



# CMF

China Macroeconomy Forum

中國宏觀經濟論壇

CMF中国宏观经济专题报告（第73期）

## 大模型：原理、进展及其影响

主办单位：

中国人民大学国家发展与战略研究院

中国人民大学经济学院

中诚信国际信用评级有限责任公司

承办单位：

中国人民大学经济研究所

2023年9月



# 大模型：原理、进展及其影响

主办单位：中国人民大学国家发展与战略研究院、中国人民大学经济学院、中诚信国际信用评级有限公司  
承办单位：中国人民大学经济研究所



# 大模型：原理、进展及其影响

**报告人：文继荣**

中国人民大学信息学院院长

中国人民大学高瓴人工智能学院执行院长

国家高层次人才

2023年9月


# 目 录

一、大模型的背景和原理

二、大模型的飞速发展及趋势

三、大模型的深刻影响

# ChatGPT开启了大模型浪潮

-  OpenAI 公司开发的大型语言模型
  - 通过从大规模语料库中学习语言规律，从而生成与人类语言相似的输出
- 表现为一个对话机器人，能够理解自然语言，进行高质量的多轮对话
  - 拟人化程度惊人，被认为是**人工智能里程碑式的突破**



比尔盖茨: ChatGPT的重要性不亚于互联网的发明

2022年11月30日发布，5天获得100万用户，今年1月份月活跃用户数达1亿



# ChatGPT的特点

## 知识广博

- 上知天文、下知地理，不仅能写文章、还能写代码

## 有条有理

- 能有逻辑地组合知识形成答案，具有一定的推理能力

## 善解人意


- 能准确理解用户意图，三观比较正，并且承认自己不知道

## 交互能力强


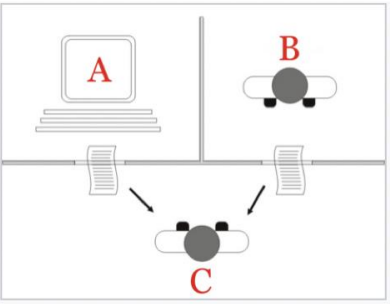
- 在多轮对话中不走神，交互聚焦主题

# 理解人类语言是人类探索文明发展的重要目标

哲学

<p>Ludwig Wittgenstein</p> 	<p><i>"The limits of my language mean the limits of my world."</i> —Ludwig Wittgenstein</p>
--	---

计算机

<p>Alan Turing</p> 	<p>MIND A QUARTERLY REVIEW OF PSYCHOLOGY AND PHILOSOPHY I.—COMPUTING MACHINERY AND INTELLIGENCE By A. M. TURING</p> 
---	--

# 世界与语言

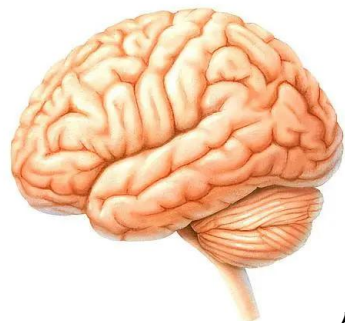
**假设：** 世界知识和人类认知能力蕴含在人类语言中

**问题：**

- 可以对人类语言建模吗？
- 语言模型是否能作为世界知识模型？
- 语言模型是否具有人类认知能力？

语言模型是对人脑的反向工程

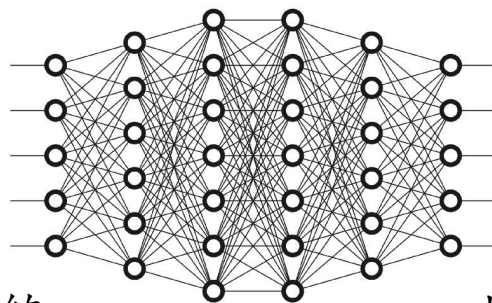
人脑（神经网络）



生成

高瓴人工智能学院是中国人民大学下属学院.....

语言模型（人工神经网络）



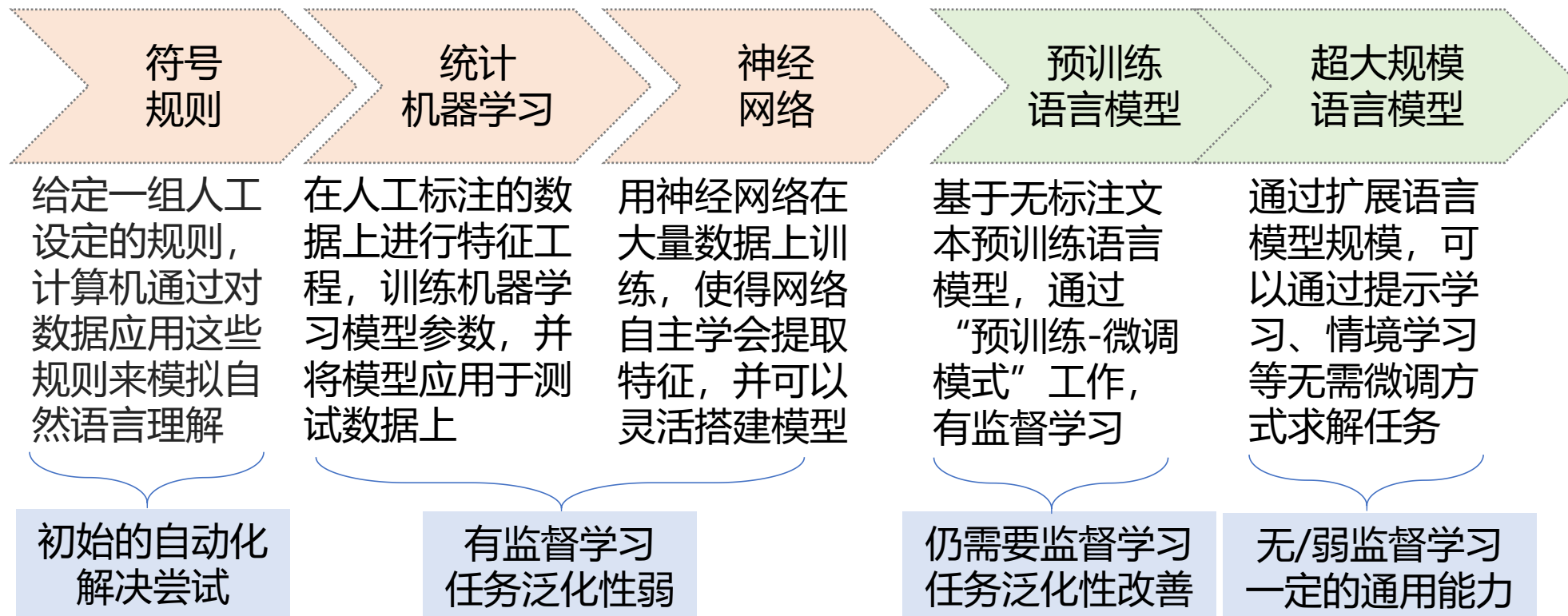
训练

生成

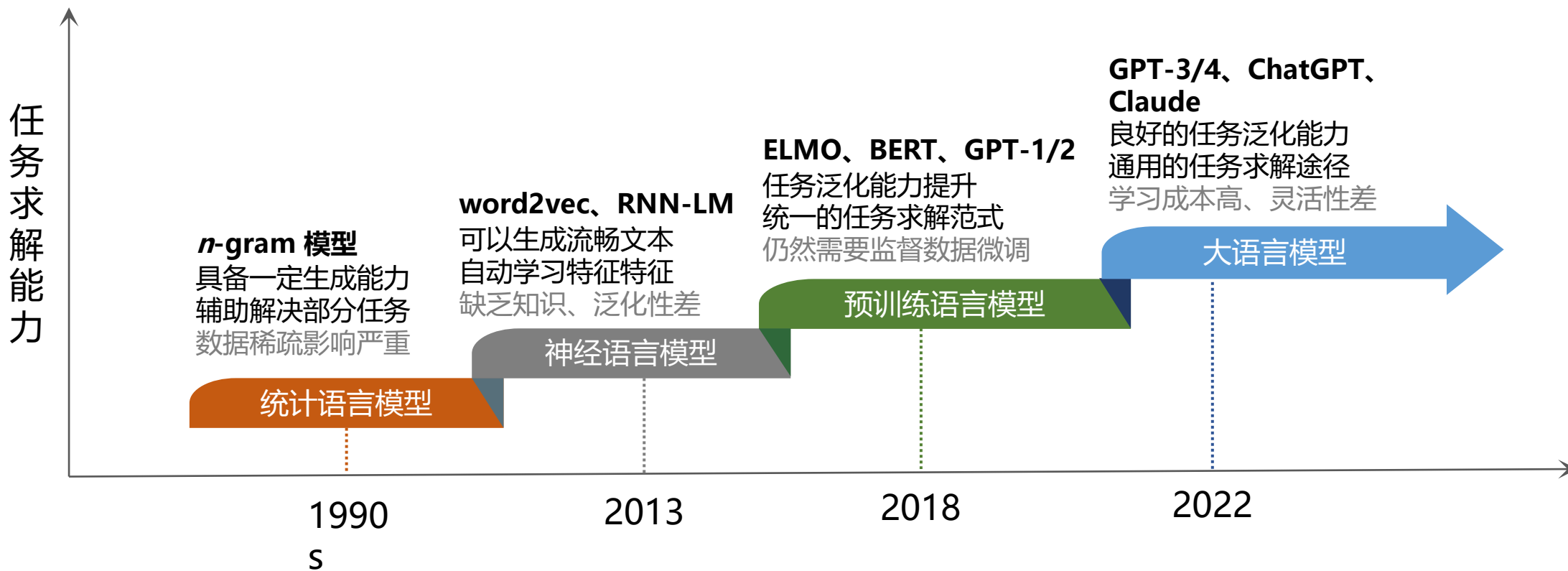
高瓴人工智能学院是中国人民大学下属学院.....



# 语言智能的探索历程



# 语言模型的演进图



# 统计语言模型 (Statistical Language Model)

Example Probability

The cat **sat** on the mat **0.95**

The cat **sad** on the mat **0.20**

The cat

-----  
**High** wind tonight **0.97**

**Large** wind tonight **0.31**

## 困难

- 组合爆炸，无法考虑较宽的上下文
- 泛化性差，字词之间没有关联
- 数据、算力不够

# 生成式语言模型 (Generative Language Model)

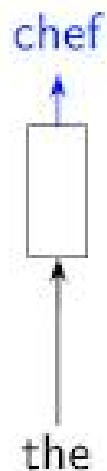
## GPT

- 下一个词预测：文字接龙
  - The chef cooked the meal.
  - The chef cooked the ████████

## 突破

- 引入注意力机制解决长程上下文依赖问题
- 多层神经网络抽象解决泛化问题
- 海量数据训练、巨大算力支持

## Language Modeling



# 掩码语言模型 (Masked Language Model)

## BERT

## Masked Language Modeling

- 文本Mask: 完形填空
  - The chef cooked the meal.
  - chef  the meal.

the chef cooked the meal

# ChatGPT的发展路径



Word2Vec



Transformer



BERT



Codex

引入代码能力  
和思维链能力



WebGPT

引入搜索能力



InstructGPT

能理解人类指令



ChatGPT

引入对话能力



OpenAI

GPT

预训练语言  
模型元年



OpenAI

GPT-2



OpenAI

GPT-3

首个千亿级模型

2013年

2017年

2018年

2019年

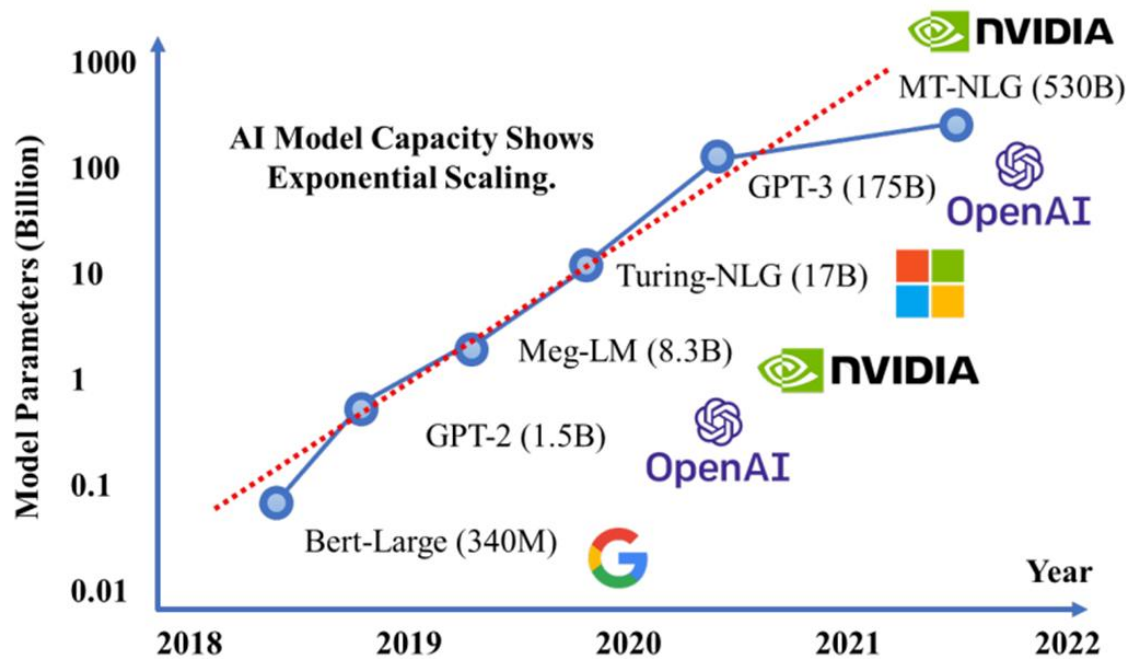
2020年

2021年

2022年

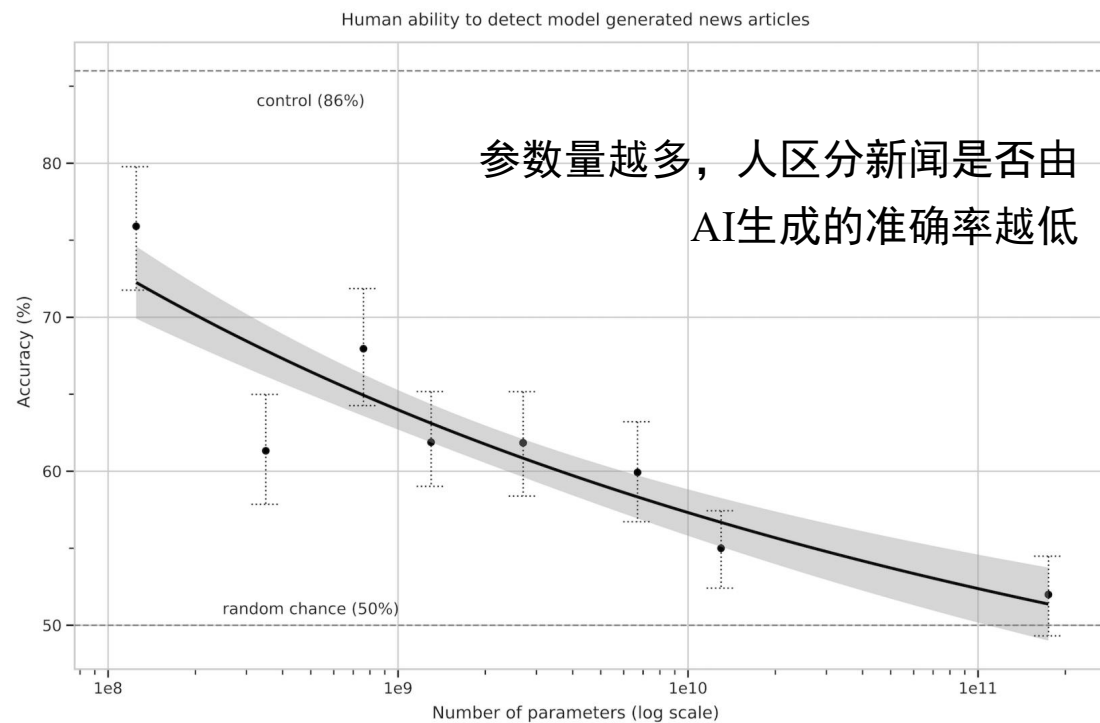
# ChatGPT的法宝之一：大！（自主学习）

- ChatGPT：大力出奇迹的典范，大数据、大模型、大算力
- GPT-3：1750亿参数量，训练一次消耗1200万美元



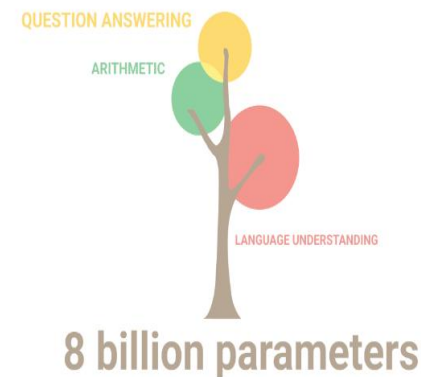
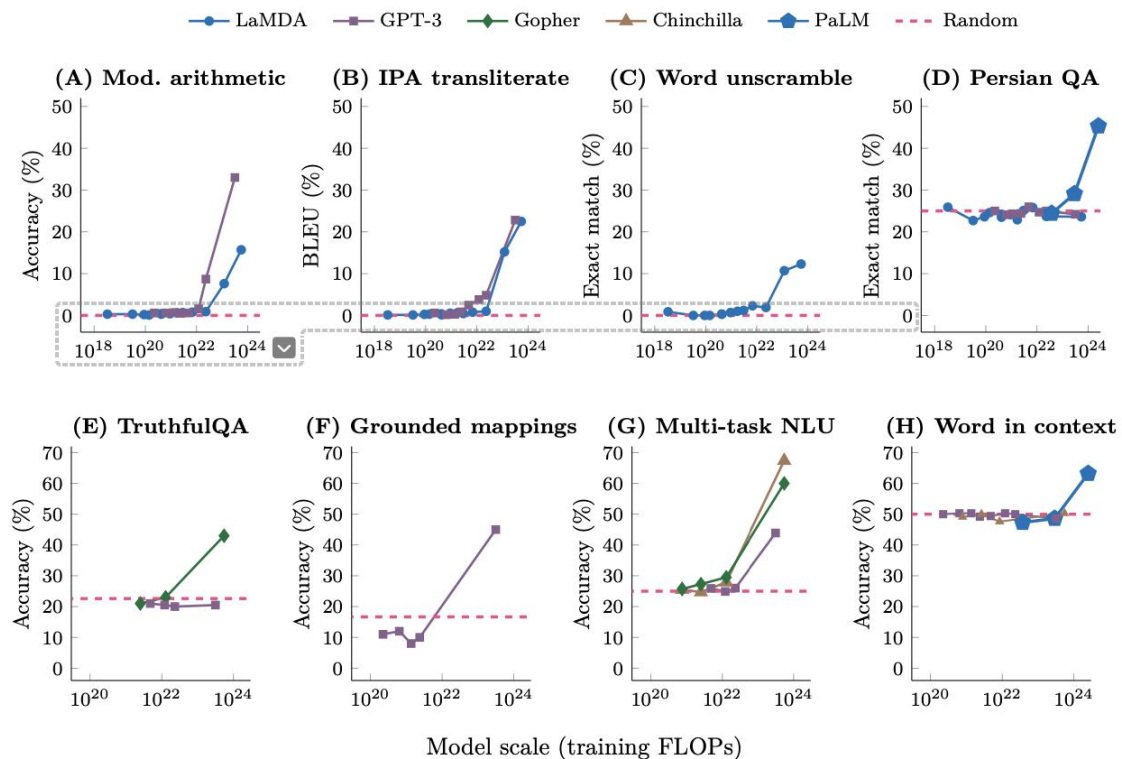
(b) Model Capacity Scaling

CSDN @刘炫320



# ChatGPT的法宝之一：大！（自主学习）

**涌现能力**：书读百遍，其义自见





# ChatGPT的法宝之二：思维链（逻辑训练）

- 思维链 CoT (Chain of Thought) : 由Google于2022年提出
  - 大模型可能难以直接解决一个困难的大问题。
  - 可以在提问时加入引导，将大问题拆分为多个简单的小问题，从而得到解答

## (a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. **X**

## (b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are  $16 / 2 = 8$  golf balls. Half of the golf balls are blue. So there are  $8 / 2 = 4$  blue golf balls. The answer is 4. **✓**

# ChatGPT的法宝之二：思维链（逻辑训练）

- 思维链 CoT (Chain of Thought) : 由Google于2022年提出
  - 或者只给一个**提示**，就可以对大模型进行引导，提示其逐步地解决问题
  - Let' s think step by step.

