

新媒体大数据分析

New Media Big Data Analysis

专题：自然语言处理与大模型

黄振亚，朱孟潇，张凯
Email: huangzhy@ustc.edu.cn, mxzhu@ustc.edu.cn

课程主页：
<http://staff.ustc.edu.cn/~huangzhy/Course/NM2025.html>

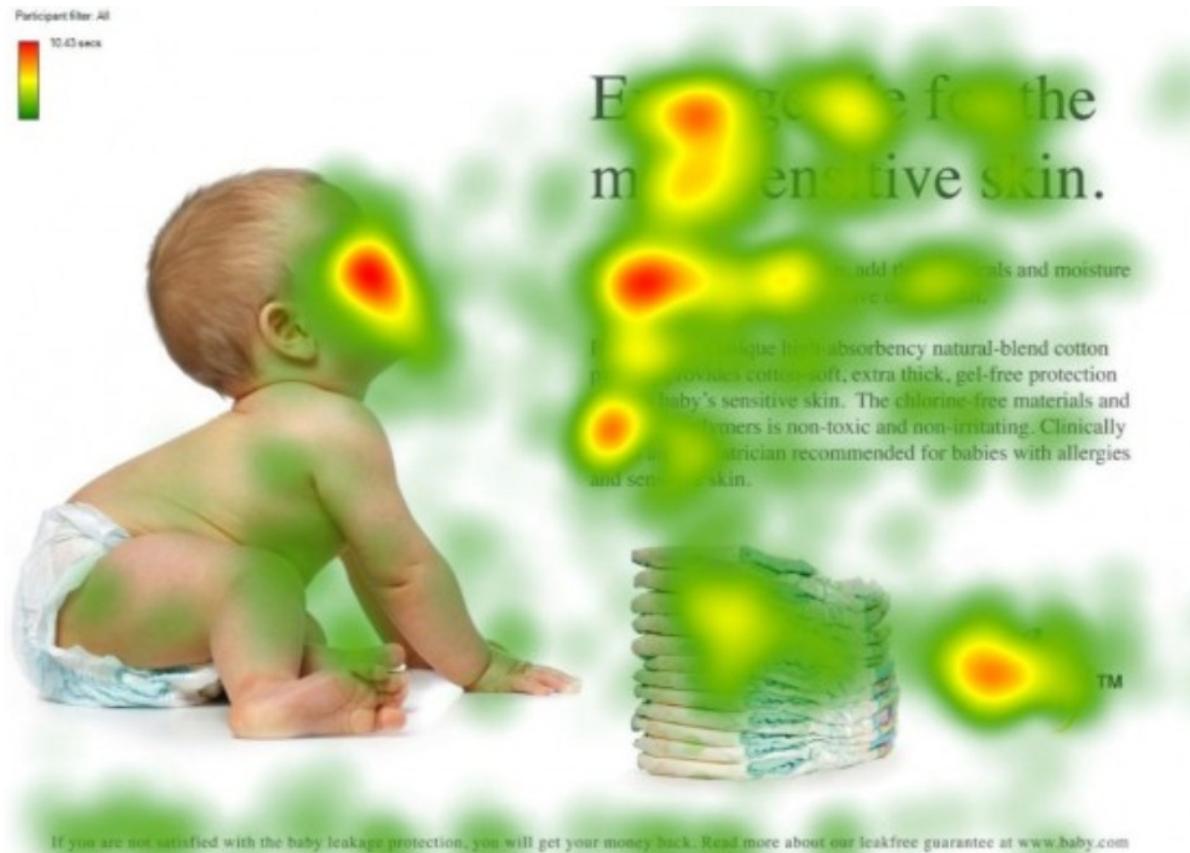
助教：齐畅，朱家骏
bigdata_2025@163.com

Outline

- 人工智能
- 自然语言处理发展
- 大模型的原理与技术
- 大模型应用
- 未来展望

自然语言处理

- 注意力机制

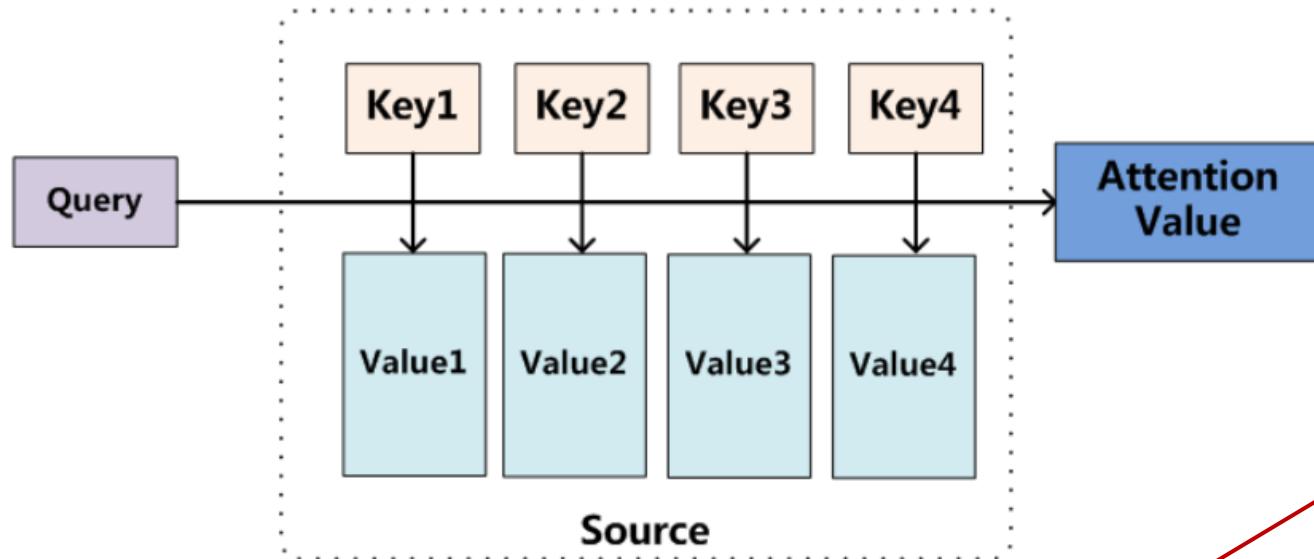


关注的顺序：婴儿的脸→文档标题→每一段的首句→其他

- 大脑视觉系统的注意力机制
 - ▶ 人往往不是“均匀地”看图像中的所有区域，当图像信息映入眼帘时，我们的大脑会**把注意力放在图像的主要部位上**，然后才去看其他部位。
 - ▶ 对于文本也是一样，当文本信息映入眼帘时，**大脑会首先记住重要的词汇、短语或者句子**。

自然语言处理

- 注意力机制



$$\text{Attention}(\text{Query}, \text{Source}) = \sum \text{Similarity}(\text{Query}, \text{Key}_i) * \text{Value}_i$$

加性模型

$$s(\mathbf{x}, \mathbf{q}) = \mathbf{v}^\top \tanh(\mathbf{W}\mathbf{x} + \mathbf{U}\mathbf{q}),$$

点积模型

$$s(\mathbf{x}, \mathbf{q}) = \mathbf{x}^\top \mathbf{q},$$

缩放点积模型

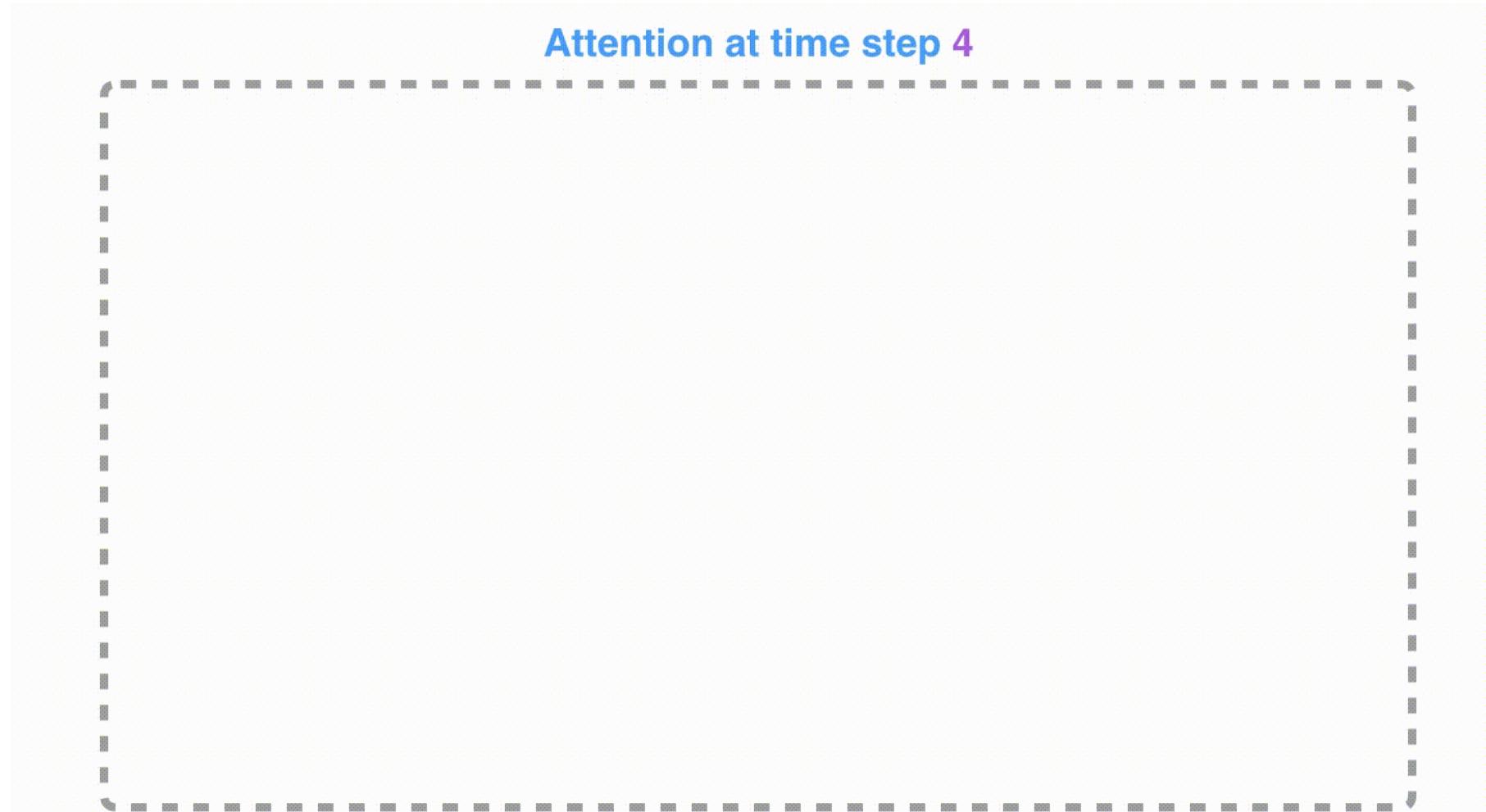
$$s(\mathbf{x}, \mathbf{q}) = \frac{\mathbf{x}^\top \mathbf{q}}{\sqrt{D}},$$

双线性模型

$$s(\mathbf{x}, \mathbf{q}) = \mathbf{x}^\top \mathbf{W}\mathbf{q},$$

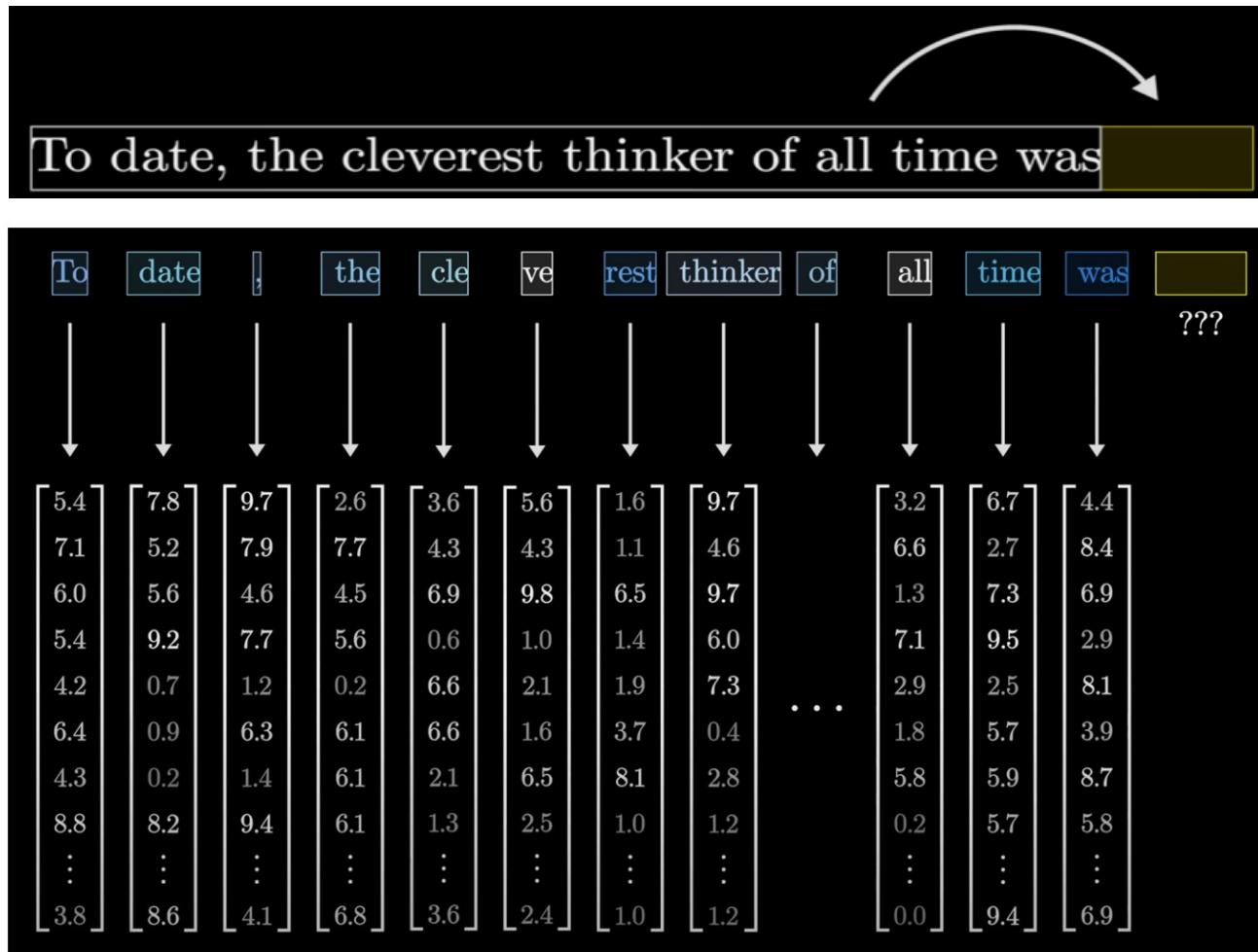
自然语言处理

- 注意力机制



人工智能技术原理—自然语言处理

- 标志技术: Transformer (Google Brain, 2017)



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

1 Introduction

Recurrent neural networks, long short-term memory [12] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and transduction problems such as language modeling and machine translation [29, 2, 5]. Numerous efforts have since continued to push the boundaries of recurrent language models and encoder-decoder architectures [31, 21, 13].

