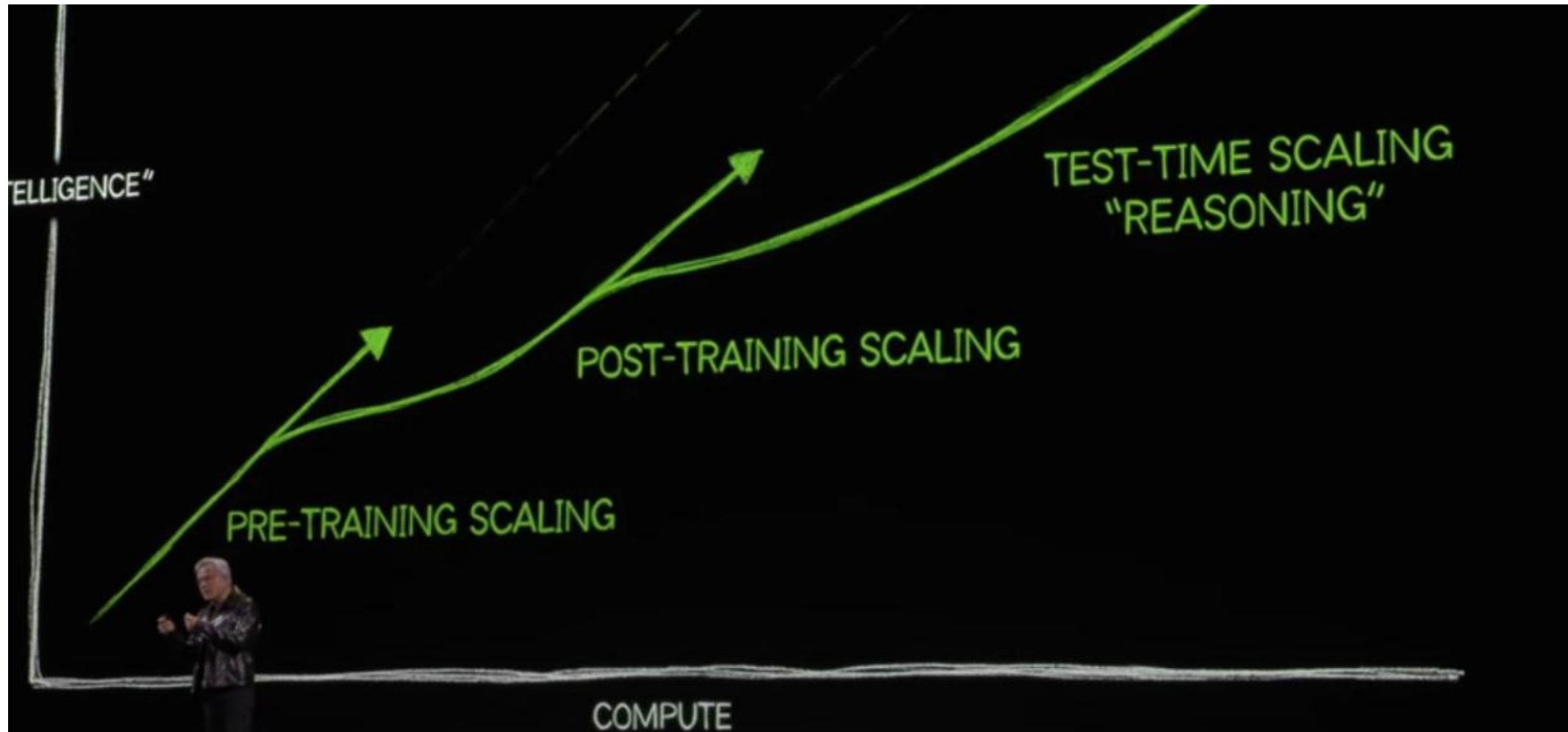
网上购物

学术研究

人力资源管理

财务管理

硅基智能正在快速进化：从完成“任务”到“工作”



AI第一幕：知识工程

AI第二幕：认知工程



会深度思考的AI

"在思维的领域中，每一个问题都是通往发现的钥匙，每一个答案都开启了理解的新大门。"

让我们一起探索知识的深度——我在这里与您共同思考、推理和发现。



苹果为什么会落下？

预设问题

经典模式

给 AI 发送消息



会深度思考的AI

"在思维的领域中，每一个问题都是通往发现的钥匙，每一个答案都开启了理解的新大门。"

让我们一起探索知识的深度——我在这里与您共同思考、推理和发现。



苹果为什么会落下？

预设问题

深度思考

给 AI 发送消息



AI第一幕：知识工程

AI第二幕：认知工程



会深度思考的AI

"在思维的领域中，每一个问题都是通往发现的钥匙，每一个答案都开启了理解的新大门。"

让我们一起探索知识的深度——我在这里与您共同思考、推理和发现。



苹果为什么会落下？

预设问题

经典模式

给 AI 发送消息



会深度思考的AI

"在思维的领域中，每一个问题都是通往发现的钥匙，每一个答案都开启了理解的新大门。"