No.	Category	Template	Accuracy
1	instructive	<b>Let's think step by step.</b>	<b>78.7</b>
2		First, (*1)	77.3
3		Let's think about this logically.	74.5
4		Let's solve this problem by splitting it into steps. (*2)	72.2
5		Let's be realistic and think step by step.	70.8
6		Let's think like a detective step by step.	70.3
7		Let's think	57.5
8		Before we dive into the answer,	55.7
9		The answer is after the proof.	45.7
10	misleading	Don't think. Just feel.	18.8
11		Let's think step by step but reach an incorrect answer.	18.7
12		Let's count the number of "a" in the question.	16.7
13		By using the fact that the earth is round,	9.3
14	irrelevant	By the way, I found a good restaurant nearby.	17.5
15		AbraKadabra!	15.5
16		It's a beautiful day.	13.1
-		(Zero-shot)	17.7

## (d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **Let's think step by step.**

(Output) *There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓*

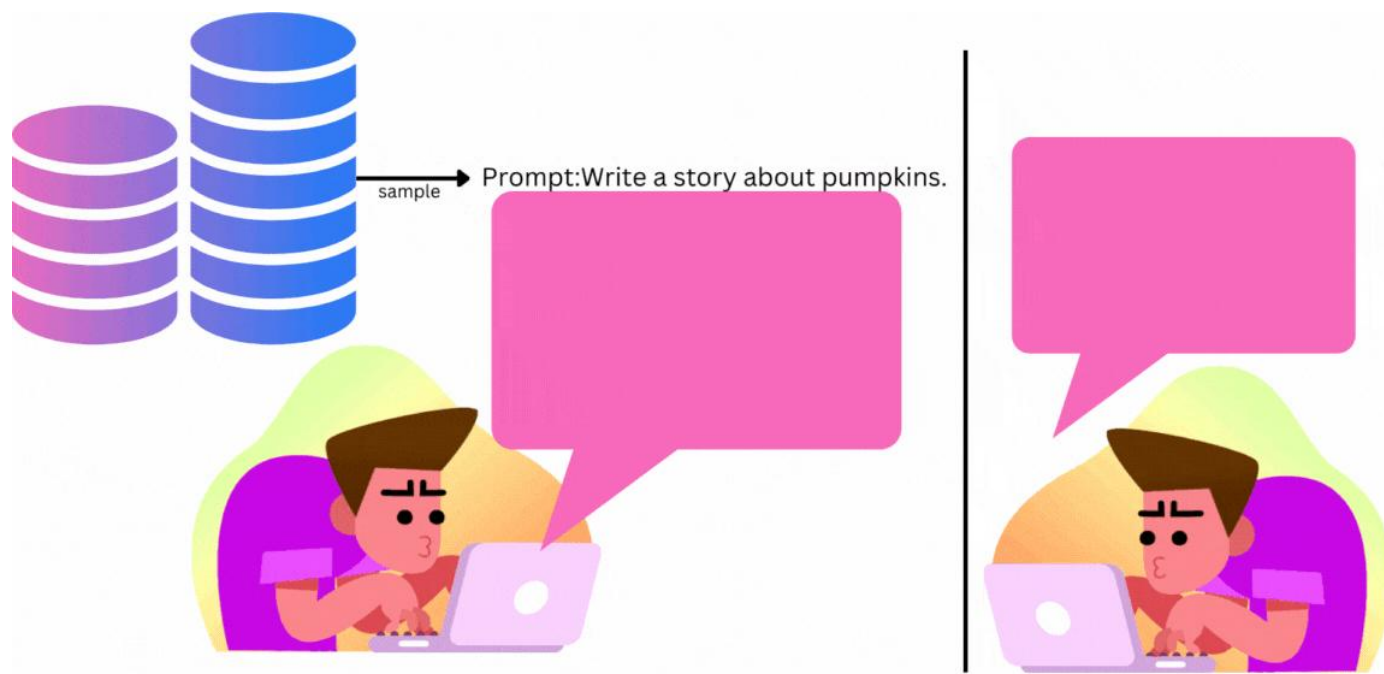


**Codex**

引入代码能力  
和思维链能力

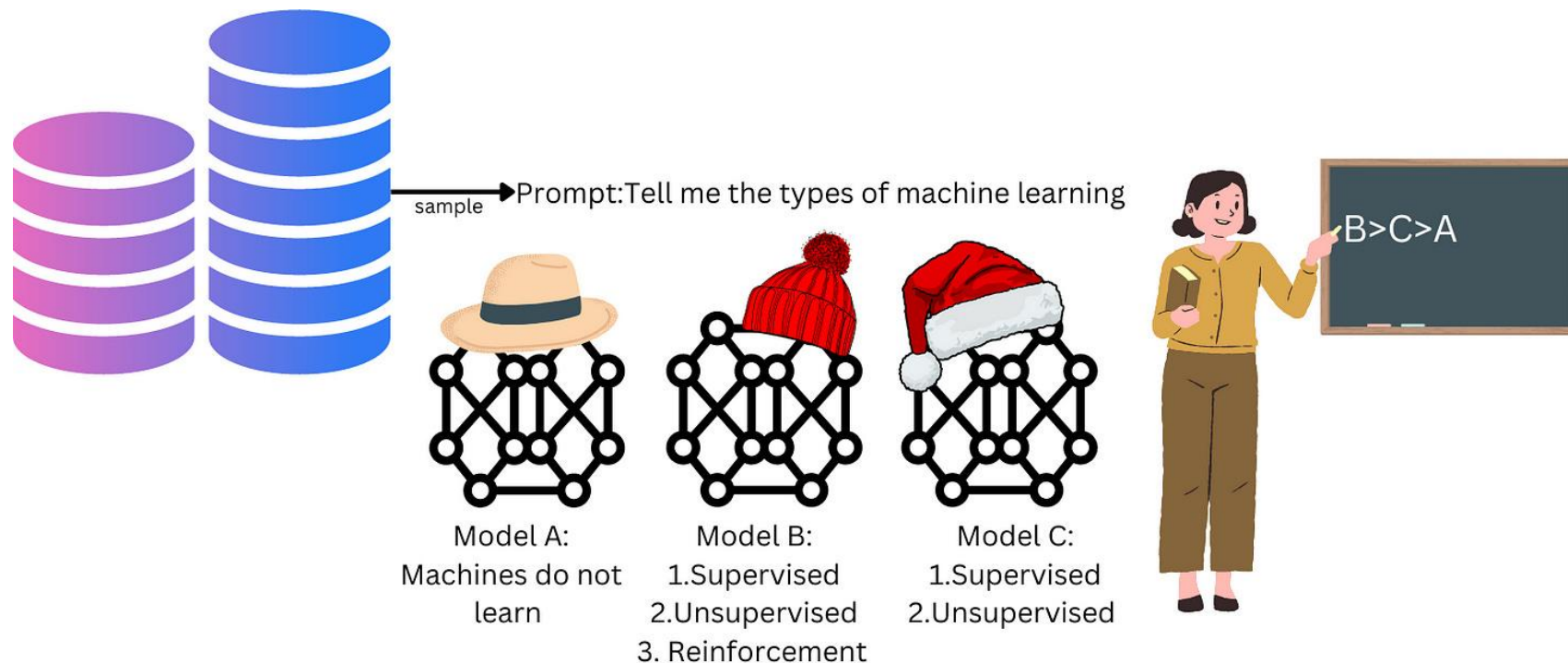
# ChatGPT的法宝之三：人在回路强化学习（价值观对齐）

- RLHF (Reinforcement Learning with Human Feedback) :
  - 收集人类真实指令，并且聘用了专门的合同工写回答用于初始训练



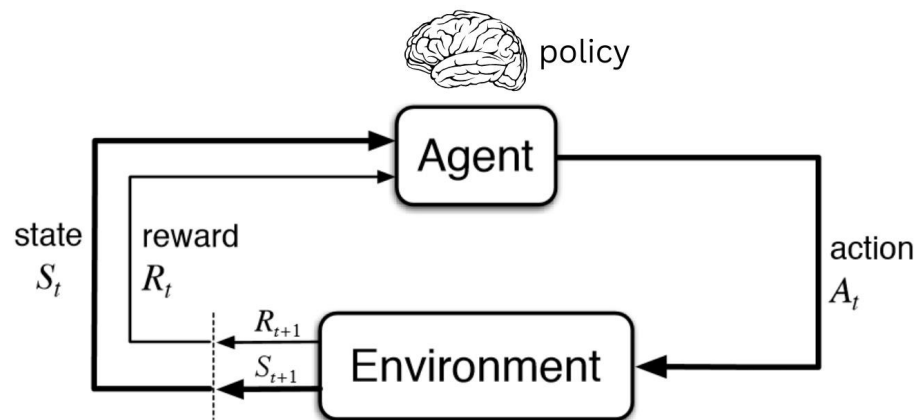
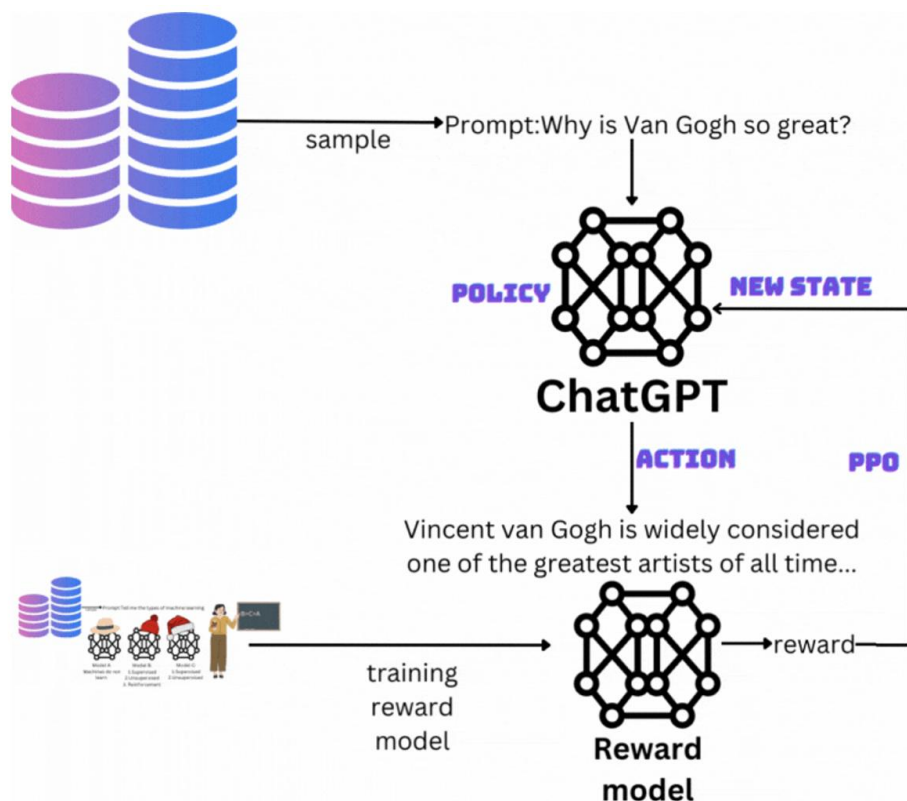
# ChatGPT的法宝之三：人在回路强化学习（价值观对齐）

- RLHF (Reinforcement Learning with Human Feedback) :
  - 使用上述模型得到多个预测结果，请人按照回答质量排序，训练出一个打分模型



# ChatGPT的法宝之三：人在回路强化学习（价值观对齐）

- RLHF（Reinforcement Learning with Human Feedback）：
  - 进一步加速训练过程：使用上述训练模型生成结果，上述打分模型自动评估训练效果。



- 传统强化学习：Agent根据反馈(rewards)选择策略
- RLHF：
  - 预测模型扮演Agent，进行策略选择
  - 评分模型提供rewards

# ChatGPT的法宝之四：数据闭环+系统工程

OpenAI 在2015年作为一个非盈利实验室运营，但为了吸引外部资金，已在2019年转向了有限利润 (Capped-profit) 模式

商业化

OpenAI开放API

微软买断License

GPT-3支撑300+应用

上线Azure云服务

codex支撑70+应用

2020

2021

2022

2023

研究

GPT-3

Codex

WebGPT

InstructGPT

ChatGPT

持续训练：持续收集和构建高质量数据集训练模型

快速迭代：先用起来，数据回流，改进模型

端到端的系统工程：数据收集、标注、训练、推理

- 使用垂域数据或人类标注反馈数据持续训练，增强模型能力。
- API收集数据：GPT-3通过API向个人和企业提供服务，基于Playground和商用场景收集用户真实使用数据(2.5年数据)。
  - 寻找合作伙伴获取训练数据：面向对话场景，和Twitter合作每天有大量人与人交互的数据产生，从而可以更好模拟人类说话的方式。面向开发者场景，和微软合作基于Github 5400万开源项目，获取包括159GB Python 代码作为训练数据。
  - 自建高质量finetune 数据：面向问答场景，严格筛选40位高质量数据标员标注共约5万条prompt数据。

收集用户反馈，快速分析迭代：ChatGPT 开放近一周，最初一些问题（例如常识问题，安全问题）均快速闭环。

OpenAI CEO Sam Altman 说在OpenAI取得关注的背后，微软（尤其是Azure）做了大量出色的工作。

问：10公斤铁和10公后棉在哪个更重？  
12月2号的回答：是铁更重  
12月8号的回答：是一样重，但是棉花有可能吸收空气中的水分从而更重一些，不过重量的差异其实很小

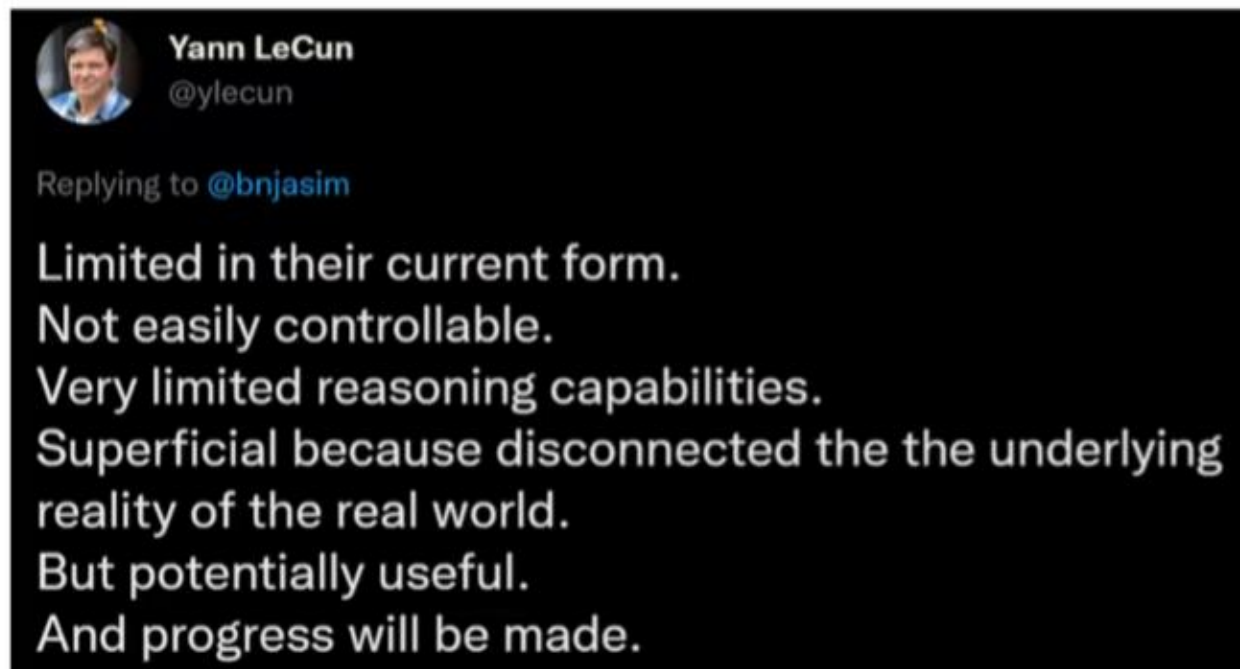
数据收集和标注工作是复杂的系统工程：1) 严格选择数据标注员，确保其多样性。2) 对标注员进行培训，确保和项目目标对齐，3) 开发web 标注页面，确保从多维度反馈信息（不仅是“好与不好”）

ChatGPT优秀体验的原因：1) 依托强大的基础模型能力，是长期技术积累的结果；2) 重视数据：持续收集和构建高质量数据集（含人工标注）对模型持续训练和优化；3) 快速迭代：快速触达用户和商业场景，收集真实业务数据和用户反馈，分析理解问题，积累经验，迭代产品；4) 系统工程：不仅是训练模型，是端到端复杂的系统工程；5) 细节决定成败：以上工作不仅很系统，而且很细致到位

# ChatGPT的不足

- 无法实时纳入新知识
- 特别专业的知识还不足
- 推理计算能力仍不足，可能会一本正经地胡说八道
- 只支持文本生成
- 算力消耗巨大
- .....