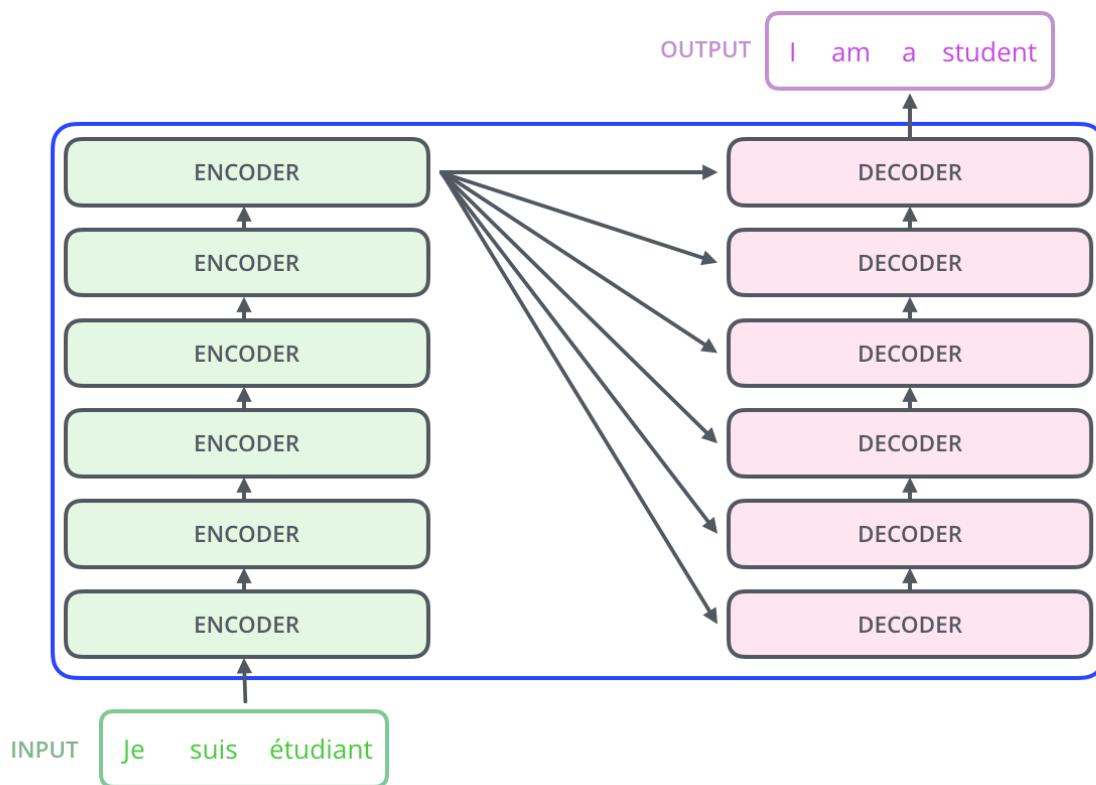
*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

神经网络方法

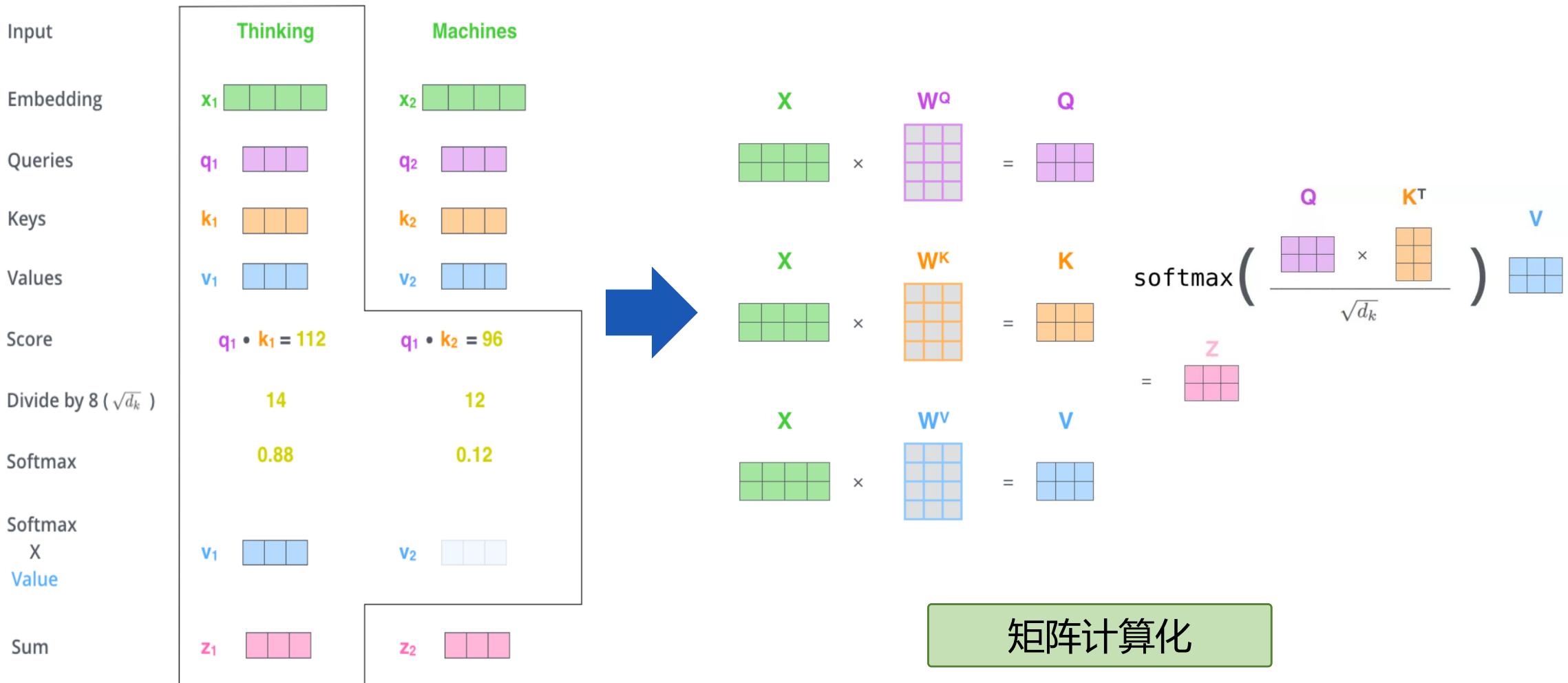
□ Transformer



Transformer

- Transformer Unit

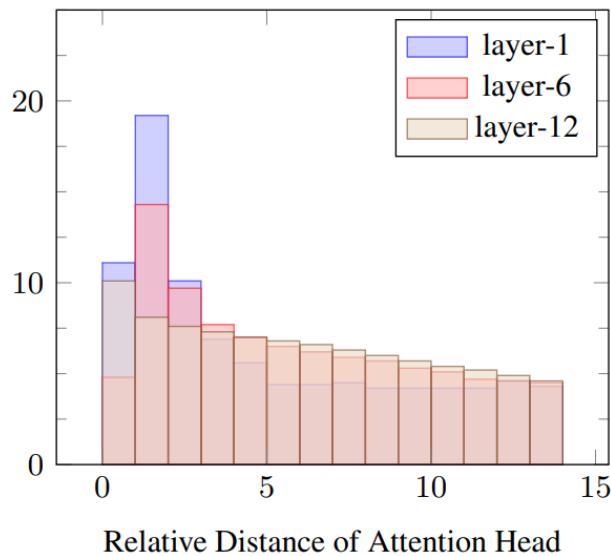
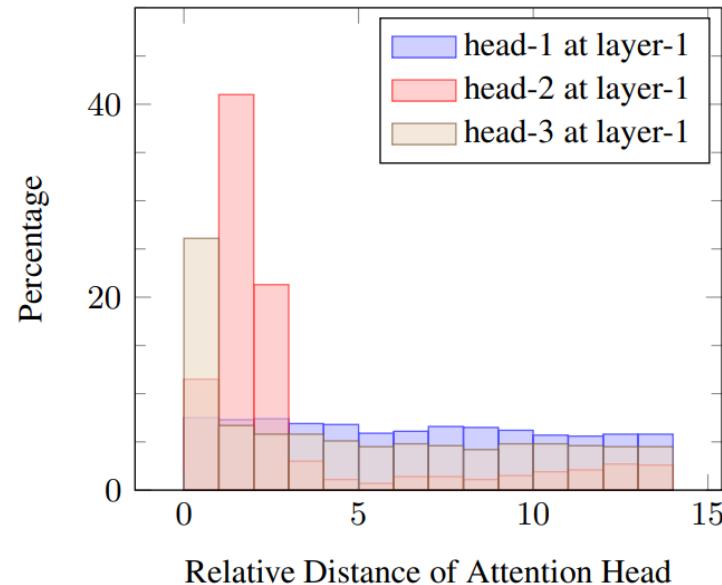
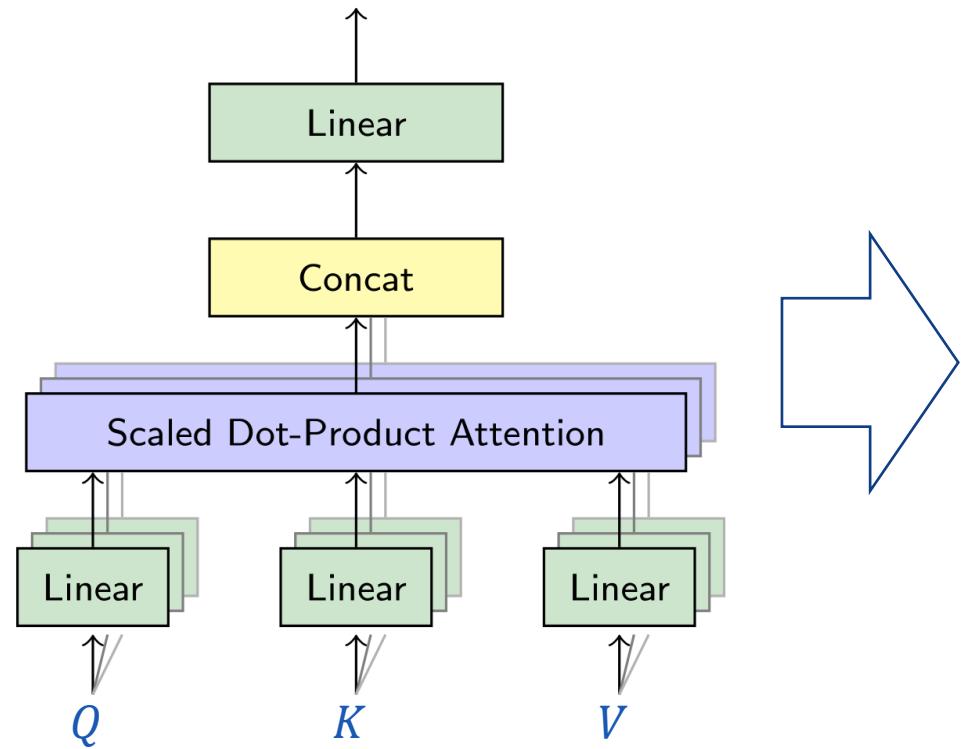
- 将输入的query、key和value沿着隐层维度切分为h个子集，分别进行注意力操作。