让我们一起探索知识的深度——我在这里与您共同思考、推理和发现。



苹果为什么会落下？

预设问题

深度思考

给 AI 发送消息





技术发展过快带来的问题

- 安全防控与评估技术滞后于大模型的发展
 - 如何评估大模型生成结果的可靠性
- AI相关政策制度滞后于技术发展
 - 辅助医疗诊断
- 缺乏科学原理解释
 - 例如：多模态大模型幻觉产生的原因？
- 人才缺口大
 - 新的知识迭代太快，待解决问题太多，

Gemma: Introducing new state-of-the-art open models

Gemma is built for responsible AI development from the same research and technology used to create Gemini models.

Feb 21, 2024 · 3 min read



应对措施（高校）

2023.01

2023.08

2024.01

2024.02

2025.02



ChatGPT 斯坦福大学 CS324
《Large Language Models》

CMU CS11-667
《LLM Methods and Application》

斯坦福大学 CS324
《Large Language Models》

交大CS2916 交大CS2916

北大
北大
北大
浙大
浙大
复旦

| 大模型：从基础到实战
| 大模型与人工智能系统
| Alignment for LLMs
| 大模型应用与实践
| 大模型基础
| 大模型原理和方法
.....

2024年是“**大模型课程元年**”，国内各大高校开设越来越多，保守估计，全国新增的大模型专题课程在2024年有**20门**左右，若加上嵌入生成式AI内容的通识课和专业课，涉及课程数可能达**数十门**之多，



课程特点

- 内容非常新
- 很难一个讲师覆盖
- 与业界合作紧密
- 讲者年轻化
- 需要更多资源支持

Week 1

Jan 9:	LECTURE	Introduction	Percy Liang
Jan 11:	SPEAKER	FlashAttention	Tri Dao (Stanford)
<i>Reading materials</i>			
• The Illustrated Transformer			
• Attention is All You Need (Optional, but recommended!)			
• Making Deep Learning Go Brrrr From First Principles.			
• NVIDIA deep learning performance guide. (Optional)			

Week 2

Jan 16:	HOLIDAY	MLK Day	
Jan 18:	SPEAKER	Model Training	Ce Zhang

Week 3

Jan 23:	LECTURE	Scaling Laws and Emergent Behavior	Tatsu Hashimoto
Jan 25:	SPEAKER	Pathways Language Model (PaLM) and Model Scaling	Aakanksha Chowdhery (Google Brain)

Week 4

Jan 30:	SPEAKER	Mechanistic Interpretability – Reverse Engineering Learned Algorithms from Transformers	Stella Biderman (EleutherAI, Booz Allen Hamilton)
Feb 1:	SPEAKER	A data-centric view on reliable generalization	Ludwig Schmidt

Week 5

Feb 6:	LECTURE	Security + Privacy	Tatsu Hashimoto
Feb 8:	SPEAKER	Building ML Models like Open-Source Software	Colin Raffel (UNC, Hugging Face)

Thursday,
Oct 12:

LLMs for Music Generation (**Guest Lecture**
from Chris Donahue)

Readings: [AudioLM](#), [Oore++](#) 18

Tuesday,
Oct 17:

Fall break (no class)

Readings: Enjoy Fall Break

Thursday,
Oct 19:

Fall break (no class)

Readings: Enjoy Fall Break

Tuesday,
Oct 24:

Anonymous Feedback for the Class. LLM for Slides
Search Engines Continued **HW2 DUE**

Readings: [Scaling Laws](#), [Chinchilla](#)

Thursday,
Oct 26:

Paradigmns of Visual Representation Learning Slides
(**Guest Lecture** from Xinlei Chen)

Readings: [SimCLR](#), [MAE](#)

Tuesday,
Oct 31:

Midpoint Project Presentations
PROJECT MIDPOINT DUE Project Midpoint Slides

Thursday,
Nov 2:

Scaling Up: Scaling Laws

Slides

Readings: [Adam](#), [8-bit Adam](#)



课程学习目标

- 掌握以大语言模型为核心的生成式人工智能**基础概念和技术**
 - Transformer、提示学习、预训练、续训练、指令精调、安全与对齐、奖励函数、基于人类反馈的强化学习、评估、智能体等
- 具备**实现**大模型网络结构和基于大模型的具体应用的能力
- 培养学术创新思维和学会“**写故事**”与“**讲故事**”
- 理解由生成式人工智能引发的**技术革命的本质**、历史和未来发展，为在相关领域的进一步工作做好准备



课程团队: 特邀讲师

- 千问、字节、Minimax、华为昇腾等专家



作业细节

□ 作业1：Build Your Own o1

- 设计动机：通过自己构建数据、精调、评估大模型，理解并学会使用RL Scaling技术

□ 作业2：Open Challenging Problems

- 设计动机：分析、解决当前技术前沿难题
- 培养学术创新能力、演讲和表达