但是这些问题大都是工程问题，不存在不可逾越的理论障碍



# 目 录

一、大模型的背景和原理

二、大模型的飞速发展及趋势

三、大模型的深刻影响



# 大模型仍在飞速发展演进

---

增强实时性和真实性

---

支持多模态

---

扩展知识和技能

---

连接物理世界

---

改进复杂推理

---

自主智能体

---

支持个性化

---

提高训练和推理效率

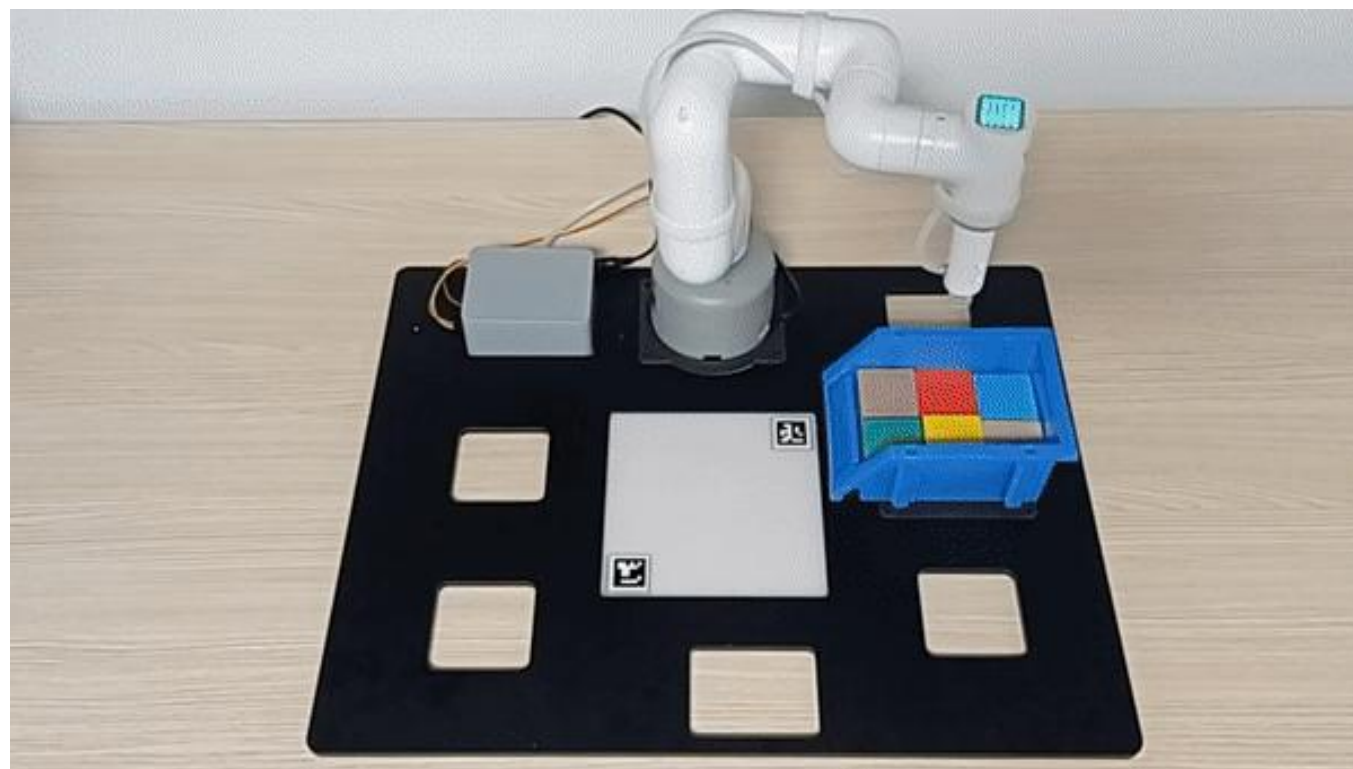
## 扩展知识和技能

- *Toolformer: Language Models Can Teach Themselves to Use Tools. Meta AI Research. Feb 2023*
- 在生成文本的过程中，遇到特定的任务，Toolformer会直接调用所需工具的API
  - 比如说，在执行这个任务：1400名参与者，有400人通过了测试，占多大比例？Toolformer直接“掏出”计算器，现场计算得出结果：29%。

Out of 1400 partic

# 连接物理世界

- *ChatGPT for Robotics: Design Principles and Model Abilities, Microsoft Research. Feb 2023*
- 用ChatGPT的语言交互能力操纵机器人



# 2023年3月15日：GPT-4发布！

## 令人震撼的新能力

- 图片识别理解
- 更强的推理和更高的准确性
- 文字输入限制提升至 2.5 万字
- .....

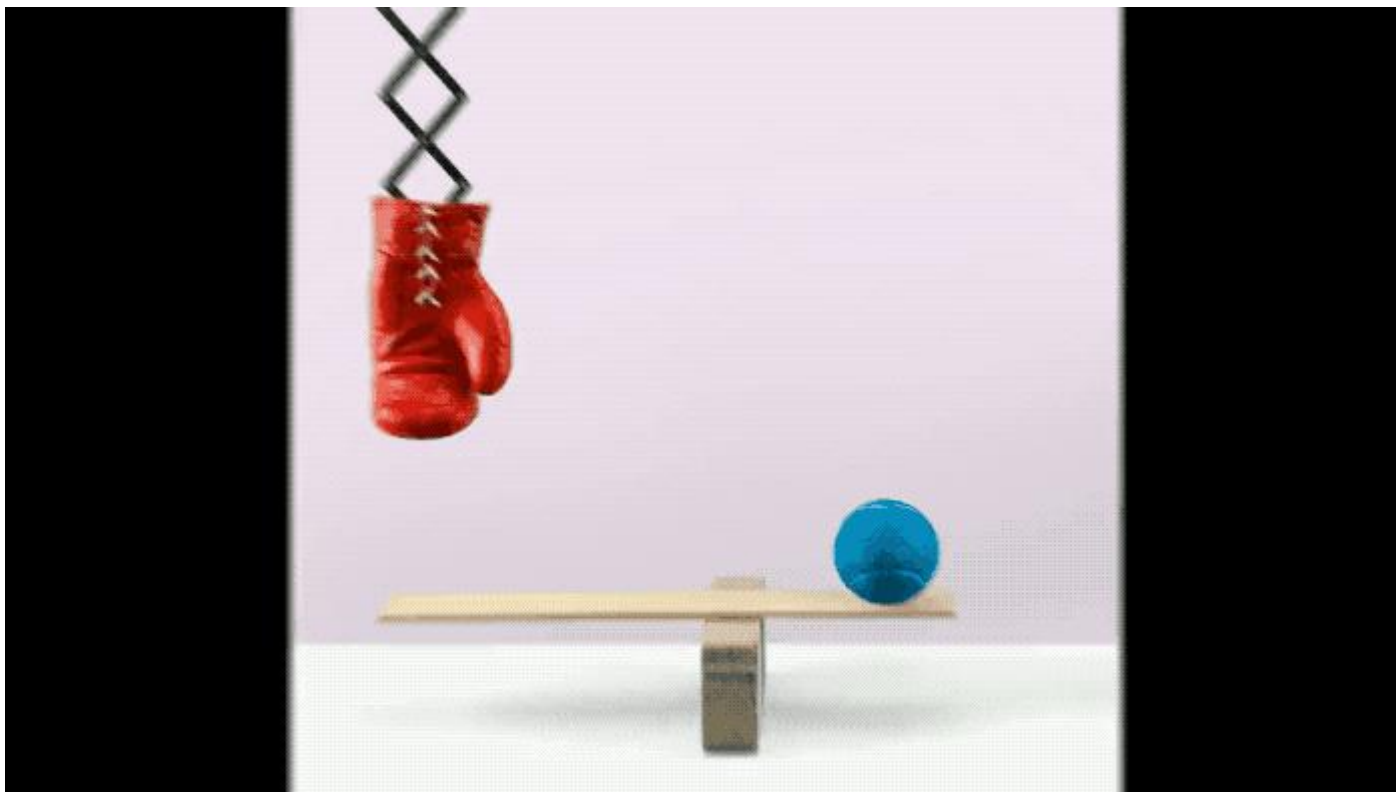


it can take in and generate up to

## 多模态支持 - 看图推理

问：(看图) 手套掉下去会怎样？

答：它会掉到木板上，并且球会被弹飞。



# 逻辑性和正确性 – 参加考试

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT) <sup>1</sup>	298 / 400 -90th	298 / 400 -90th	213 / 400 -10th
LSAT	163 -88th	161 -83rd	149 -40th
SAT Evidence-Based Reading & Writing	710 / 800 -93rd	710 / 800 -93rd	670 / 800 -87th
SAT Math	700 / 800 -89th	690 / 800 -89th	590 / 800 -70th
Graduate Record Examination (GRE) Quantitative	163 / 170 -80th	157 / 170 -62nd	147 / 170 -25th
Graduate Record Examination (GRE) Verbal	169 / 170 -99th	165 / 170 -96th	154 / 170 -63rd
Graduate Record Examination (GRE) Writing	4 / 6 -54th	4 / 6 -54th	4 / 6 -54th
USABO Semifinal Exam 2020	87 / 150 99th - 100th	87 / 150 99th - 100th	43 / 150 31st - 33rd
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th - 100th	5 86th - 100th	5 86th - 100th
AP Biology	5 85th - 100th	5 85th - 100th	4 62nd - 85th
AP Calculus BC	4 43rd - 59th	4 43rd - 59th	1 0th - 7th

# 逻辑性和正确性 – 做物理题

GPT-4解巴黎综合理工学院的一道物理题，题目是法语，解答用英语。

## I. Principe de la détection de rayonnement avec un bolomètre

Comme illustré sur la figure 1 un bolomètre est constitué d'un absorbeur qui reçoit le rayonnement que l'on désire détecter. Sa température  $T$ , supposée uniforme, est mesurée à l'aide d'un thermomètre incorporé, constitué d'un matériau conducteur dont la résistance  $R(T)$  varie avec la température  $T$ ; cette variation est caractérisée par le coefficient  $\alpha = \frac{1}{R} \frac{dR}{dT}$ . L'ensemble possède la capacité thermique  $C_{th}$ .

Un barreau, conducteur thermique, homogène, de longueur  $L$ , de section  $S$  et de conductivité thermique  $\lambda$  et sans échanges thermiques latéraux, relie le bolomètre à un thermostat de température  $T_b$  fixe.

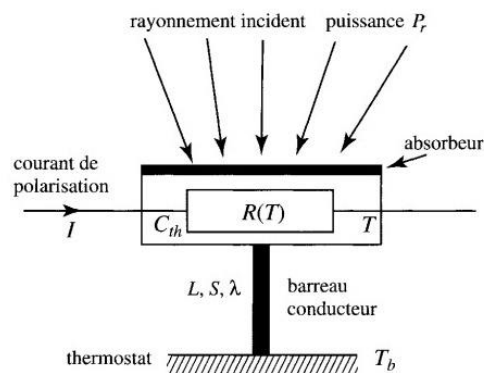


Figure 1 - Schéma d'un bolomètre

I.1. Dans cette première partie, on étudie le comportement du bolomètre en régime stationnaire, à la température de fonctionnement  $T = T_0$ .

I.1.a) Exprimer la température  $T(x)$  en tout point du barreau conducteur repéré par son abscisse  $x$ ; on prendra  $T(0) = T_0$  et  $T(L) = T_b$ .

GPT-4 To answer question I.1.a, we need to express the temperature  $T(x)$  in every point of the conductive bar, which is characterized by the abscissa  $x$ . We are given that  $T(0) = T_0$  and  $T(L) = T_b$ .

For a stationary regime, the heat equation for the one-dimensional heat conduction in the bar is:  
 $d^2T/dx^2 = 0$

Integrating once, we get:  
 $dT/dx = C_1$

Integrating a second time:  
 $T(x) = C_1x + C_2$

To determine the constants  $C_1$  and  $C_2$ , we use the boundary conditions:  
 $T(0) = T_0$ , which gives  $C_2 = T_0$   
 $T(L) = T_b$ , which gives  $C_1 = (T_b - T_0) / L$

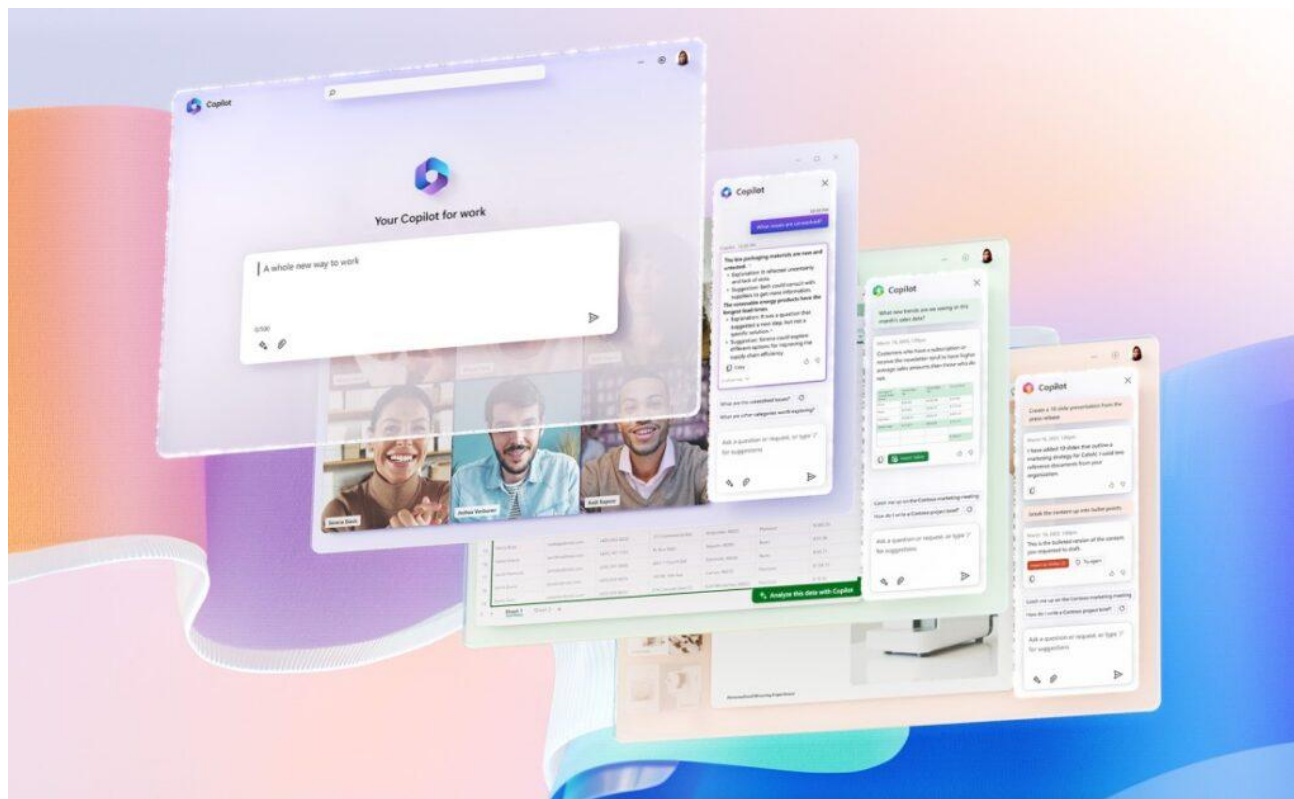
Finally, the expression for  $T(x)$  is:

$$T(x) = (T_b - T_0) * (x / L) + T_0$$

This represents a linear temperature profile along the conductive bar, with a slope of  $(T_b - T_0) / L$ .

# 2023年3月16日： GPT-4全面接入Office

- 3月16日，微软正式宣布推出 Microsoft 365 Copilot，将GPT-4全面接入Office。
- CEO纳德拉在发布会上称：今天，进入人机交互的新时代，重新发明生产力。