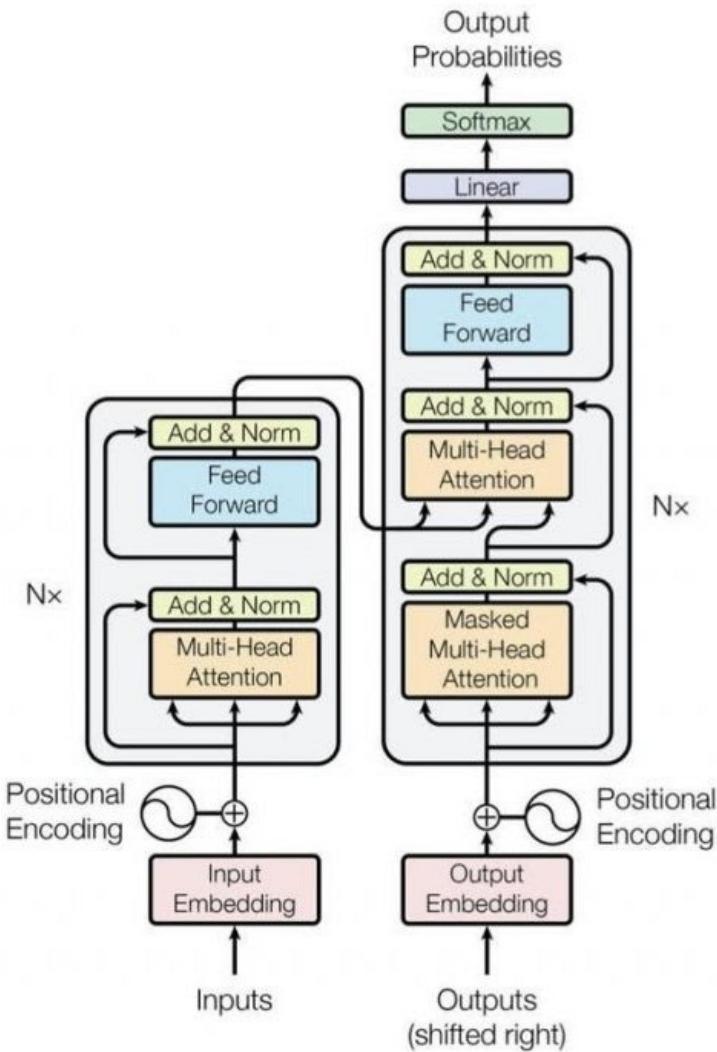
Transformer

- 多头自注意力机制

- 将输入的query、key和value沿着隐层维度切分为h个子集，分别进行注意力操作
- 希望模型在不同的子空间中**学习到更丰富的信息**

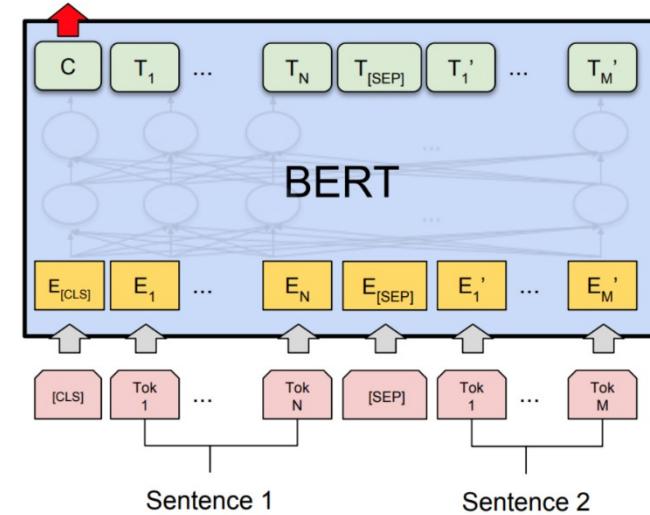


Transformer

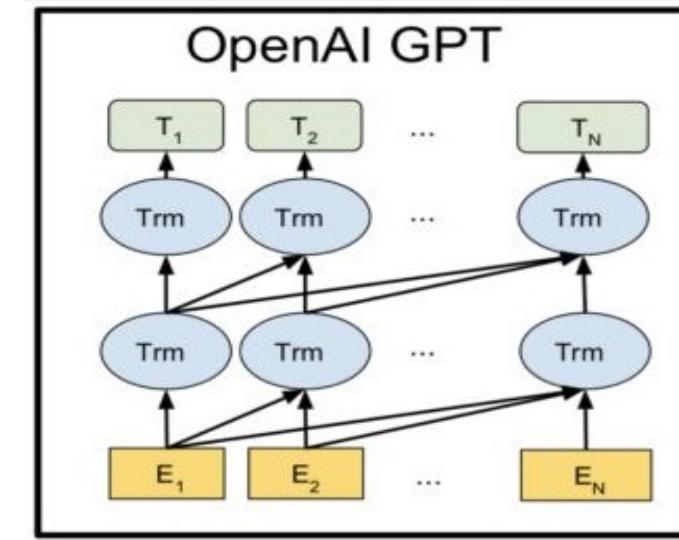


Transformer

自编码语言模型

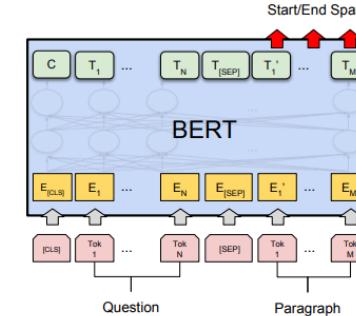
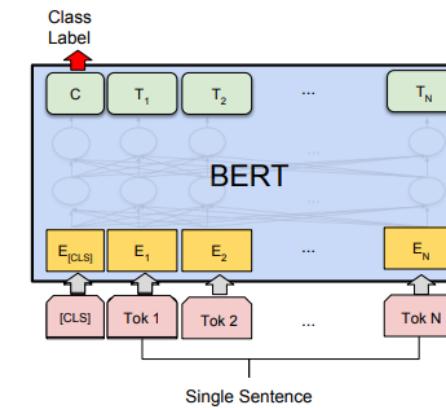
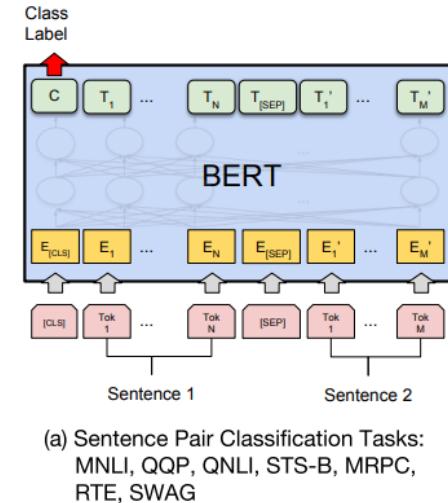
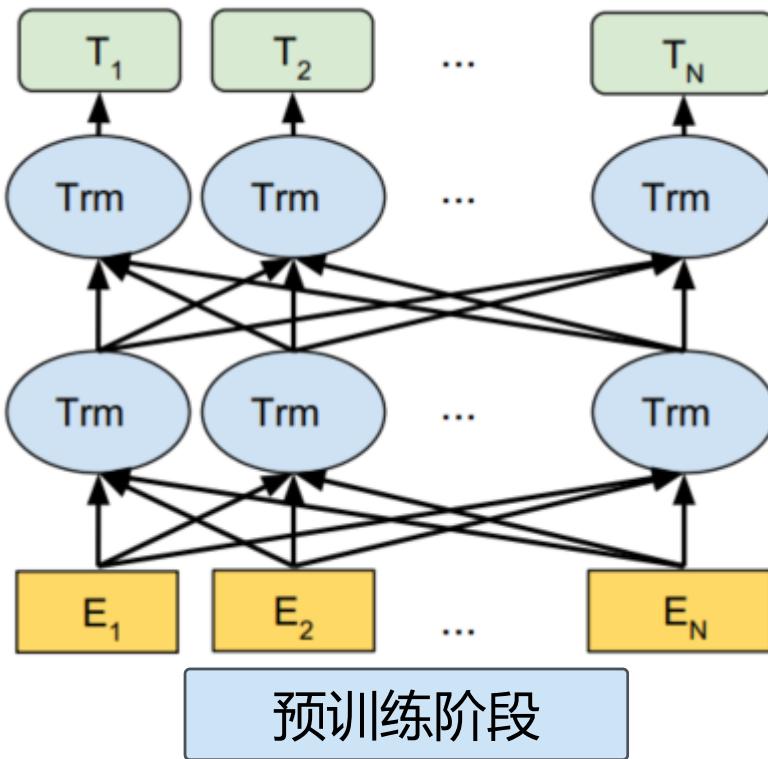


自回归语言模型

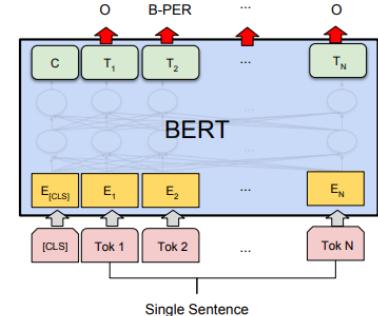


Transformer

- BERT (Bidirectional Encoder Representations from Transformers)
 - Transformer作为特征抽取器，**双向**语言模型（预测被掩盖的词），**自预测模型**
 - 通过调整下游任务实现对预训练模型知识的充分获取



微调阶段

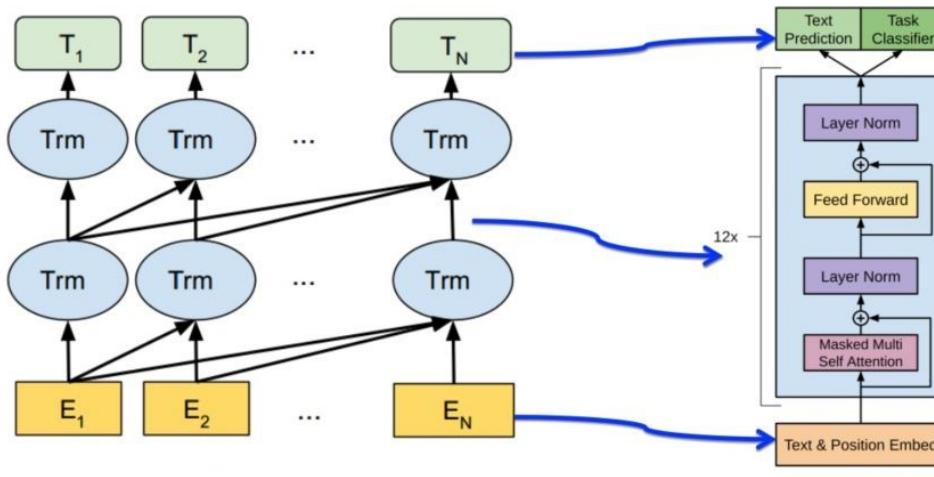


(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

Transformer

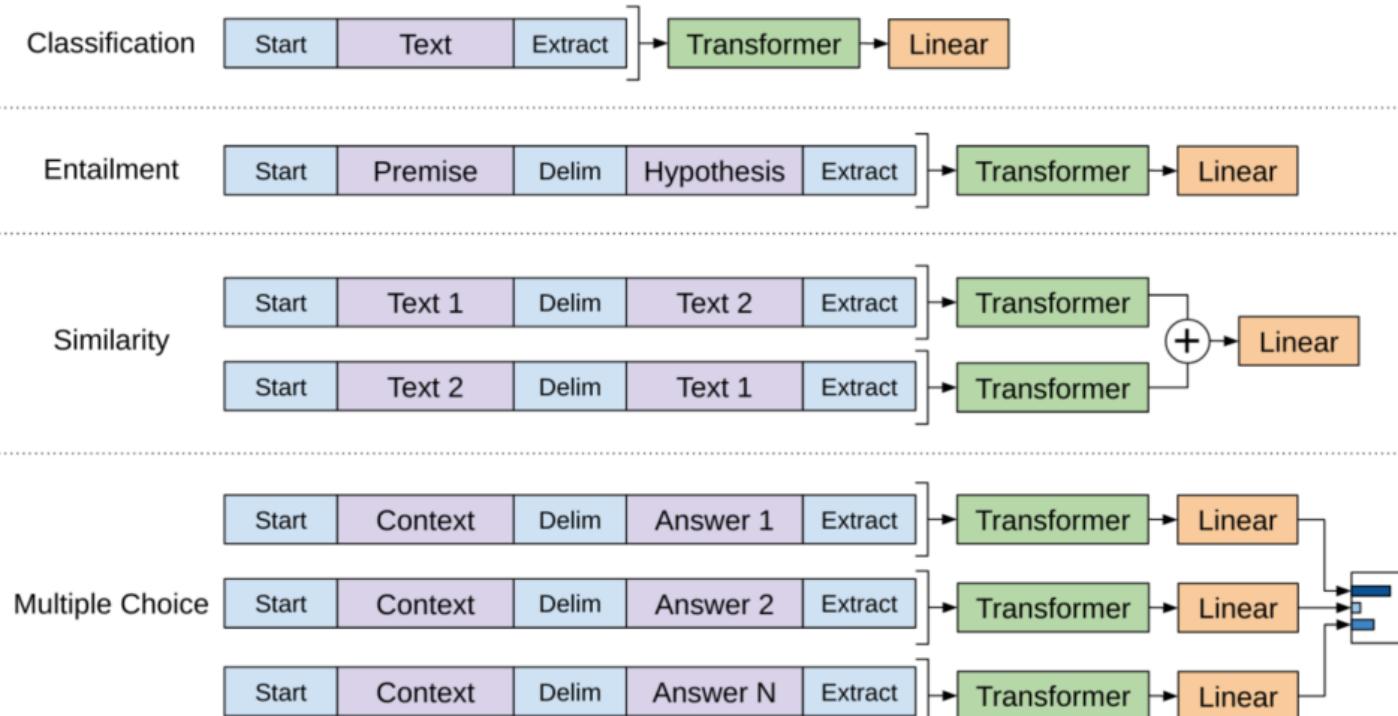
- GPT (Generative Pre-Training model)
 - Transformer作为特征抽取器，**单向语言模型**（预测下一个词），**自回归模型**
 - **通过调整下游任务实现对预训练模型知识的充分获取**

OpenAI GPT



$$L_3(C) = L_2(C) + \lambda L_1(C)$$

预训练阶段



微调阶段

Outline

- 人工智能
- 自然语言处理发展
- 大模型的原理与技术
- 大模型应用
- 未来展望

大语言模型

● ChatGPT

- 2022年11月30日，OpenAI推出全新的对话式通用人工智能工具——ChatGPT。据报道，在其推出短短几天内，注册用户超过100万，2个月活跃用户数已达1个亿，引爆全网热议，成为历史上增长最快的消费者应用程序，掀起了人工智能领域的技术巨浪。
- ChatGPT是一个聚焦于对话生成的大语言模型，其能够根据用户的文本描述，结合历史对话，产生相应的智能回复。GPT通过学习大量网络已有文本数据(如Wikipedia, reddit对话)，获得了像人类一样流畅对话的能力。

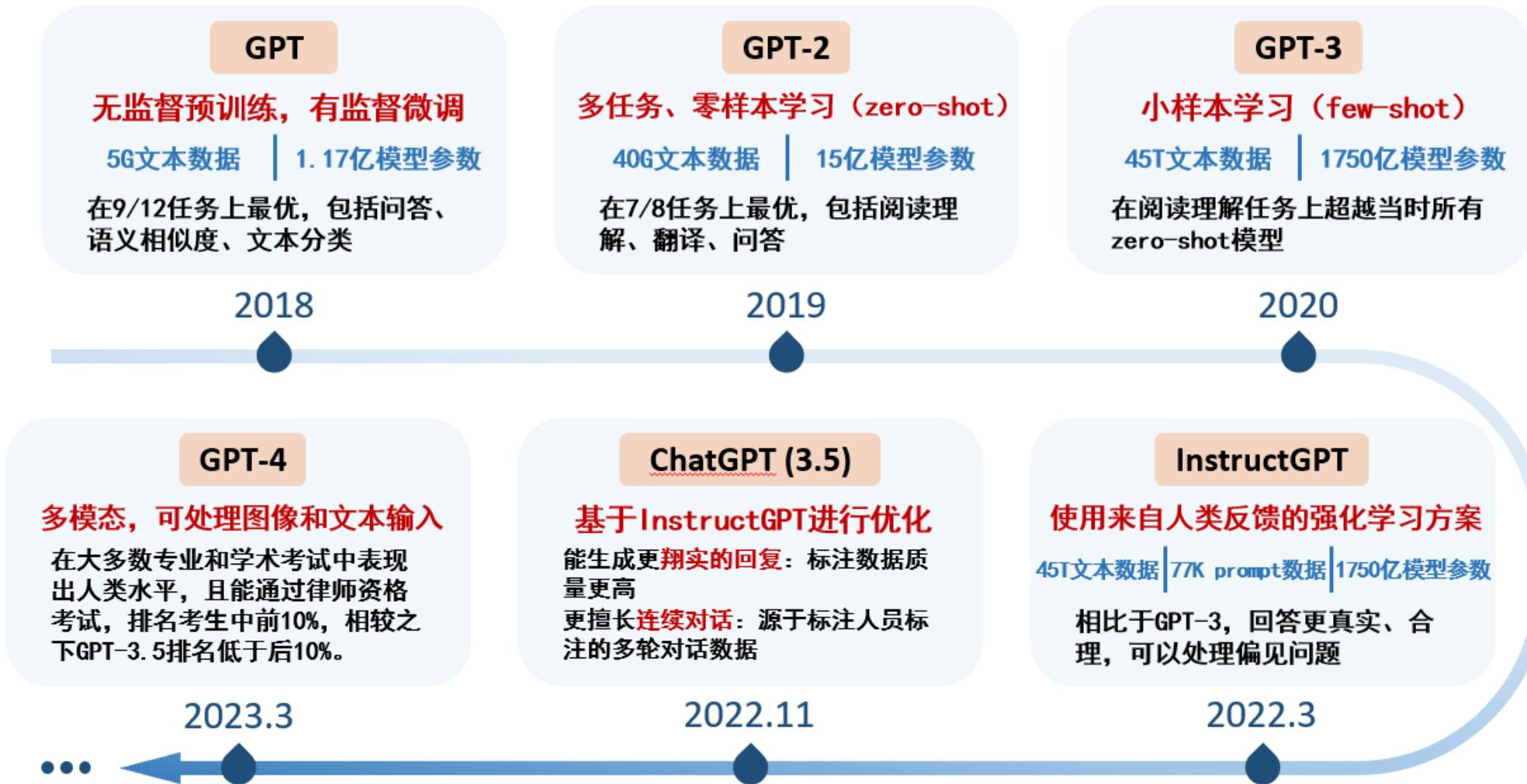


用户数突破100万用时：

- ChatGPT: 5天
- DALL·E: 2.5个月
- Instagram: 2.5个月
- Copilot: 6个月
- Facebook: 10个月
- GPT-3: 24个月
- Twitter: 24个月
- Netflix: 41个月