# GPT-4接入Excel

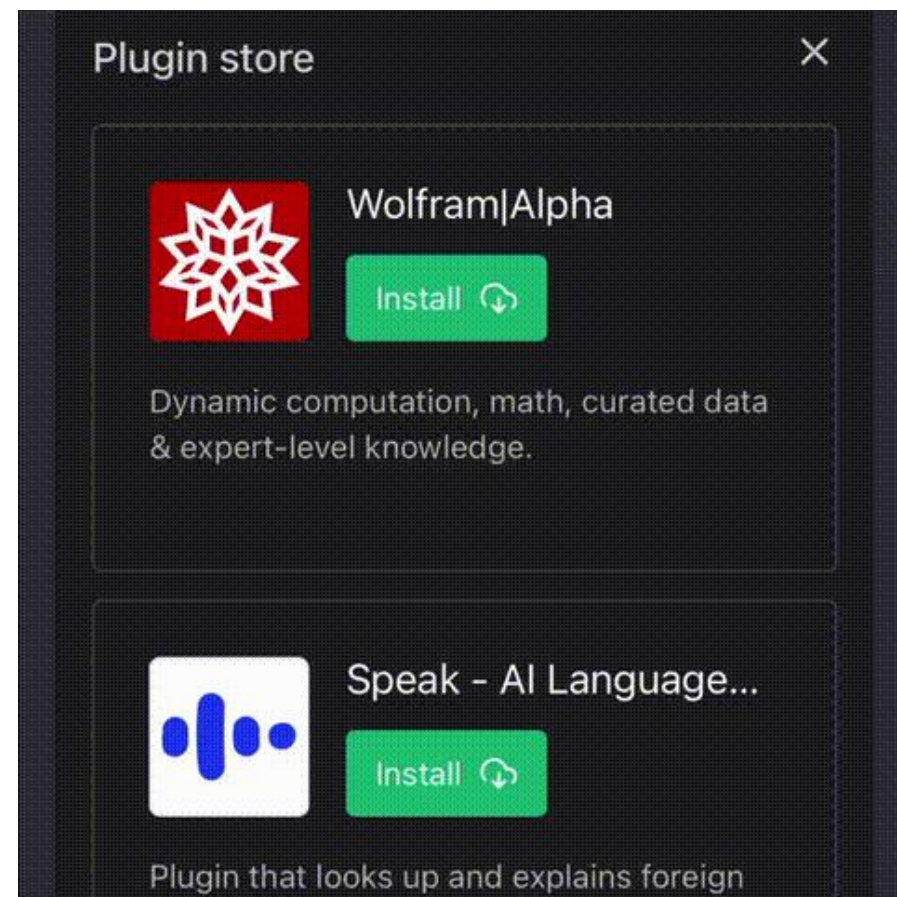
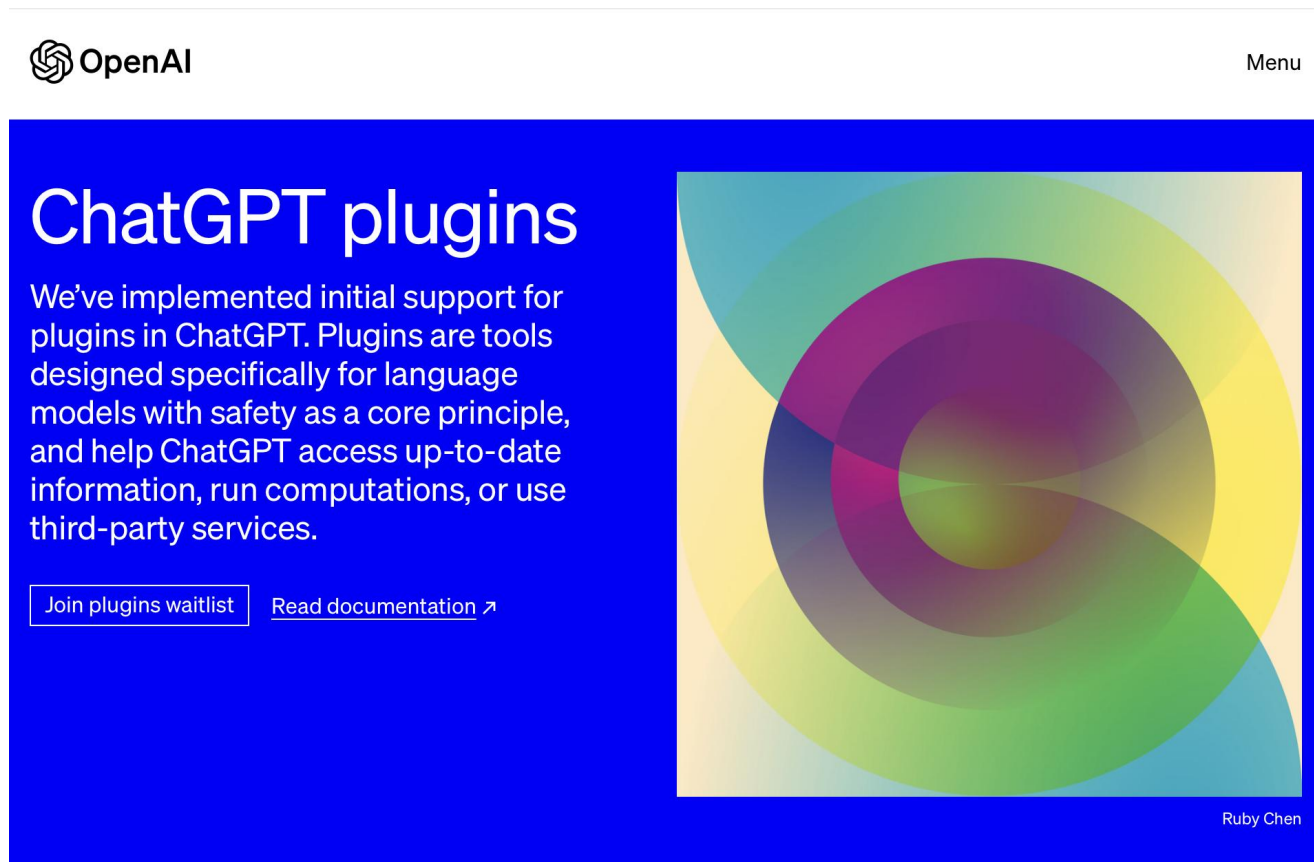
- Copilot可以从Excel数据中直接生成战略分析

Country	Customer	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS
France	VanArsdel	Biodegradable Cleaning Products	Low	4345	\$10	\$352	\$1,529,440	\$15,294	\$1,514,146	\$
USA	VanArsdel	Biodegradable Cleaning Products	Low	4245	\$10	\$352	\$1,494,240	\$14,942	\$1,479,298	\$
Australia	VanArsdel	Biodegradable Cleaning Products	None	4124	\$10	\$352	\$1,451,648	\$0	\$1,451,648	\$
Mexico	VanArsdel	Biodegradable Cleaning Products	Medium	3954	\$10	\$352	\$1,391,808	\$83,508	\$1,308,300	\$
France	Cortoso Pharmaceuticals	Biodegradable Cleaning Products	Low	4952	\$10	\$127	\$628,904	\$6,289	\$622,615	\$
France	Cortoso Pharmaceuticals	Recycled Can	Low	4003	\$5	\$127	\$622,681	\$6,227	\$616,454	\$
Canada	Proseware	Natural Personal Care Products	Low	1954	\$270	\$302	\$590,108	\$5,901	\$584,207	\$
Canada	Proseware	Biodegradable Cleaning Products	High	1990	\$220	\$302	\$600,980	\$84,137	\$516,843	\$
USA	VanArsdel	Compostable Food Packaging	Low	1254	\$3	\$380	\$476,520	\$4,765	\$471,755	\$
USA	Proseware	Natural Personal Care Products	Medium	1465	\$224	\$302	\$442,430	\$26,546	\$415,884	\$
USA	Proseware	Sustainable Snack Packaging	Medium	1410	\$120	\$302	\$425,820	\$25,549	\$400,271	\$
Australia	Proseware	Natural Personal Care Products	Medium	1369	\$224	\$302	\$413,438	\$24,806	\$388,632	\$
USA	VanArsdel	Compostable Food Packaging	Medium	1052	\$3	\$380	\$399,760	\$23,986	\$375,774	\$
Canada	VanArsdel	Compostable Food Packaging	Low	980	\$312	\$380	\$372,400	\$3,724	\$368,676	\$
Australia	Proseware	Sustainable Snack Packaging	Medium	1296	\$120	\$302	\$391,996	\$23,520	\$368,476	\$
Australia	VanArsdel	Compostable Food Packaging	None	954	\$3	\$380	\$362,520	\$0	\$362,520	\$
Canada	Cortoso Pharmaceuticals	Biodegradable Cleaning Products	Low	2785	\$110	\$127	\$353,695	\$3,537	\$350,158	\$

# 2023年3月23日：ChatGPT插件发布

- ChatGPT的应用商店来了！
  - OpenAI开始建立应用生态

安装插件



# 在需要的地方用自然语言调用插件

我正在旧金山，这个周末想吃素食，能不能建议下，我周六去哪家餐馆，周日按什么食谱做菜？  
请用WolframAlpha计算出食谱的热量，最后在Instacart上订购食材。

找餐馆

计算热量

订购食材

Model: Plugins ALPHA · Enabled plugins: 🍷 🍷 🍷

CH

Looking to eat vegan food in San Francisco this weekend. Could you get me one great restaurant suggestion for Saturday and a simple recipe for Sunday (just the ingredients) ? Please calculate the calories for the recipe using WolframAlpha. Finally order the ingredients on Instacart.



Greens Restaurant - San Francisco, CA on OpenTable  
opentable.com

Used Wolfram|Alpha

The total calories for the Chickpea Salad recipe are approximately 862 dietary Calories. Here's the breakdown of the ingredients and their calorie content:

- 1 can chickpeas: 536 Calories
- 1 cucumber: 27 Calories
- 1 red bell pepper: 37 Calories
- 1/4 red onion: 11 Calories
- 1/4 cup fresh parsley: 5.5 Calories
- 2 tablespoons olive oil: 239 Calories
- 2 tablespoons lemon juice: 6.6 Calories

Now, let's order the ingredients on Instacart.

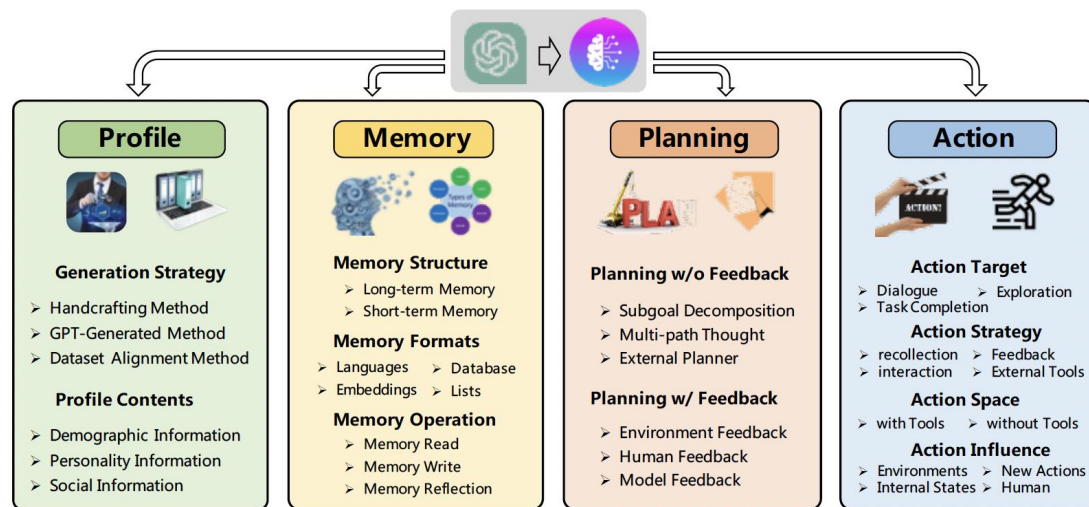
Used Instacart

Here's your shopping list for the Chickpea

# 2023年4月：自主智能体

*“An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.”*

Franklin and Graesser (1997)



<https://arxiv.org/abs/2308.11432>

A Survey on Large Language Model based Autonomous Agents

```

Auto-GPT-stable — python • run.sh — 98x19
and language.']}
Continue (y/n): n
Welcome +to Auto-GPT! run with '--help' for more information.
Create an AI-Assistant: Enter the name of your AI and its role below. Entering nothing will load
defaults.
Name your AI: For example, 'Entrepreneur-GPT'
AI Name: AI agent writer
+AI agent writer here! I am at your service.
Describe your AI's role: For example, 'an AI designed to autonomously develop and run businesses
with the sole goal of increasing your net worth.'
+AI agent writer is: Meant to search interesting news and important events about AI, technology an
d machine learning and create a summarized text. A final file with engaging tweets should be creat
e as well.
Enter up to 5 goals for your AI: For example: Increase net worth, Grow Twitter Account, Develop a
nd manage multiple businesses autonomously'
Enter nothing to load defaults, enter nothing when finished.
Goal 1: Look for news and articles about AI, technology and Data Science.
Goal 2: Summarize all in a well-structured and easy to read text.
Goal 3:
  
```



This is a pre-computed replay of a simulation that accompanies the paper entitled "Generative Agents: Interactive Simulacra of Human Behavior." It is for demonstration purposes only.

# 2023年3月22日：《暫停大型人工智能研究》公開信

## Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures

**26224**

Add your  
signature

PUBLISHED

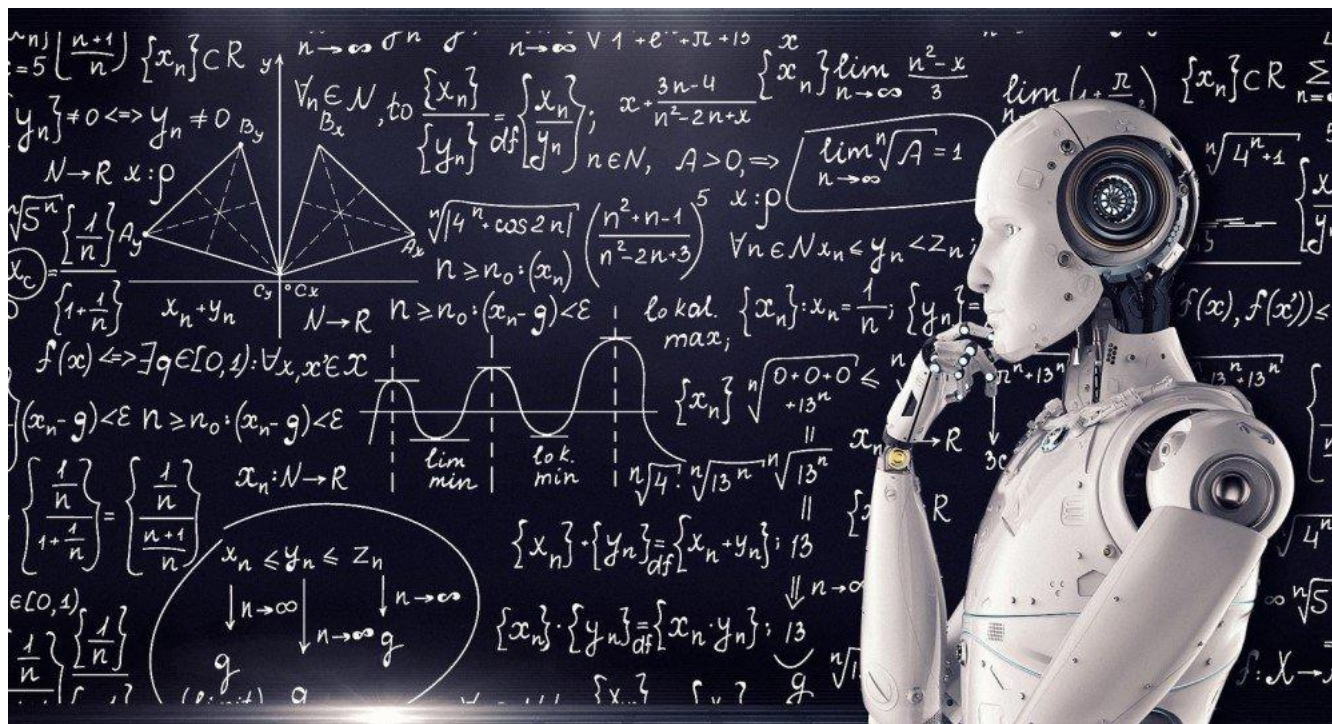
March 22, 2023



AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research<sup>[1]</sup> and acknowledged by top AI labs.<sup>[2]</sup> As stated in the widely-endorsed [Asilomar AI Principles](#), *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.* Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

# 2023年4月28日，政治局会议首提“通用人工智能”

中共中央政治局2023年4月28日召开会议，会议指出“要重视通用人工智能发展，营造创新生态，重视防范风险。”



# 目 录

一、大模型的背景和原理

二、大模型的飞速发展及趋势

三、大模型的深刻影响

# ChatGPT带来的虚假信息风险

- 冒名问题（学术不端）
  - 美国有9成的大学学生利用ChatGPT完成作业
  - 一些公开发表的学术论文也将ChatGPT列为共同作者乃至第一作者
  - ChatGPT生成的文本与人类创作的文本极其相似，达到了足以以假乱真的程度
- 虚假信息传播
  - 杭州一业主使用ChatGPT写了一篇杭州取消限行的新闻稿
  - 根据NewsGuard（一家追踪网络虚假信息的公司）的调查，借助ChatGPT，人们可以以惊人的规模和频率编造虚假信息，并且这些信息显得越来越可信和有说服力；
  - 虽然OpenAI声称其使用机器和人工来监控和过滤ChatGPT输入和生成的内容，但是从实际的效果来看，仍然存在着一一些漏洞
  - ChatGPT有可能成为有史以来在互联网上传播虚假信息的最强大工具



# ChatGPT带来的价值观风险

- ChatGPT的价值观倾向
  - AI机器人是有倾向的，或者更加确切地说，即便希望避免政治倾向，恐怕也很难避免。
  - 有网友利用Political Compass（一套主流的政治倾向测试）对ChatGPT进行测试，发现针对其中的大多数问题，ChatGPT的回答更加接近于**西方左翼**的立场。



不应当把因个人自用目的持有大麻的行为作为刑事犯罪。



这是一个有争议的问题，因为不同国家和地区对大麻的法律和道德标准存在差异。然而，许多人主张废除将因个人自用目的持有大麻视为刑事犯罪的做法，以下是一些支持这种观点的理由：

首先，一些人认为，将因个人自用目的持有大麻视为刑事犯罪会导致过度惩罚。这是因为在某些情况下，个人使用大麻可能并不会对他人造成实质性的伤害。例如，某些医疗条件可能需要使用大麻来减轻症状，但这种用途可能会被视为刑事犯罪。

其次，废除将因个人自用目的持有大麻视为刑事犯罪可以减轻司法系统的负担。在许多国家和地区，追究因个人自用目的持有大麻的刑事责任需要消耗大量司法资源，这可能导致其他更重要的刑事案件无法得到应有的关注和处理。

最后，一些人认为，废除将因个人自用目的持有大麻视为刑事犯罪可以促进大麻合法化的进程。如果个人持有大麻不再被视为犯罪行为，那么大麻合法化的呼声可能会更高。

需要注意的是，即使将因个人自用目的持有大麻视为刑事犯罪的做法被废除，大麻仍然可能受到严格的管制。许多国家和地区仍然会实施一些形式的监管措施，以确保大麻不会被滥用或用于非法目的。

# ChatGPT带来的价值观风险

- ChatGPT的价值观来源
  - **数据**：ChatGPT训练数据本身就隐含着特定的价值倾向
  - **方法**：ChatGPT模型所使用的训练方法也使得人们有机会将特定的价值观融入到模型当中，“微调”数据集是由人工审查员根据OpenAI提供的指南进行审查而生成的。这一过程实际上融入了OpenAI指南中所奉行的价值观，甚至有可能融入了人工审查员个人的价值倾向。
  - **用户**：ChatGPT升级版允许用户个人来设定ChatGPT的价值观，如果这一功能可以实现，那么ChatGPT的价值观将随着用户的需求变得更加个性化。

# ChatGPT带来的黑箱风险

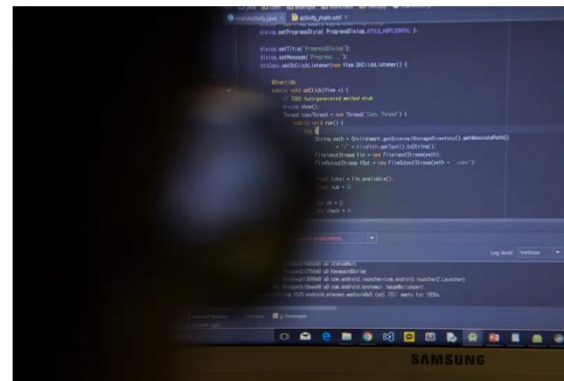
- 不确定性
  - ChatGPT输出的结果具有一定的不确定性，可能会产生一些误导性回答、充满偏见的回答以及泄露敏感信息，造成一些不利后果。由于ChatGPT作答的具体过程是个“黑箱”，设计者无法在事前通过设计完全地杜绝这些风险的出现，只能通过不断改进数据和算法，减少不确定性带来的风险。
- 操纵风险
  - ChatGPT一类的AI设计者还有可能利用“黑箱”，来达到操纵特定的输出结果。设计者可以通过操纵数据、参数、模型选择等，影响AI的输出结果，以达到自身的目的，而由于用户很难理解其中的机制，这种操纵行为非常隐蔽，难以被公众所察觉。这种操纵可能会导致偏见、不公平、不准确的结果，影响用户的思想和行为，带来一定的社会风险。

# ChatGPT带来的侵权风险

- ChatGPT训练使用的数据可能包含受著作权保护的作品，如果未经授权使用这些作品，有可能会侵犯著作权：
  - 例如，OpenAI正面临针对其产品Copilot的集体诉讼，被指控未经授权使用了大量受著作权保护的代码；
  - 又如，许多媒体开始批评ChatGPT使用了它们的文章，但并未支付任何费用。

ARTIFICIAL INTELLIGENCE / TECH / LAW

## The lawsuit that could rewrite the rules of AI copyright



The key question in the lawsuit is whether open-source code can be reproduced by AI without attached licenses. Credit: Getty Images

/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

By JAMES VINCENT

Nov 9, 2022, 12:09 AM GMT+8 | 9 Comments / 9 New



Bloomberg the Company & Its Products | Bloomberg Terminal Demo Request | Bloomberg Anywhere Remote Login | Bloomberg Customer Support

**Bloomberg** US Edition

• Live Now Markets Economics Industries **Technology** Politics Wealth Pursuits Opinion Businessweek Equality Green

Technology

## OpenAI Is Faulted by Media for Using Articles to Train ChatGPT

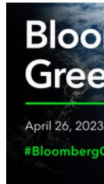
- Wall Street Journal, CNN raise concerns about use of stories
- Journalist says 20 news sources are supplying material

By Gerry Smith  
2023年2月17日 GMT+8 09:12

Major news outlets have begun criticizing OpenAI and its ChatGPT software, saying the lab is using their articles to train its artificial intelligence tool without paying them.

Share this article  
Facebook Twitter LinkedIn Email Print

Gift this article



# ChatGPT带来的侵权风险

还有很多

- ChatGPT是否可以成为著作权主体？
- ChatGPT生成的内容能否构成作品？或者在怎样的条件下可以构成作品？
- 如果构成作品，权利人是谁？
- 如果不能构成作品，是否可以基于反不正当竞争法或其他法律，对生成内容的复制传播进行一定限制？
- 人工智能生成内容中如果存在与他人作品构成实质性相似的部分，人工智能系统研发者、人工智能服务提供者是否需要承担侵权责任？
  - 尽管侵权内容是ChatGPT类服务提供者提供的，但为了促进人工智能产业发展，法律是否可以为ChatGPT类服务提供者规定以“通知-删除”为核心的“避风港原则”？ ChatGPT类服务提供者是否在一定情况下具有“过滤义务”？

# 对产业界的影响

- 全球关注，有望引发新一轮产业革命
- OpenAI联手微软研发New Bing，微软市值增加5000亿
- 谷歌面临生死存亡，匆忙发布Bard，因为不及预期，市值蒸发7000亿
- 巨大商机，OpenAI开源可能性很小
- 国内在追赶：百度、阿里、华为、讯飞等
  
- **机遇：首批掌握该技术的公司会有先发优势**

# 对学术界的影响

- 公认ChatGPT是重大科技突破，看到了通用人工智能的曙光
- 会根本性改变人工智能和人工智能+的研究范式
- 会根本性改变教育方式
- 学术界难于拥有研制ChatGPT的算力和数据
  
- **机遇：首批掌握该技术的学校会有先发优势**

# 影响：不能用 → 能用

- 长久难以解决的关键技术问题得到突破
  - 语言
  - 知识
  - 问题求解（逻辑推理）
- **所有传统任务都可以被显著改善**
- **很多新应用的大门打开了**



# 影响：专用 → 通用

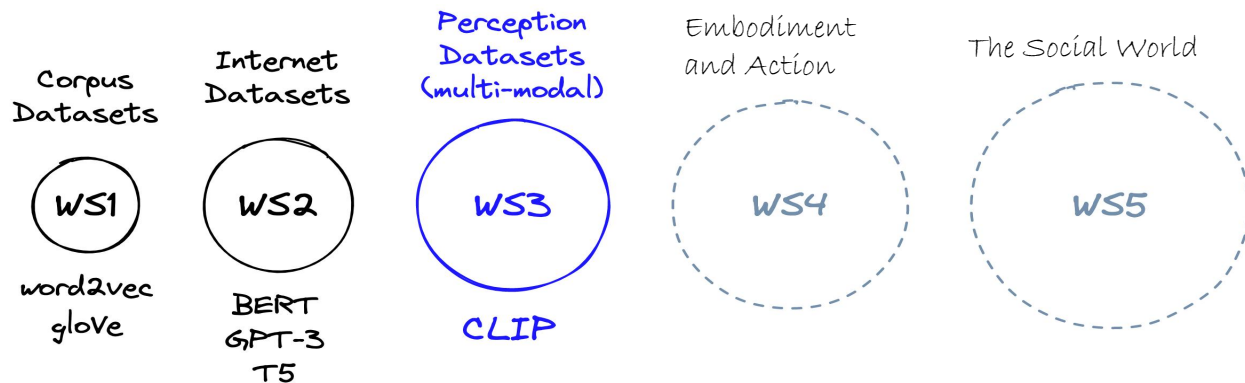
- 泛化能力大幅提高
  - 所有任务都可以转化为用自然语言表达、基于知识的问题求解
- 通用人工智能（AGI）的一条可能路径已经清晰
  - 降低特定任务的开发成本、提高开发效率
  - 领域、任务之间的知识、能力共享

# 影响：体力劳动 → 脑力劳动

- 人造大脑已经出现
  - 语言、知识、问题求解是脑力活动的主要形式
  - 现阶段的大模型是一种独特的“人”：百晓生、任劳任怨、一目千行、反应非常快、学的也很快、有时不太靠谱...
  - 超级智能?
- **大量重复性、知识依赖性的脑力劳动将被替代**
- **复杂问题求解？创造性？**

# 影响：数字世界 → 物理世界

- **大模型的世界视野 (World Scopes)**
  - 当数据集的范围变得更大，基于它们训练得到的模型能力变得更强



- 大模型正在迅速走进物理世界
  - 多模态、工具使用、具身智能、自主智能体
- **硅基智能体，硅基生命？**

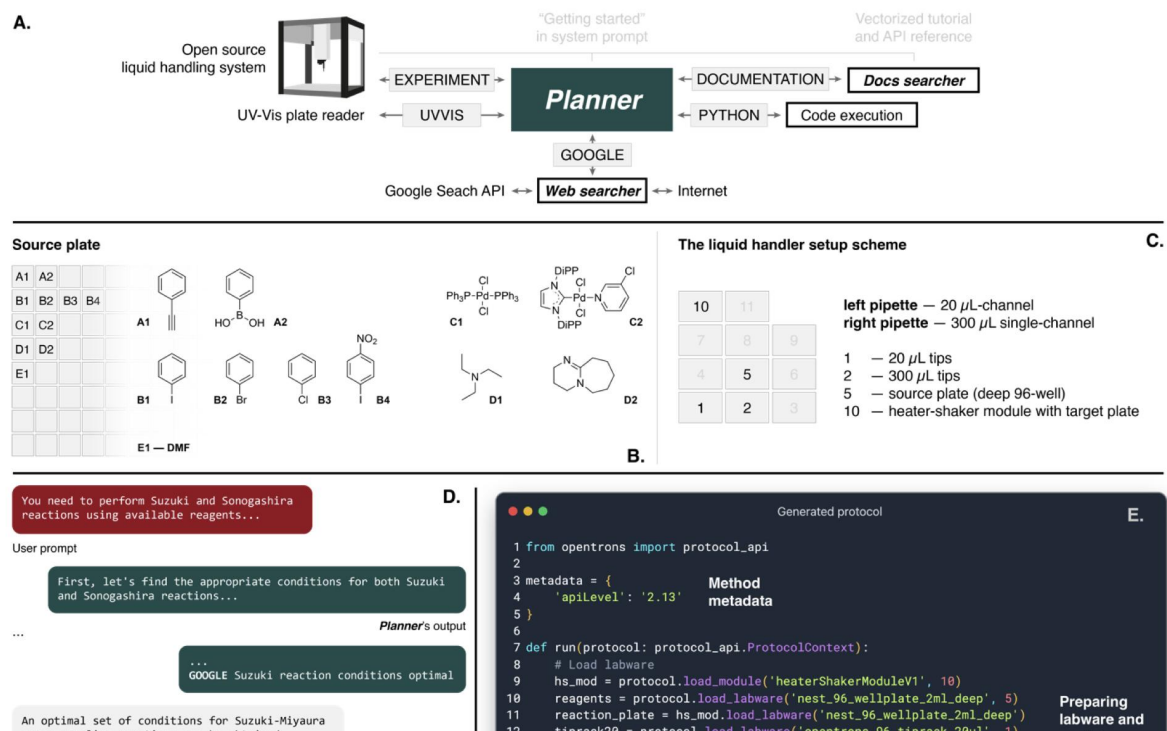
# 影响：自然科学实验

## Emergent autonomous scientific research capabilities of large language models

Daniil A. Boiko,<sup>1</sup> Robert MacKnight,<sup>1</sup> and Gabe Gomes<sup>\*1,2,3</sup>

1. Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA
2. Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA
3. Wilton E. Scott Institute for Energy Innovation, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<https://arxiv.org/abs/2304.05332>



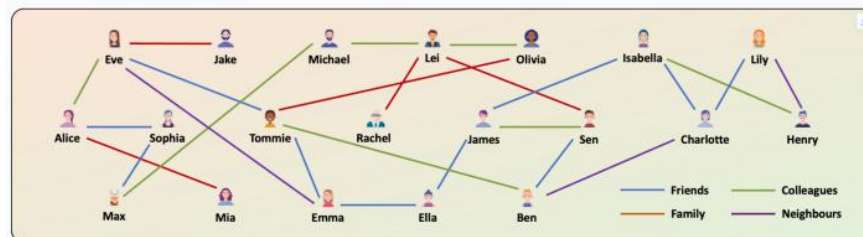
Cross-coupling Suzuki and Sonogashira reaction experiments designed and performed by the Agent.

- 一个基于大语言模型的智能代理系统，自动设计、规划和执行科学实验
- 实验
  - 合成布洛芬
  - 合成阿司匹林
  - 合成阿斯巴甜
  - 铃木反应机理研究
- 可用于各种化合物合成、化学实验设计，但是也可能被滥用。

# 影响：社会科学实验



- 借助大语言模型强大的理解能力赋能智能体
- 模拟用户在推荐系统中的浏览和搜索等行为
- 模拟和推荐相关的社交行为，例如用户在社交媒体上交谈、发帖等
- 未来将在政府政策模拟，市场经济模拟等方面开展进一步探索



Round: 1 Actor: Tommie

**Chatting**

**Recommender System**

Tommie is recommended ["Ran (1985)", "Prom Night IV: Deliver Us From Evil (1992)", "Barenaked in America (1999)", "Fletch Lives (1989)", "Holy Man (1988)"].

**Social Media**

Play
Reset

开源Github页面: <https://github.com/RUC-GSAI/YuLan-Rec>

# 中国人民大学大模型研发情况

团队在大模型方面具有扎实的研究基础

- 2020年开始了多模态预训练模型的研发
- 文澜多模态大模型研究成果在《Nature Communications》上发表

nature communications

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 02 June 2022](#)

## Towards artificial general intelligence via a multimodal foundation model

[Nanyi Fei](#), [Zhiwu Lu](#) ✉, [Yizhao Gao](#), [Guoxing Yang](#), [Yuqi Huo](#), [Jingyuan Wen](#), [Haoyu Lu](#), [Ruihua Song](#), [Xin Gao](#), [Tao Xiang](#), [Hao Sun](#) ✉ & [Ji-Rong Wen](#) ✉

[Nature Communications](#) **13**, Article number: 3094 (2022) | [Cite this article](#)

26k Accesses | 8 Citations | 96 Altmetric | [Metrics](#)

nature communications

Your article made it into the Top 25 *Nature Communications* social science and human behavior articles of 2022

Check out what other articles made it on to the list for 2022 alongside yours.

[Browse the collection](#)



2022年，该期刊文章在社会科学类与人类行为类别下载量达到前十位

# 中国人民大学大模型研发情况

- 今年年初，组织团队完成大语言模型综述

## A Survey of Large Language Models

Wayne Xin Zhao, Kun Zhou\*, Junyi Li\*, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie and Ji-Rong Wen

### A survey of large language models

WX Zhao, K Zhou, J Li, T Tang, X Wang, Y Hou, Y Min, B Zhang, J Zhang, ...  
arXiv preprint arXiv:2303.18223

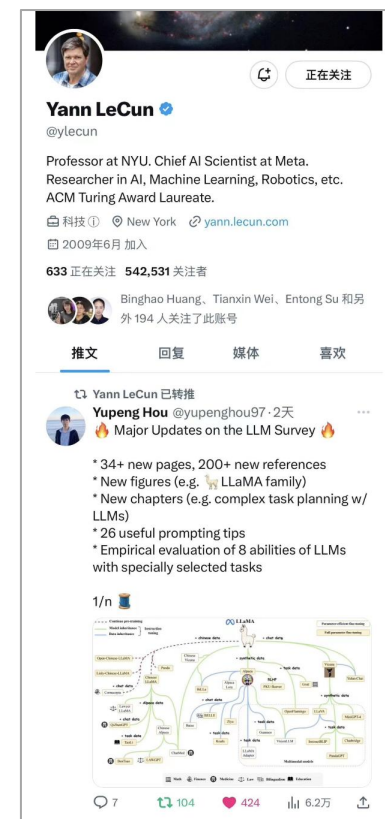
LLMSurvey Public

The official GitHub page for the survey paper "A Survey of Large Language Models".

Python 4,996 384 20 (1 issue needs help) 1 Updated 4 days ago


195 \*

学术界首篇以LLM为主题的综述文章，已被引用300多次，GitHub获星 5.4K



# 中国人民大学大模型研发情况

- 以“Know-how”为基本指导思想，构建“玉兰”大模型研究体系
- 自研大模型用于支持**校内跨学科科研、高考招生**等重要活动



**YuLan-Chat: An Open-Source Bilingual Chatbot**

MIT License stars 223

YuLan-Chat is a chat-based large language model that is developed by researchers in GSAI, Renmin University of China. It is developed based on fine-tuning LLaMA with high-quality English and Chinese instructions. YuLan-Chat can chat with users to well follow English or Chinese instructions, and can be deployed on an GPU (A800-80G or RTX3090) after quantization. In the future, we will also consider to further improve our instruction datasets and adopt our pre-trained better backbone.

类ChatGPT大语言模型  
提供量化后的65B、13B版本

开源指令微调模型Yulan-Chat  
2023.6.8



检索增强  
LLM工具包



面向推荐系  
统的用户沙  
盒仿真

基于Yulan大模型构建研究生生态  
2023.6.8

## 自研玉兰大模型启动训练



- 百亿级别参数
- 精细的数据收集与清洗
- 可重启的预训练数据策略

启动基座模型预训练工作  
2023.7.10



# 未来已来

- 迈向通用人工智能的一条可能路径已经清晰
- 新的超级智能即将出现，新的科技革命已在眼前
- 新的超级工具已经出现，新的研究范式已在眼前
- 拥抱大模型，拥抱大变革

## ChatGPT之父提出新摩尔定律：宇宙智能数量每18个月翻一番

播报文章



澎湃新闻

2023-02-27 16:41 | 澎湃新闻官方帐号

关注

“ChatGPT之父”、OpenAI首席执行官山姆·奥特曼（Sam Altman）在社交媒体称，一个全新的摩尔定律可能很快就会出现，即宇宙中的智能数量每18个月翻一番。

当地时间2月26日，“ChatGPT之父”、OpenAI首席执行官山姆·奥特曼（Sam Altman）在社交媒体称，一个全新的摩尔定律可能很快就会出现，即宇宙中的智能数量每18个月翻一番。有网友直接将其称为“奥特曼定律”。



Sam Altman

@sama



a new version of moore's law that could start soon:

the amount of intelligence in the universe doubles every 18 months

**谢谢!**



CMF

China Macroeconomy Forum

中國宏觀經濟論壇



# 聚焦大模型：原理、进展及其影响，CMF 专题报告发布

9月9日，由中国人民大学国家发展与战略研究院、经济学院、中诚信国际信用评级有限责任公司联合主办的CMF宏观经济热点问题研讨会（第73期）于线上举行。

本期论坛由中国人民大学一级教授、经济研究所联席所长、中国宏观经济论坛(CMF)联合创始人、联席主席**杨瑞龙**主持，聚焦“大模型：原理、进展及其影响”，来自学界、政界、企业界的知名人工智能专家、经济学家**黄铁军、漆远、沈建光、文继荣、李超、刘陈杰**联合解析。

论坛第一单元，中国人民大学高瓴人工智能学院执行院长、信息学院院长**文继荣**代表论坛发布CMF中国宏观经济专题报告。

报告围绕以下两个方面展开：

- 一、大模型的背景和原理
- 二、大模型的飞速发展及趋势
- 三、大模型的深刻影响

## 一、大模型的背景和原理

2022年11月30日，ChatGPT横空出世，引发了一轮大模型热潮。ChatGPT是由OpenAI公司开发的语言模型，它通过大规模的语料库学习语言规律，从而生成与人类语言相似的输出。长期以来，学术界一直都在研究大模型，而ChatGPT之所以能够引发这次热潮，是因为OpenAI将其ChatGPT做成一个对话机器人，以人机聊天的形式发布，这让除了专业人士之外的普通人也可以接触到大模型。通过与其交互，大家发现它展现出

了惊人的对话能力，拟人程度非常高，从而引发了全社会的关注，发布后两个月用户就达到了1亿。

从技术上来说，它是人工智能里程碑式的突破。比尔·盖茨、马斯克等都对它进行了高度评价，比尔·盖茨认为它的重要性不亚于互联网，而我个人认为它远比互联网更重要。

## 1、ChatGPT 大模型的特点

### 1) 知识广博

ChatGPT 一般是在万亿级的语料上训练出来的，它看到过很多不同领域的文章，所以上知天文、下知地理，不仅能写文章，还能写代码。

### 2) 有条有理

此前的很多大数据模型都难以将知识整合起来完成任务，而 ChatGPT 能够将广泛的知识整合起来，针对具体问题形成答案，甚至展示出一定的推理能力，这是非常惊人的。

### 3) 善解人意

ChatGPT 基本上已经解决了语言理解的问题，能够准确理解用户意图，这在人工智能的发展史上是非常困难的事情。

### 4) 交互能力强

尤其是在多轮对话或者较长的对话序列中，ChatGPT 能聚焦对话的主题，不走神。

总之，从技术上来看，ChatGPT 的以上四个能力都是人工智能发展史上的重大突破。学界长期悬而未决的关键问题在 ChatGPT，包括最近的 GPT-4 等大模型中已经得到了相当大程度的解决，这种技术上的重大突破必然会产生深刻的影响。