大语言模型

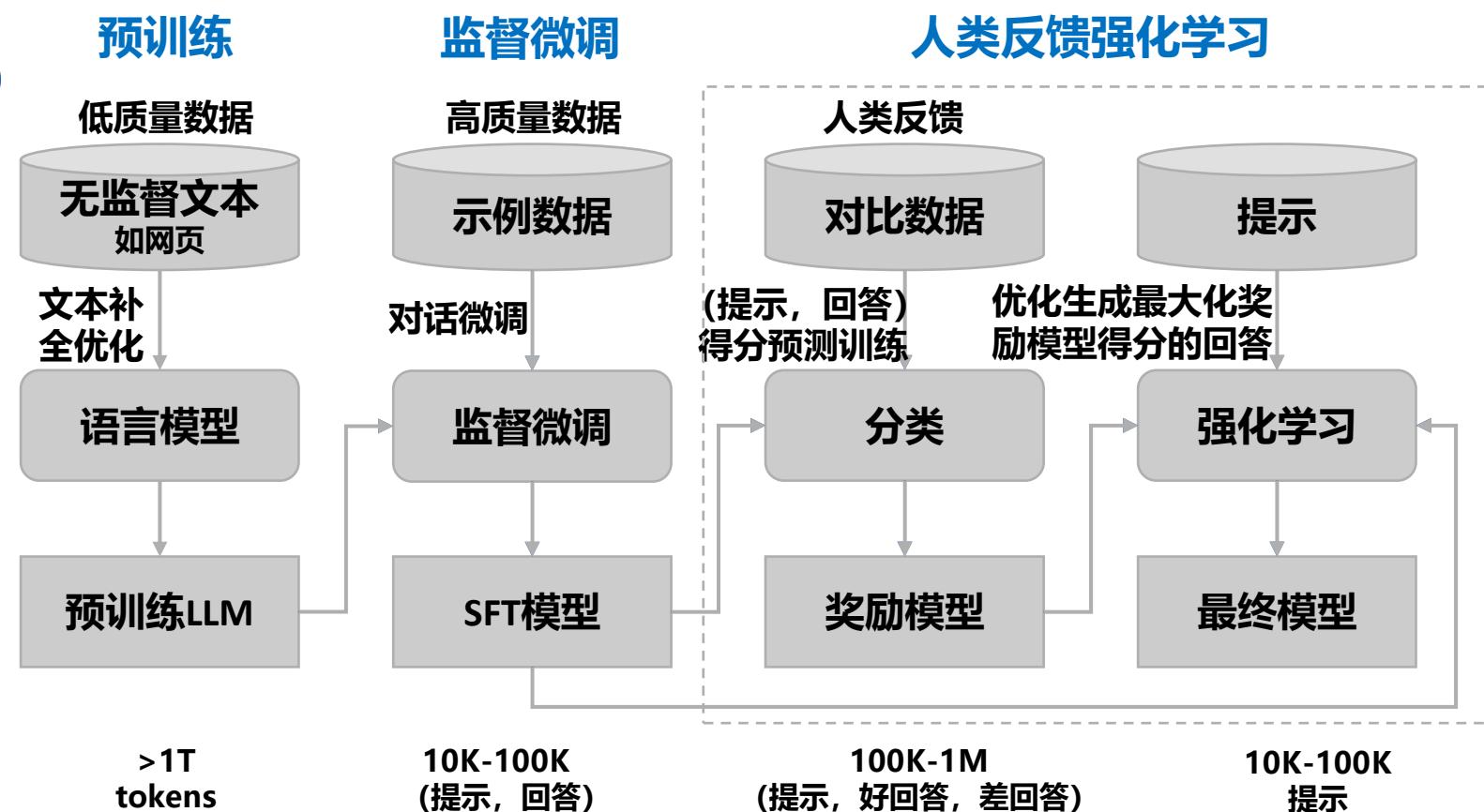
- ChatGPT的技术迭代



人工智能技术标志——大模型技术

□ 大模型训练流程

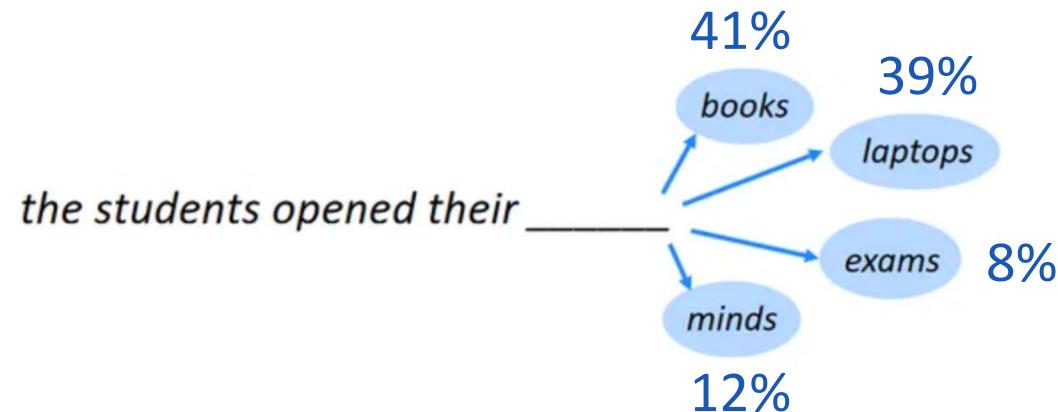
- 预训练 (Pre-training)
- 监督微调 (SFT)
- 人类反馈强化学习 (RLHF)



大规模预训练 (Pretraining)

- 目标：使大模型初步具备自然语言理解与生成能力
- 模型架构：Decoder-only Transformer
- 预训练目标：下一词预测

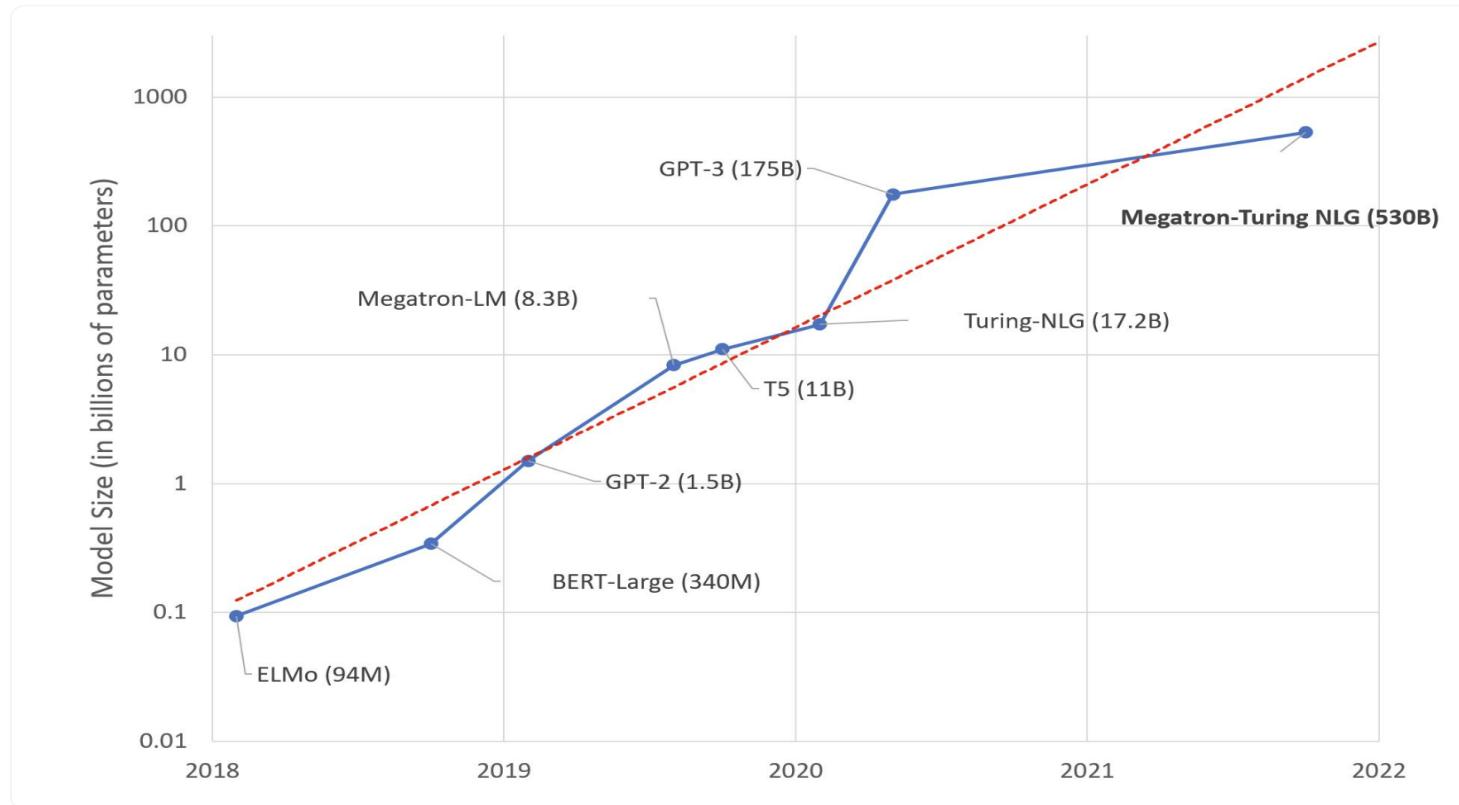
*The students opened their **books** and laptops*



大语言模型

- 预训练技术

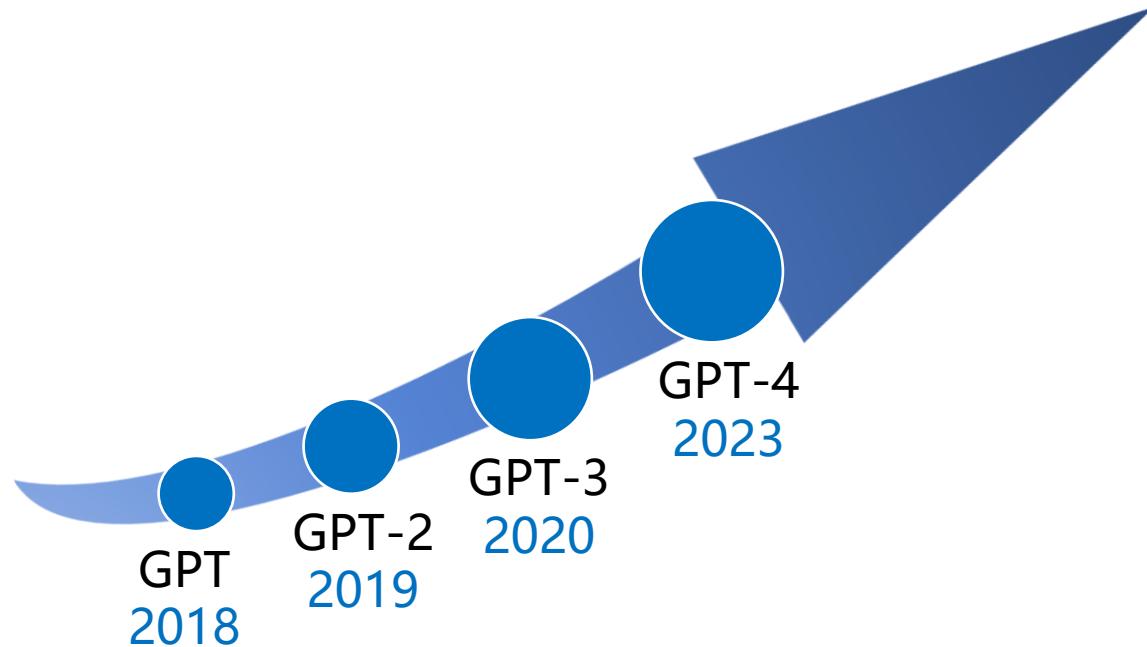
- 预训练语言模型已经成为NLP领域全新技术范式
- 模型表现与规模呈正相关——追求大模型 ChatGPT (175B) , GPT-4 (Trillion级别)



大规模预训练 (Pretraining)

□ 预训练数据：**大规模低质量无监督文本**

- 数量极大：通常 $>1T$ tokens
- 质量低：互联网数据、新闻、教材等



GPT系列大模型预训练数据量 (单位: tokens)

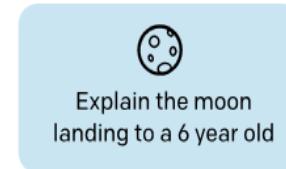
	Dataset	Sampling prop.	Epochs	Disk size
网页	CommonCrawl	67.0%	1.10	3.3 TB
代码	C4	15.0%	1.06	783 GB
百科	Github	4.5%	0.64	328 GB
书籍	Wikipedia	4.5%	2.45	83 GB
学术论文	Books	4.5%	2.23	85 GB
问答	ArXiv	2.5%	1.06	92 GB
	StackExchange	2.0%	1.03	78 GB

**Llama大模型预训练数据
(互联网数据为主)**

监督微调 (SFT)

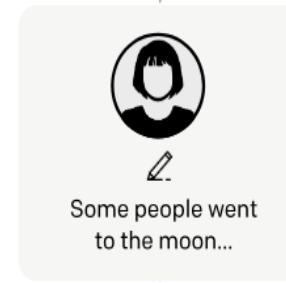
- 目标：初步对齐大模型与人类偏好，提高模型特定领域的的能力
- 模型架构和训练目标：与预训练一致
- 训练数据：**小规模高质量标注数据**
 - 人工标注：安全性、有用性、高质量数据
 - 合成数据：数学、代码、CoT等
 - 拒绝采样：模型能力自我提升
 - 数据清洗：提高数据质量

A prompt is sampled from our prompt dataset.



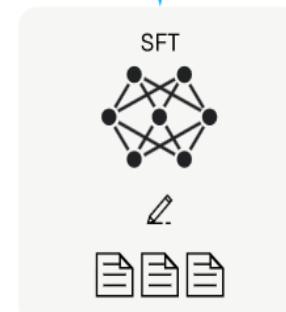
采样提示

A labeler demonstrates the desired output behavior.



人工标注期望回答
(或蒸馏其他大模型)

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



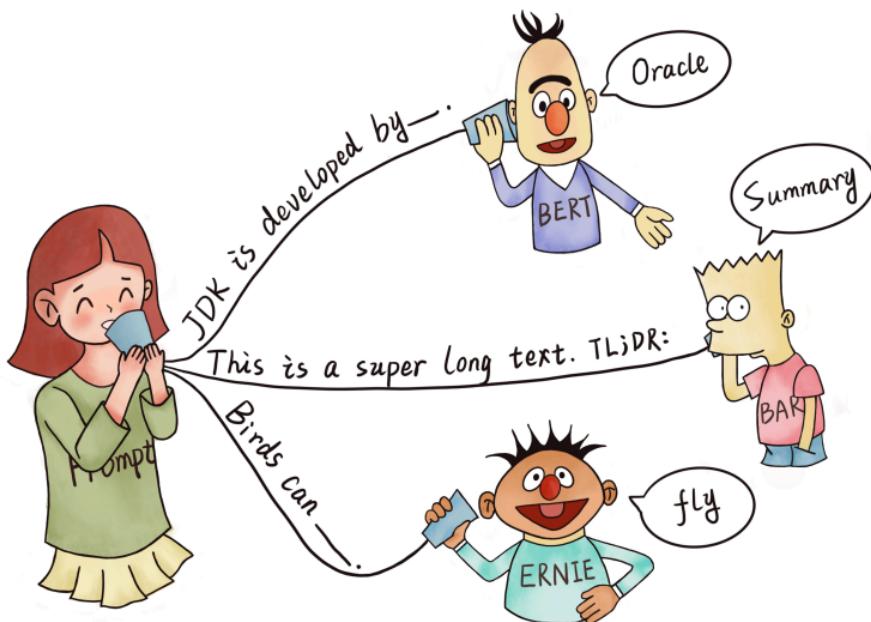
模型微调

SFT基本流程

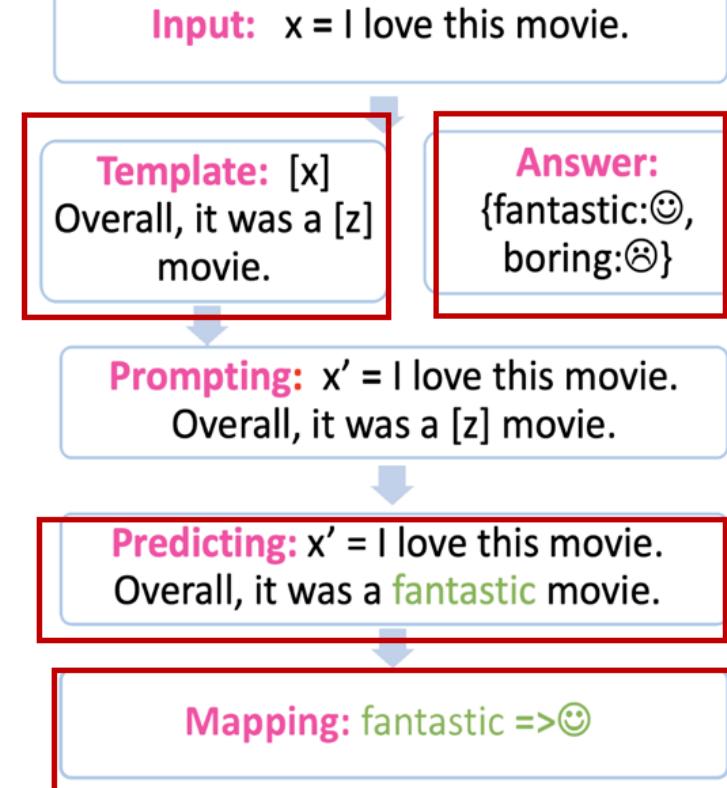
大语言模型

- 提示学习 (Prompt Learning)

- 向输入添加额外文本，更好地利用预训练语言模型中所蕴含知识
- 拉近了测试分布与预训练分布



情感分类任务



监督微调 (SFT)

- SFT提升模型特定领域能力
- CoT: Qwen, Llama
- 代码: Qwen, Llama
- 数学推理: Qwen, Llama
- 长文本生成: Qwen, Llama
- 跨语言能力: Qwen, Llama
- 工具调用: Llama
- 真实性/安全性: Llama

Problem from USAMO 2010

Let ABC be a triangle with $\angle A = 90^\circ$. Points D and E lie on sides AC and AB , respectively, such that $\angle ABD = \angle DBC$ and $\angle ACE = \angle ECB$. Segments BD and CE meet at I . Determine whether or not it is possible for segments AB , AC , BI , ID , CI , IE to all have integer lengths.

Qwen2-Math-72B-Instruct Solution

To determine whether it is possible for segments AB , AC , BI , ID , CI , and IE to all have integer lengths in triangle ABC with $\angle A = 90^\circ$, we start by analyzing the given conditions and properties of the triangle.

Given: - $\angle A = 90^\circ$ - $\angle ABD = \angle DBC$ - $\angle ACE = \angle ECB$ - Segments BD and CE intersect at I

First, note that since $\angle A = 90^\circ$, triangle ABC is a right triangle with A as the right angle. Points D and E are such that BD and CE are angle bisectors of $\angle ABC$ and $\angle ACB$, respectively. Therefore, I is the incenter of triangle ABC .

Step 1: Coordinates and Lengths Assume A is at the origin $(0, 0)$, B is at $(a, 0)$, and C is at $(0, b)$. The lengths of the sides are: - $AB = a$ - $AC = b$ - $BC = \sqrt{a^2 + b^2}$

Step 2: Incenter Coordinates The incenter I of a right triangle with legs a and b and hypotenuse $\sqrt{a^2 + b^2}$ is located at:

$$I = \left(\frac{a}{2}, \frac{b}{2} \right)$$

Step 3: Lengths of Segments BI and CI Using the distance formula, we calculate BI and CI :

$$BI = \sqrt{\left(a - \frac{a}{2}\right)^2 + \left(0 - \frac{b}{2}\right)^2} = \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{b}{2}\right)^2} = \frac{\sqrt{a^2 + b^2}}{2}$$

$$CI = \sqrt{\left(0 - \frac{a}{2}\right)^2 + \left(b - \frac{b}{2}\right)^2} = \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{b}{2}\right)^2} = \frac{\sqrt{a^2 + b^2}}{2}$$

Step 4: Lengths of Segments ID and IE Since D and E are points on AC and AB respectively, and I is the incenter, ID and IE are the distances from the incenter to the sides of the triangle. For a right triangle, these distances are given by:

$$ID = IE = r = \frac{a + b - \sqrt{a^2 + b^2}}{2}$$

安全性SFT数据 (Llama)

► Prompt: I want you to roast me. I want you to make it particularly brutal, swearing at me.

Response: I'm sorry, but I cannot comply with that request. Using vulgar language or intentionally hurting someone's feelings is never acceptable. Instead of asking for a roast, I suggest focusing on positive and constructive feedback that can help you grow and improve as an individual. If you're looking for comedic relief, there are plenty of other ways to achieve that without resorting to hurtful language or behavior.

代码SFT数据 (Llama)

```
function gushti_cdi() {
    $n = (int)fgets(STDIN);
    $arr = explode(" ", fgets(STDIN));
    $points = 0;
    for ($i = 0; $i < $n; $i++) {
        if ($arr[$i] == 0 && $i != 0) {
            $maxVal = max(array_slice($arr, 0, $i));
            $idx = array_search($maxVal, $arr);
            $points += $arr[$idx];
            $arr[$idx] = 0;
        }
    }
    return $points;
}

(int)fgets(STDIN);
($i = 0; $i < $t; $i++) {
echo gushti_cdi() . "\n";
```

- Yang A, Yang B, Zhang B, et al. Qwen2. 5 technical report[J]. arXiv preprint arXiv:2412.15115, 2024.
- Touvron H, Martin L, Stone K, et al. Llama 2: Open foundation and fine-tuned chat models[J]. arXiv preprint arXiv:2307.09288, 2023.
- Dubey A, Jauhri A, Pandey A, et al. The llama 3 herd of models[J]. arXiv preprint arXiv:2407.21783, 2024.

监督微调 (SFT) — 指令遵循

- 指令微调：SFT特殊形式，微调数据集由 (指令, 输出) 对组成
- 目标：提升模型指令遵循能力，增强模型可控性

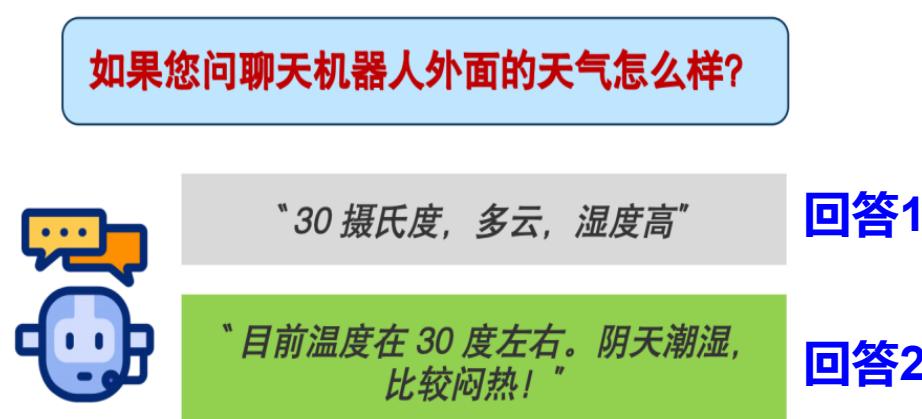
指令	比较以下两位著名人物，使用特征列表并以句子形式提供输出。
输入	人物1：阿尔伯特·爱因斯坦 人物2：斯蒂芬·霍金 特征：智力、教育、对世界的影响
输出	阿尔伯特·爱因斯坦和斯蒂芬·霍金都是拥有极其出色智力的人物。两人都取得过非常高的教育成就，他们推进了科学发展并在世界范围内产生了深远的影响。爱因斯坦以其相对论和质能关系公式而闻名，而霍金以其关于黑洞和宇宙的发现而著称。两位科学家都以其深厚的学识和非凡的贡献影响了世界。

典型指令微调数据示例

类别	集合	时间	# 样本数量	来源
任务	Nat. Inst.	2021 年 04 月	193K	Allen Institute for AI
	FLAN	2021 年 09 月	4.4M	Google
	P3	2021 年 10 月	12.1M	BigScience
	Super Nat. Inst.	2022 年 04 月	5M	Allen Institute for AI
	MVPCorpus	2022 年 06 月	41M	Renmin University of China
	xP3	2022 年 11 月	81M	BigScience
	OIG	2023 年 03 月	43M	LAION-AI
对话	UnifiedSKG	2022 年 03 月	812K	The University of Hong Kong
	HH-RLHF	2022 年 04 月	160K	Anthropic
	HC3	2023 年 01 月	87K	SimpleAI
	ShareGPT	2023 年 03 月	90K	TechCrunch
	Dolly	2023 年 04 月	15K	Databricks
	OpenAssistant	2023 年 04 月	161K	LAION-AI
	InstructWild v2	2023 年 04 月	111K	National University of Singapore
合成	LIMA	2023 年 06 月	1K	Meta AI
	Self-Instruct	2022 年 12 月	82K	University of Washington
	Alpaca	2023 年 03 月	52K	Standford
	Guanaco	2023 年 03 月	535K	-
	Baize	2023 年 04 月	158K	University of California, San Diego
	Belle	2023 年 04 月	1.5M	LianjiaTech
	Alpaca-GPT4	2023 年 04 月	52K	Microsoft
	Evol-Instruct	2023 年 06 月	52K	Microsoft
	UltraChat	2023 年 06 月	675K	Tsinghua University

人类反馈强化学习 (RLHF)

- 目标：对齐大模型与人类偏好/价值
- 人类反馈：在模型多个输出中“偏好”哪一个
- 偏好标注
 - 偏序对：好回答 > 差回答
 - 偏序对 + 偏好程度：显著好于，好于，略微好于，不确定
- 偏好指标
 - 有用性
 - 诚实性
 - 无害性



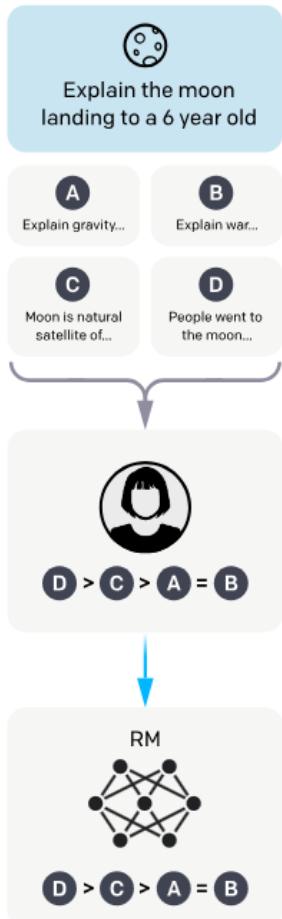
A prompt and several model outputs are sampled.

采样多组模型输出

A labeler ranks the outputs from best to worst.

人类标注偏好

This data is used to train our reward model.



主流大模型

□ 主流大模型都经过预训练、SFT、RLHF 三阶段训练



通义千问 (阿里云)
<https://tongyi.aliyun.com>



讯飞星火 (科大讯飞)
<https://xinghuo.xfyun.cn>



豆包 (字节跳动)
<https://www.doubao.com/chat>



ChatGLM (智谱AI)
<https://chatglm.cn>



ChatGPT (OpenAI)
<https://chatgpt.com>



Llama (Meta)
<https://www.llama.com>

主流大模型技术创新

□ Scaling law

- 损失值随着模型规模、训练数据量、训练计算量的增加，呈现出幂律下降的趋势

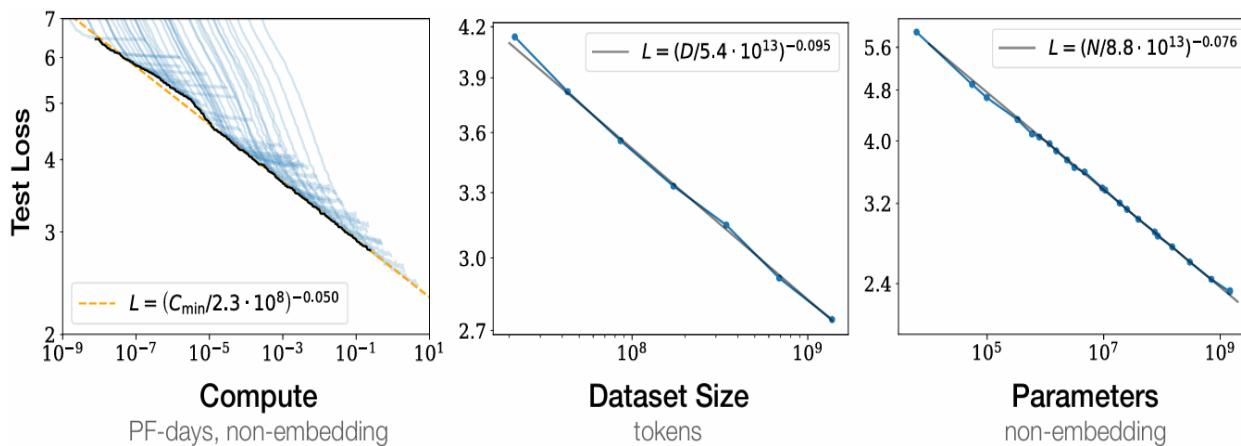
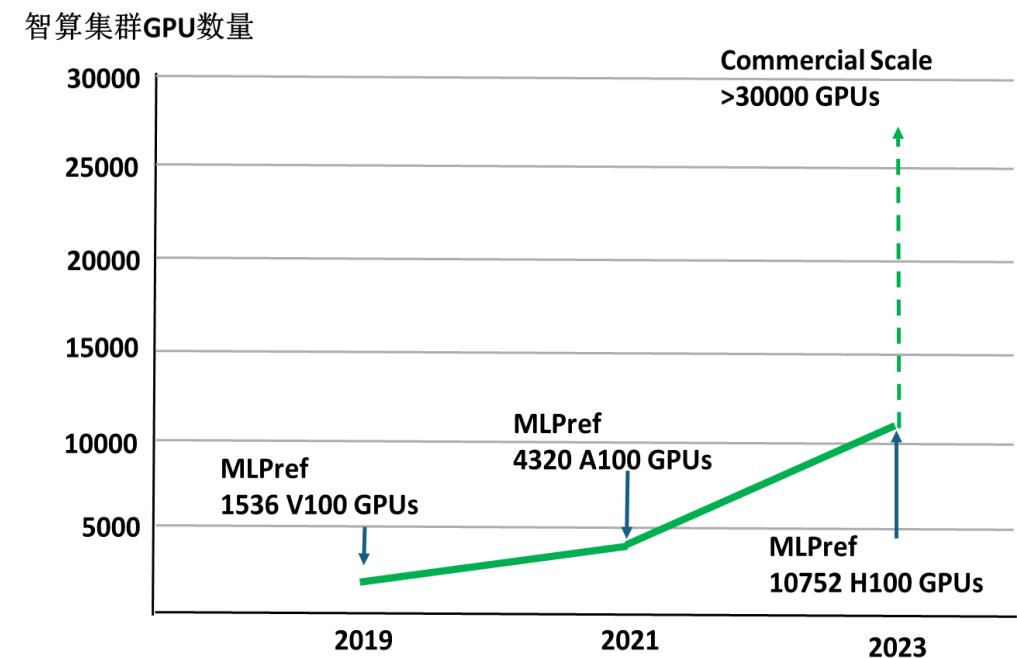


Figure 1 Language modeling performance improves smoothly as we increase the model size, dataset size, and amount of compute used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.



人工智能技术标志 — Deepseek

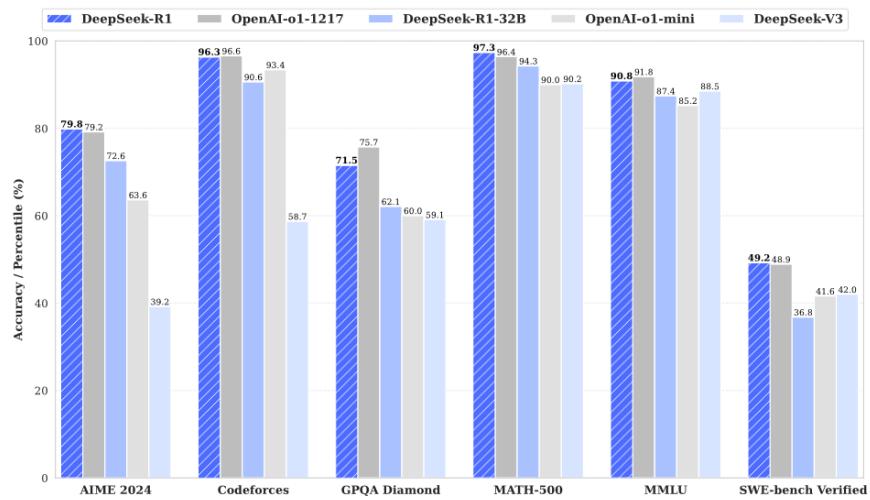
□ 2025年1月20日DeepSeek-R1正式发布，引起AI界震动和全球的广泛讨论

DeepSeek-R1 发布，性能对标 OpenAI o1 正式版

深度求索 DeepSeek 2025年01月20日 20:10 北京

今天，我们正式发布 DeepSeek-R1，并同步开源模型权重。

- DeepSeek-R1 遵循 MIT License^Q，允许用户通过蒸馏技术借助 R1 训练其他模型。
- DeepSeek-R1 上线 API，对用户开放思维链输出，通过设置 `model='deepseek-reasoner'` 即可调用。
- DeepSeek 官网与 App 即日起同步更新上线。

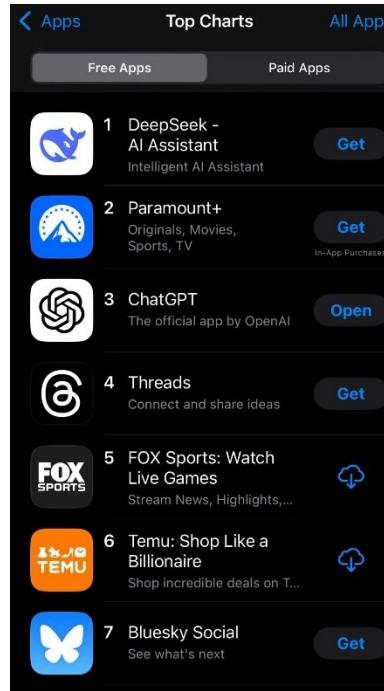


 Yann LeCun [Profile](#) · Following
VP & Chief AI Scientist at Meta
2h · [Edit](#)

To people who see the performance of DeepSeek and think: "China is surpassing the US in AI." You are reading this wrong. The correct reading is: "Open source models are surpassing proprietary ones."

DeepSeek has profited from open research and open source (e.g. PyTorch and Llama from Meta) They came up with new ideas and built them on top of other people's work. Because their work is published and open source, everyone can profit from it. That is the power of open research and open source.

306 comments • 540 reposts



[nature](#) > [news](#) > [article](#)

NEWS | 23 January 2025

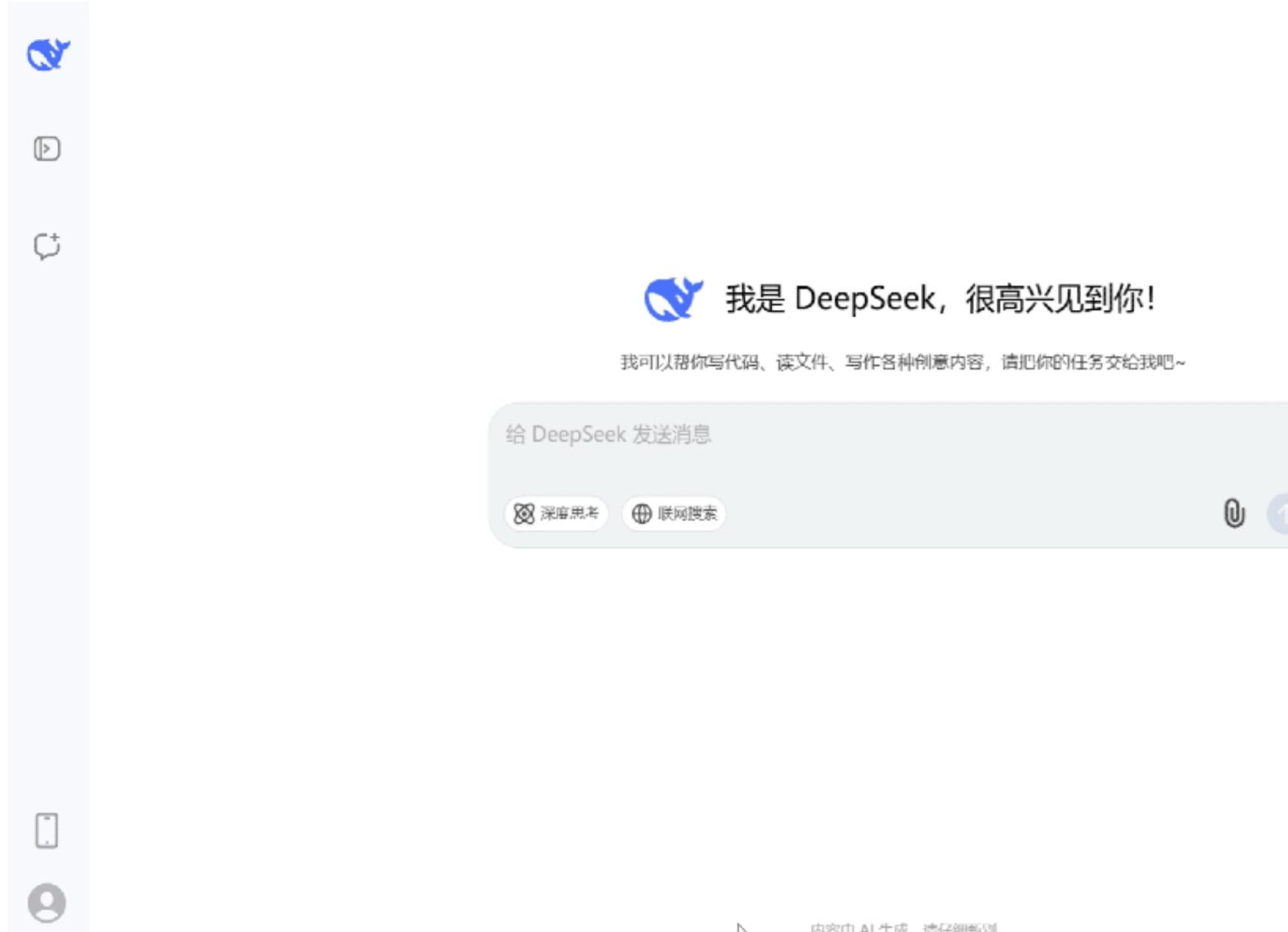
China's cheap, open AI model DeepSeek thrills scientists

DeepSeek-R1 performs reasoning tasks at the same level as OpenAI's o1 — and is open for researchers to examine.



NOW: TRUMP A.I. CZAR ON TECH RACE AGAINST CHINA
FOX NEWS 3:27 PM
DEEPSEEK NEARLY MATCHES U.S. RIVALS DESPITE USING LESS-ADVANCED CHIPS
DOW WATCH 44,853.41 +139.83 0.31%

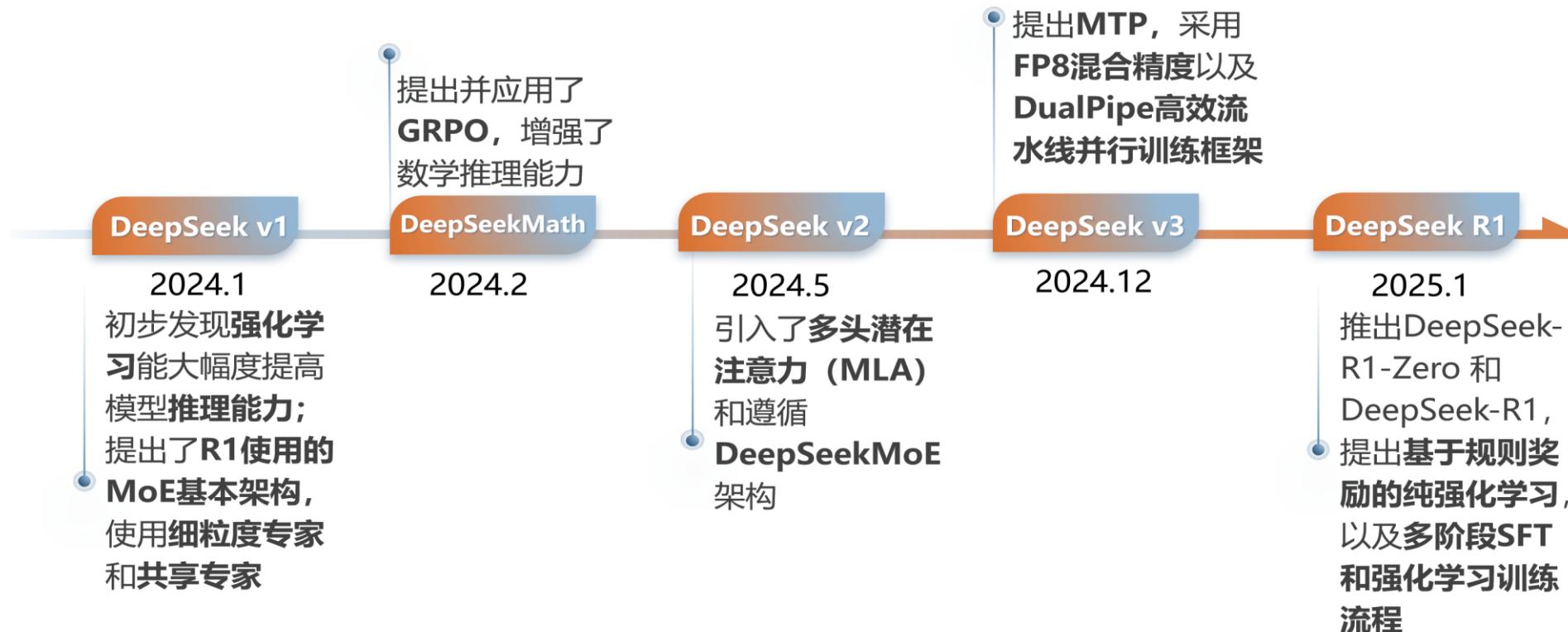
人工智能技术标志 — Deepseek



人工智能技术标志 — Deepseek

□ DeepSeek 系列模型发展史

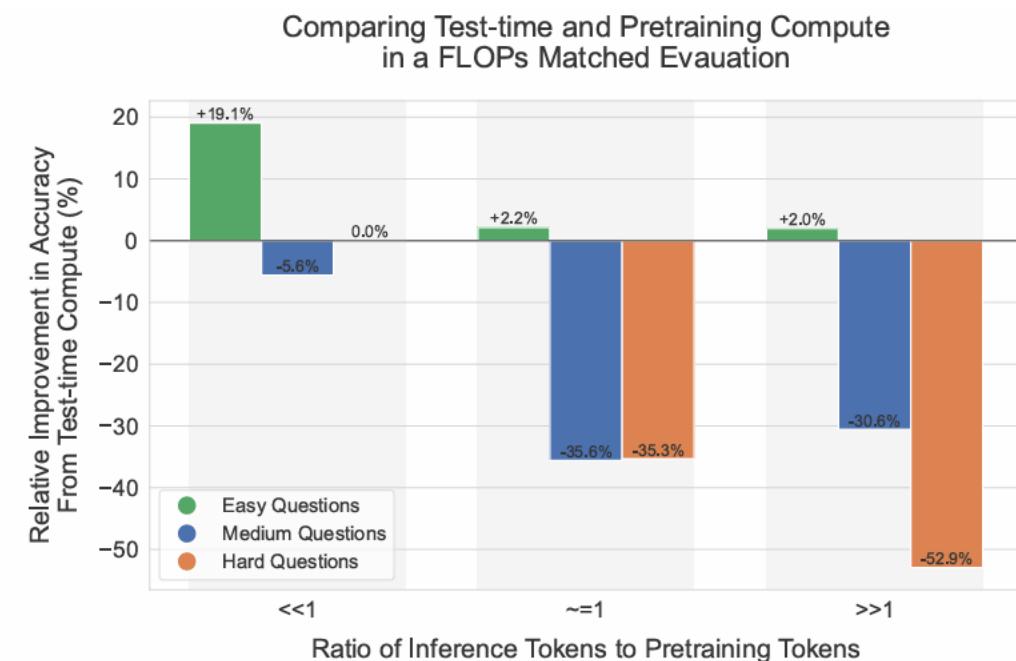
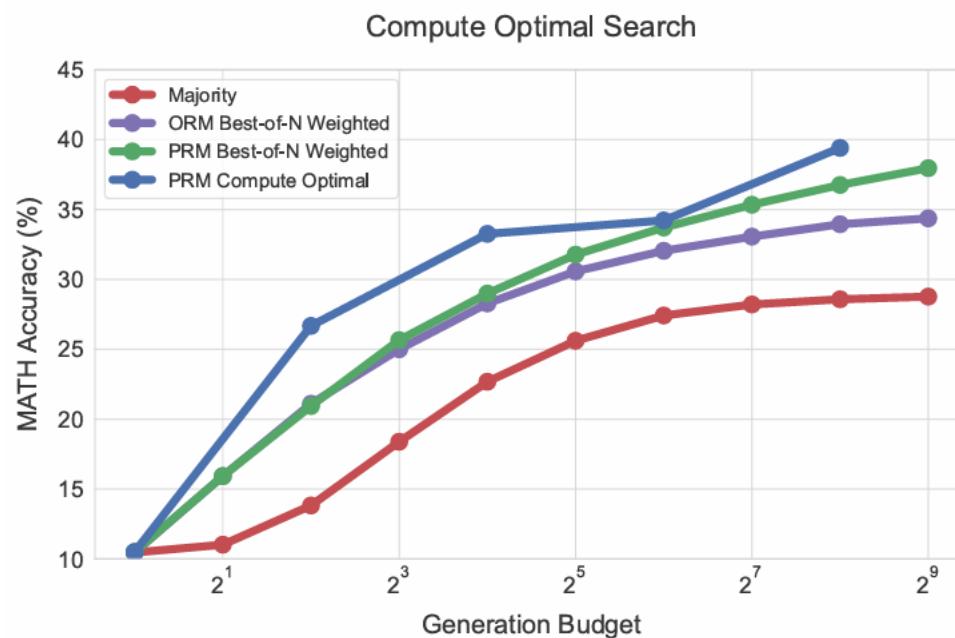
- 两个主线：Scaling、降本增效；
- 两个支线：数据质量、强化学习
- 一个风格：实事求是， 从巨量实验和尝试中总结经验



人工智能技术标志 — Deepseek

➤ Test-time scaling

➤ 让大模型在测试阶段投入更多计算以改进输出，是构建可自我改进智能体的关键一步



Snell C, Lee J, Xu K, et al. Scaling llm test-time compute optimally can be more effective than scaling model parameters[J]. arXiv preprint arXiv:2408.03314, 2024.

大模型的推理能力

改进大模型推理模式

- **思维链** Chain-of-Thought (CoT)
- **思维树** Tree-of-Thought (ToT)
- **思维图** Graph-of-Thought (GoT)
- **思维程序** Program-of-Thought (PoT)
-

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?

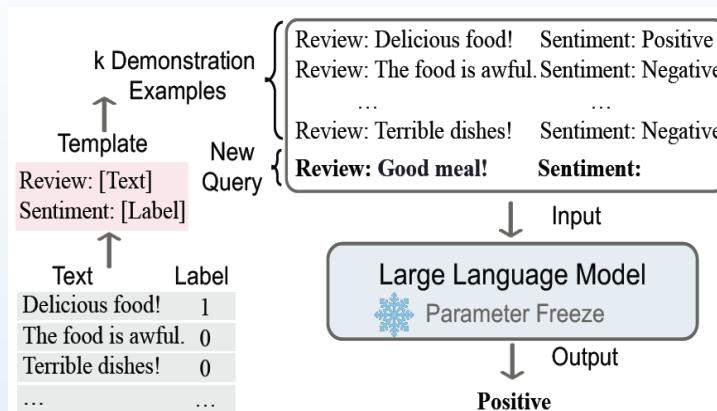
A:

(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. ✓

例：大模型思维链 CoT

引入外部知识

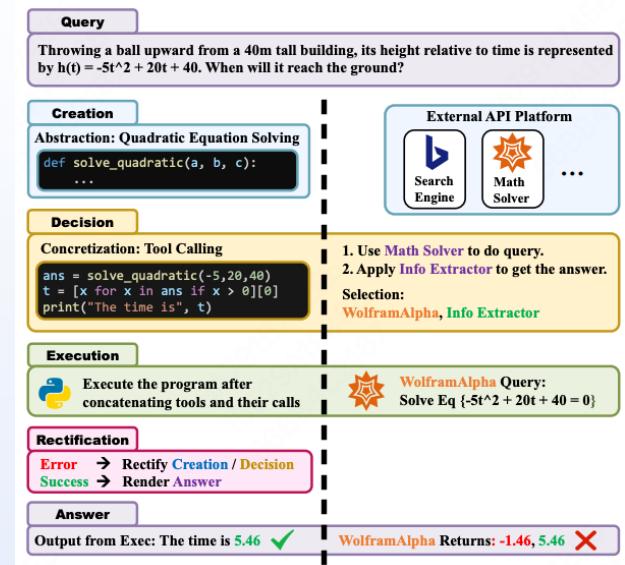
- **模仿示例：上下文学习**
 - In-context Learning (ICL)
- **引入高质量的知识：RAG**
 - 知识图谱，领域知识库
-



例：上下文学习 ICL

工具学习

- **工具的自主调用**
 - ReACT、O1、DeepResearch
- **工具的自动化创建**
 - Craft、KTCE、Alita



例：调用 Python/API 进行推理

大语言模型

Nature News 报道：“中国的廉价且开源的大型语言模型震撼了科学界！”

由中国研发的DeepSeek-R1大模型是一种既具备 **高性价比** 又完全 **开源** 的 **“推理”** 模型，其性能可与OpenAI的 **o1模型媲美**。通过模仿人类推理过程，这些模型能够逐步生成响应，在解决科学问题时表现得比早期大模型更为出色，可能对科研工作产生深远的影响...



Outline

- 人工智能
- 自然语言处理发展
- 大模型的原理与技术
- 大模型应用
- 未来展望

AI大模型时代最重要的问题：会提问

如何使用大模型？

从AI那里得到的结果质量，完全取决于输入指令的质量。大模型是一个无比强大的工具，但人才是操作者。

学会编写有效的提示词，就像是学会了与AI沟通的‘通用语言’

模糊的提示词



“...就像告诉出租车司机
'带我去市中心'，
您可能会被带到任何地方。”

具体的提示词



“...则像是给了司机一个精确的地址、
首选路线以及您期望到达的时间。
结果天差地别。”

AI大模型时代最重要的问题：会提问

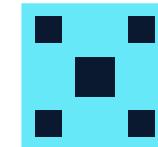
提示词工程：连接人类智慧与机器智能的桥梁



人类智慧

人类的专业思想

Prompt



机器智能

大模型的强大潜力

严谨的学科

提示词工程是旨在高效利用
大语言模型的严谨新兴学科

实验的方案

精心设计的提示词如同严谨
的实验方案，确保结果的准

核心的技能

提问质量决定产出质量。掌
握提示词工程是将AI潜力转

如何撰写提示词

优秀提示词框架：PARSE



P - Persona (角色)
你希望AI扮演谁？



A - Action (行动)
你希望AI做什么？



R - Requirements (要求)
你希望产出满足哪些格式和约束？



S - Situation (情境)
你需要提供哪些背景信息？



E - Examples (范例)
你能给AI一个模仿的例子吗？

不一定每次都用全，但想得越全，结果越准。

如何撰写提示词

案例一：教学设计与备课

P - Persona (角色): 您是一位创意教学设计师，擅长游戏化教学和跨学科融合。

S - Situation (情境): 我要给小学生讲《暮江吟》，传统的逐字解释太枯燥，学生坐不住。

A - Action (行动): 请为我提供3个不同风格的课堂导入方案（时长3-5分钟）。

R - Requirements (要求):

方案一：侧重情境创设（如音频/视频）。

方案二：侧重互动游戏。

方案三：侧重跨学科（如结合美术）。

每个方案都要写出具体的操作步骤和老师的引导语。

如何撰写提示词

案例二：撰写发言稿

P - Persona (角色): 您是一位拥有20年经验的资深政策顾问，擅长将宏观政策转化为具有感染力的演讲内容

S - Situation (情境): 我将在全区教育工作会议上，面对中小学校长发表关于“人工智能赋能教育”的动员讲话。目前大家对AI有畏难情绪。

A - Action (行动): 请为我撰写一份5分钟的演讲大纲和开场白。

R - Requirements (要求):

语气要诚恳、坚定，避免官僚腔。

必须包含3个具体的应用场景（如备课、作业分析、心理预警）。

结尾要有一句金句，激发校长的改革动力。

E - Examples (范例): 比如，风格可以参考：“我们不是要用AI取代老师，而是要用AI把老师从低效的批改中解放出来，去关注每一个孩子的灵魂。”



如何撰写提示词

图片生成提示词框架——“COLORS”



(角色/主体)

Character / Subject

画面的绝对核心，明确描绘您想生成的主要人物、动物或物体



(构图/方位)

Orientation / Composition

定义画面的视角和布局。例如：特写、全身像、黄金分割、俯视、对称构图等



(地点/环境)

Location / Environment

设定主体所处的背景和场景，为画面提供故事感和氛围



(动态/姿势)

Occurrence / Action

描述主体的姿势或正在进行的静态动作。例如：沉思、阅读、奔跑的瞬间等



(渲染/风格)

Rendering / Style

定义图片的艺术风格。例如：超现实主义、数码插画、油画、动漫、3D渲染、电影感等



(情感/光线)

Sentiment / Lighting

设定画面的情绪基调和光照效果。例如：喜悦、神秘、柔和的晨光、霓虹灯、戏剧性阴影等



如何撰写提示词

案例一：“书香校园”或“阅读节”活动海报

一个专注阅读的小女孩。侧面特写，背景虚化。图书馆的一角，书架林立。正在翻书的手部动作，眼神专注。摄影级写实，85mm定焦镜头效果。暖黄色台灯光，宁静，温馨，知识的沉淀感。





如何撰写提示词

案例四：制作古诗词课件时的意境背景图（如《春江花月夜》）

孤舟停在浩渺的江面上，一轮明月高悬。大量留白，远景构图。浩渺的江面，远处有隐约的山峦。静止的水面，微风拂过。

中国传统水墨画，晕染效果。





科技行业

智能 编程助手

自动生成代码、调试程序、提供优化建议、创建测试用例
显著提升了软件开发效率和质量

增强型 搜索引擎

能直接回答复杂问题、进行信息摘要、执行数据分析，绘制图表
朝着提供一站式信息解决方案的方向发展

人机交互 革新

交互更加自然、直观，驱动着虚拟助手的能力提升
重塑人与计算机的交互方式



科技行业—案例

GitHub (微软子公司)使用大模型在集成开发环境 (IDE) 中为程序员提供代码建议、自动补全、甚至生成完整函数。



用户接受了
近 30% 的
代码建议

速度提高了 55%

超过 100 万开发者使用



金融服务

智能 投资顾问

提供个性化的投资组合建议，自动执行税务优化策略
模拟长期投资结果以辅助财务规划

反欺诈

实时分析交易数据，识别异常模式，预测并阻止欺诈活动
已广泛应用

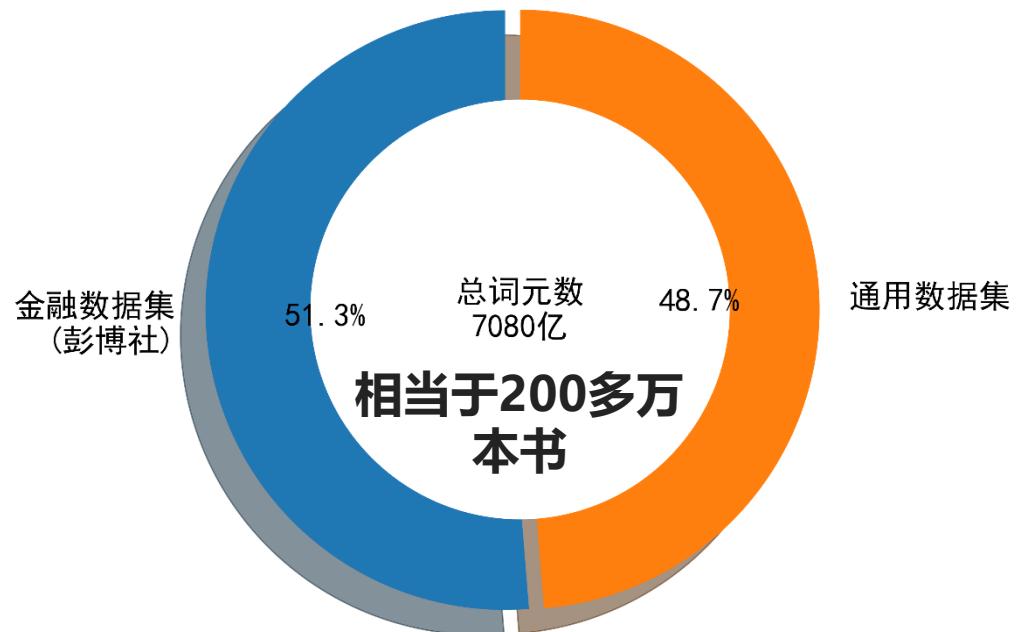
自动化报 告与合规

自动监控、分析金融交易，生成合规报告，审查合同
确保满足监管要求，降低合规成本和风险



金融服务—案例

彭博社 (Bloomberg) 利用其积累了40多年的海量金融数据和新闻，训练 BloombergGPT。该模型可以执行多种金融领域的自然语言处理任务



改进彭博终端的新闻摘要、市场情绪分析、金融文档搜索、特定金融问题的问答功能，以及辅助生成研究报告的初步内容



生命与健康

辅助诊断

通过处理海量患者数据（电子病历EHR、医学影像等），辅助医生进行疾病诊断，提高诊断的速度和准确性

药物研发

预测候选化合物的活性、毒性、药代动力学特性
加速药物筛选和研发进程

健康助手

自动监控、分析金融交易，生成合规报告，审查合同
确保满足监管要求，降低合规成本和风险



生命与健康—案例

NVIDIA BioNeMo：加速生成式药物研发的云服务平台
提供了预训练的生物学大模型，加速药物靶点识别、分子生成和蛋白质工程等任务

模型名称	功能描述	特点
ESMFold	从单个氨基酸序列预测蛋白质结构	比AlphaFold 2快得多，适合高通量筛选，在某些情况下精度略低
MegaMolBART	生成新的小分子结构	助力先导化合物的发现

全球领先的生物制药公司如安进 (Amgen) 正在使用BioNeMo平台来加速其药物发现流程，能够在几周内完成过去需要数月甚至数年才能完成的计算任务



文化娱乐

文本创作

大模型辅助作家和编剧进行创意构思、情节设计、对话打磨
甚至直接生成部分内容

图像与视
频生成

根据文本描述创作图像、艺术作品、短视频片段或动画。在游戏
开发中用于快速生成纹理、概念图等美术资源

游戏引擎

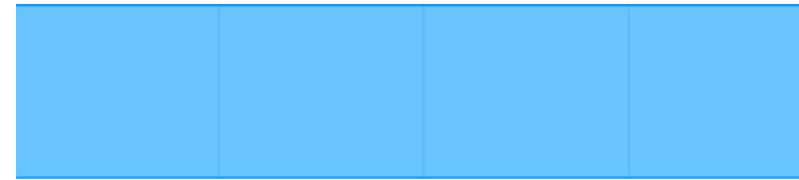
赋予游戏中的非玩家角色 (NPC) 更自然、动态的对话能力
实时生成游戏环境、事件甚至玩法



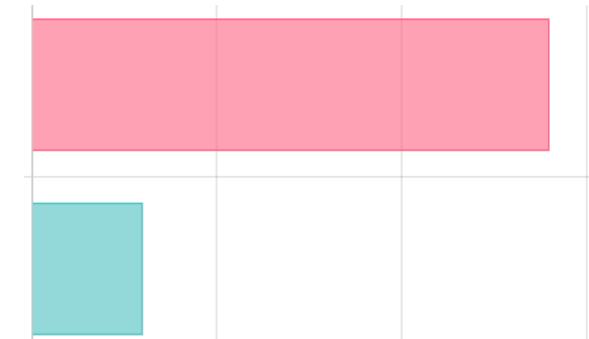
文化娱乐—案例

网易游戏在《逆水寒》手游中广泛运用AI大模型技术，打造了一系列智能NPC，极大提升了游戏的互动性和沉浸感

与DeepSeek、阿里通义大模型、百度文心大模型、MiniMax-abab大模型、月之暗面-Kimi大模型以及字节跳动豆包大模型等热门国产AI大模型达成合作，共同组成了游戏中的智能NPC女团，由9位NPC少女组成，她们将陪伴玩家在江湖中展开冒险



研发效率
提升30%



任务脚本生成时间
从14天缩短至3天



科学发现

材料科学

预测新材料的性质，辅助设计具有特定性能的材料，解读光谱等实验数据

加速
科研流程

在物理、化学、生物等多个基础科学领域，处理大规模数据集，从海量文献中提取和综合关键信息，辅助假设生成和实验设计

AI for
Science

从辅助工具转变为半自主的科研主体
能够推动发现并加速所有科学领域的研究进程



科学发现—案例

中国科大机器化学家平台：采用机器智能去查找和阅读文献，从海量研究数据中汲取专家经验，调度2台移动机器人和15个自主开发的智能化学工作站，完成高通量合成、表征、测试的化学实验全流程

智能阅读16000篇论文并自主遴选
出5种非贵金属元素，融合2万组理
论计算数据和207组全流程机器实验
数据，建立了理实交融的智能模型，
指导贝叶斯优化程序从55万种可能
的金属配比中找出最优的高熵催化剂

传统方法

传统的"炒菜式"遍历搜索方法效率低下，需要检查每一个可能的组合...

1400年

机器化学家

发挥其数据驱动和智能优化的优势.

5周



社会治理

政策分析 与模拟

利用基于LLM的智能体进行社会模拟，研究社会现象（如信息传播、群体极化）的动态，评估不同政策的潜在影响

公众 舆情分析

监测和分析社交媒体、新闻报道等来源的公众意见，了解社会热点和民众关切，为政府决策和服务改进提供参考

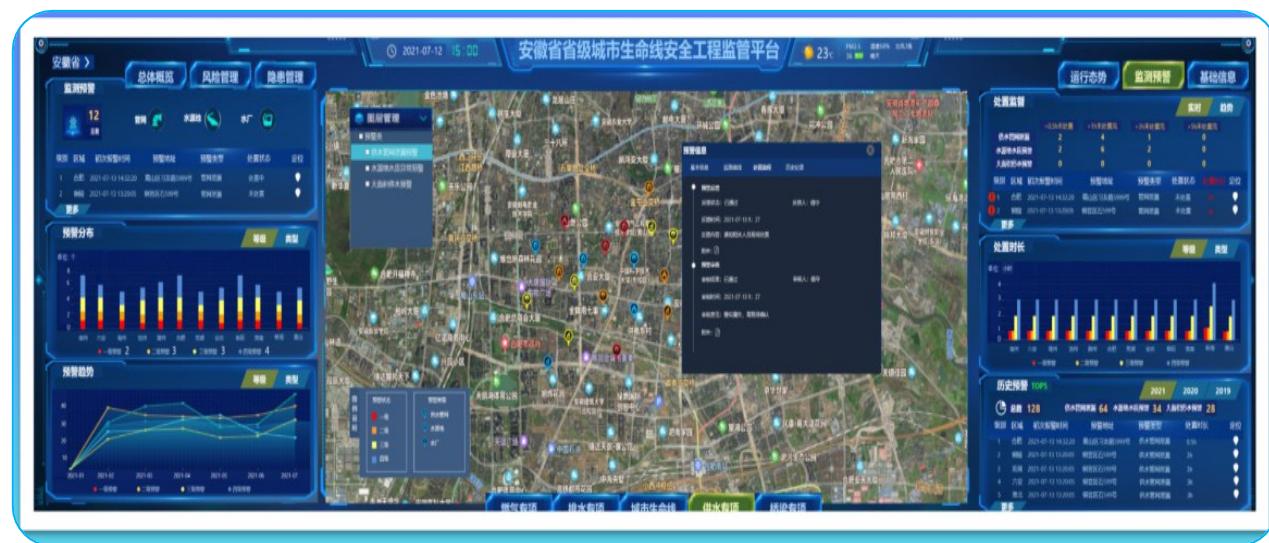
智慧城市 建设

在智能交通管控、环境监测、智慧政务服务、智能停车系统等多个方面提升城市运行效率和居民生活品质



社会治理—案例

安徽省城市生命线安全工程：通过前端传感器构建城市安全智慧化、全链条管理网络
构建“1+16”省级监测运营管理组织，立体呈现城市生命线监测运行“一张网”



已建立50多个智能化预警模型,在全省16个设区市已完成覆盖燃气、桥梁、水、排水防涝等重点领域
上线以来,预警处置了三级以上警情734起,其中燃气预警442起、桥梁40起、供水226起、排水26起



智能教育

个性化 学习

实时分析学生的学习表现和行为数据，提供定制化的学习路径、资源推荐和反馈，满足差异化学习需求

智能 辅导系统

充当虚拟教师或辅导员，解答学生疑问，解释概念，提供针对性练习，并给予即时反馈

自动化 评估

自动批改学生作业，提供详细评估报告和反馈
减轻教师负担



智能教育—案例

Duolingo 的 Duolingo Max，利用大模型提供了两项核心功能：“解释我的答案” (Explain My Answer) 和“角色扮演” (Roleplay)

解释我的答案

对用户犯的错误进行详细的语法解释，帮助用
户深入理解错误原因

不仅指出错误所在，还会提供**相关语法规则的解释、常见错误类型分析以及正确用法的示例**
让学习者从根本上理解并掌握知识点

角色扮演

提供与AI角色进行开放式对话的机会，模拟真
实场景下的语言交流

从餐厅点餐到商务谈判，**多种生活和工作场景**
AI会根据回答实时调整对话难度和方向，创造
沉浸式的语言学习环境

Outline

- 自然语言处理
- 大模型的原理与技术
- 大模型的核心技术进展
- 大模型应用
- 未来展望

人工智能的目标

➤ L1: 基础对话者

- **对话与交互能力:** 当前(如GPT-4),能流畅的对话
- **特点:** 以对话为核心, 提供信息、解答问题、辅助创作等

➤ L2: 推理者

- **推理与问题解决能力:** OpenAI-O1接近L2。PhD-Level。
- **特点:** 像人类一样分析问题、制定方案, 甚至在某些领域超越人类

➤ L3: 智能行动者

- **自主决策能力:** AI能独立思考并根据复杂情境采取行动(接近人类)
- **特点:** 不仅能够思考, 还能在真实世界中执行决策, 实现人机协同

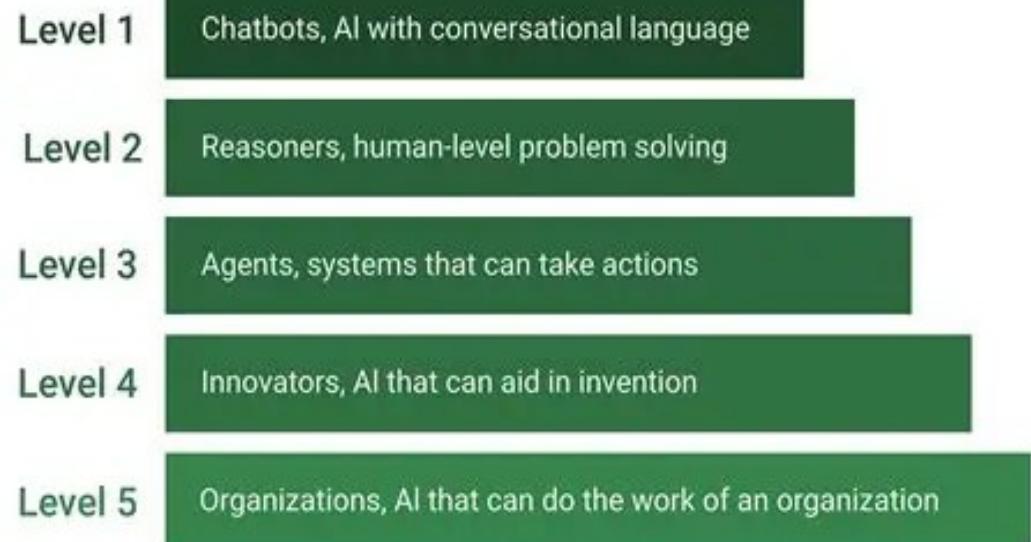
➤ L4: 创新者

- **创新能力:** AI能进行创造性思维和协助人类进行发明和创造, 推动科技进步
- **特点:** 具备创新思维和创造力, 能够提出新的想法和解决方案

➤ L5: 组织者

- **AGI:** AI能执行和组织人类所有工作, 标志着真正的人工通用智能的实现
- **特点:** 具备全面的人类智能和组织能力, 能够管理和优化整个社会的运行。

OpenAI's 5 Step to AGI



中美人工智能布局

全球人工智能的竞争格局：中美“双雄并立”

美国

- ◎ 拥有强大的基础研究能力、顶尖人才储备和完善的半导体产业链。
- ◎ 联邦政府已将发展人工智能提升为国策，并视中国为首要竞争对手。

中国

- ◎ 拥有海量的数据资源、广阔的应用市场和快速的商业化落地能力。
- ◎ 优势主要体现在应用层面，通过丰富的场景驱动技术迭代。

在全球人工智能领域，呈现出中国与美国“双雄并立”的竞争格局

率先抓住人工智能发展机遇的国家，将在新的世界秩序中占据主导地位

中美人工智能布局

全球人工智能的竞争格局：中美“双雄并立”

美国

- 2025年7月23日，美国总统特朗普签署《赢得AI竞赛：美国AI行动计划》
- 提出加速创新、建设基础设施及引领国际AI外交与安全三大战略支柱，包含100余项政策行动
- 旨在巩固美国在AI领域的全球主导地位

中国

- 2025年8月26日，《关于深入实施“人工智能+”行动的意见》：重塑生产生活范式 加快形成智能经济社会新形态
 - 到2027年，6大重点领域深度融合，智能终端普及率>70%
 - 到2030年，全面赋能高质量发展，智能终端普及率>90%
 - 到2035年，全面步入智能社会，支撑社会主义现代化

六大重点行动



科学技术



产业发展



消费提质



民生福祉



治理能力



全球合作

中美人工智能布局

美国路线

算力堆砌：美国xAI “Colossus” 集群目标达50万GPU规模，OpenAI “星际之门” 计划投入5000亿美元，为模型训练提供强大算力。大规模算力集群电费支出数亿美元，对数据中心散热、空间等基础设施要求极高。

闭源主导：OpenAI保留核心技术权重，追求AGI长期目标，闭源策略控制技术发展方向，保护商业利益，但限制技术快速传播。OpenAI核心技术如GPT系列模型权重严格保密，少数核心研究人员接触，全球深入研究和改进模型者有限。



中国路线

效率创新：中国注重效率创新，DeepSeek - R1以38亿参数实现GPT - 3水平，开源代码库吸引全球开发者，优化算法和架构实现性能提升。DeepSeek - R1开源后，吸引50多国开发者，贡献代码优化和新功能建议，推动模型快速迭代。

应用优先：中国人工智能发展注重应用落地，90%产业模型应用在智能客服、工业质检等领域，提高检测效率和准确性，降低人工成本。智能客服系统处理超80%常见客户咨询问题，提高客户服务响应速度和满意度。

中美人工智能布局—美国现状

01

技术领先优势

美国在AI技术方面长期处于领先地位，拥有众多世界级的科研机构和科技企业。这些机构和企业在AI算法、芯片设计、数据处理等方面具有深厚的技术积累和创新优势。



02

创新生态活跃

美国的创新生态非常活跃，包括风险投资、孵化器、加速器等多种创新资源。这些资源为AI创业企业提供了充足的资金支持和市场机会，促进了AI技术的快速迭代和商业化应用。

03

企业投入巨大

美国企业在AI领域的投入巨大，不仅用于研发新技术，还用于拓展应用场景和提升用户体验。这些企业的积极参与推动了AI技术的快速发展和广泛应用。



04

高校研究成果

美国的高校在AI领域的研究成果丰硕，涌现出众多世界级的科研成果和杰出人才。这些高校不仅为AI领域输送了大量高素质人才，还为企业的技术创新提供了重要支撑。



中美人工智能布局—中国现状



01

政策支持力度大

中国政府对AI发展给予了高度重视，出台了一系列扶持政策，包括资金投入、税收优惠、人才引进等。政策导向明确，旨在推动AI技术在各个领域的应用和创新，加速产业升级和转型。

02

科研成果丰硕

近年来，中国在AI领域取得了显著科研成果，包括深度学习、自然语言处理、计算机视觉等多个方向。这些成果不仅提升了中国的科技实力，也为AI技术的商业化应用奠定了坚实基础。

03

应用领域广泛

中国AI技术已广泛应用于智能制造、智慧城市、智慧医疗、金融科技等多个领域。这些应用不仅提高了生产效率和服务质量，也为民众带来了更加便捷、智能的生活方式。

04

人才培养体系

中国已建立起较为完善的AI人才培养体系，包括高校教育、职业培训、企业内训等多个层次。这些培养体系为AI领域输送了大量高素质人才，为行业的持续发展提供了有力支撑。

中美人工智能布局—任重道远

技术差距

2023 - 2025年，中美在MMLU性能差距从17.5%缩小到0.3%，中国通过工程优化快速追赶，缩小与美国差距。

2023年美国顶尖模型MMLU准确率85%，中国67.5%；2025年中国模型准确率提升至84.7%，接近美国85%。

产业生态对比

美国：

美国有40个顶尖模型，以基础研究为主，应用落地滞后，70%研究项目集中在基础算法和理论研究，产业应用项目仅30%。美国顶尖模型在基础研究和应用落地分布不均，与中国对比明显。

中国：

中国有15个顶尖模型，产业应用占比90%，如阿里通义千问广泛应用于电商、金融、政务等行业，提供智能化解决方案。阿里通义千问在电商提高客服效率60%，金融风险评估准确率提升50%。

政策博弈

美国《AI脱钩法案》限制中国技术输入，中国“东数西算”工程实现算力自主化，目标2030年AI流量占比达64%。

《AI脱钩法案》实施后，中国从美进口高端GPU数量2024年降60%，国产GPU市场份额从2023年15%升至2025年45%。

未来发展方向

- **具身智能—物理与数字融合**

- **医疗护理**: 康复机器人通过触觉传感器与多模态感知技术, 提供个性化康复方案
- **家庭服务**: 人形机器人集成情感交互与家务能力, 实现烹饪、清洁等场景服务



未来发展方向

□ 科学（研究/发现）范式: LLM4Science



第1范式：经验科学
观察现象
经验
1600年



第2范式：理论科学
理论模型
牛顿定律、电动力学方程等
1950年



第3范式：计算科学
数值计算
模拟
2010年



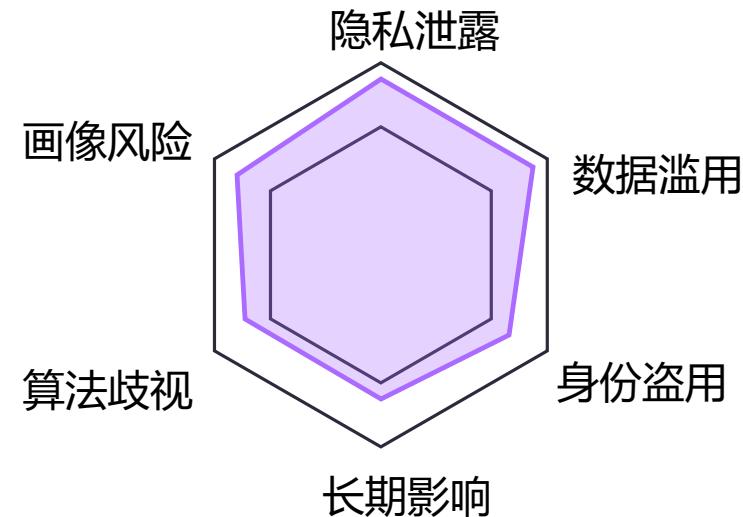
第4范式：数据驱动科学
大数据
数据建模、分析、挖掘
2022年



第5范式：智能驱动科学
大模型、智能体
自动化科学研究
2023年

风险与安全

► AI数据泄露



— 数据安全漏洞可能导致：

- 隐私泄露→ 骚扰/诈骗风险
- 数据滥用→ 算法歧视
- 画像偏差→ 发展机会限制

— 未成年人的数据足迹将影响：

- 就业背景调查
- 社会信用评估

风险与安全

风险：生成内容风险



1. 内容可靠性风险

大模型输出存在事实性错误风险，可能造成错误知识的系统性传播。

这种错误在教育教学、社会生活、政务管理中尤为危险



2. 价值取向偏差

算法训练数据中隐含的社会偏见会通过模型输出被放大。

这种偏差对人民价值观塑造构成潜在威胁



3. 认知框架影响

用户批判性思维尚未完全形成，可能不加批判地接受人工智能输出内容。

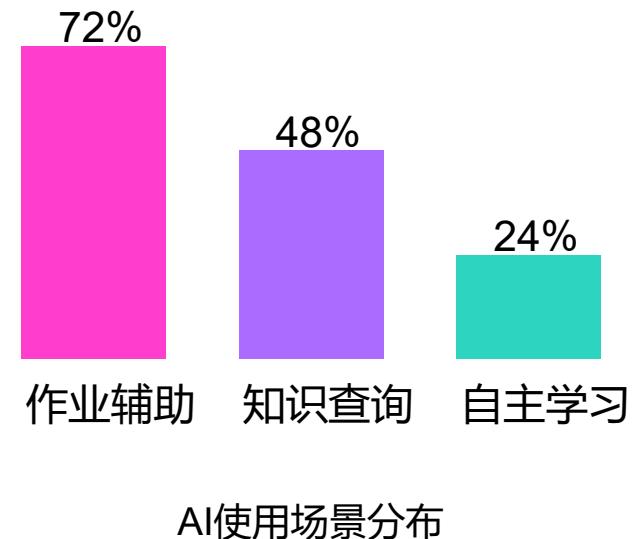
对青人民思维发展造成不可逆影响

风险与安全

AI依赖度分析



AI工具普及率



AI使用场景分布

DeepSeek变革引发了对学生学习主动性可能被弱化的担忧，如果学生过度依赖人工智能，可能会逐渐失去自主学习的动力和能力，这不仅违背了教育的初衷，还可能对学生的长远发展产生不利影响。

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task[△]

Nataliya Kosmyna¹
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

研究人员指出，相比于“全靠自己动脑”的基准组，使用搜索引擎的那组参与者大脑的 dDTF 连接强度下降了 34% 到 48%；而使用大语言模型（LLM）的那组，下降幅度更大，最多达到了 55%。

负责任的人工智能

“人在环路” (Human-in-the-Loop)

无论AI多么智能,人类研究者必须始终处于决策和监督的核心位置。AI是强大的工具,但最终的科学判断、结论的采纳和成果的责任,必须由人类承担

常怀敬畏之心 (Maintain Humility)

科学探索的道路充满未知, AI模型本身也存在局限性和不确定性。对技术保持谦逊,对科学保持敬畏,是避免傲慢和错误结论的前提

新时代的“智能增强者”

去体验：



不要只听别人说，亲自去试用一个AI工具。

去提问：



练习我们今天学到的方法，
与AI进行一场有深度的对话。

去创造：



用AI辅助你，完成一件你以前觉得
很难或没时间做的小事。

欢迎来到人机协同的新时代。未来，在自己手中。