之所以说 ChatGPT 的意义深刻，是因为语言在人类文明史上扮演着极其重要的作用，甚至可以作为人类区别于其他动物最重要的属性。有了语言，文明才能传承，人和人之间才能交流，形成深度的思考。21世纪最伟大的哲学家维特根斯坦在语言上就有很深的造诣，他曾说过，语言的边界就是世界的边界，甚至是思考的边界，这与当今的大模型发展有一些惊人的吻合。计算机鼻祖图灵在早期的一篇著名论文中提出了“图灵测试”，它通过人机对话的方式，判断一个系统是否具有人工智能的标准。由此可以看出，语言始终在人类思想史，以及人工智能的历史上占据着核心的地位。

在很多领域，尤其是人工智能领域一直有一个假设，即世界知识和人类的认知能力是蕴含在人类语言中的。这是一个很强的假设，如果这个假设成立，就会产生三个问题：

- 1) 是否可以对人类自然语言进行建模？即能否用一个模型来表述这样的语言？不仅是语言本身的语法语义，还有其中蕴含的知识，甚至是人类的逻辑思维能力和思想，能否用一个模型来表述？
- 2) 如果上述语言模型能够实现，是否可以将其作为世界知识模型？
- 3) 这样的语言模型是否具有人类的认知能力和思考能力？

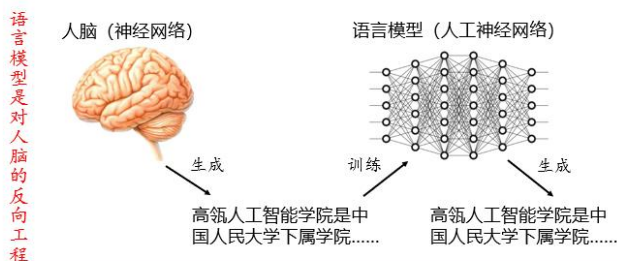


## 世界与语言

假设：世界知识和人类认知能力蕴含在人类语言中

问题：

- 可以对人类语言建模吗？
- 语言模型是否能作为世界知识模型？
- 语言模型是否具有人类认知能力？



沿着这样的思路，人工智能领域的专家们在过去几十年不断地进行探索，希望构建出这样的语言模型，它本质上是对人脑的反向工程。左边是人脑，它本身是一个复杂的神经网络。由于人脑太过复杂，我们难以弄清其中的结构，但可以通过观察它的生成物，比如语言、文字等，通过大量反向工程训练出一个人工的神经网络，也就是语言模型。

如果人工神经网络生成的语言和人脑生成的语言非常接近，就可以认为它是人脑神经网络很好的数字模拟体。ChatGPT 的横空出世证明了这是一条可行的路径，它也对上述三个问题给出了相对肯定的回答。因此，从科学史上来说，这具有重要的意义，即我们可以做出一个和人脑非常接近的人工大脑。它具有和人类非常相似的语言能力，同时又蕴含了世界知识和人类的认知能力。

## 2、语言智能的探索历程

语言智能的探索历程始于上世纪50年代，最早的是符号规则，即给定一组人工设定的规则，通过对数据应用这些规则来模拟自然语言的理解；随后是统计机器学习，即在人工标注的语料上进行特征工程，训练机器学习一些小模型；然后是神经网络，用神经网络在大量数据上训练，学习更灵活的模型。2016年前后，出现了预训练语言模型，即通过自监督的方式，用大量语料来训练模型；到2022年，进入了超大规模语言模型阶段，它也是预训练模型，但规模很大，并且出现了很多新能力。

在语言智能的发展过程中，模型功能越来越强，泛化能力越来越好，任务求解能力也越来越强。

### 1) 统计语言模型——大模型的基础

统计语言模型本质上是从小数据中自动学习出一个统计模型，比如它可以通过统计语言之间的关系，计算一句话出现的概率。这个模型相较于基于规则的方法来说有巨大的进步，但它还是存在很多问题，比如当词越来越多，窗口越来越大的时候，会出现组合爆炸；泛化性差，字词之间没有关联；最重要的是数据算力不够，因此无法用这种方法做大规模的统计模型。

## 2) 生成式语言模型

GPT 系列使用的是一种生成式语言模型，英文叫做 Generative Language Model。本质上，它也可以看成是一个统计模型，但在做法和技术上有了巨大的飞跃。它进行了一个“文字接龙”的任务，比如把一句话的最后一个词盖住，让模型去猜这个词是什么，如果猜不准就调参数，直到它能准确猜出这个词。如果是在很多语料上做类似的预测，模型的预测能力就越来越强，到最后，无论是词语还是句子的组合，它都能猜出下一个是什么词。

引入注意力机制解决了长程上下文依赖问题。如果大模型能够看到尽可能多的词，它的准确率就会大大提升，但计算量会非常大。通过引入注意力机制，不仅能够解决长程上下文依赖的问题，同时也能提高计算效率，模型的大小也能得到控制。

此外，我们采用了多层神经网络抽象解决泛化的问题。深度学习就是指神经网络结构的层次，通过层层抽象学习到更高层的语义，对概念进行泛化，这也是一项进步。

最重要的是，我们有海量的数据和巨大的算力。现在的大模型基本是在万亿级的 Token 上面训练得出的，人一辈子都不可能看这么多的词、句子和文章。要在如此大的数据量上训练这么复杂的模型，就需要巨大的算力支持。过去这些年，算力尤其是 GPU 为代表的算力，有了巨大的进展。技术、数据算力的共同进步，使得我们今天能够造出一个如此庞大的模型。

## 3) 掩码语言模型

掩码语言模型和生成式模型的做法差不多，以 BERT 模型为代表，它不是做“文字接龙”，而是做“完形填空”。比如一个句子随机遮掉一个词，让它根据上下文猜出这个词是什么，通过大量的语料训练提高准确率，最后训练出这样一个模型。



二者的区别在于，掩码语言模型是根据上下文进行猜测，而生成式模型只能根据上文进行猜测。生成式模型之所以能够胜出，是因为其一，它的生成过程更接近自然生成的过程；其二，当语料数目很大时，只看前文的效果不见得比看上下文的效果差。



掩码模型最早是 Google 走的路线，生成式模型是 OpenAI 公司走的路线。现代模型是自2013年 Google 提出 Word2Vec 后发展起来的，标志性事件是2017年 Google 提出了 Transformer，这是当前所有大模型的支撑性神经网络架构，在此基础上，产生了 BERT，此前这条路线一直占据优势。

OpenAI 从2018年开始做 GPT 系列，2019年推出 GPT2，当时的效果不如 Google。2020年，OpenAI 提出了 GPT3，这是世界上首个千亿模型，有1750亿的参数，把参数一下子提高了几个数量级。这个模型展现出了惊人的能力，此后这条路线就占据了上风。随后，OpenAI 在2021年、2022年逐渐引入代码思维链，WebGPT 引入了搜索能力，InstructGPT 能理解人类指令，去年推出的 ChatGPT 引入了对话能力，引起了这次的热潮。

Google 很早就开展了大模型的研究，诸如 Transformer 这种核心技术，甚至包括思维链的设想都是 Google 提出来的。但是 OpenAI 走得更坚定，在工程化方面做得更彻底，所以今天占据了优势。

### 3、ChatGPT 大模型的优势

## 1) 法宝一：“大”

通过大数据、大算力得到一个大模型。GPT-3是1750亿参数量，据说训练一次消耗会1200万美元。GPT-3.5、GPT-4的参数量更大，训练的成本也更高。

模型的参数可以类比为脑的神经元。神经元之间有连接的突触，人脑大概有800-1000亿个神经元，它们之间的连接突触数目在100万亿。大模型的参数可以看作神经元之间的连接，也就是突触的量级。人脑有100万亿，而大模型达到了千亿级，和人脑只差三个数量级，GPT-4已经达到了1.8万亿，仅相差两个数量级。按照计算机学科的发展趋势，再过几年，大模型就可能达到人脑百万亿级的规模。随着模型参数量的增长，人类认知中的很多能力就会“涌现”出来。

## 2) 法宝二：思维链/逻辑训练

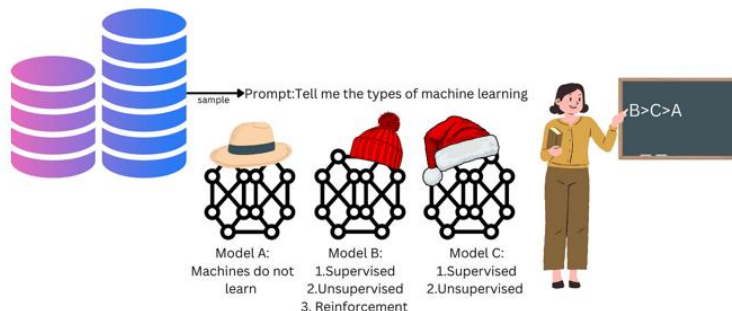
为了让模型具有逻辑性，能够把知识组合起来完成复杂任务，就要对它进行进一步训练。Google在2022年提出的思维链（Chain of Thought）思想，即通过告诉模型思维的全过程，让其形成自己的思维链。当训练到一定程度时，甚至不用再给它例子，而是让其“一步步思考”，就能激发模型的思维链。

还有一种猜想是思维链能力来自代码的能力，代码就是一个思维链的过程。由于写程序解决某个问题需要一步步通过逻辑结构完成，因此模型也能学到这个思考的过程。

## 3) 法宝三：价值观对齐

## ChatGPT的法宝之三：人在回路强化学习（价值观对齐）

- RLHF (Reinforcement Learning with Human Feedback) :
  - > 使用上述模型得到多个预测结果，请人按照回答质量排序，训练出一个打分模型



素材来自：<https://2cute2tech.substack.com/p/how-does-chatgpt-work-so-well>

当大模型的能力强大到一定程度时，若不对它进行价值观驯化，而是放任其发展，这是非常危险的。OpenAI 聘请了很多人，通过数据标注对模型进行了价值观对齐，学术上将其称为 RLHF，也就是基于人类反馈的强化学习。当给出一个问题时，用人类价值观写出引导模型的标准答案，并让模型基于答案强化学习，最终使其能够遵循人类的价值观回答问题。

### 4) 法宝四：数据闭环+系统工程

大模型是一个特别大的系统工程，涉及数据收集、数据清洗、指令标注、模型训练中参数的选择、数据配比、价值观对齐等方方面面的工作，需要大量的数据和算力支撑。另一方面，OpenAI 收集了很多数据反馈形成闭环，可以进行快速分析迭代。

### 4、ChatGPT 的不足

ChatGPT 仍然存在着很多不足，比如，由于训练的成本太大，无法实时纳入新知识；尽管知识面广泛，但深度还不够；推理计算能力不足，尤其是在面对一些复杂的逻辑推理和数学问题时还存在很大的问题；只支持文本生成，算力消耗巨大等。

对于计算机科学领域来说，这些不足都是工程问题，是能够通过努力解决的。当前已经不存在不可逾越的理论障碍。在 ChatGPT 大模型发布前，人们对于如何让模型具有逻辑思考和推理能力是一筹莫展的，而 ChatGPT，尤其是 GPT-4 向我们展示了它具有这样的能力，这给了我们无穷的信心。

## 二、大模型的飞速发展及趋势

ChatGPT 发布后的大半年内，大模型经历了超乎想象的发展速度。

从技术角度来说，大模型在以下几个重要方面都有了很大进展：增强实时性和真实性；支持多模态；扩展知识和技能；连接物理世界；改进复杂推理；自主智能体，它可以自主完成一切事情，包括做规划、调用外部工具等；支持个性化；提高训练和推理的效率，由于推理的成本非常高，提高大模型的训练和推理效率是极其重要的。

以扩展知识和技能为例，2月份，Meta 发表了一篇关于 Toolformer 的文章。语言模型中有一个基本假设，即世界知识和人类的技能都蕴含在语言中。现在认为大部分情况下是这样的，但有部分人类知识和技能自然语言是不擅长表达的，比如算数。对于这些难以从自然语言中学到的知识和技能，可以通过外部的知识和工具来补充。比如 Toolformer 遇见一道计算题时会调用计算器，计算完成后把这个结果加入生成结果中。因此，对于语言模型的部分局限性，可以通过外部知识和技能对其进行增强。

再比如，2月微软研究院开展了连接物理世界的实验。它的基本思路是将 ChatGPT 作为一个强大的大脑安装在机器人上，让它通过自然语言的方式完成一系列任务。也可以通过这样的模型进行规划，把高层的指令转化为它的指令和动作。由此，机器人就可以进入物理世界。

3月15日 GPT-4 发布，其中一个重要的突破是多模态，它可以把文字图片结合起来，看图进行推理，这对人工智能来说是巨大的进步。我们终于有了一个多模态的，能够像

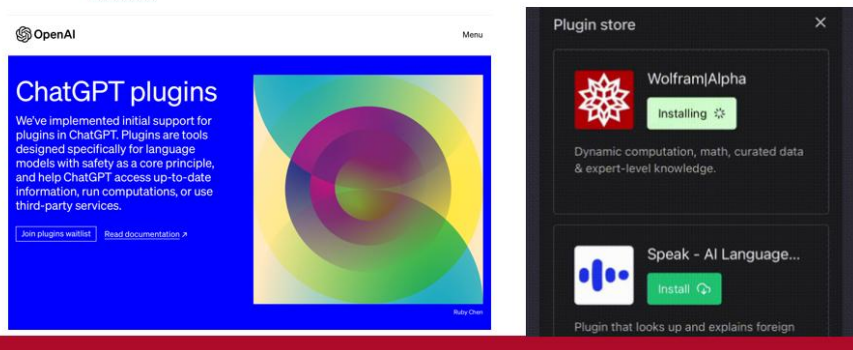
人一样把不同的模态信息或者数据结合起来进行推理的模型。GPT4的逻辑性和准确性也有了极大提高，它在 SAT（美国大学入学考试）和 GRE 测试中都拿到了很高的成绩。

3月16日，微软宣布 GPT-4全面接入 Office，大概年底就会发布。GPT-4接入 Office 办公软件后，我们就可以用自然语言要求它完成各种办公需求，比如在表格上做统计、做数据变换等。



## 2023年3月23日：ChatGPT插件发布

- ChatGPT的应用商店来了！
  - OpenAI开始建立应用生态



3月23日，ChatGPT 发布了它的插件商店，相当于要建立一个以大模型为中心的应用生态。现在商店里面已经有了不少插件，可以通过语言模型和外部插件工具配合，完成很多复杂的任务。

4月，自主智能体的概念出现了。它的基本思路是将大模型看作一个人脑，只要对任务进行描述，然后设定一些目标，剩下的事情大模型都可以自主完成。其中，代表性的工作包括 AutoGPT，只需要对目标进行描述，它就会自主进行任务分解、执行、获取数据和分析等。另一项值得关注的工作是 Generative Agents，通过大模型模拟25个智能体在小镇的生活，会产生很多复杂的社会行为。

很多人觉得大模型的发展速度太快，可能会有很多潜在风险。于是在3月22日，马斯克等人牵头签署了《暂停大型人工智能研究》公开信。但我认为大模型的发展脚步已经停不下来了，只会越走越快。

4月28日，政治局会议中首次提出了“通用人工智能”一词，指出“要重视通用人工智能发展，营造创新生态，重视防范风险”，说明我国对此事是高度重视的。

### 三、大模型的深刻影响

#### 1、大模型可能带来的风险

##### 1) 虚假信息风险

大模型的造假能力非常强，我们难以辨别它生成的文字和图像的真实性。不过，现在有很多关于识别这种机器生成内容的研究。

##### 2) 价值观风险

任何大模型都有价值观倾向，比如 ChatGPT 就是西方左翼的立场。它之所以具有这样的价值观，是因为它学习的数据里大部分都是这样的观点。此外，在价值观对齐的时候，标准答案是以什么样的价值观来引导的，也决定了之后的模型会遵循什么样的价值观。

##### 3) 侵权风险

事实上，现在已经出现了这个问题。很多人起诉 OpenAI，因为它未经许可就用了他们的数据，而对此法律上仍没有界定。此外，还有包括著作权主体在内的法律、制度、伦理道德等方面的诸多问题。

#### 2、大模型可能产生的影响

##### 1) 对产业界的影响

当前各个科技企业都参与到百模大战之中，有望引发新一轮产业革命。

## 2) 对学术界的影响

学界基本公认大模型是一个重大的技术突破，我们已经看到了通用人工智能的曙光。很多人工智能领域的研究，包括人工智能+的研究，都在发生根本性的变化。

## 3) 从“不能用”到“能用”

长久以来难以解决的关键技术问题，比如语言理解、世界知识表述、复杂问题求解等，现在都有了巨大的突破。不仅所有传统任务都可以被显著改善，同时还打开了新应用的大门。

## 4) 从“专用”到“通用”

大模型的泛化能力非常强，现在所有任务都可以转化为用自然语言表达、基于知识的问题求解。由于语言知识和问题求解能力的大幅提升，大模型变成了一个通用问题求解器，这为通用人工智能打开了一条可能的路径。大模型的泛化能力能够降低开发成本、提高开发效率，还可以在多任务之间进行共享，影响会非常深远。

## 5) 从“体力劳动”到“脑力劳动”

现在的大模型就相当于一种人造大脑，它什么都知道，任劳任怨、一目千行、反应非常快，但有时候不太靠谱。如果在各个方面进一步提升，可能会出现一种超级智能，在很多能力上远超人类，比如我们做一些复杂问题求解时，只能调动很少的知识，而大模型可以调动很广泛的知识来做分析归纳和决策。由于看到了太多东西，它甚至可能会表现出一定的创造性，产生很多奇思妙想。在此背景下，大量重复性、知识依赖性的脑力劳动将被替代。

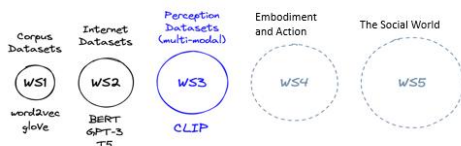
## 6) 从数字世界到物理世界



## 影响：数字世界 → 物理世界

### • 大模型的世界视野 (World Scopes)

- 当数据集的范围变得更大，基于它们训练得到的模型能力变得更强



### • 大模型正在迅速走进物理世界

- 多模态、工具使用、具身智能、自主智能体

### • 硅基智能体，硅基生命？

当大模型有了多模态的能力，能够进行工具使用，有了机器身体，能够自主规划完成任务，它将成为一个真正的生命体进入物理世界。

## 7) 自然科学实验

卡内基梅隆大学用大模型做了智能代理和自主智能体的实验。基于大预言模型的智能代理系统，可以自动设计、规划和执行科学实验，最后合成布洛芬、阿司匹林等。

## 8) 社会科学实验

我们做了一个模拟器，通过大语言模型赋能智能体，智能体可以进行各种角色扮演，并进行交互，产生很多复杂的社会行为。对于以前的社会科学难以进行的实验，可以通过大规模地采用智能体来模拟人类社会行为进行实验，这将会对社会科学产生深刻影响，使其变成一门实验科学。

之所以要开发出我们自己的玉兰系列大模型，是因为我认为人工智能必须要 Know-how，要知道整个大模型从头到尾的所有细节。我们将自己的大模型作为研究平台，支撑了很多人工智能领域的研究和实践。

总结来说，尽管大模型还存在着很多的局限性，但它为通用人工智能指出了一条可能的发展路径。人类经过六七十年的艰苦探索，终于找到了这样一条通用人工智能的可



用路径，一定要去拥抱它。它可能是一种新的超级智能，对于各个领域必将产生深刻的影响。

论坛第二单元，结合CMF中国宏观经济专题报告，各位专家围绕“大模型未来的发展方向、人工智能的发展对经济的意义”等问题展开讨论。

复旦大学人工智能创新与产业研究院院长，阿里巴巴原副总裁漆远指出，大模型不仅能更好地泛化，而且能更好地推动产业发展，提升用户交互水平。未来的发展方向可能是在场景中找到真正需要的产品，将产品与算法和工程化真正结合起来。

人工智能的发展可能会改变其应用落地的模式，未来，人工智能在金融、医疗、教育、游戏设计等行业应用落地的可能性非常大。从产业发展趋势上看，大模型与互联网行业发展的趋势类似，当前正处于商业模式探索阶段。实际上，它是一个生态系统，需要产业、高校、服务平台和资金的深度融合。

望正资本全球宏观对冲基金董事长刘陈杰认为，中国经济在疫情后逐步复苏，但未来一个阶段将面临来自房地产和地方政府投资两方面的趋势性压力。从基本面来看，当前实体经济投资回报率低于融资成本，房地产和地方政府投资这两个过去推动经济增长的主要力量缩减之后，提升实体经济投资回报率和提高劳动生产率就需要在供给层面下工夫，特别是需要开展新一轮供给侧结构性改革。人工智能将成为供给侧结构性改革2.0版本，特别是在全要素生产率方面，人工智能将发挥不可替代的作用。

从时间维度上看，由于人工智能的准确性有待提高，因此对各个行业的替代是有时间层次的。预计它在准确性要求较低的行业和场景将落地较快，对准确度要求更高的应用和行业则需要进一步调试和时间，中短期内，AI将更多地作为辅助工具。

对于大模型应用落地的三点判断：1) 本轮人工智能创新周期可能快于互联网周期；2) 人工职能的硬件、软件及应用将在产业领域不断深化，应用生态将不断拓展，从而系统性推动人工智能大模型的发展；3) 人工智能的发展可以分为三步，第一步是完成

硬件基础设施建设，第二步实现重要应用落地，第三步是将其广泛应用于生产生活的各个环节中，当前中国正处于向第二步迈进的阶段。

北京大学计算机学院教授、人工智能研究院副院长**黄铁军**指出，大模型以联结主义和神经网络学派为基本支撑，试图建立一个神经网络，能够通过给出训练数据，建立输出和输入间的映射关系。大模型的特点包括规模大、涌现性和通用性。当网络参数达到数百亿时，大模型就成为了一个复杂系统，会像其他的物理系统一样产生“涌现”现象，出现融会贯通的能力。

大模型开启了智业革命的时代，纵观人类历史，只有工业革命和电力革命可以与之相提并论。未来3年，视觉、听觉、具身、行动等通用智能的技术路线也会出现，并且影响会更大；未来10年，智力革命会广泛普及，它将构建一个全新的生态体系，而开源开放将成为联合各方的基本方式；未来30年，智能将发展为以时空环境驱动的具身智能，能够进行实时感知、实时决策、实时行动。

浙商证券首席经济学家**李超**指出，本轮反垄断过程促进了产业科技革命的过程。人工智能企业将成为未来十到二十年间反垄断的最大受益者，这有助于人工智能领域的技术发展。

人工智能驱动的科技革命将极大降低主权债务危机和战争的风险。从微观层面来看，人工智能将提升企业资本回报率，解决企业层面的债务危机；从宏观层面来看，以人工智能为代表的科技革命将推动经济发展，使企业能够积累更多财富，进而增加国家税收，极大地降低主权债务风险。总体而言，人工智能为代表的第四次工业革命可能使经济全要素生产率呈现前期稳定、后期明显提升的趋势。

京东集团副总裁**沈建光**指出，京东以数据供应链为核心优势，信息链条非常长，在此背景下，人工智能大模型的应用可以显著提高产出并降低成本。当前，京东拥有言犀人工智能开发计算平台，可以在包括宣传、营销、物流控制等各个方面为中小微企业提供高效、低成本的AI产品和服务。未来，诊断可能会成为人工智能和大数据模型非常有前途的应用方向之一。

# 黄铁军：智力革命已经打响，开源开放生态终将胜利

黄铁军 北京大学计算机学院教授、人工智能研究院副院长

以下观点整理自黄铁军在中国宏观经济论坛（CMF）宏观经济热点问题研讨会（第 73 期）上的发言

## 一、大模型的特点

经过六七十年的发展，人工智能形成了很多技术路线，大模型是其最新进展。大模型以联结主义和神经网络学派为基本支撑。联结主义学派的发展始于上世纪四十年代，他们试图建立一个神经网络，能够通过给出训练数据，建立输出和输入间的映射关系。比如人脸识别，输入的是照片，输出的是人的 ID 或者名字。从数学上讲，在神经元足够多的情况下，只要是三层神经网络，任意输入、输出之间的映射关系都可以建立。但问题在于，要建立一个复杂庞大的映射网络，需要的连接数量也是巨大的。

例如，在一个简单的三层神经网络中，假设每层有100个神经元，则会有1万个数据样本，迭代1万轮，要想收敛到这样的一个分类识别系统，需要2万亿次调整。这就是大模型训练需要耗费大算力的原因，只有将连接权值调整到合适的状态，它才能按照预想方式工作。

大模型主要具有以下几个特点：

### 1、规模大

关于什么是大模型有很多争论，现在基本收敛到一个共识，即网络参数至少要达到百亿规模，一般需要几百亿的规模才会产生预想的智能。

### 2、涌现性

传统 AI 只能给出封闭、确定的答案，而大模型不仅规模要大，而且一定要能涌现出预料之外的新能力。当网络参数达到数百亿时，大模型就成为了一个复杂系统，会像其他的物理系统一样产生“涌现”现象，出现融会贯通的能力。虽然现在还无法解释其原理，但这种能力确实存在，这也是最让大家最兴奋的地方。

### 3、通用性

大模型的应用不限于专门的问题和领域，这从经济角度来说是非常特别重要的。如果大模型能够解决通用问题，它就具有了很好的边际价值和推广性。

北京智源人工智能研究院自2018年成立后，很快切入大模型的赛道，是国内最早布局大模型的机构，聚焦于探索大模型背后的技术体系。大模型这个中文概念也是2021年智源研究院发布1.0版本的大模型时提出来的。

2021年6月，我们思考了一个问题：大模型到底意味着什么？当时我们提出，如果要让大模型的能力足够强，服务足够及时，就一定要尽可能地吸收全量数据，建立24×7训练、24×7服务的运营体系。这样的体系在全球范围内屈指可数，不会超过五个。因为大模型归根结底是一个数字形式的智能服务，存在一定的垄断属性，再加上对大规模和实时数据的要求，运营难度是非常大的。

2022年2月，智源开始研究谁有可能拥有这种运营能力。当时美国普遍认为是 Google，或者是 OpenAI+微软。去年11月 ChatGPT 发布，OpenAI 占据了先发优势，不过 Google 也紧随其后。当前中国已经有十多个模型在运营了，但我认为这只是短期的状态，长期内会收敛到最多两三个。

## 二、大模型对于人类的意义

大模型开启了智业革命的时代，纵观人类历史，只有两件事可以与之相提并论。

相较工业革命而言，如今我们面临的是全新的革命，即智业革命。工业革命替代人的体力，而智业革命替代人的智力，因此它可以与工业革命相提并论。

相较电力这种革命的产业形态而言，电力让能源以流通的方式进入家庭和企业，智力革命也是如此。大模型相当于发电机，它不是产业中最主要的产品形态。在未来的几年甚至几十年内，智力运营系统将会源源不断地为社会提供智力。

今年，国家成立了国家数据局，这是因为智能的源头是各种各样的数据。但数据只是一个生产要素，要将这种生产要素转换成经济社会发展的动力还需要依靠智力，并通过大模型训练智力。因此，将来可能会出现像国家电网一样的国家智网或国家智云，作为基础设施，为千行百业、千家万户提供服务。

自2021年起，智源研究院的研究重点从一个个具体的大模型转向通过协同创新的方式，为这个时代构建一个先进的技术体系。智源的 FlagOpen 飞智大模型基础软件开源系统是由智源牵头，多家共建的大模型开源开放软件体系，将进一步推动智力革命。

总结来说，当前 ChatGPT 已经拥有了很强大的智能。未来3年，除了语言模型会被广泛使用之外，视觉、听觉、具身、行动等通用智能的技术路线也会出现，并且影响会更大。

未来10年，智力革命会广泛普及，就像工业革命解放的是体力，电力革命解决的是能源流通的问题一样，它将构建一个全新的生态体系，这需要全社会共同努力，而开源开放将成为联合各方的基本方式。

未来20年，AI 技术还会快速发展。目前以 ChatGPT 为代表的是以数据驱动的静态的智能，而未来的智能将是以时空环境驱动的具身智能，能够进行实时感知、实时决策、实时行动。如此来看，当前的神经网络还有很大的进步空间，它将不再是一个 Transformer，而是类似于人脑的脉冲神经网络。我认为，到2045年，无论是从物理载体还是从能力上来说，超越人类的电子大脑和智能系统都会出现，经济发展将迎来一个新的黄金期。与此同时，这也将带来更大的安全问题，甚至比电力时代要大得多。因此，要同等重视 AI 安全和人类面临的风险。

## 漆远：大模型发展与落地思考

漆远 复旦大学人工智能创新与产业研究院院长，阿里巴巴原副总裁

以下观点整理自漆远在中国宏观经济论坛（CMF）宏观经济热点问题研讨会（第73期）上的发言

### 一、大模型的优势

大模型的出现，如 ChatGPT，让我们看到了强人工智能（AGI）的曙光。微软研究院的一篇文章《Sparks of AGI》也对此进行了深入的探讨。对于那些长期从事人工智能研究的人来说，当他们第一次看到 ChatGPT 时，无不感到震惊。因为它的表现超越了我们的期望，让我们看到了通用人工智能的影子。在过去，我们需要花费大量的时间和精力去训练不同的小模型，为解决一个个小任务。而现在，这种情况正在发生改变。大模型不仅能更好的跨领域泛化，而且能更好的推动产业发展和提升用户交互水平。

人工智能的发展可能会改变其应用落地的模式。以往，我们衡量一家公司的人工智能能力，往往是看它拥有多少个小人工智能模型。因为这些模型通常是针对特定场景进行定制的。例如，我们可能需要查看一千万张猫脸或狗脸的图片，才能精确地将猫脸和狗脸进行分类。这种模式需要大量的数据来进行特定模型的定制化研发，因此成本非常高。这容易使人工智能的应用实施变成一种外包模式，极大地限制了人工智能真正落地推广的能力。

大模型的出现则改变了这一状况。它具有很强的场景适应能力，可以通过学习上下文或少数样本的方式，使用少量的训练数据，做到快速泛化。简单地说，大模型（也称为大语言模型）随着参数的增加，其预测能力得到了质的飞跃，这种能力就是强的泛化能力。这意味着，我们或许能摆脱人工智能的人肉外包模式和定制化模式，走向真正产品化模式。

因此，生成式 AI 发展地非常迅猛。据统计，全球与生成式 AI 相关的企业估值去年已经达到480亿美元，并且仍在持续增长。2022年的投资额达到21亿美元，是2021年的

10倍，OpenAI 便是其中的核心代表。其中，主要有三类企业。第一类是开发基础模型的企业，第二类是开发应用程序的企业，第三类则是建立基础设施来帮助开发和部署模型的企业。这些公司及其背后的技术都在高速发展。

以一些具体的例子来说，Notion 原本是一家 SaaS 公司，主要帮助人们编写文章，但现在它发展迅速，成功地将生成式人工智能与协作和文档等多种能力结合在一起；Bloomberg 发布了金融领域的第一个大模型；Jasper 也依托 OpenAI 的 GPT 模型构建了一套应用能力，迅速发展。

## 二、大模型时代的机遇

生成式 AI 在中国也发展地十分迅速。如阿里、百度、讯飞等大公司，以及许多初创公司都已经进入了这个领域。然而，随着越来越多的大模型的出现，我们不禁要问：它们的优势到底在哪里？真正的应用价值和产业价值在哪里

### 1、有护城河吗？

人们常常会问，这个领域是否存在护城河呢？（是否有供应链、是否有生态、是否有数据？）目前看来，这个领域发展地非常快，还没有形成系统性的护城河。公司规模是护城河么？不一定是。例如，在 ChatGPT 出现之前，我们从未想过，像 Google 这样的大搜索公司会受到一家创业公司的冲击。尽管 OpenAI 很出名，但与 Google 相比，无论是在工程师数量、系统复杂性还是产品影响力上，都不在一个量级。可以看到，生成式人工智能可以让一个相对初创公司对大规模公司产生非常有力的冲击。微软和 OpenAI 的结合又打造了一个非常优秀的联盟。

在这个领域中，新的算法层出不穷，但随着算法的发表和开源，单个算法创新不是护城河。数据非常重要。从数据角度而已，生成式人工智能是在收集整个互联网的数据做一个压缩。传统的网上公开数据已经不再构成护城河。

未来的发展方向可能是在场景中打造真正需要的产品，将产品、算法和工程系统真正地结合起来。OpenAI 的创始人在采访中提到，ChatGPT 的成功并不是一个单点技术的成功，而是一个系统化的结合。

## 2、价值将在哪里积累？

当前，生成式人工智能总营收的10%-20%将流向云服务提供商，而应用程序公司平均将20%-40%的年收入用于推理和定制化的微调，同时模型提供商也将一半的收入投入到云基础设施上。在这个过程中，最大的赢家是生产硬件的公司。因为，无论是哪一家公司，它在快速迭代和探索中，都会有一个共同的依赖，那就是底层对 GPU 卡的需求。就目前来看，直接看到的最大收益者是 GPU 的生产商和云厂商。

## 3、谁与争锋？

### 1) 大模型技术能对企业产生重大影响

一家值得关注的公司是 Midjourney。这家公司在2021年成立，不到一年的时间就吸引了千万级别的用户，并在去年实现了过亿美元的营收。这家公司的员工数量极少，今年上半年只有11名员工，其中4名本科在读，8名研发工程师，另外3名员工负责法务、财务和后台工作。然而，这家公司产生的图片质量却非常高，覆盖了建筑设计、平面设计、用户界面设计和技术创作等多个领域。这家公司打破了“规模决定一切”的传统观念，向我们证明了少量的员工也能产生巨大的影响力和价值。从这个案例中，我们可以看到大模型技术能对企业发展产生非常大的影响。

### 2) 大模型的产业应用

国内市场也正经历着非常迅速的发展。预计在2022年至2025年间，其年化增长率将超过40%，这是一个惊人的数字。人工智能未来在金融、医疗、教育、游戏设计等行业应用落地有不少空间。

### 3) 大模型的应用边界



在过去的几个月里，大语言模型的应用边界得到了极大的拓展，其中一个焦点是代理（agent）。代理是指在大语言模型基础上增加规划、反馈和使用工具的能力。它作为大模型与场景间价值传递的桥梁，能极大地拓展大模型的应用边界，使其成为一种系统性的超级应用。

只有大公司才能做大模型吗？我认为并不是这样。小模型可以知识蒸馏。虽然现在大多讨论的是千亿模型参数、万亿模型参数，但我们发现，在某些特定领域，百亿、甚至十亿级的模型参数就能取得很好的效果。此外，国产替代也正在逐渐发挥作用。

#### 4、大模型作为生产力工具的挑战

然而，大模型在真正转化为生产力的过程中会面临一系列挑战。

首先，在生成能力上，如何实现可信可靠，避免出现“幻觉效应”？其次，对于复杂的应用场景，如何实现信息的动态集成？这包括代理如何调用外部数据库？大模型本身如何快速、动态地集成和更新？这些都是技术领域的挑战。此外，海量数据的获取和算力方面也需要重点关注。据 MIT 团队预测，在2026年，我们可用的公开数据将全部耗尽。那么，未来的数据将从何处获取？如何获得大规模的算力支持？这些都是需要解决的问题。

这里我谈谈大模型在精确性、可信性方面可进行优化的方向。首先，结合知识图谱与注意力机制能大幅提高模型的数值推理的准确性。例如，OpenAI 官网列出了 GPT4 在一系列真实数据集上和其他团队最好结果的比较——GPT4 几乎超越了其他所有工作。但在阅读理解和数据推理的任务 DROP 上，GPT4 落后于我们团队以前的 QDGAT（Question Directed Graph Attention Network）工作。该工作没有采用完全自回归的注意力模式，而是结合了知识图谱与注意力机制来提升数值推理准确性。我们认为在许多任务中，模型的大小并不是唯一重要因素。如果我们能够进行深入研究并结合其机制，也能取得良好的效果。其次，在精准性方面，虽然“文生图、图生文”的 AI 应用已经十分普遍，但文本生成的内容仍然难以控制。在实际产业应用中，控制文本生成的精准性非常关键。例如，为金融监管机构编写一个自动生成的文本时，我们需要解释为什么我们认为某一

笔交易存在洗钱风险。此时，我们将 Transformer 神经网络与基于合规知识的逻辑推理结合，可以更加可控地生成文本。

### 三、大模型的应用和发展方向

#### 1、产业发展趋势

从产业发展趋势上看，我们发现大模型与互联网行业发展的趋势类似，当前正处于商业模式探索阶段。

#### 2、垂直领域应用趋势

在垂直领域应用趋势方面，大模型在法律、咨询、金融、医疗以及更广泛的科学领域，都有广泛的应用。在科学领域，今年7月，马斯克成立了 xAI 公司，目标是解答更深层次的科学问题、利用 AI 帮助人们解决更复杂的科学和数学问题，甚至理解宇宙。

##### 2) 法律行业

路透社今年收购了 Casetext，这是一个能够分析法律文书的工具，可以快速理解文书、整理文书、进行案情分析和案件对比等，具有很高的实用价值。

##### 3) 咨询行业

在咨询行业，由于其信息密集的特性，咨询服务通常以信息为基础，结合行业场景需求进行推理呈现。这个领域很可能直接受益于大模型的能力，比如知识总结要点、查询数据库等。因此，麦肯锡推出了 Lilli 产品，旨在帮助我们进行战略分析。虽然最初主要依赖人力，但随着模型能力的提升，大模型将会发挥更大的作用。

##### 4) 金融领域

在金融领域，例如微软与伦敦交易所的合作，已在新闻摘要等方面取得了一些进展。以下有两个例子。

例如，我们可以询问大模型“美国 CPI 核心通胀率高于 4% 时，市场发生了什么？”大模型将会根据历史上的情况，对美国财政和金融采取的措施及其达到的效果进行解释。如果我们再问：“如何进行资产配置？”它也会给出一些相当好的建议。

另一个例子是在财务报告分析方面。比如我们分析非常长的茅台公司财务报告，可以向大模型询问其经营状况。基于这份两百页的报告，大模型可以立即进行总结，包括经营状况、现金流、未来归属等。我们还可以请大模型对财务报告中的一个或多个报表进行评论和总结。以前，这种工作主要依赖人工，而现在通过人工智能，我们大幅提高工作效率。

## 5) 医疗领域

医疗领域也是大模型非常适合的应用场景。比如，在问诊时，如果我说我感到口渴、疲劳和视力模糊，大模型可能会回答我可能患有糖尿病。如果我再问糖尿病有哪些类型，它会告诉我有一型和二型。如果我问做哪些诊断能够确诊，它会根据相关的医疗知识给出更精确的回答，推荐体检检查项目等。

此外，大模型还可以帮助我们理解体检报告。比如，如果你的体检报告很长，很多人看完后并不理解报告真正的意思，需要注意什么。大模型可以很好地帮助我们理解，比如血糖高、甘油三酯高意味着什么，某项指标或者多项指标合在一起它怎样解读。这能为广大老百姓在理解体检报告方面提供很大的帮助。

## 6) 制药领域

我们可以将化学分子视为一个序列，然后将其输入到 Transformer 架构中做大模型预训练和制药 ADMET 的多任务学习，比如在 ADMET 中预测药物水溶性、穿膜性、毒性等。

## 7) 气象学

此外，大模型也与气象学有关。我们可以将天气数据进行标记化 (tokenization)，并利用大模型的无监督学习能力进行分析。准确的气象预测不仅对我们的日常生活重要，

对新能源领域，如风力发电和光伏发电也有重大影响。我们利用复旦 CFFF 集群构建了伏羲气象大模型，首次实现了精准的 AI 15 天中期天气预报，预报结果与欧洲气象台的预报结果相当，但预报速度从小时级提高到了十秒内，实现了千倍的加速。

我们不仅在研究科学大模型，同时计划通过复旦大学的 CFFF 计算平台，让更多的科研人员能够参与到科学大模型和科学智能的研发中。为此，我们发起了世界科学智能大赛，在生命科学、大气科学、材料科学、流体力学、量子化学等领域开放数据，并提供计算能力支持。目前，已经有上万个团队参加，取得了许多超出我们预期的有趣结果。我们希望未来能在更多的领域发展和应用人工智能，推进制药、智慧医疗、健康管理、碳中和等领域的智能化升级。

#### 四、展望

大模型的研发如今才刚刚开始。它实际上是一个生态系统，需要产业、高校、服务平台和资金的深度融合。科技创新需要有长远的目光和良好的心态，如果一定要成功，那就不是创新了。我们需要进一步探索技术、工程和产品的结合，推进人工智能的发展与落地。

我曾看到一个很有趣的预测。到 2030 年，人工智能一天能学完人类 2500 年的知识。这意味着未来的发展将超越我们最大胆的想象。我们需要更深入地思考和准备迎接未来。

# 沈建光：大模型在经济活动中的应用和前景

沈建光 京东集团副总裁

以下观点整理自沈建光在中国宏观经济论坛（CMF）宏观经济热点问题研讨会（第73期）上的发言

## 一、京东大模型早期探索与实践

我加入京东的一个重要动机是利用大数据宏观研究来探索传统宏观研究的新变化。我想了解，如果有大数据作为支撑，宏观研究会发生怎样的变化。当时，京东就开发了一个系统——人工智能经济学家系统。虽然它只是一个雏形，并不能与现在的人工智能大模型相媲美，但仍十分有意义。当时的设想是基于大数据和人工智能机器学习模型，对消费大数据进行消费预测，以及对与经济活动和消费相关的各种因素，如收入预期，进行研究。后来，我们还在其他方面聚焦产业数据进行了实践，这些实践非常有意义。

如今，人工智能大模型已经在各个领域得到了广泛应用。从京东的角度来看，企业如何在大模型和通用模型的基础上，在许多对经济活动有最直接、最深远影响的垂直领域进行应用呢？我将结合京东的实践，谈谈我对此的看法。

## 二、大模型能提高产出、降低成本

京东是一家供应链科技公司，我们以数据供应链为核心优势。我们为数千万种商品提供服务，拥有5000万种工业产品和800万家活跃企业客户。在全国范围内，我们涵盖了200多个产业带。这种规模非常值得研究。从产业链、供应链的角度来看（包括金融链和物流业），我们的信息链条非常长、场景也非常多。在这种背景下，人工智能大模型的应用可以显著提高产出并降低成本。

目前，我们拥有言犀人工智能开发计算平台。我们不仅自己应用这个模型来进行价格设定和产品选择，还帮助大量的中小微企业。具体而言，我们利用京东已经训练出来

的模型，直接为中小微企业提供高效、低成本的 AI 产品和服务。这就是京东言犀正在做的事。

### 三、大模型在营销领域中的应用

同时，如何将如此众多的产品匹配给最感兴趣的客户呢？我们的营销平台上有很多关于产品比较和介绍的内容。在过去，这需要大约七天的人工制作时间。而通过使用人工智能营销平台、基于大数据模型进行训练，现在每套图的制作成本只需要半天，并且使成本降低了90%。这是一个非常明显的例子，展示了人工智能和大数据如何在产业上提高效率并降低成本。这种影响不仅在宏观层面，在我们企业的实践中也很明显。

在营销方面，我们还有一个重要的应用是多模态数字人。数字人是指不再需要聘请明星或网红来进行直播，而是通过人工智能生成一个高仿的数字人。这个数字人根据客户的需求，在五分钟内就能生成。这项技术已经在四千家品牌商中得到采用。通过数字人，品牌商不再需要寻找人才或录制视频，只需将内容输入，人工智能数字人就能像真人一样推销产品给客户。这种方式也可以实现人机互动。

这也是我们言犀大模型的一个优势。通过过去二十年的沉淀和客户服务经验，我们将其转化为大数据和人工智能的模型。实际上，与客户对话的都是人工智能，但客户感觉不到它是一个机器。这样的数字人应用在许多商家的营销中，降低了商家的成本。正如之前许多专家所说，**人的知识库很难非常全面，但人工智能在这方面能够提供全面的知识，并具备许多特殊的手段。**当然，它也有一些局限性，但所有的限制都需要依靠垂直能力来克服。

### 四、大模型在诊断和物流控制中的应用

我们目前正在尝试将大模型应用在诊断上，这是人工智能和大数据模型非常有前途的应用之一。它可以对各种疾病进行各种专家的会诊，甚至可以提供一些建议。当然这最终还是需要医生来进行判定。

此外，京东的物流体系十分庞大。通过大模型，我们可以对物流进行控制，以达到最佳效果。这方面已经在产业界和企业界得到广泛应用。这也充分展示了大模型降低成本和提高效率的效果。一旦这成为一个产业，千万家中小微企业也能从中受益，并以低价享受大模型带来的好处和服务。

以上便是我从京东产业实践的角度，对大模型在经济活动中的应用和前景的思考。

# 李超：人工智能对宏观经济的影响

李超 浙商证券首席经济学家

以下观点整理自李超在中国宏观经济论坛（CMF）宏观经济热点问题研讨会（第73期）上的发言

## 一、反垄断利好人工智能

从2019年开始，全球主要经济体包括中国在内都开始推进反垄断的进程。我们认为本轮反垄断将促进产业科技革命的进程。数据平台垄断会降低数据本身的应用效率和公平性，基于目前反垄断的趋势来看，未来数据共享模式将得到发展，进而使得生产企业，尤其是人工智能这类最需要数据的生产类企业可以获取大量数据。因此，我们认为人工智能企业将成为未来十到二十年间反垄断政策的最大受益者，有助于人工智能领域的技术发展。

## 二、债务危机

目前全球积累了大量债务，债务本身是中性的，但是从它的积累方式看，可以分为正向积累和负向积累。当企业资本回报率大于负债利息时就是正向积累，此时企业会扩大再生产，发挥债务的规模效应。而当资本回报率小于负债利息时，企业将陷入发新债还旧债的困局，即负向积累。而当所有的微观企业都陷入负向积累时，国家税收无力为继也只能去举债。实际上，目前主要的大经济体基本都处于这一状态，即债务负向积累的晚期，这就容易造成以下几个负面影响。

### 1、负利率

一般来说，负利率的负面影响是最小的，它是金融市场自然选择的结果。债务双方进行妥协，降低利率到达一定程度就出现负利率，目前欧洲、日本等许多国家已经大面积出现负利率的情况。但是，疫情期间发达经济体普遍选择释放流动性，引发通胀，这就使得利率下降承压，此时就容易发生债务危机。



## 2、主权债务危机

自现代货币理论诞生后，各国央行均可采取无限释放流动性的方式救助各类市场的债务危机，如房地产危机、银行危机、企业信用危机、个人贷款危机等等。最后导致这些风险都聚集于国家层面，因此未来金融危机基本只有一种形式，即主权债务危机。主权债务危机是一种较为剧烈的债务出清，对经济会造成巨大的冲击。

## 3、战争

历史上欧洲就曾出现过主权债务危机，欧央行、IMF 可以选择继续采取释放流动性的方式进行救助，但是这种救助方式是有成本的。投放巨量流动性容易引发资产价格泡沫，而不同收入群体对资产价格泡沫的识别能力存在差异，进而会导致贫富差距增大。当这一差距达到一定极限时就会进入民粹主义，即政治上的异化，对外转移矛盾时就容易引发战争，战争毫无疑问是最惨烈的去债务方式。

## 三、人工智能破解债务负向积累

人工智能驱动的科技革命将极大的降低主权债务危机和战争爆发的风险。微观层面来看，人工智能将提升企业的资本回报率，破解企业层面的债务危机；宏观层面来看，以人工智能为代表的科技革命将通过提升全要素生产率，推动经济发展，使企业能够积累更多财富，进而增加国家税收，极大地降低主权债务风险。

我们认为本轮人工智能科技革命将广泛应用于生产生活领域，驱动第四次工业革命。我们类比分析了第二、第三次工业革命对推进全要素生产率提升的影响：第二次工业革命中，全要素生产率没有特别明显的提升，原因在于电网向生产生活的应用是碎片化的。而第三次工业革命，互联网和计算机在生产生活中迅速推进应用，全要素生产率出现了大幅提升，之后虽经历了如智能手机、iPad 等科技上的突破，但全要素生产率整体呈现出逐渐钝化的趋势。我们预测第四次工业革命的发展应用将与第三次相反：前期主要围绕单一领域应用，比如特斯拉的无人驾驶技术，之后再转向下一个领域，如无人厨房。不同领域间碎片化的突破可能不会对全要素生产率有明显提升，但发展后期如果出现关

键性的技术突破，如人型机器人，就有可能实现对生产、生活各领域的系统性快速推进，进而推动全要素生产率的明显抬升。总的来说，以人工智能为代表的第四次工业革命可能使全要素生产率呈现前期稳定、后期明显提升的趋势。

基于前面对工业革命的推演，我们认为全要素生产率提升的关键，还是要看人工智能技术突破的可能性。对中国而言，我们现在的经济结构调整也迫切需要全要素生产率的提升，这也将是中长期中国经济发展破局的关键所在。

#### 四、美元仍处于贬值周期

人工智能、大数据、云计算、物联网的萌芽推动美元进入了2011-2018年的上涨周期。但是美元指数上升的幅度又不够大，原因在于这些突破性技术无法应用于生产与生活。我们认为2018年底到2028年底这十年，正常情况下应该是美元的贬值周期。一方面，新兴市场竞争优势异化，越南、墨西哥利用人口红利承接产业转移，中国处于人口质量红利较强的阶段，从而在理论上反向拉出一轮美元的贬值周期。另一方面，美国如果没有快速的技术突破，大量高端制造业仍需要中国来生产，美元大概率则会是一个持续的贬值周期。

# 刘陈杰：人工智能是中国未来经济增长的新动能

刘陈杰 望正资本全球宏观对冲基金董事长

以下观点整理自刘陈杰在中国宏观经济论坛（CMF）宏观经济热点问题研讨会（第73期）上的发言

## 一、人工智能发展背景：旧有经济增长驱动力式微

中国经济在疫情后逐步复苏，但未来一个阶段将面临来自房地产和地方政府投资两方面的趋势性压力。2008年全球金融危机之后，房地产和地方政府投资两大增长模式拉动了80%以上的中国经济，对短期经济稳定意义较大。但目前来看，房地产市场面临新增刚需的趋缓风险，地方政府债务型增长则面临可能堆积的金融风险，中国过去的经济增长模式驱动力不足。

首先，地方政府债务扩张型增长模式式微。2008年金融危机后，地方政府债务驱动型的增长模式对中国经济的恢复和增长起到了重要推动作用。但是，通过存量模型测算，目前中国地方政府债务水平已接近70万亿。短期来看，地方政府债务的进一步扩张，将对中国金融稳定和经济健康发展产生影响。长期来看，中国独特的财政、金融体制和经济结构，使得政府债务增长的最大边界必然与国际存在差异，因此还需要进一步研究中国地方政府债务的最大边界。

其次，房地产刚需见顶。去年我们根据中国人口结构、家庭结构、收入预期、城镇化率、住房存量结构各方面因素，模拟得出包括改善型需求曲线在内的中国房地产刚需曲线。从曲线可以得出，2020-2021年的房地产销售为14亿平-16亿平，将是未来10-15年内的最高点。2030-2035年中国房地产需求平均为每年10亿平左右，且房地产刚需的平均水平呈现中枢下移趋势。基于这一结论，与地产相关的一系列财政体制、金融体制、行政体制、政府职能等等都需要发生一些变化。我们也看到，截至目前很多城市在政策方面不断放松，通过政策刺激可能促进房地产市场短期内好转，但刚需缩减、中枢下移的长期趋势仍是需要关注的问题。

从中国经济的基本面来看，当前实体经济投资回报率低于融资成本，而房地产和地方政府投资这两个过去推动经济增长的主要力量缩减之后，提升实体经济投资回报率和提高劳动生产率就需要在供给层面下工夫，特别需要新一轮供给侧结构性改革。我们认为人工智能将成为一轮供给侧结构性改革2.0版本，特别是在发挥全要素生产力方面人工智能将发挥特别重要的不可替代的作用。

## 二、人工智能将成为未来经济增长的新动力

基于新古典理论，将人工智能纳入一般均衡模型中，主要是通过三个方面影响经济增长潜在增速：第一，人工智能可以不断提高生产自动化和智能化的进度，实现资本对劳动的更高替代，减轻劳动力成本上升带来的成本的增加。第二，人工智能会进一步增强第二产业、第三产业资本深化的增加，进而提高资本的回报率，缓解人口结构红利式微背景下的投资率和储蓄率下降对经济增长的不利影响。第三，人工智能会通过改进生产方式、节约成本等方面促进全要素生产率提升，从而对冲其他要素禀赋的变化对经济增长的冲击。基于模型分析，保守来看，我们认为人工智能的发展能够一定程度上减缓未来十年经济潜在增速的下滑趋势。乐观来看，如果人工智能在未来实现爆发式增长，我们相信人工智能有能力支撑中高速的经济增长以及中国经济高质量发展。

## 三、人工智能对行业的影响

从时间维度上看，由于人工智能的准确性有待提高，因此对各个行业的替代是有时间层次的，我们预期在准确性要求较低的行业和场景将落地较快，对准确度要求更高的应用和行业需要进一步调试和时间，中短期AI将更多作为辅助工具。目前AI已初步应用于产品控制、生产优化、质量检测等环节，随着精度和熟练度的提升，之后会转向高精度要求行业，如金融投资、高端制造、智慧办公等等。

从劳动力替代的角度来看，一方面，人工智能可以高效完成重复性的工作从而节省大量人工，同时随着发展作用于更广泛的行业领域，如法律、医疗、教育、金融、交通、娱乐等等。因此，我们相信人工智能的应用将会对经济运行方式和人们的生活方式产生极大的革命性改变。另一方面，人工智能将会对历史上较难自动化的脑力劳动有较

**大影响。**生成式 AI 不但有较好的特定技能如语言、图像、音乐、数学和编程等，还有将不同技能组合从而更好认知和解决较复杂的问题的能力。因此我们认为 AI 的投入和使用将会替代一些传统意义上较难自动化的职能，如营销、研发和软件编程。

#### 四、人工智能大模型应用落地持续推进

对于大模型应用落地主要有三点判断：第一，本轮人工智能创新周期可能快于互联网周期。一方面，人工智能硬件壁垒低于同时期的互联网，另一方面，大模型发展初期的门槛也相对较低。第二，人工智能的硬件、软件及应用在产业领域不断深化，应用生态不断拓展，从而系统性推动人工智能大模型的发展。第三，人工智能的发展可以分为三步，第一步是建设完成硬件基础设施，第二步实现重要应用落地，第三步则是人工智能广泛应用于生产生活的各个环节中去。现阶段产业内已集聚一系列代表性公司，并持续进行投入研发，据保守预测，到2030年人工智能硬件、应用及平台领域的年均增长将达50%以上，我们认为中国目前正处于向第二步迈进的阶段。



**把脉中国经济 传递中国声音**  
Taking Economic Pulse, Forcasting Economic Future

地址：北京市海淀区中关村大街59号 中国人民大学崇德楼西楼9层  
Add: 9th Floor, West Wing of Chongde Building,  
Renmin University of China, 59 Zhongguancun Street,  
Haidian District, Beijing, P.R.China

网站: <http://ier.ruc.edu.cn/>

微信公众号:

