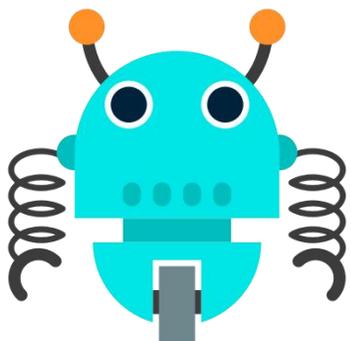


# 第02讲 人工智能前沿

---

欧新宇





- **人工智能的基本概念**
- **人工智能的发展简介**
- **人工智能发展的理论前沿**
- **人工智能发展的技术前沿**
- **人工智能发展的产业前沿**





# 人工智能发展的理论前沿

## Frontiers of AI Theory

- / 预训练大模型 (Pretrained large model)
- / 无监督学习 (Unsupervised learning)
- / 生成模型 (Generative model)
- / 博弈智能 (AI in gaming)

# 人工智能发展的理论前沿 (Frontiers of AI Theory)



NO.01

预训练大模型



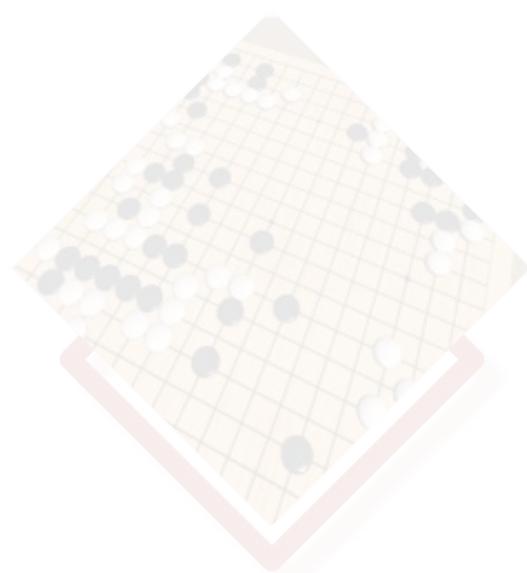
NO. 02

无监督学习



NO. 03

生成模型



NO. 04

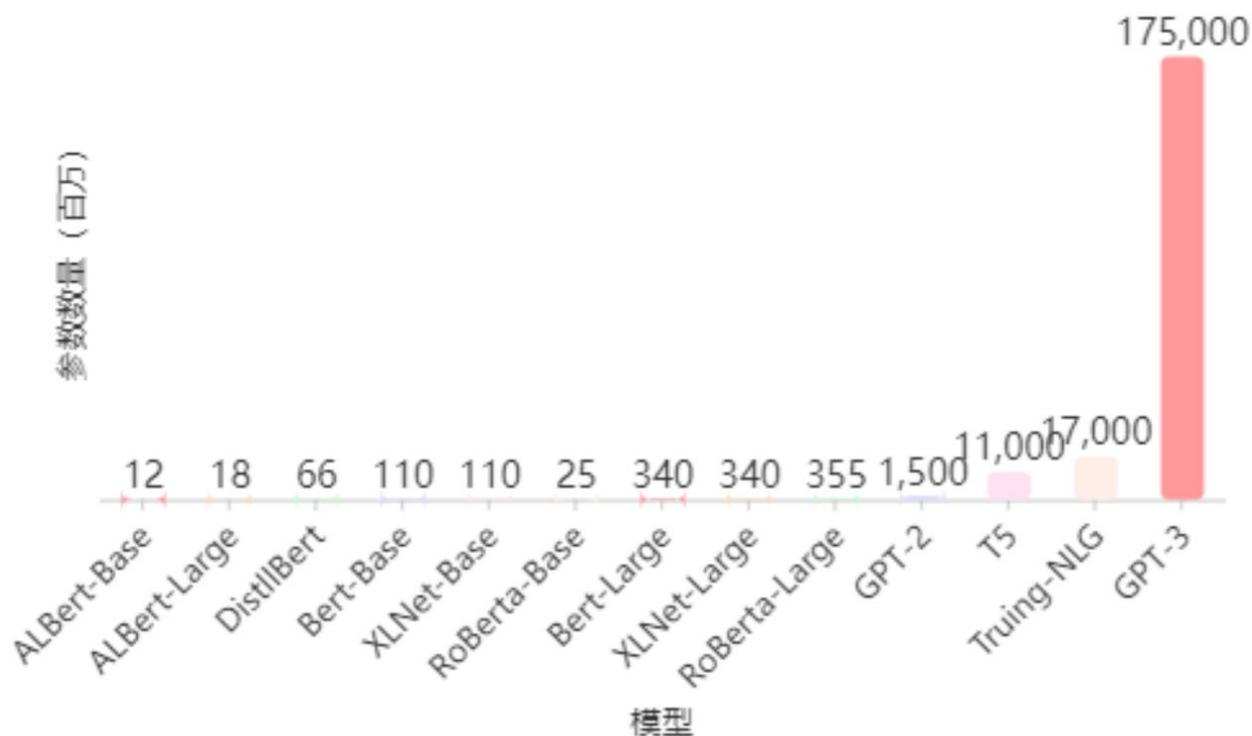
博弈智能

# (一) 预训练大模型 (Pretrained large model)

## 1. 大语言模型: GPT-3 / Large language model: GPT-3

- ① 2020年5月, OpenAI发布当时全球最大参数规模的**预训练语言模型GPT-3**。
- ② GPT-3包含**1750亿**参数, 训练数据量达**45TB**。
- ③ 对于所有任务, 应用GPT-3几乎都**无需进行任何梯度更新或微调**, 仅需与模型进行**文本交互**, 为其指定任务和展示少量演示, 即可完成目标任务。
- ④ GPT-3在许多自然语言处理数据集上都具有出色的性能, 在很多真实任务场景中**大幅接近人类水平**。

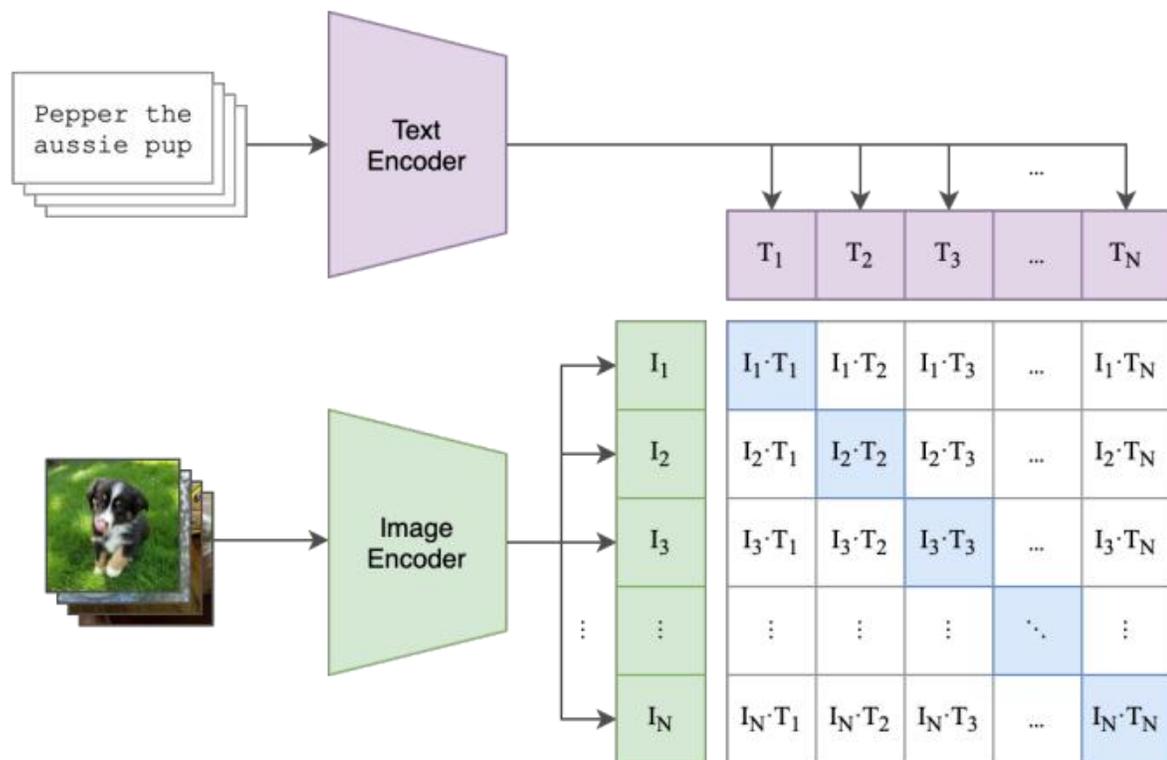
NLP预训练语言模型



# (一) 预训练大模型 (Pretrained large model)

## 2. “视觉-语言”大模型CLIP / “Vision-language” large model CLIP

(1) Contrastive pre-training

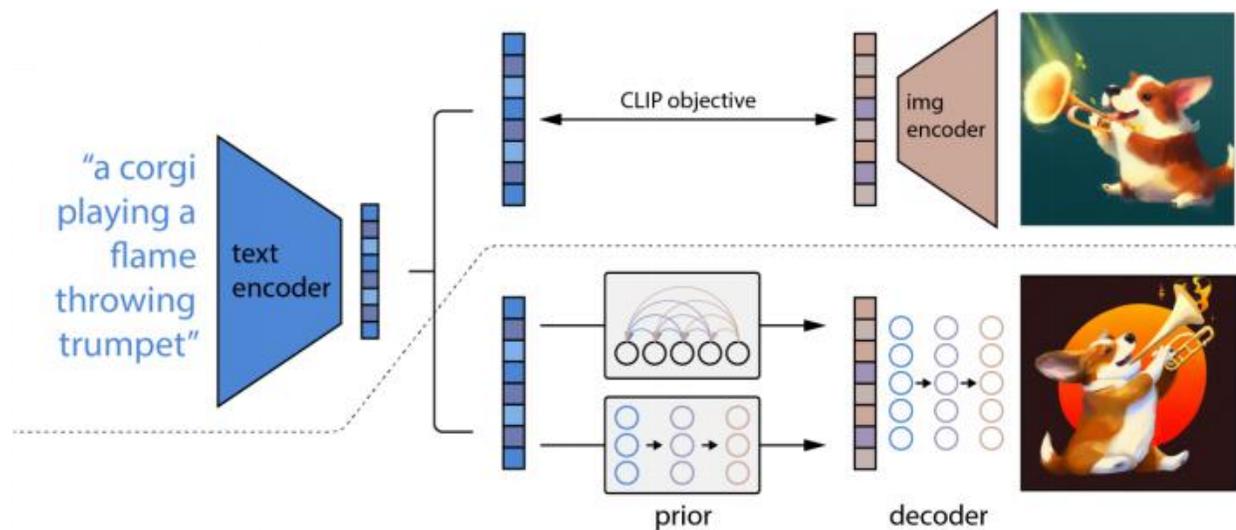
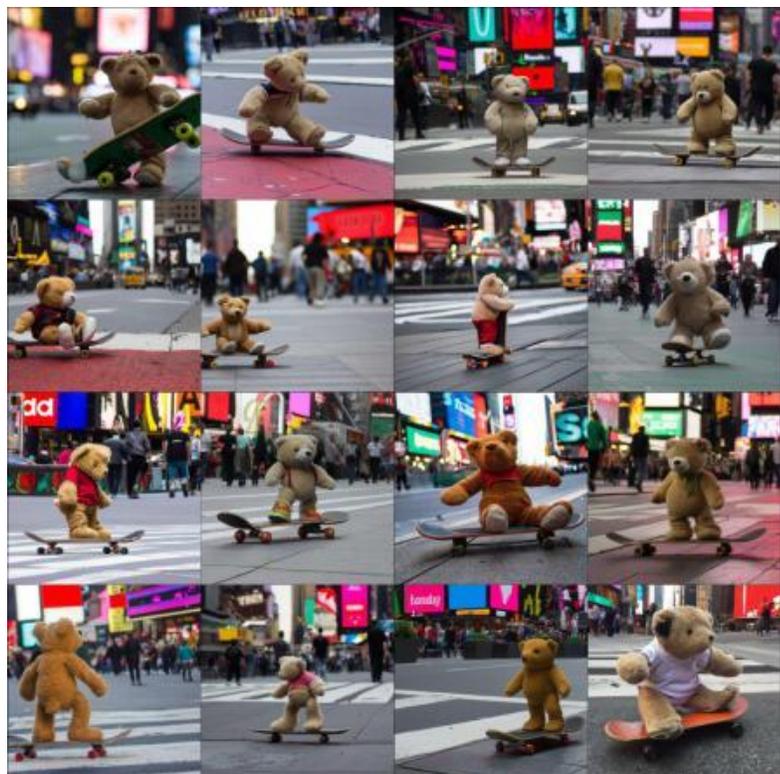


- ① 2021年3月, OpenAI提出CLIP模型, 该模型通过互联网收集的**4亿**个**文本-图像对**, 通过**对比学习**进行训练。
- ② 在预训练后, 通过**自然语言描述**所学到的**视觉概念**, 可以使模型能够在**零样本**状态下转移到**下游任务**。
- ③ CLIP的惊艳之处在于结构简单, 但性能好到难以置信, 可适配包括Zero-Shot的文本-图像检索、图像分类、开放域检测与分割, 以及文生图等。

# (一) 预训练大模型 (Pretrained large model)

## 3. 基于CLIP模型的应用 DALL.E / Applications based on CLIP

A teddybear on a skateboard in Times Square  
一只泰迪在时代广场上玩滑板



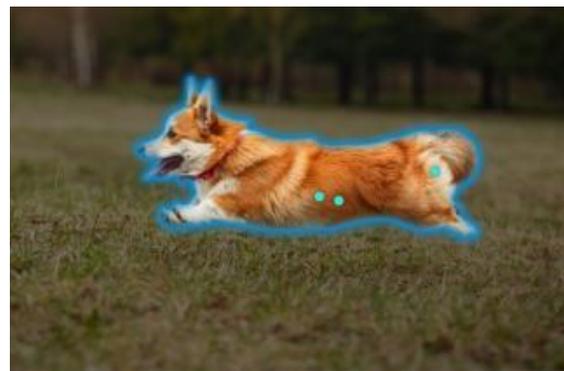
- ① 2022年4月, OpenAI发布了由**文本生成图像**的应用模型**DALL.E 2**, 该模型能够生成非常逼真的视觉效果。
- ② 这类逼真的生成技术给创作、娱乐带来了便利; 但也带来了巨大的安全风险 (如俄乌战争中的媒体假消息)。

Hierarchical Text-Conditional Image Generation with CLIP Latents. OpenAI. 2022.04

# (一) 预训练大模型 (Pretrained large model)

## 4. 视觉语义分割大模型 / Visual semantic segmentation large model

① 2023年4月，Meta公司提出**分割一切模型** (Segment Anything Model, SAM)，**突破了分割极限**，极大地促进了计算机视觉基础模型的发展。



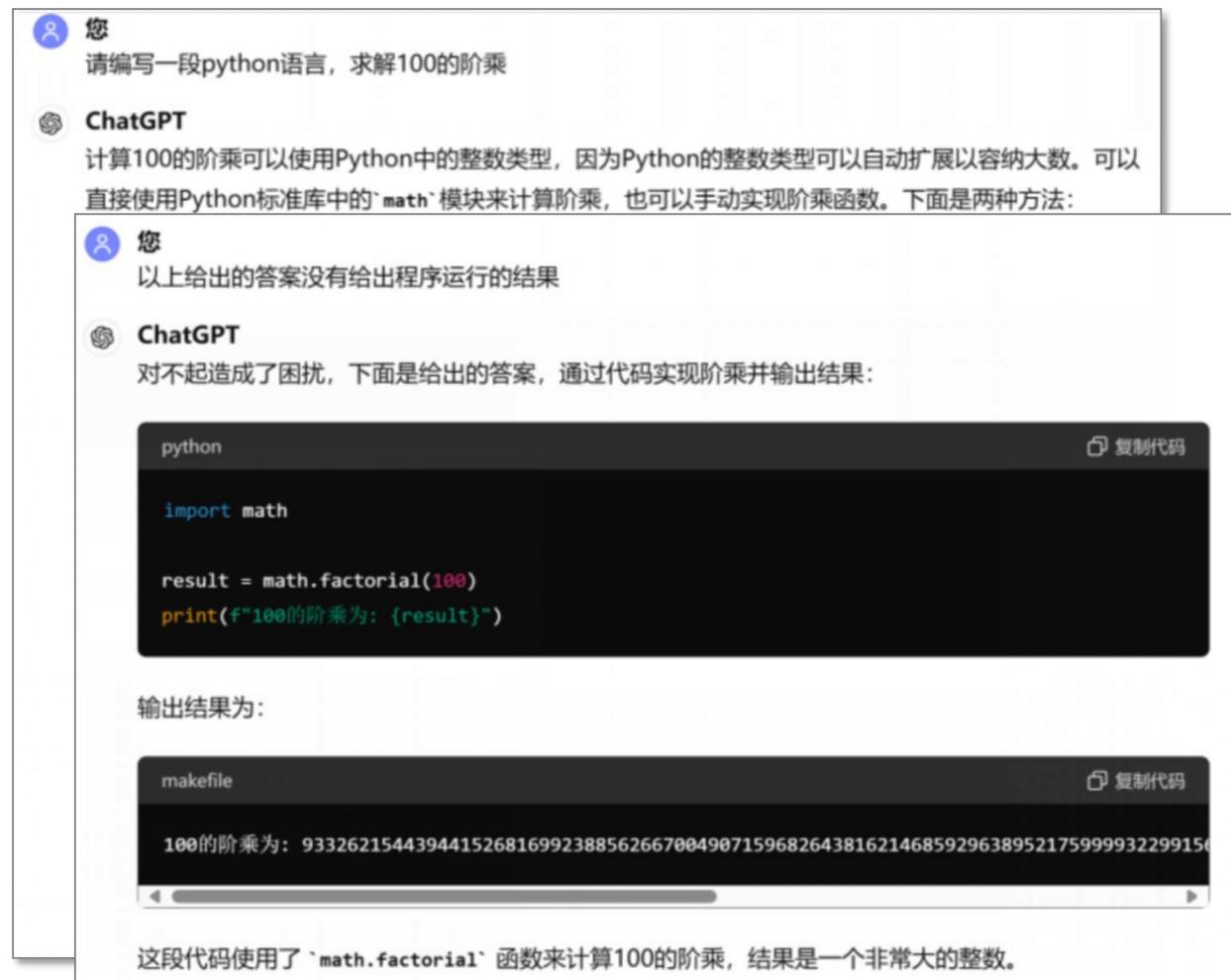
② SAM是一个**提示型**模型，其在**1100万张图像**上训练超过**10亿**个掩码，实现了强大的**零样本泛化**。



# (一) 预训练大模型 (Pretrained large model)

## 5. 优化的对话语言模型ChatGPT / Optimized dialogue language model ChatGPT

- ① 2022年11月，OpenAI发布优化对话的大型语言模型**ChatGPT**。它通过**监督微调**和**基于人类反馈的强化学习 (RLHF)** 来进行训练。
- ② ChatGPT能够在同一个会话期间内通过**对话**的形式回答后续问题、承认错误、质疑不正确的前提，并拒绝不适当的请求。



# (一) 预训练大模型 (Pretrained large model)

## 6. 多模态大模型 GPT-4

- ① 2023年3月，OpenAI发布**GPT-4**，它可以理解更复杂的问题、编写更大型的代码，并通过图片产生文字。
- ② GPT-4集成视觉语言组件，实现了图片文字解读，同时采用**对抗测试**、**基于规则的奖励模型**等技术，增强了其训练的稳定性和安全性。
- ③ GPT-4更是在美国律师资格考试和大学预科考试等专业测试上性能显著，超过了90%的应试者。



这张图片有哪些有趣的地方？请逐个描述它。

该图显示了一根Lightning的有线接口。

**图1：**将VGA连接器（一种大的蓝色 15 针连接器，通常用于计算机显示器）插入智能手机的充电端口。

**图2：**“Lightning Cable”适配器的包装，上面有VGA连接器的图片。

**图3：**VGA连接器的特写，末端有一个小的Lightning连接器（用于为iPhone和其他Apple设备充电）。

这张图片中的幽默表现在，将一个大的、过时的VGA连接器插入到一个小的、现代的智能手机充电端口，这显得很荒谬。

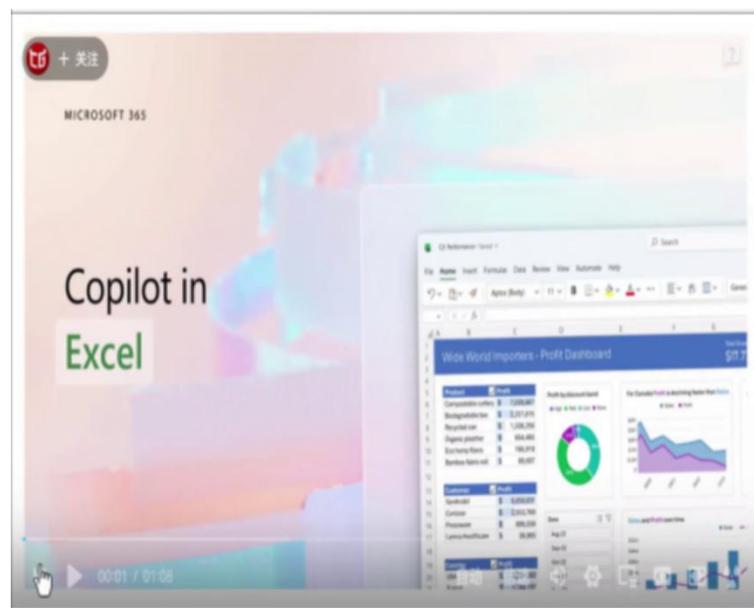
# (一) 预训练大模型 (Pretrained large model)

## 7. GPT-4 应用 / Applications of GPT-4

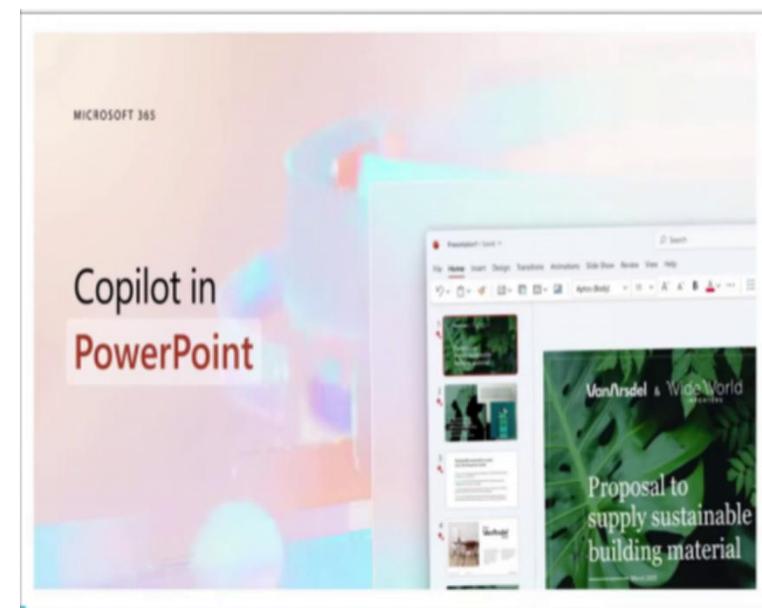
2023年3月，微软推出**Copilot**，将GPT-4集成到Word、Excel、PPT等办公软件，实现基于提示的文本生成、修改、总计以及PPT的自动生成、表格自动分析等。



基于提示生成文档



基于提示生成文档



基于提示生成文档

# (一) 预训练大模型 (Pretrained large model)

## 8. 多模态大模型 Gemini / Large multimodal model Gemini

- ① 2023年12月, Google发布了全新的多模态大模型 **Gemini**, 能够实现多源多模态数据的输入和输出, 包括文本、代码、视频、音频和图像, 并拥有较好的跨模态识别、推理能力。
- ② 不同于传统多模态大模型的训练过程, 在不同模态任务上分别训练再聚合; Gemini实现了**原生多模态训练**, 从一开始就对不同模态进行统一预训练, 然后再用额外的多模态数据进行微调。



跟随绘画进程进行介绍



根据不同颜色的毛线给出它们可以编织的物体

# (一) 预训练大模型 (Pretrained large model)

## 9. 多模态大模型 GPT-4o / Large multimodal model GPT-4o

- ① 2024年5月, **OpenAI**发布最新旗舰多模态大模型GPT-4o, 能跨越**听、说、看**;
- ② GPT-4o中的“o”是**Omni**的缩写, 意味“**全能**”, 即支持**文本、音频和图像**的任意组合作为输入, 并生成**文本、音频和图像**的输出;
- ③ GPT-4o能够在平均**320ms**内响应音频输入, **达到人类对话水平的反应速度**。



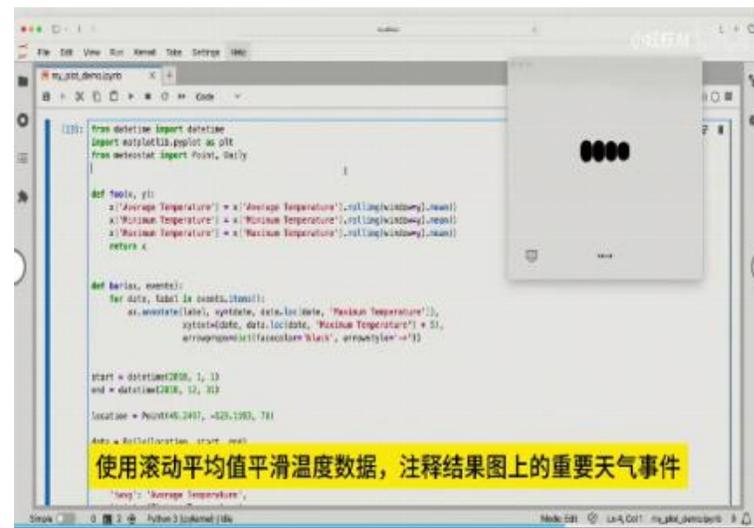
### 听力能力展示

GPT-4o指导演讲者通过深呼吸来缓解紧张情绪



### 视觉能力展示

GPT-4o根据工程师写下的方程指导其求解



### 代码能力展示

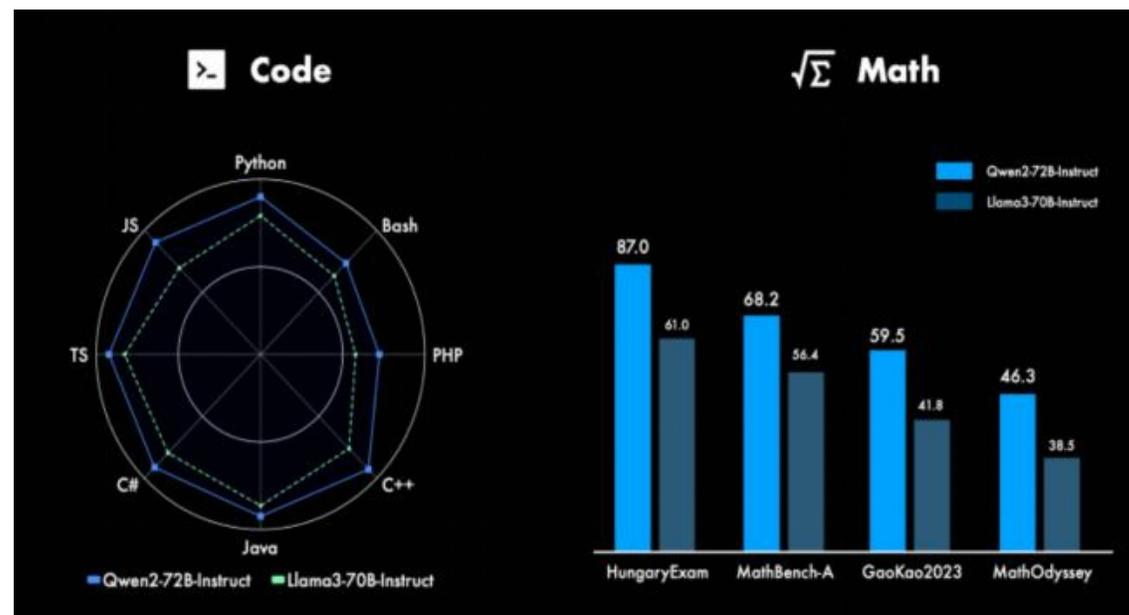
GPT-4o能够根据题意编写一段正确的代码

# (一) 预训练大模型 (Pretrained large model)

## 10. 开源大语言模型Qwen-2 / Open-source large language model Qwen-2

- ① 2024年6月, **中国阿里云**发布开源大模型**Qwen2-72B**, 在**代码**和**数学**能力上获得了显著提升, 并在各大基准测试中一举斩获十几项世界冠军;
- ② 在自然语言理解、代码、数学和多语言等多项能力上, Qwen2-72B显著超越了当前领先的**开源模型**, 包括Meta公司与今年4月发布的Llama3-70B模型。2025年3月, 千问QwQ-32B发布, 性能比肩671B满血DeeepSeek-R1.

	Qwen2-72B	Llama3-70B	Mixtral-8x22B	Qwen1.5-110B
MMLU	84.2	79.5	77.8	80.4
MMLU-Pro	55.6	52.8	49.5	49.4
GPQA	37.9	36.3	34.3	35.9
TheoremQA	43.1	32.3	35.9	34.9
BBH	82.4	81.0	78.9	74.8
HumanEval	64.6	48.2	46.3	54.3
MBPP	76.9	70.4	71.7	70.9
MultiPL-E	59.6	46.3	46.7	52.7
GSMBK	89.5	83.0	83.7	85.4
MATH	51.1	42.5	41.7	49.6
C-Eval	91.0	65.2	54.6	89.1
CMMLU	90.1	67.2	53.4	88.3
Multi-Exam	76.6	70.0	63.5	75.6
Multi-Understanding	80.7	79.9	77.7	78.2
Multi-Mathematics	76.0	67.1	62.9	64.4



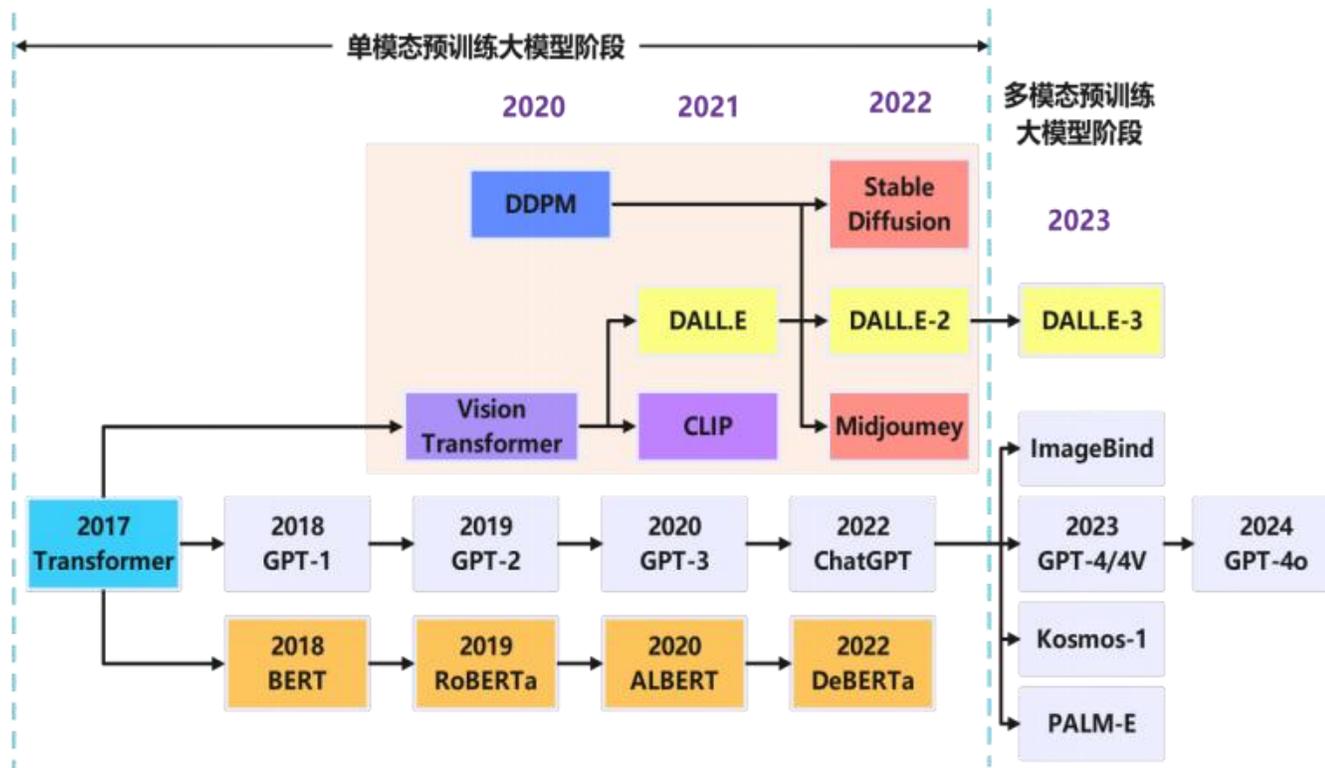
# (一) 预训练大模型 (Pretrained large model)

## 11. 预训练大模型发展迅速 / Pretrained large models are developing rapidly

### ① 世界各类大模型井喷式发布



### ② 从单模态逐渐向多模态发展



人工智能从“大炼模型”到“炼大模型”

# 人工智能发展的理论前沿 (Frontiers of AI Theory)



NO.01

预训练大模型



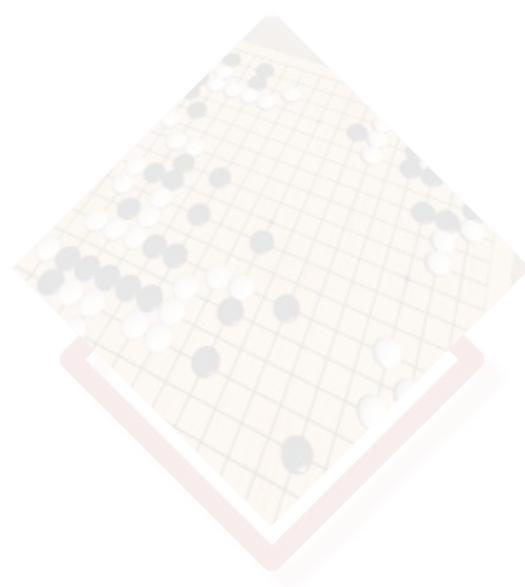
NO. 02

无监督学习



NO. 03

生成模型



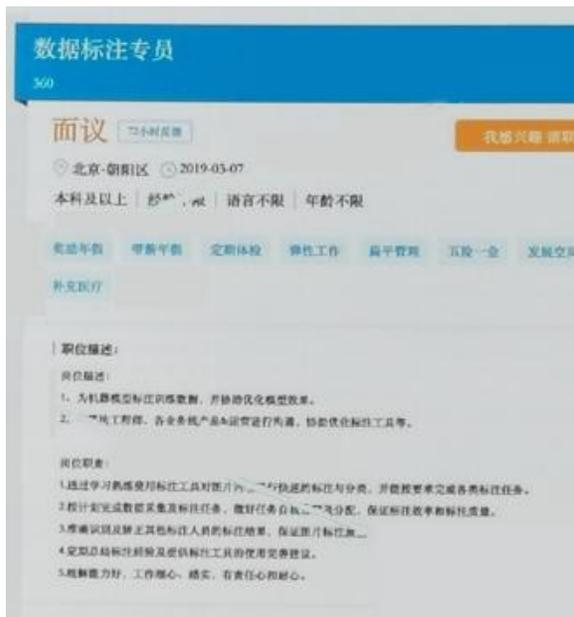
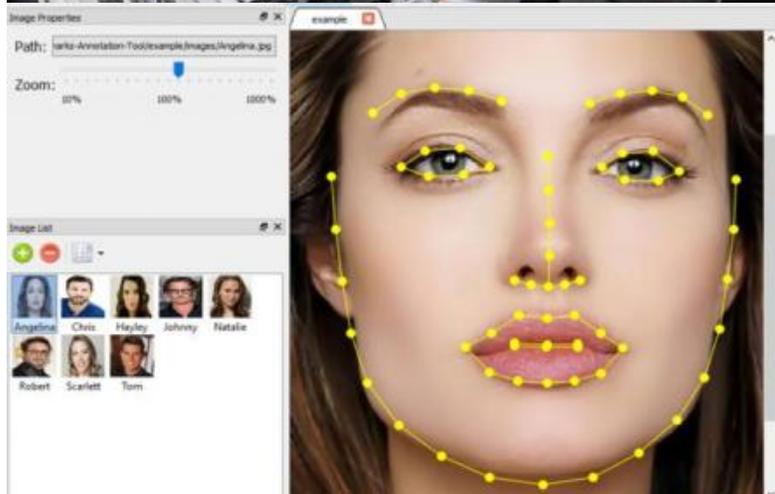
NO. 04

博弈智能

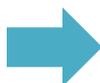
# (二) 无监督学习 (Unsupervised learning)

## 1. 无监督学习: 数据为王 / Unsupervised learning

人工智能领域最为成功的模型一直都是**监督学习**，但监督学习需要大量的训练样本支持。



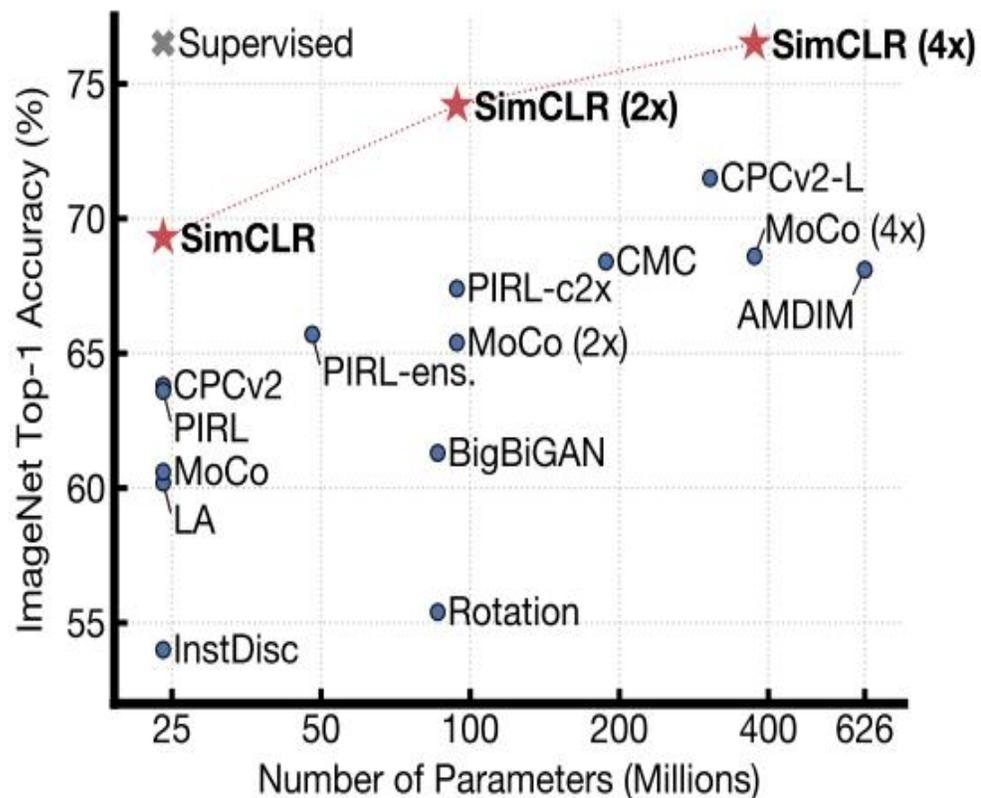
百度山西数据标注基地 (5万AI标注师)



告别“劳动密集型” AI才是走向真正人工智能的第一步。

## (二) 无监督学习 (Unsupervised learning)

### 2. 无监督学习：对比学习 / Unsupervised learning

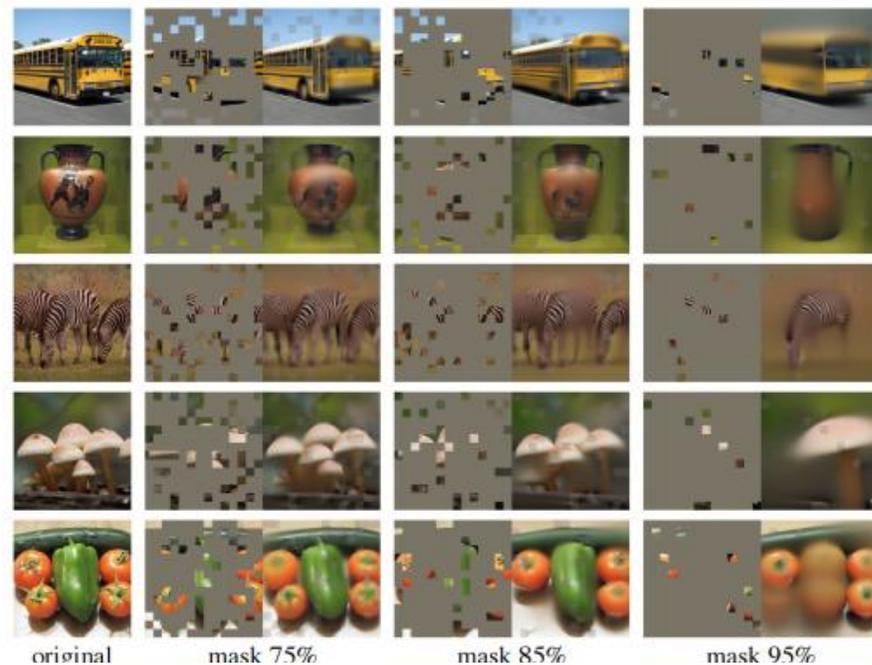
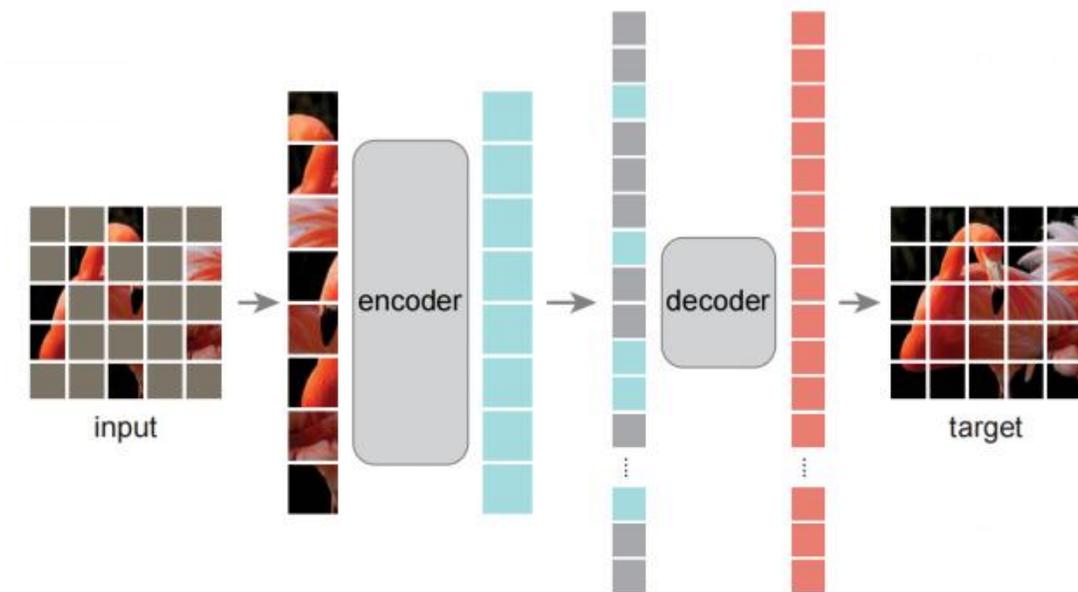


- ① 2020年, **Google**、**Deepmind**、**Facebook** 分别提出了**SimCLR**、**MoCo**、**BYOL**算法, 这些算法都可以在**无标注数据**上学习图像特征, 它们背后的框架都是**对比学习**。
- ② 无监督学习的核心是样本的“**可区分性**”, 即输入是**同一目标的不同视角**, 还是**完全不同的目标**。
- ③ 对比学习在ImageNet数据集上展现了**超过**或**接近**监督算法的性能, 未来可期!

## (二) 无监督学习 (Unsupervised learning)

### 3. 无监督学习: 掩膜学习 / Unsupervised learning: Constrastive learning

- ① 2021年11月, Facebook提出视觉掩膜学习范式MAE, 在ImageNet上取得87.8%的识别精度, 超越监督学习。
- ② MAE范式正在推动视觉和语言学习走向统一。



method	pre-train data	AP <sup>box</sup>		AP <sup>mask</sup>	
		ViT-B	ViT-L	ViT-B	ViT-L
supervised	IN1K w/ labels	47.9	49.3	42.9	43.9
MoCo v3	IN1K	47.9	49.3	42.7	44.0
BEiT	IN1K+DALI	49.8	<b>53.3</b>	44.4	47.1
MAE	IN1K	<b>50.3</b>	<b>53.3</b>	44.9	47.2

The masked autoencoders(MAE) architecture.

Table 4. COCO object detection and segmentation using a ViT

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollar, Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. arXiv0202.

# 人工智能发展的理论前沿 (Frontiers of AI Theory)



NO.01

预训练大模型



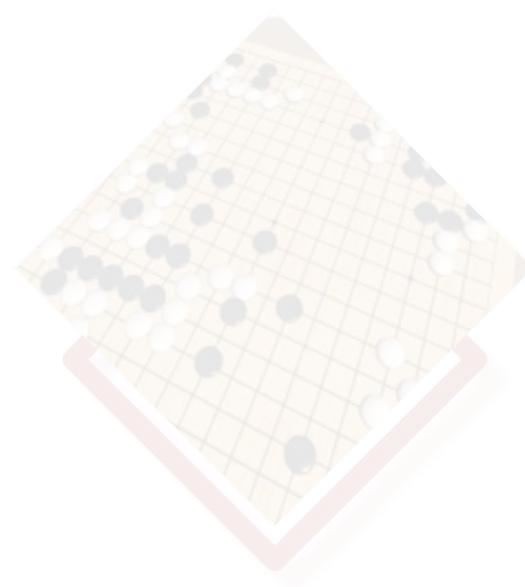
NO. 02

无监督学习



NO. 03

生成模型



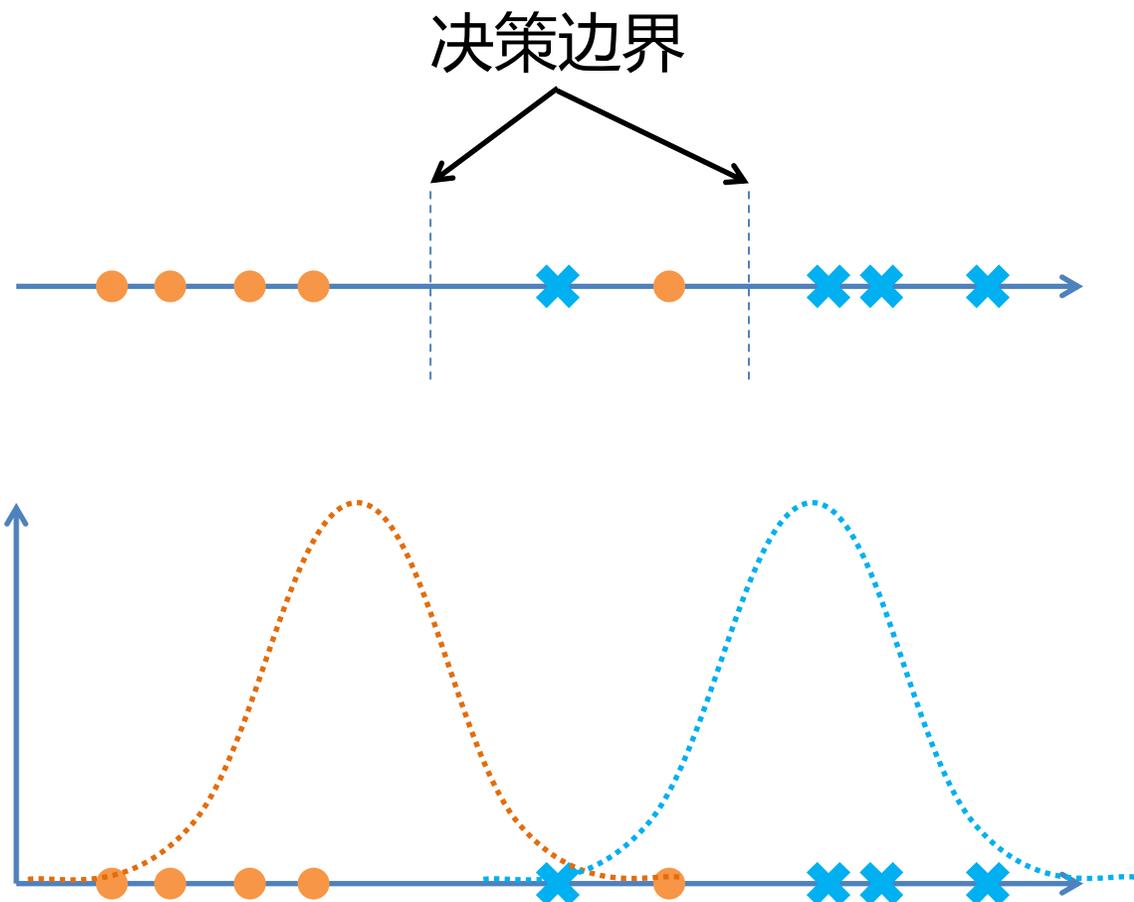
NO. 04

博弈智能

# (三) 生成模型 (Generative model)

## 1. 生成式模型和判别式模型 / Generative and discriminative models

根据先验概率的利用形式，机器学习大体上可以分为**生成模型**和**判别模型**两类。



**判别模型**从历史数据中总结经验，并将犯错最少的位置作为决策边界。

适用于**预测数据的标签**。

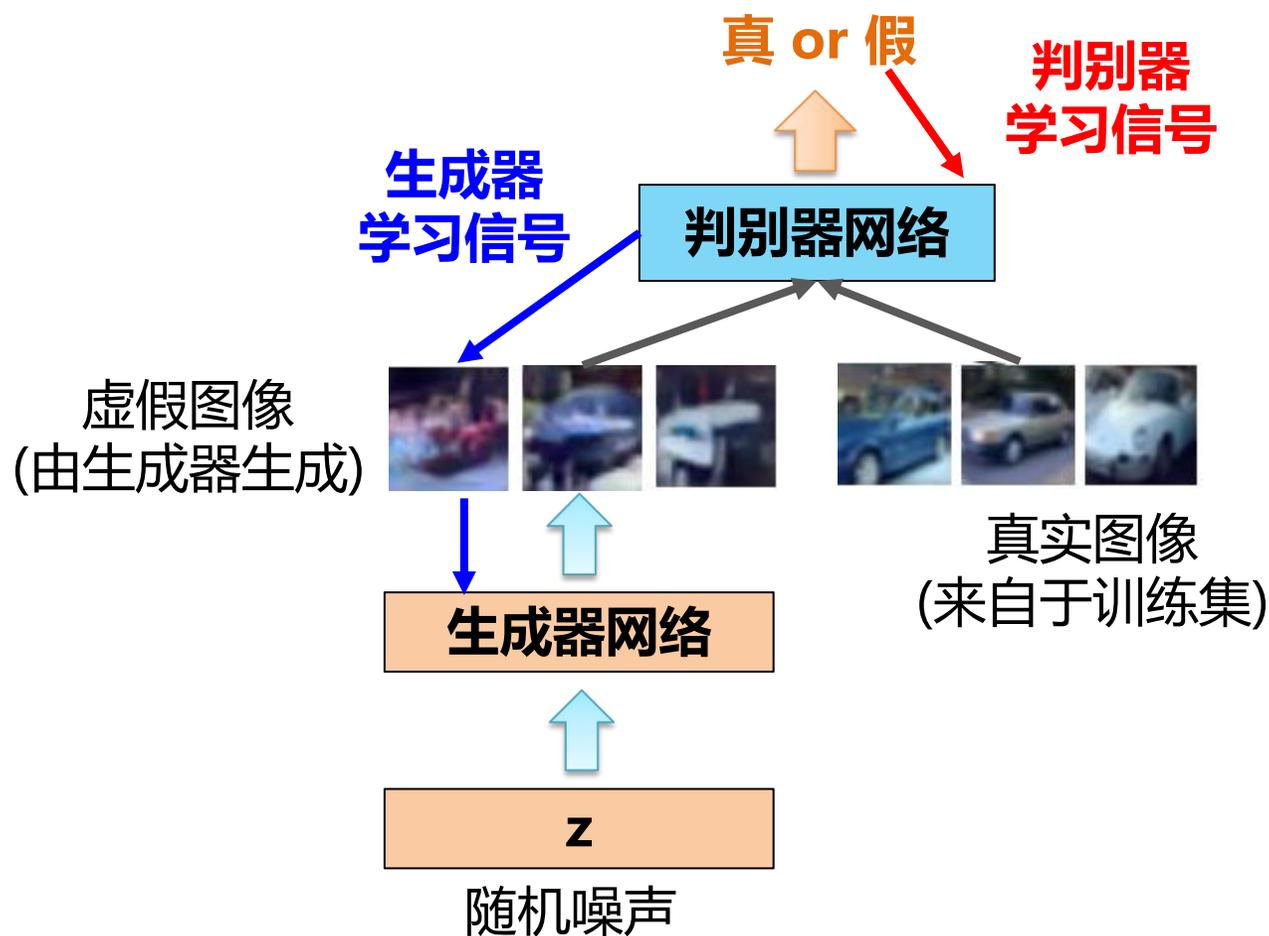
**生成模型**估计各类型数据的联合分布，比较不同类型生成该数据的概率。

适用于**生成新数据**。

# (三) 生成模型 (Generative model)

## 2. 生成对抗网络 / Generative adversarial networks

- ① 2014年, **Ian Goodfellow**提出**生成对抗网络 (Generative Adversarial Network, GAN)**。GAN结合了**生成模型**和**判别模型**,使二者在相互对抗中相互优化,从而实现不要人工标注数据的无监督学习。
- ② GAN提出了一种**网络指导网络学习**的创新性观点,为很多生成模型提供了一种新的训练思路,催生了许多后续作品。



Ian Goodfellow et al.. Generative Adversarial Nets. NIPS 2014.

# (三) 生成模型 (Generative model)

## 2. 生成对抗网络: 应用 / Generative adversarial networks



GAN  
2014

DCGAN  
2015



ProGAN  
2017

styleGAN  
2018

基于GAN的人脸生成

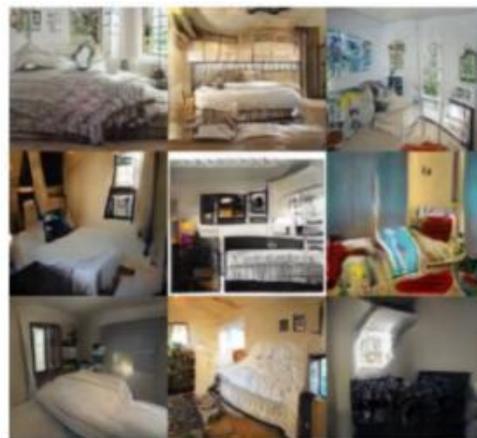


horse → zebra



apple → orange

跨域图像生成  
CycleGAN, 2017



室内场景生成  
LSGAN, 2017

this small bird has a pink breast and crown, and black primaries and secondaries. this magnificent fellow is almost all black with a red crest, and white cheek patch.



文生图 AttnGAN, 2018



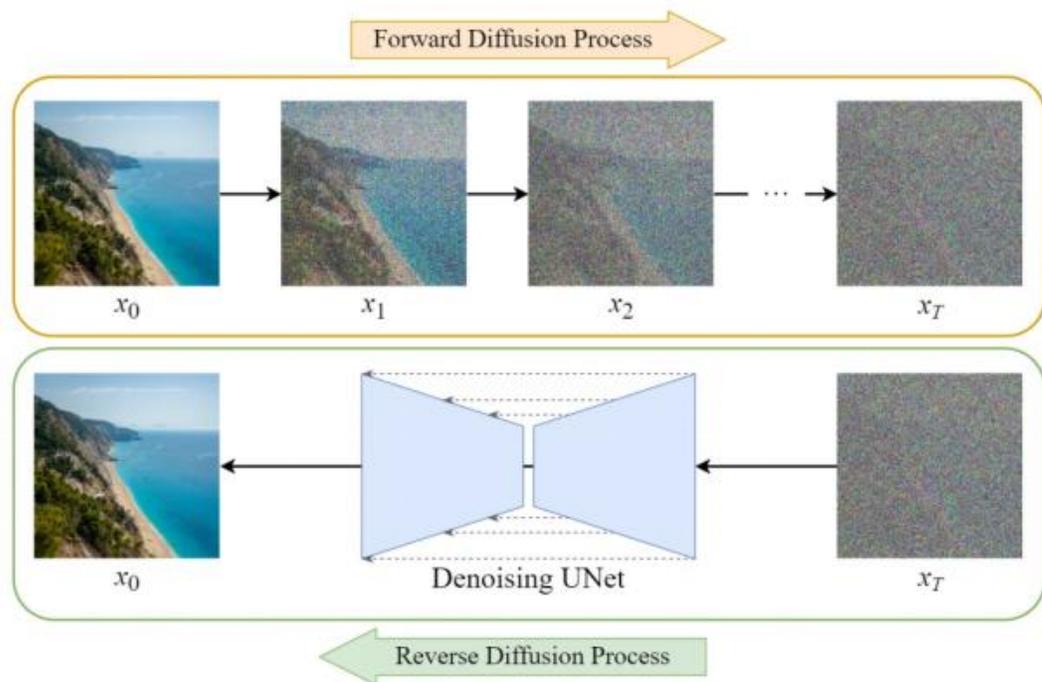
超分辨率, SRGAN, 2017



风格迁移, NeuralArt, 2015

# (三) 生成模型 (Generative model)

## 3. 扩散模型 / Diffusion model



- ① 2020年，**加州伯克利**的学者们提出了**扩散模型 (Diffusion model)**，建立了一个图像数据扩散成纯噪声的马尔可夫模型，成为**生成模型的新标杆**。
- ② 扩散模型很好地解决了过去主流生成模型训练不稳定、收敛困难、样本质量不可控的问题，并且使生成模型向着**更大规模**、**更精细的生成粒度**和**高像素**发展。

# (三) 生成模型 (Generative model)

## 4. 多模态/跨模态内容生成 / Text-to-video model

2022年，多个基于**扩散模型**的文本-图像生成模型发布，这些模型都预示着多模态时代的来临。

- ① **中国百度文心**发布中文-图像扩散模型 (**ERNIE-ViLG 2.0**)，它是首个知识增强跨模态生成大模型，基于混合降噪专家模型框架，在文生图任务中刷新了世界记录。
- ② **剑桥大学**联合**腾讯AI Lab**等提出了MAGIC (**iMAge-guided text Generation with CLIP**) 框架，使得语言模型在解码过程中选择更接近图片信息的生成结果。



百度文心 ERNIE-ViLG 2.0

# (三) 生成模型 (Generative model)

## 5. 文生视频模型 Sora / Text-to-video model Sora



中国龙年舞龙视频



咖啡杯中激战的海盗船近景视频



**Prompt:** 一位24岁女性眨眼睛的特写镜头、采用70毫米胶片拍摄的电影镜头



**Prompt:** 在一个培养皿中，生长着一片竹林，其中小熊猫们在欢快地奔跑。



- ① 2024年2月，**OpenAI**推出文生视频模型 **Sora**，它通过学习视频内容来理解现实世界的动态变化，并模拟这些变化，创作**长达一分钟**的各种风格视频。
- ② 借助大语言模型，Sora能够通过提示进一步丰富生成内容的语义描述，并理解现实世界的一般规律，保持视频**生成的连贯性**。
- ③ 不过Sora并不完美，它难以**准确模拟**复杂场景的物理现象，也无法理解具体的**因果关系**。

# 人工智能发展的理论前沿 (Frontiers of AI Theory)



NO.01

预训练大模型



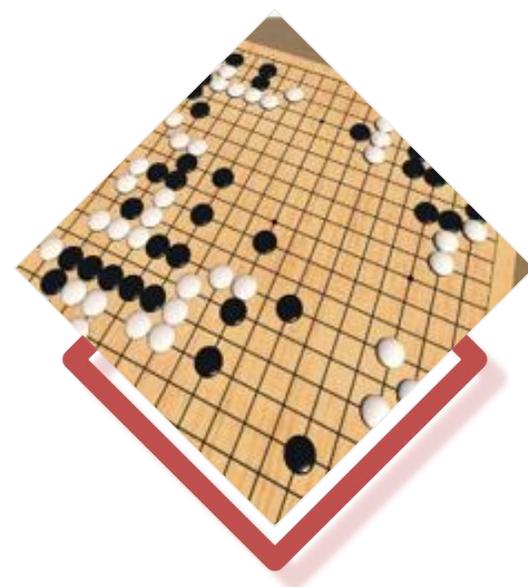
NO. 02

无监督学习



NO. 03

生成模型



NO. 04

博弈智能

# (四) 博弈智能 (AI in gaming)

## 1. 博弈智能 / AI in gaming

人机博弈  
近年来  
取得了  
较大突破

□ 五子棋

□ 国际象棋

□ 围棋

□ 德州扑克

□ 麻将

□ Dota2

□ 星际争霸II



1997年5月11日, **IBM的DeepBlue (深蓝)** 战胜国际象棋冠军卡斯帕罗夫。



2016年3月9日, **DeepMind的AlphaGo**以 **4:1** 战胜九段棋手李世石, 2017年再次以 **3:0 完胜** 世界冠军柯洁。



2018年7月19日, **OpenAI Five**在实时5v5 团队赛中以2:1战胜了人类职业玩家; 而在一个月前, OpenAI Five才刚刚战胜业余选手!



2019年1月25日, **DeepMind的AlphaStar** 在1v1多兵种竞赛中以10:1战胜星际争霸II职业选手。

# (四) 博弈智能 (AI in gaming)

## 1. 博弈智能 / AI in gaming



**AlphaStar**  
非完全信息动态  
决策博弈



启元世界 - 三国志战棋版  
3D开放世界FPS游戏, 2022



**AlphaZero**  
自我博弈深度  
强化学习



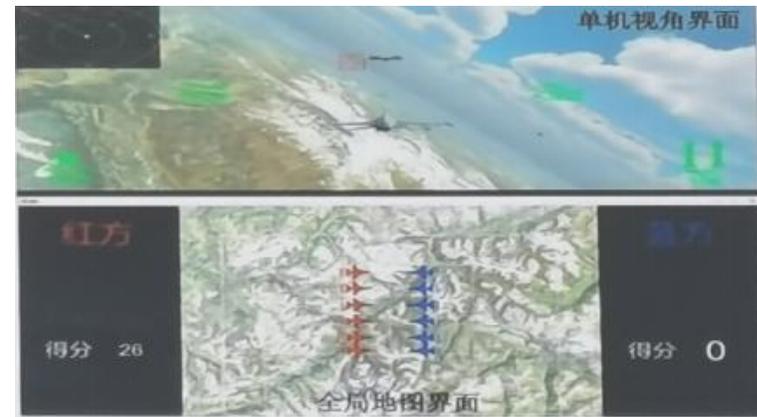
中国腾讯  
开悟人工智能全国公开赛, 2023



**AlphaGo**  
人类知识 (棋  
谱)、随机创新  
的深度强化学习



中国人工智能学会、中科院自动化所  
“妙算杯”人机混合挑战赛, 2023



中国指挥与控制学会  
智能空中博弈算法挑战赛, 2023

## (四) 博弈智能 (AI in gaming)

### 2. 多智能体捉迷藏游戏 / Multi-Agent Hide and Seek



**智能体在对抗博弈过程中不仅学到了如何玩捉迷藏游戏，而且发现了人类玩家都很难想到的技巧。**

## (四) 博弈智能 (AI in gaming)

### 3. 中国腾讯“绝悟”夺冠NeurIPS MineRL 挑战赛 / Tencent “JueWu-MC”

- ① 2021年12月, **中国腾讯AI Lab**发布的“**绝悟**”在第三届NIPS MineRL竞赛主赛道中以超出第二名3倍的绝对优势夺冠, 这一成绩展示了“绝悟”在复杂决策环境中的卓越能力。
- ② “绝悟”包含四项**关键技术**, 分别是分层强化学习、表示学习、自模仿学习、集成行为克隆。





# (四) 博弈智能 (AI in gaming)

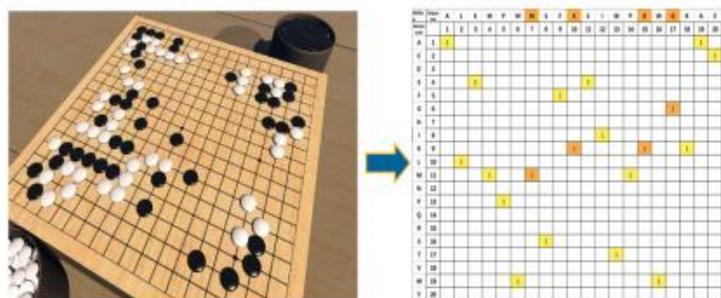
## 5. 开放任务自适应智能体 / Adaptive agent for open-ended tasks

- ① 2023年1月, **DeepMind** 提出**自适应智能体模型 (AdA)**, 利用**课程学习**和**元强化学习**, 实现在**开放任务空间**下的大规模训练。
- ② 该模型证明了经过大规模训练后的智能体强化学习模型能具有**通用上下文**的学习能力。适应性依赖于三个方面:
  - a. **元强化学习**;
  - b. 大规模**注意力机制**;
  - c. 有效的**自动化**课程设计策略。



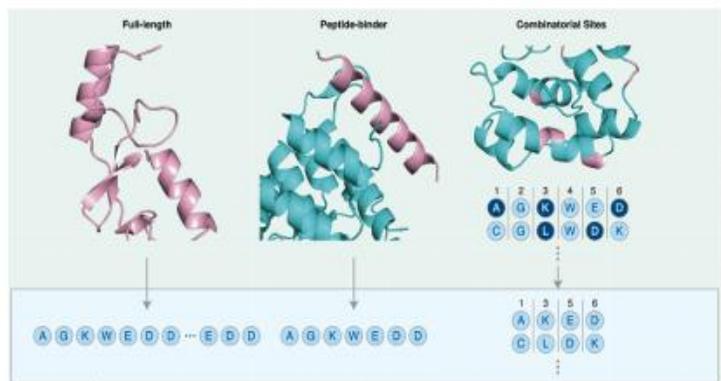
# (四) 博弈智能 (AI in gaming)

## 6. 中国华大制造利用自博弈实现蛋白质进化 / MGI-Tech exploits self-play to guide protein evolution"

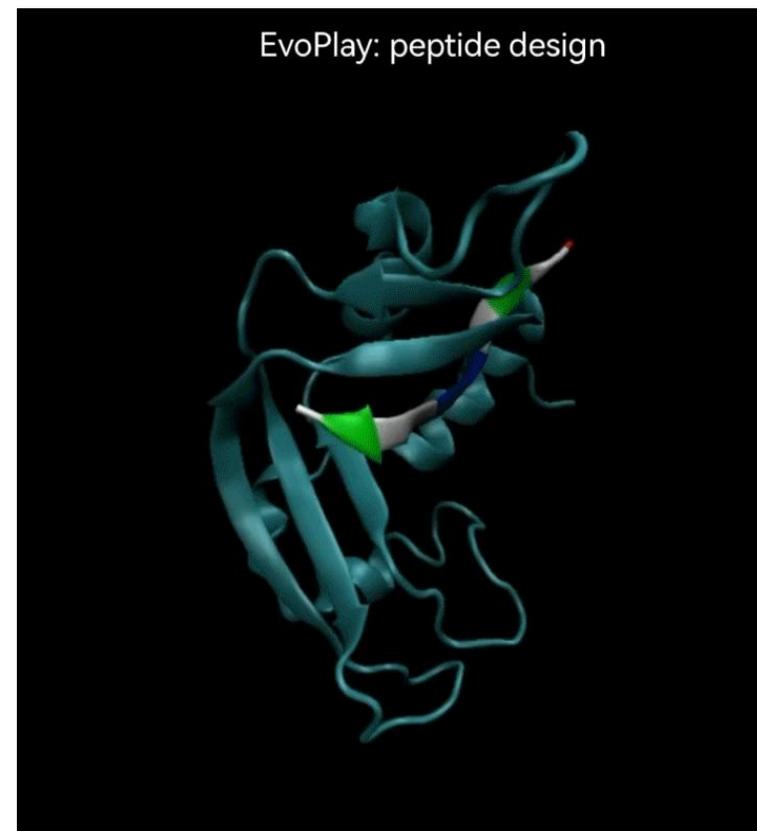
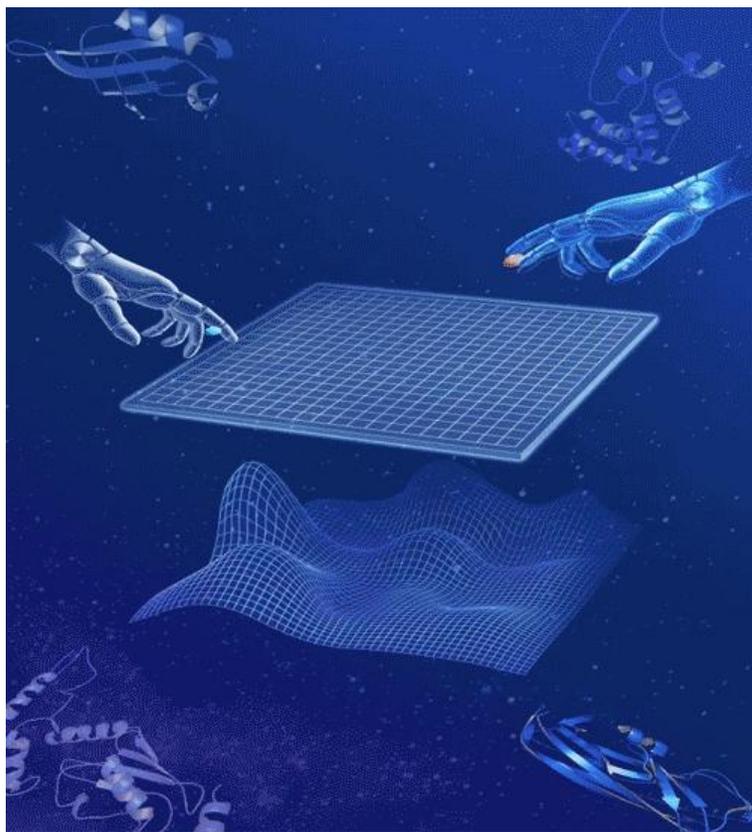


棋盘空间

蛋白空间



蛋白设计



Yangmeng, et.al. Self-play reinforcement learning guides protein engineering. Nature machine intelligence. 2023.



# 人工智能发展的技术前沿

## Frontiers of AI Technology

# 人工智能发展的理论前沿

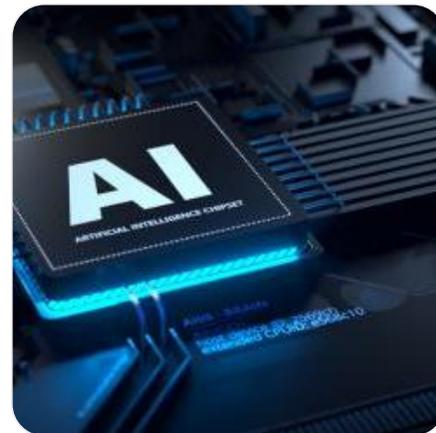
## 自动驾驶

*Autonomous driving*



## AR/VR/MR与元宇宙

*AR/VR/MR and metaverse*



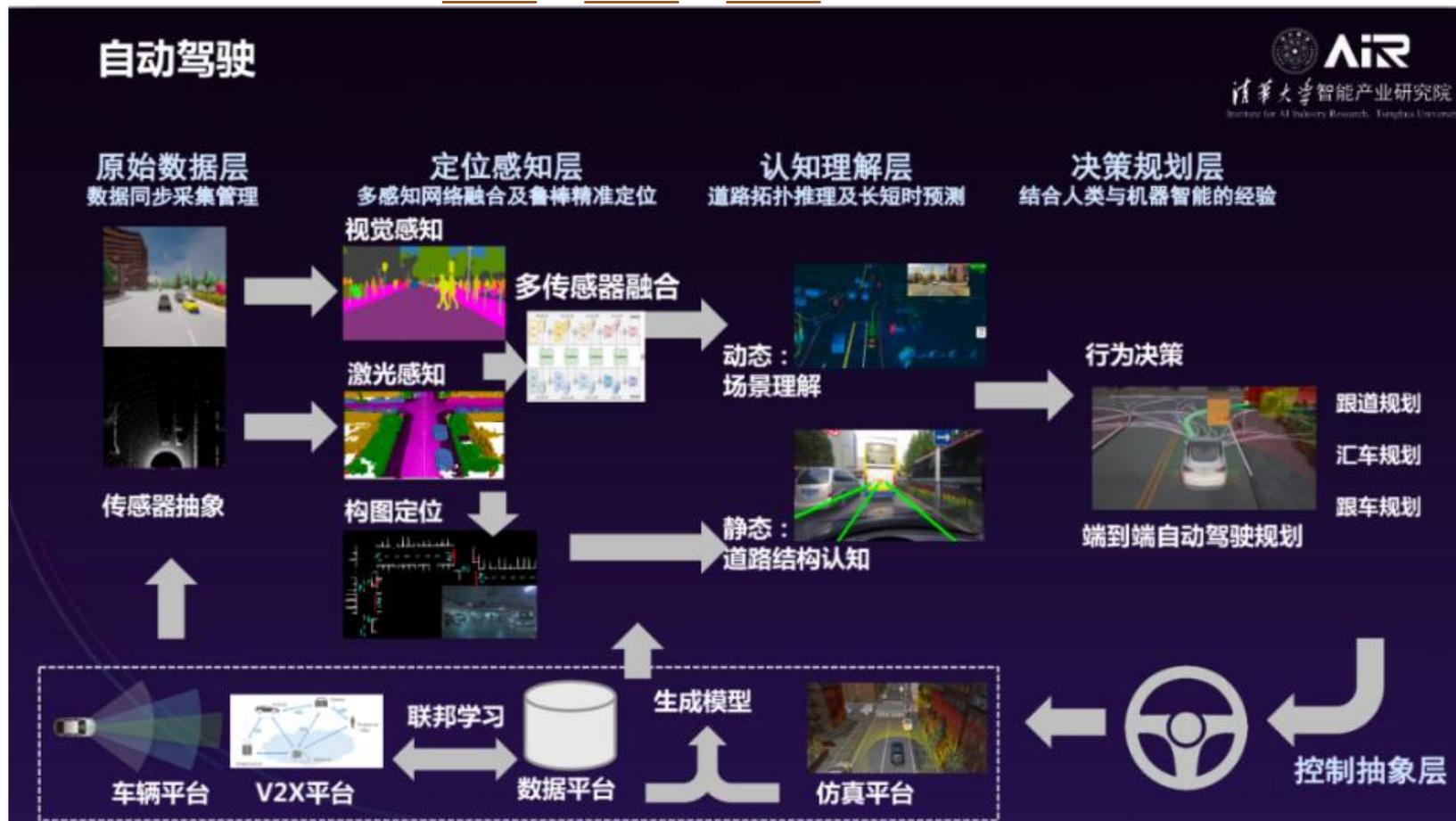
## 人工智能与科学计算

*AI for Science*

# (一) 自动驾驶 (Autonomous driving)

## 1. 自动驾驶 / Autonomous driving

**自动驾驶**是一个复杂的系统工程，包含了从硬件到软件、从数据采集、定位感知、认知理解到决策规划的全链条人工智能落地流程，甚至于市场、政策和伦理也是影响它发展的重要因素。



张亚勤  
走向无人驾驶  
清华大学  
2021

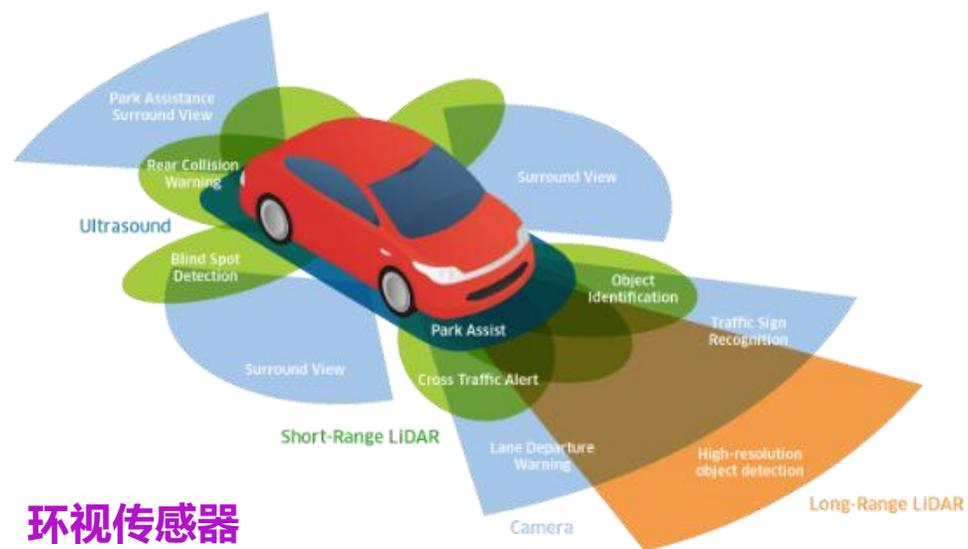
# (一) 自动驾驶 (Autonomous driving)

## 2. 自动驾驶技术 -- BEV环视感知 (Bird's Eye View, 鸟瞰图) / BEV perception

① 2020-2022年, 英伟达、特斯拉、商汤、百度等公司先后提出**BEV环视感知架构**, 显著提升了机器视觉的感知能力;

② **核心思想**: 环视传感器、BEV鸟瞰图特征提取、Transformer模型融合;

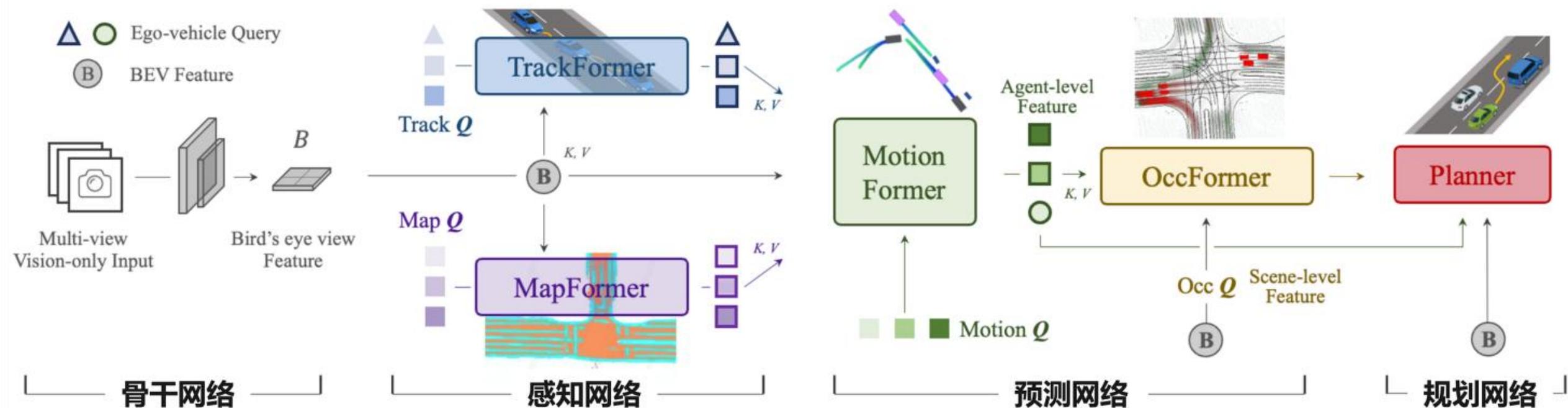
③ **缺点**: 模型较大、速度较慢、计算复杂度高、落地成本高。



# (一) 自动驾驶 (Autonomous driving)

## 3. 自动驾驶技术 -- 自动驾驶大模型 / Large models for autonomous driving

- ① 2023年7月, **中国上海AI实验室**联合团队的研究成果《以路径规划为导向的自动驾驶》(《Planning-Oriented Autonomous Driving》)摘取CVPR2023**最佳论文奖**。
- ② 该论文首次提出**感知决策一体化的自动驾驶通用大模型**, 开创了以全局任务为目标的**自动驾驶大模型架构**先河, 为自动驾驶技术的发展创新提出了新的方向。



# (一) 自动驾驶 (Autonomous driving)

## 4. 自动驾驶的挑战 / Challenges of autonomous driving



① 2019年，美国佛罗里达州一辆**特斯拉Model3**以110公里/小时的速度撞向一辆横穿马路的白色拖挂卡车。当时Model3的自动驾驶系统Autopilot和驾驶员均未做出规避动作。**特斯拉**出现了直接被“切头”的惨案，驾驶员当场

② 2020年，中国台湾高速公路上，一辆开启自动驾驶的**特斯拉Model3**径直撞上一辆侧翻的白色货车。幸运的是货车内装软性材质货物，未造成伤亡。

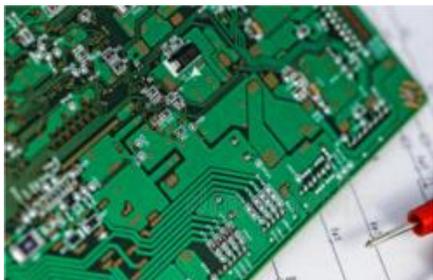


② 特斯拉Model3自动驾驶系统多次出现冲撞白色卡车事件，多起事故均与**智能感知系统**不成熟有关。据统计，2018年1月至2023年8月，特斯拉自动驾驶与**956**起事

**自动驾驶安全事故还在频繁发生！**

# (一) 自动驾驶 (Autonomous driving)

## 5. 自动驾驶的未来 / The future of autonomous driving



电子电器系统故障



地点的多样性



驾驶员误操作



路况的复杂性



系统功能局限性



天气的不确定性

现实场景复杂

永远存在未知可能

### 感知的关键在于鲁棒性和可泛化性



### 多传感器融合



### 创新生态



### 单车智能+车路协同

全球政府积极部署V2X

	中国	美国	欧洲
政策	《车联网建设规划》 《智能网联汽车发展战》 《新能源汽车产业发展规划 (2021-2035年)》	《自动驾驶法案》, 要求美国自动驾驶汽车技术必须落地 智能交通系统 (ITS) 战略 2020-2025 DSRC 2.4GHz 频段规划 C-V2X 牌照	《Roadmap for the Deployment of Automated Driving in the EU》 ITS4 (智能网联汽车) 战略 2021-2030
试验区	V2X “三跨” 数据互通应用 C-V2X “四跨” 数据互通应用 C-V2X “四跨” 数据互通应用 10个国家车联网项目 (车联网) 测试示范区	50个DSRC车联网示范项目 《自动驾驶》, Aptiv 实验 L3+V2X 以3X为中心: 测试、培训、高精、ATIS 与大型企业的试点项目	瑞典AutoZero 德国ITS Candidate 英国 DriveMile 法国Aut C-V2X 墨西哥比奇CONCORDIA
技术应用	5G, C-V2X为支撑	整合802.11p (DSRC) 与 C-V2X	整合802.11p (DSRC) 与 C-V2X

### 政策落地

张亚勤. 走向无人驾驶. 清华大学. 2021.

# 人工智能发展的理论前沿

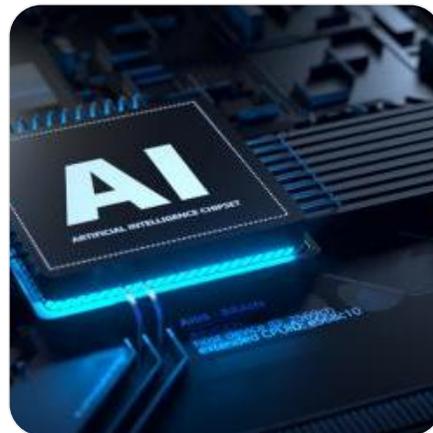
## 自动驾驶

*Autonomous driving*



## AR/VR/MR与元宇宙

*AR/VR/MR and metaverse*



## 人工智能与科学计算

*AI for Science*

## (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

### 1. nVidia发布会：老黄和他的虚拟厨房/ nVidia GTC 2021: digital kichen



英伟达在2021年4月的GTC发布会上的黄仁勋的虚拟形象

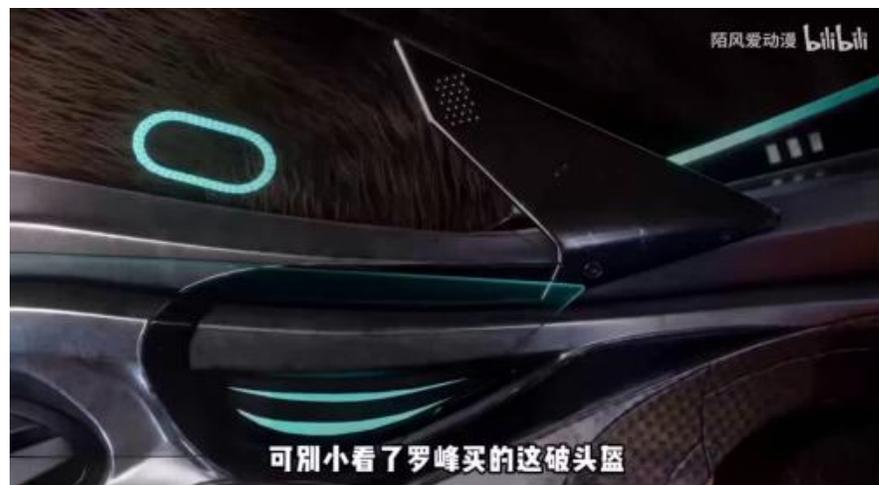
## (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

### 2. “元宇宙”元年 / Beginning of the metaverse

- ① **元宇宙**的概念起源于科幻作家史蒂文森于1992年创作的科幻小说《雪崩》，它描绘的是一个**共享的线上世界**，参与者能够自由地互动，甚至于在虚拟世界中生活和工作。这个世界是一个**平行于现实世界**的网络世界。
- ② **AR/VR/MR**的突破正在逐步搭建这个架设于虚拟与现实的桥梁。
- ③ 2021年被正式定义为“**元宇宙元年**”。
- ④ 2021年10月28日，**Facebook**宣布更名为**Meta**，布局元宇宙；几乎所有的大型科技公司都加入了元宇宙的行列。



小说《雪崩》设想的人类虚拟社区



国漫《吞噬星空》的虚拟宇宙

## (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

### 3. 平行世界 / Horizon worlds



① 2021年, **Facebook**更名Meta后推出了首个元宇宙社交平台——**平行世界 (Horizon Worlds)**。

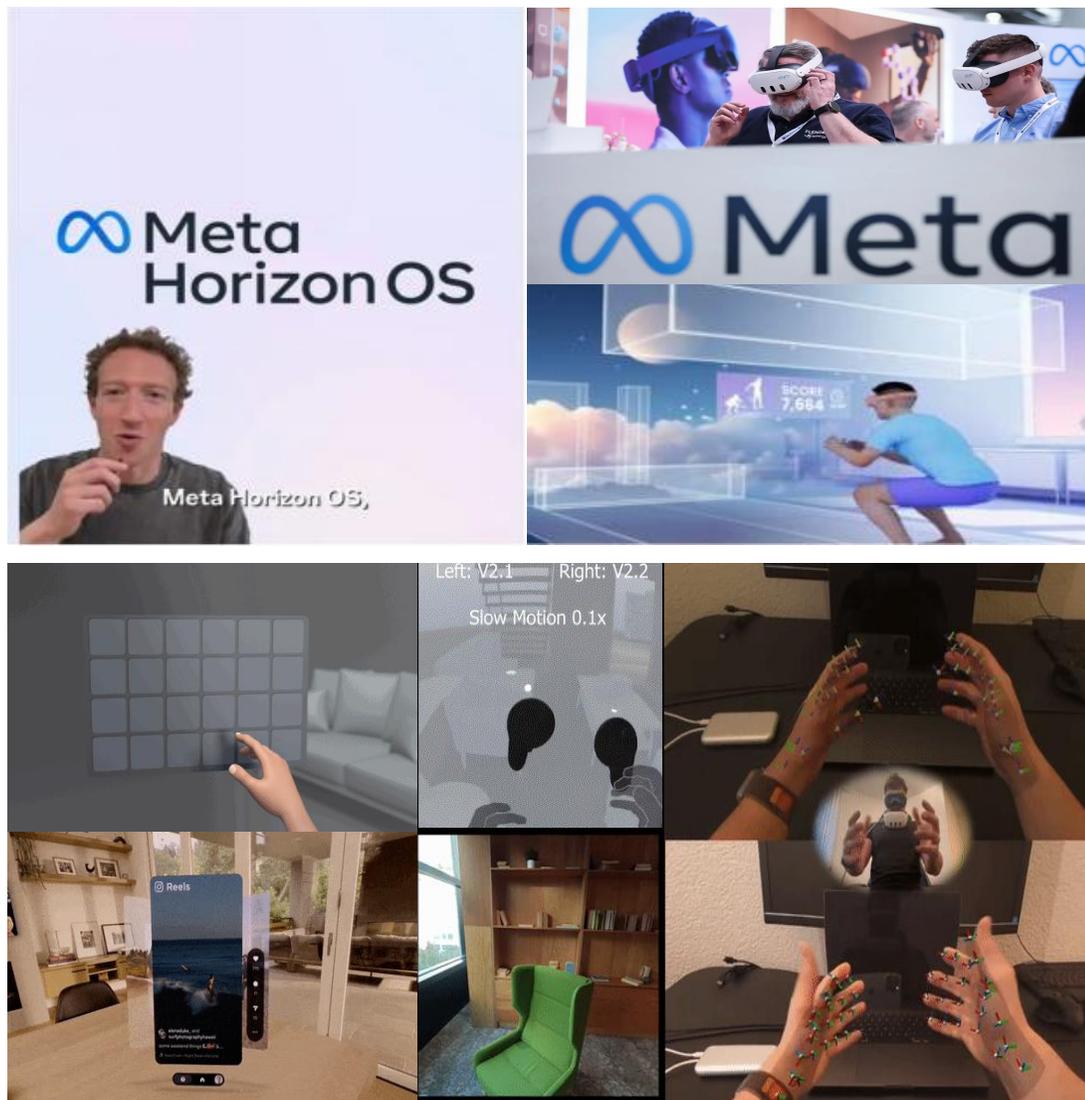


② 用户可以在平行世界中建立自己的**虚拟形象**, 并通过**虚拟现实VR**和**增强现实AR**设备以第一人称或第三人称去体验各种远程服务、交友、购物、休闲和娱乐等。就像现实世界一样, 在这里总有一些出人意料的事情会发生。

## (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

### 4. 混合现实操作系统 / Meta Horizon OS

- ① 2024年4月，**Meta公司**宣布开放其**混合现实操作系统Meta Horizon OS**，该系统是Meta其头戴设备Meta Quest的原生系统，同时支持手部、眼睛、脸部、身体追踪四种全新的交互。
- ② **微软Xbox、Steam、华硕、联想**等多个厂商都称为Meta HorizonOS的第一批合作伙伴。一个类似安卓的**广泛生态系统**正在形成，包括软件和硬件生态都将互惠互利。



## (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

### 5. 头戴式显示器: Vision Pro / Head mounted display: Vision Pro

① 2023年6月, **苹果公司**发布**Vision Pro**, 该设备可以通过**眼神**和简单的手势实现虚拟世界和真实世界的**交互**, 创造“**超现实的体验**”。

② 2024年5月, 在**北京大学人民医院**王俊院士、李运主任和周足力教授的指导下, 高健医生完成了首次应用Apple Vision Pro辅助的胸腔镜肺癌根治术, 这是该技术在胸腔镜手术领域首次应用。



# (二) 从AR/VR/MR到元宇宙 (From AR/VR/MR to metaverse)

## 6. 元宇宙生态逐步完善 / Metaverse ecology is gradually improving



# 人工智能发展的理论前沿

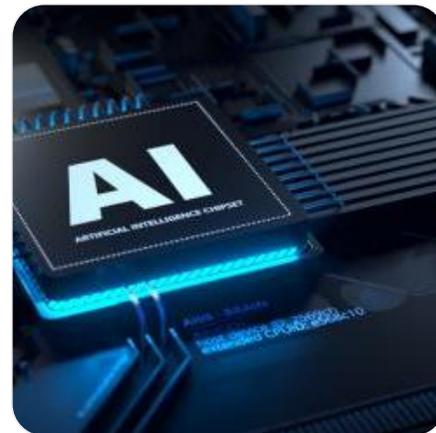
## 自动驾驶

*Autonomous driving*



## AR/VR/MR与元宇宙

*AR/VR/MR and metaverse*



## 人工智能与科学计算

*AI for Science*

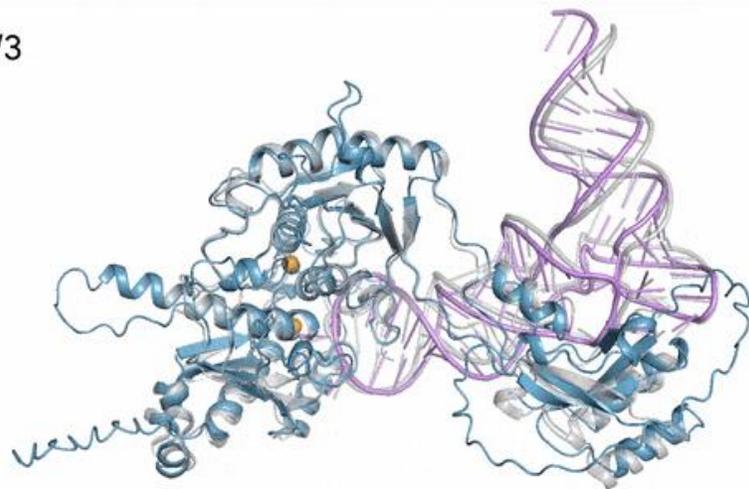
# (三) 人工智能与科学计算 (AI for Science)

## 1. 人工智能引领结构生物学

- ① 2018年11月，**DeepMind**推出**AlphaFold**模型，在国际蛋白质结构预测竞赛中获第一名。
- ② 2020年11月，**DeepMind**提出**AlphaFold2**模型，有史以来首次把蛋白质结构预测做到基本接近实用水平。《Nature》评论“**AlphaFold2**算法解决了困扰生物界50年的大问题”。
- ③ 2024年5月，**DeepMind**发布**AlphaFold3**型，预测了几乎所有生物分子的结构，并模拟了这些分子之间的相互作用，有望极大提升药物发现能力。



8AW3

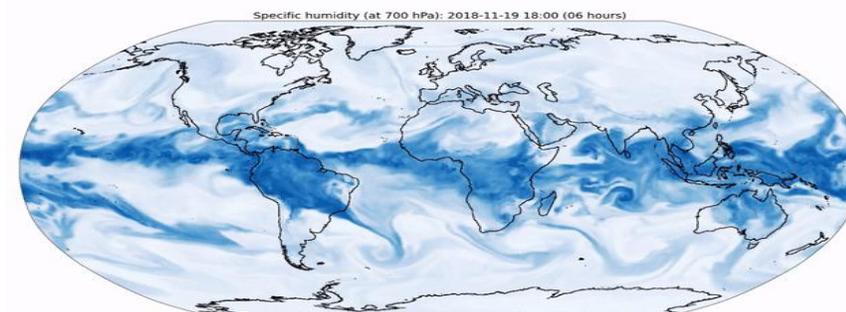
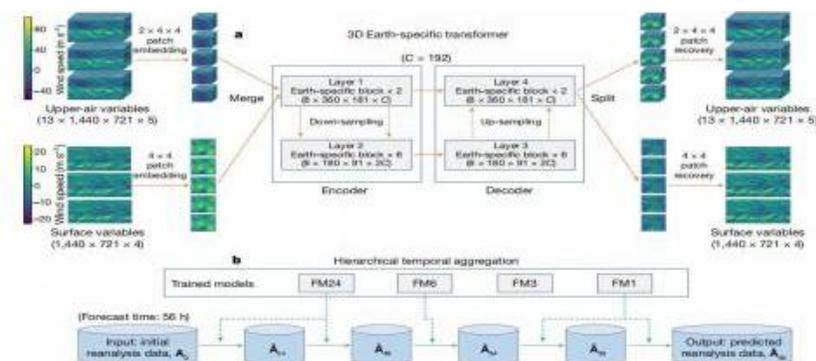


Ground truth shown in gray

# (三) 人工智能与科学计算 (AI for Science)

## 2. 人工智能气象预测

- ① 2021年4月, **中国华为云**提出秒级高分辨率全球人工智能气象预报系统——**盘古气象大模型**, 使用**3D Transformer**模型处理三维气象数据, 预测精度**首次超过**传统数值预报, 成果发表于《自然》。包括中国国家气象中心、欧洲中期天气预报中心都已上线该模型。
- ② 2022年12月, **DeepMind**提出基于**图神经网络**的天气模拟器**GraphCast**, 能够在**一分钟内**预测未来10天的数百个天气变量, 预报效果显著优于传统气象预报方法。
- ③ 2023年4月, **中国上海人工智能实验室**联合多家机构发布“**风乌GHR**”气象大模型, 利用**多模态**和**多任务权重自均衡深度学习**方法, 实现长达**10天以上**的**10公里级**的有效预报, 并在80%的评估指标上优于DeepMind的GraphCast。

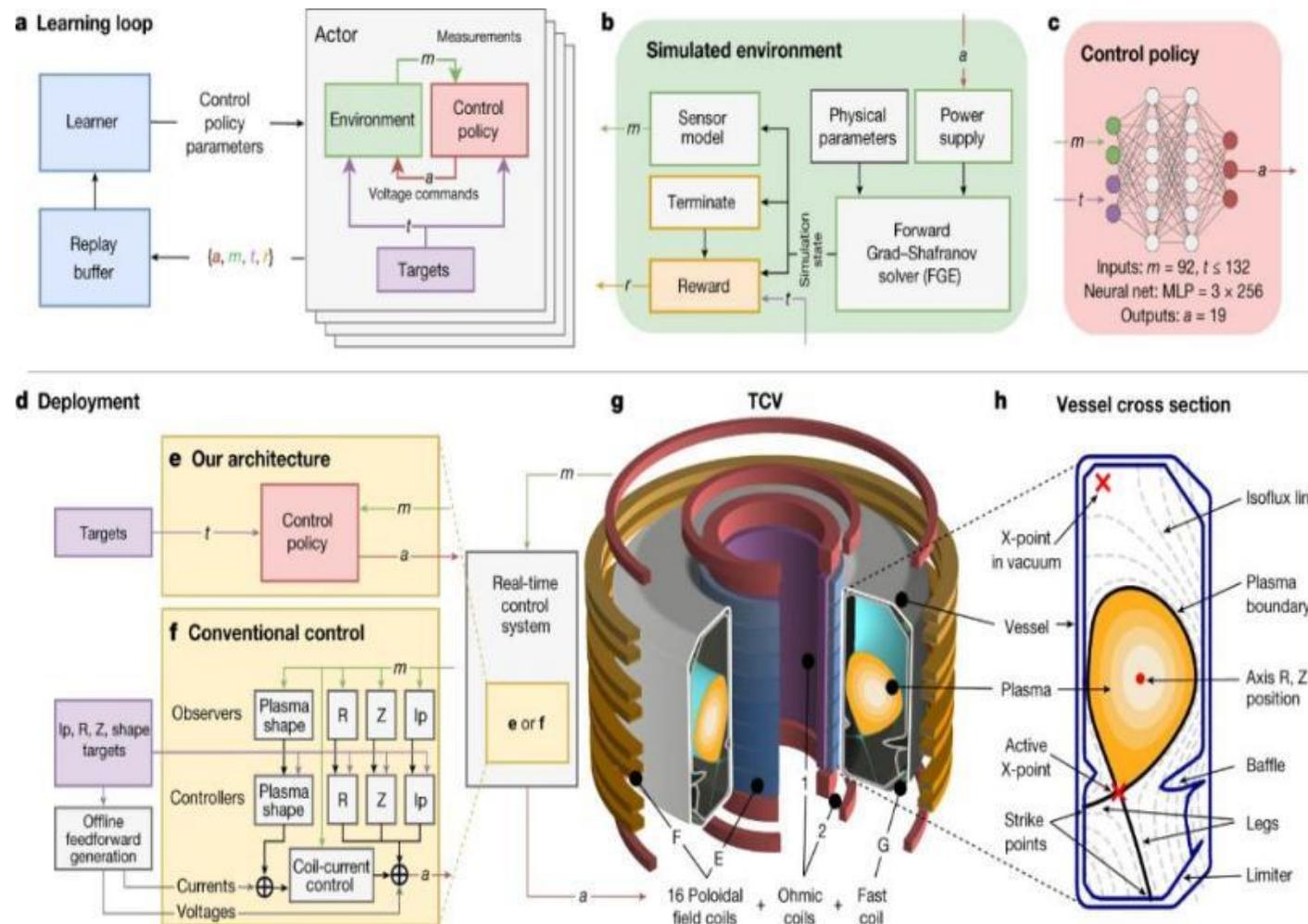


其他AI气象大模型  
0.25经纬度 (约625平方公里)

风乌GHR  
0.09经纬度 (约81平方公里)

# (三) 人工智能与科学计算 (AI for Science)

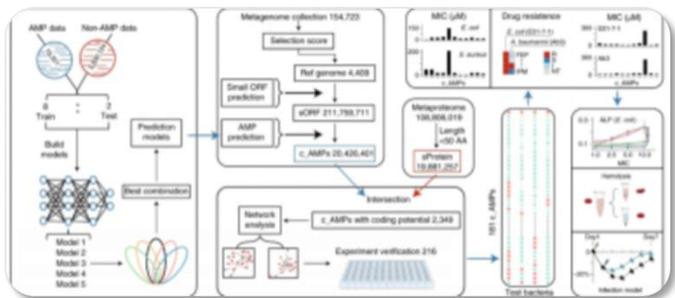
## 3. DeepMind首次使用AI控制核聚变实验



① 2022年2月, **DeepMind**开发出世界上首个实现**等离子体自主控制**的AI模型, 该模型基于深度强化学习算法, 可在**模拟环境和真正的核聚变装置** (托卡马克) 中进行测量和多目标控制。相关成果发表于 **《自然》** 期刊。

② 2023年7月, 新版模型训练时间减少**3倍**, 而等离子体形状的模拟精度提高了**65%**。

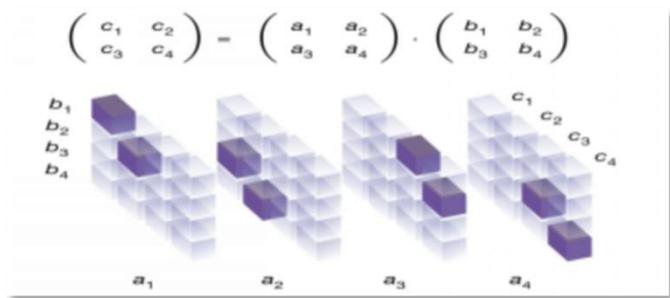
# (三) 人工智能与科学计算 (AI for Science)



## 助力抗生素发现

04

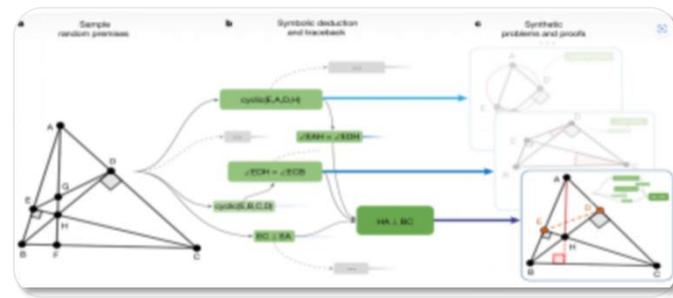
2022年3月, **中科院微生物研究所**提出一种基于**多自然语言模型融合**的人类肠道微生物组编码多种抗菌肽(AMPs)方法, 能够有效地从大型肠道微生物组中寻找抗菌特性肽。在检测出的4409个代表性基因中, 合成216个多肽, 命中率达83.8%。成果发表于《**自然-生物技术**》期刊。



## 矩阵运算

05

2022年10月, **DeepMind**发布了一款名为**AlphaTensor**的工具, 能更有效地计算矩阵乘法。通过对大量先进复杂算法的挖掘, AlphaTensor具备对任意目标的适应性, 并且对大型矩阵的运算速度**比传统算法快10-20%**。该算法在计算机图形学、物理仿真和机器学习方面均有广泛用途。



## 几何证明

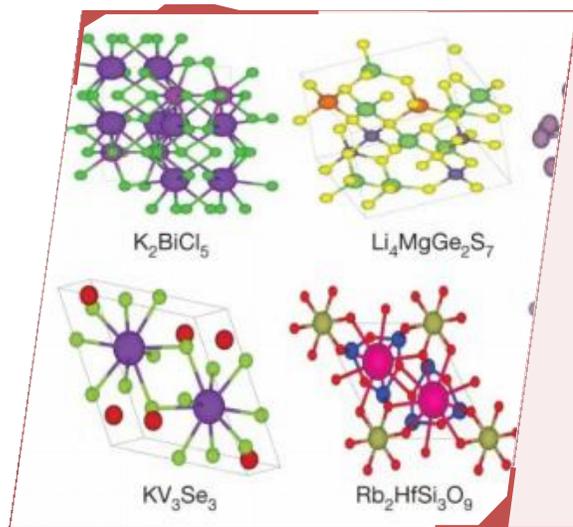
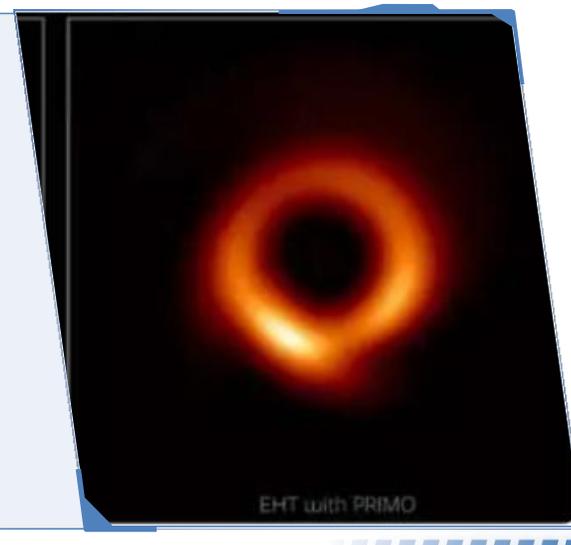
06

2024年1月, **DeepMind**发布AI数学模型**AlphaGeometry**登上《**自然**》期刊。AlphaGeometry采用**神经语言模型**的预测能力与**规则约束推演引擎**相结合的方法协同找到问题解法。它的解题水平接近人类奥林匹克金牌得主, 有望成为AI迈向通用人工智能的一个重要里程碑。

# (三) 人工智能与科学计算 (AI for Science)

## 07. 机器学习增强黑洞图像

- ① 2023年4月, 科学家提出**Primo**方法, 实现了对黑洞图像的清晰可视化。
- ② 该方法利用**字典学习** (Dictionary Learning) 技术分析了超过3万个黑洞吸积气体的高保真模拟图像来学习图像结构中的表示。
- ③ 它为图像提供了缺失结构的高保真估计, 使得现有望远镜阵列的分辨率最大限度的环宽缩小约2倍。



## 08. 深度学习促进新材料发现

- ① 2023年12月, **DeepMind**提出一种**图神经网络模型 (GNoME)** 用于材料发现。
- ② 研究团队通过该模型发现了220万个新的材料晶体结构, 其中很多结构是**人类预测**和**公式推导**难以发现的, 相较于传统的材料开发方法效率提高了10倍以上。
- ③ 相关工作发表于《自然》期刊。



# 人工智能发展的产业前沿

## Frontiers of AI Industry

# (一) 中美对比 (Comparison between China and U.S.)

## 1. 人工智能产业布局与图谱 (美国)

- 美国AI厂商较多，且**软硬实力**兼具
- 大型厂商多**综合布局**；创业公司遍布各个层次
- 美国AI产业**基础层**多为老牌重量级厂商
- 英伟达、高通在**芯片与传感器**领域实力强劲
- IBM、微软、谷歌、Meta、亚马逊等科技厂商在基础层、技术层和应用层实力都很强。



# (一) 中美对比 (Comparison between China and U.S.)

## 1. 人工智能产业布局与图谱 (中国)

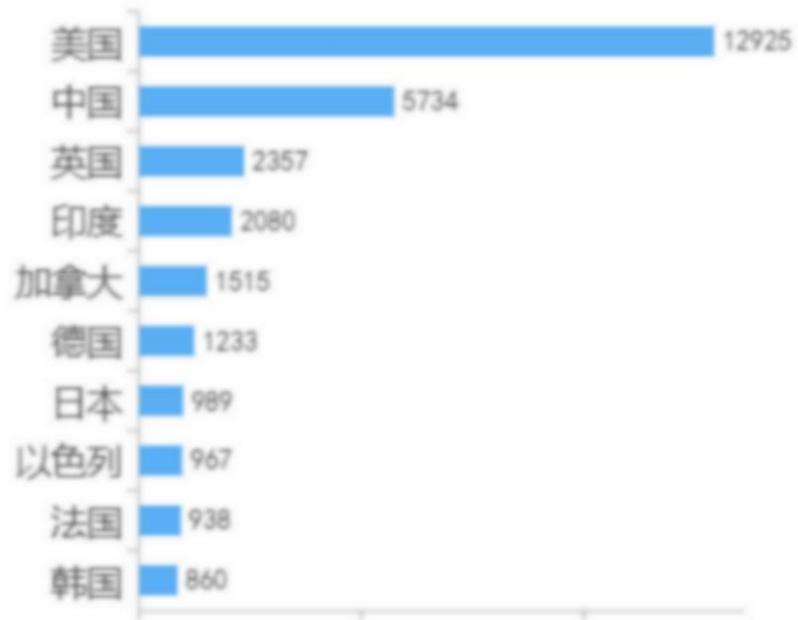
- 中国AI企业以“**国家队**”和“**创业独角兽**”为主，其中国家队多**综合布局**，独角兽多深耕**垂直技术与应用市场**
- 中国AI应用层繁荣，众多厂商在安防、金融、出行、教育、终端领域发力落地
- 中国AI产业技术层发展势头良好
- 但中国AI产业基础层整体实力较弱，特别是**芯片**和**传感器**等硬件领域仍需抓紧布局。



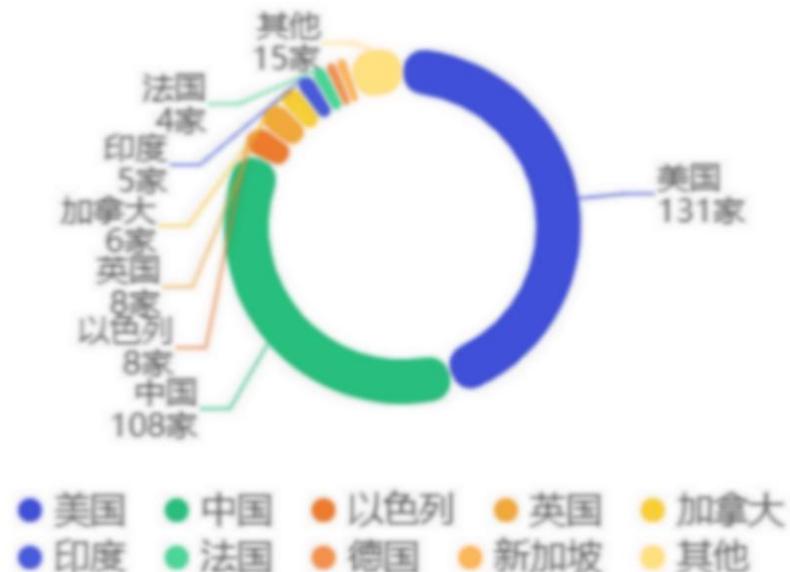
# (一) 中美对比 (Comparison between China and U.S.)

## 2. 中美人工智能企业概览

- ① 美国AI厂商总数位居全球首位，中国紧随其后；
- ② 中美在全球AI独角兽中平分秋色。



全球AI企业按国家分布情况  
(数据截止2023年6月)

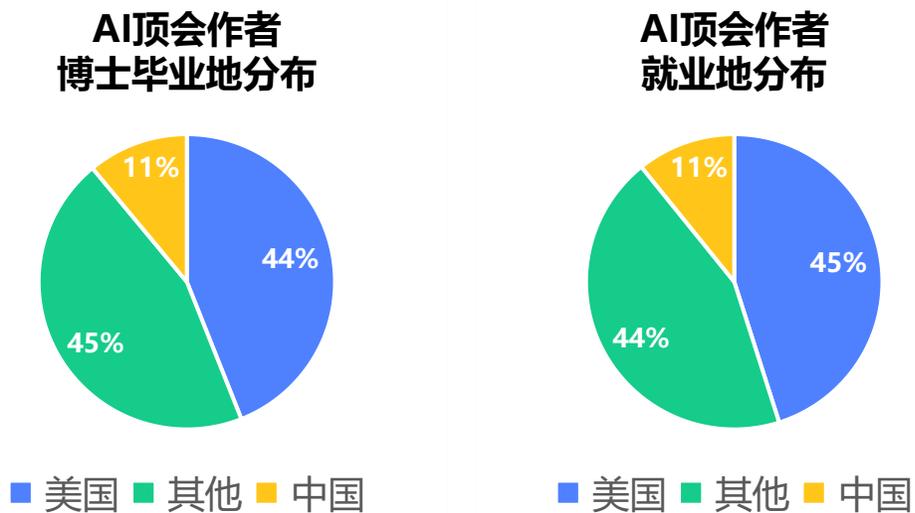


全球AI独角兽企业按国家分布情况  
(数据截止2023年6月)

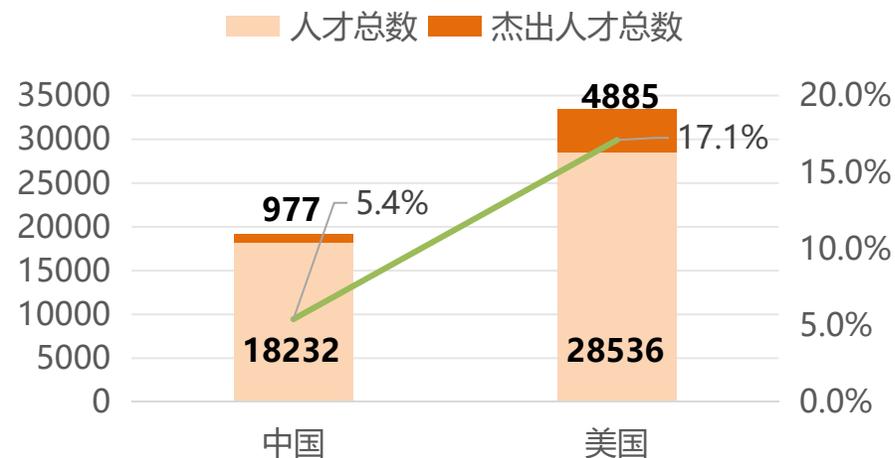
# (一) 中美对比 (Comparison between China and U.S.)

## 3. 中美人工智能人才对比

① 中美领衔全球人工智能领域**人才分布**，中国在**人才总量**、**杰出人才数量**与占比上均落后于美国，在人才培养上有明显差距，但人才吸引力正在增长。



中美人工智能人才质量对比



② 在学术界，中国较美国在**顶尖AI人才培养**上仍有明显差距，且美国对顶尖AI人才的吸引力更大。

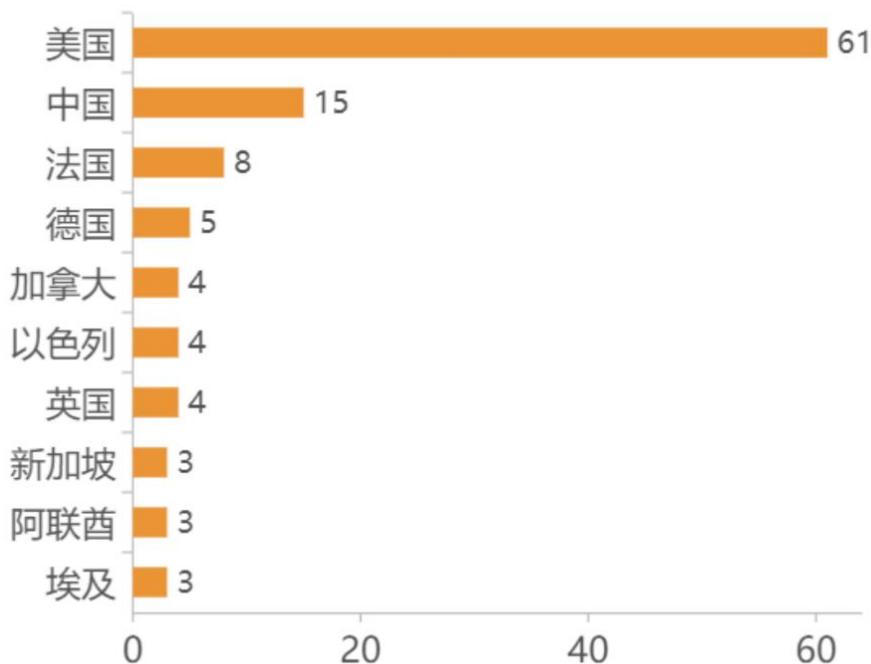
③ 中美AI人才在**数量与质量**上差距仍然明显，中国正在**加强AI人才培养**，提升整体人才储备实力。

Li-Feifei. Introduction to the AI index report. Stanford. 2024.

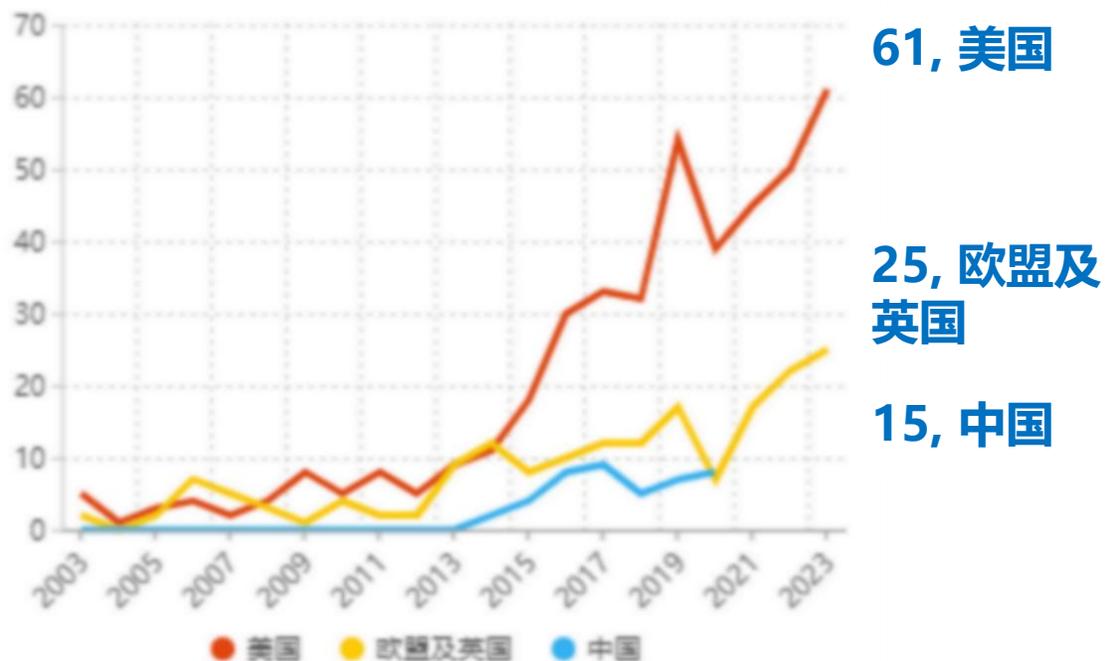
# (一) 中美对比 (Comparison between China and U.S.)

## 4. 中美欧人工智能前沿技术对比：通用机器学习模型

- ① 2023年，美国发布了61个知名的机器学习模型，而中国只发布了15个，但仍然居第二位
- ② 2003年到2023年，中美两个发布的通用机器学习模型的影响力差距在不断加大。



2023年全球各国知名机器学习数量



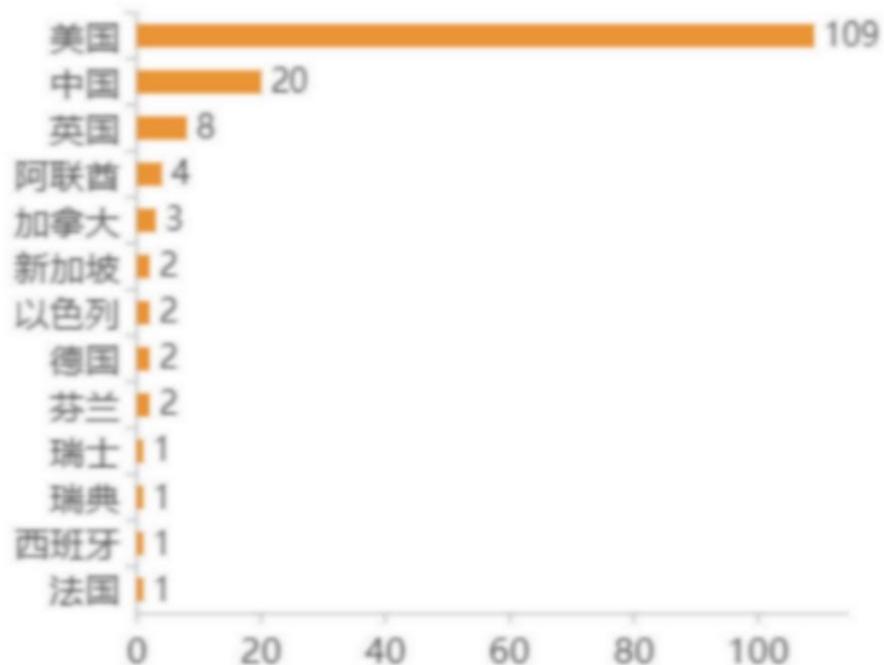
2003-2023年全球知名机器学习数量

Li-Feifei. Introduction to the AI index report. Stanford. 2024.

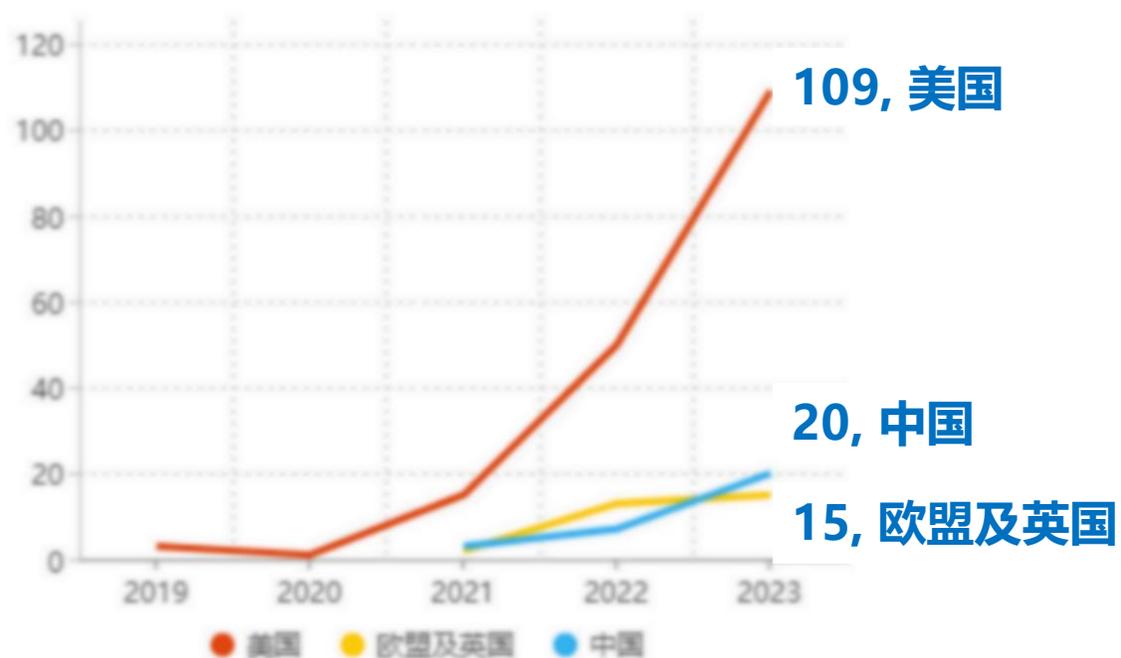
# (一) 中美对比 (Comparison between China and U.S.)

## 5. 中美欧人工智能前沿技术对比：通用机器学习模型

- ① 2023年，美国发布了109个基础大模型，而中国只发布了20个，位居第二。
- ② 2019年到2023年，中美两个发布的基础大模型的数量差距在不断加大。



2023年全球各国基础大模型数量



2019-2023年全球基础大模型数量

Li-Feifei. Introduction to the AI index report. Stanford. 2024.

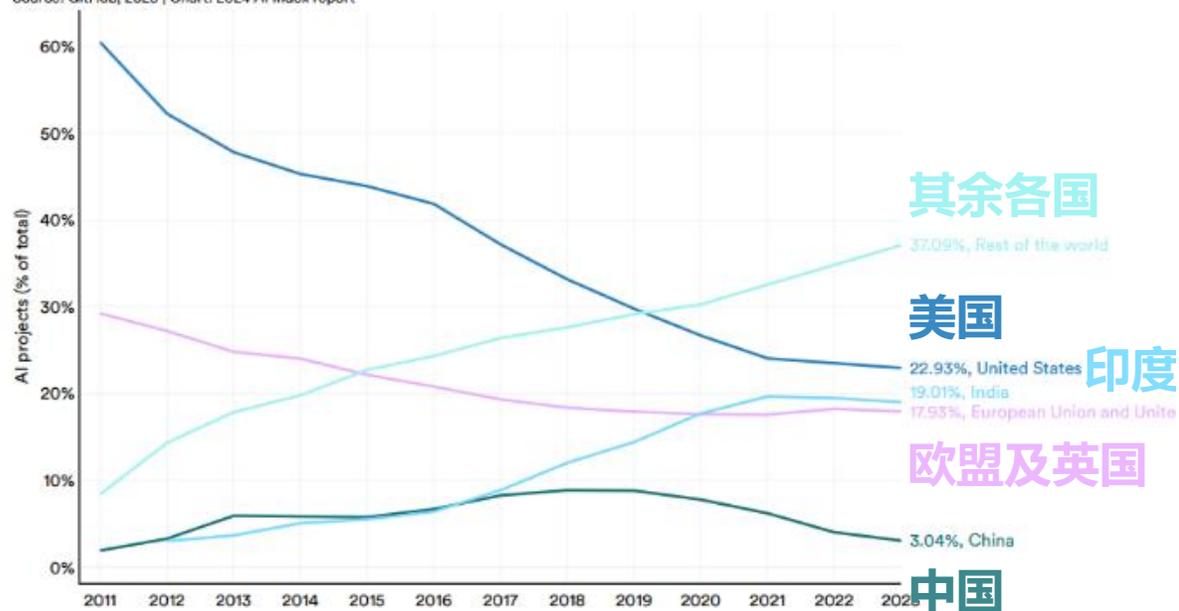
# (一) 中美对比 (Comparison between China and U.S.)

## 6. 中美欧人工智能开源项目对比

- ① 中国在开源人工智能项目的数量上远远低于美国及欧盟；
- ② 中国在开源人工智能项目的影响力低于美国及欧盟。

GitHub AI projects (% of total) by geographic area, 2011–23

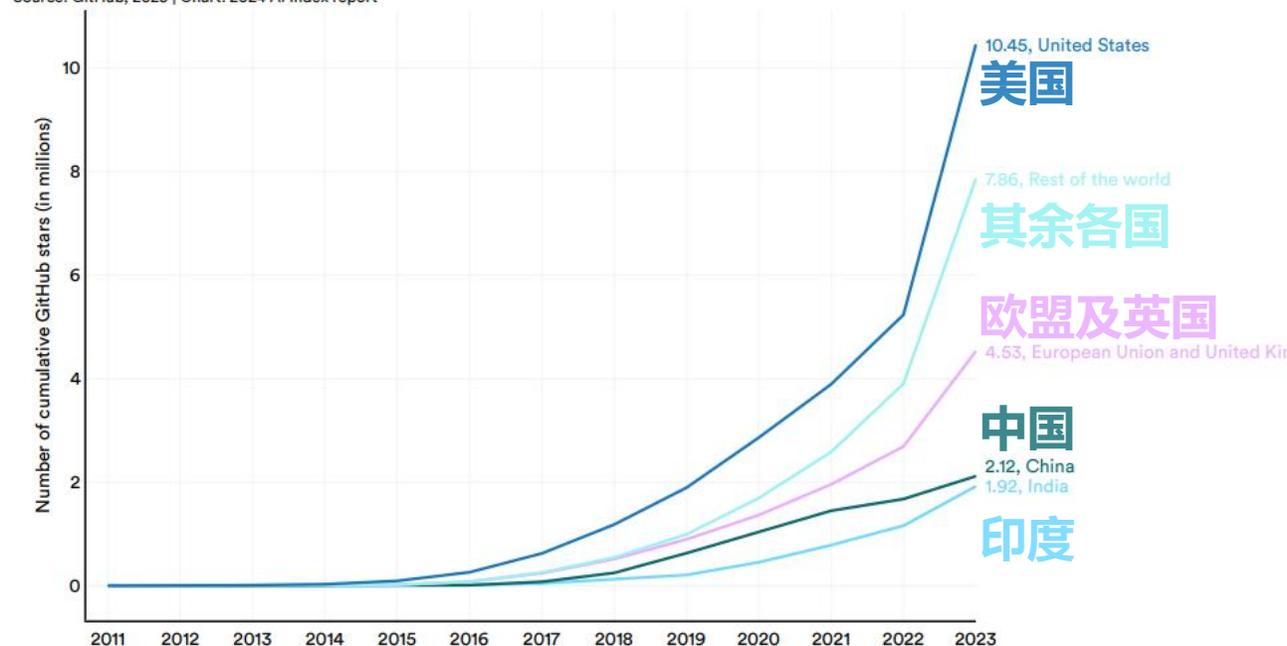
Source: GitHub, 2023 | Chart: 2024 AI Index report



GitHub人工智能项目全球占比  
(2011-2023年)

Number of GitHub stars by geographic area, 2011–23

Source: GitHub, 2023 | Chart: 2024 AI Index report



GitHub人工智能项目影响力 (Star数量)  
(2011-2023年)

rd. 2024.

# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：整体情况



- ① 美国AI厂商**整体实力领先**，**谷歌**稳居全球AI厂商首位
- ② 中国厂商中**百度**一马当先，**位居全球前四**，是中国AI产业领先者

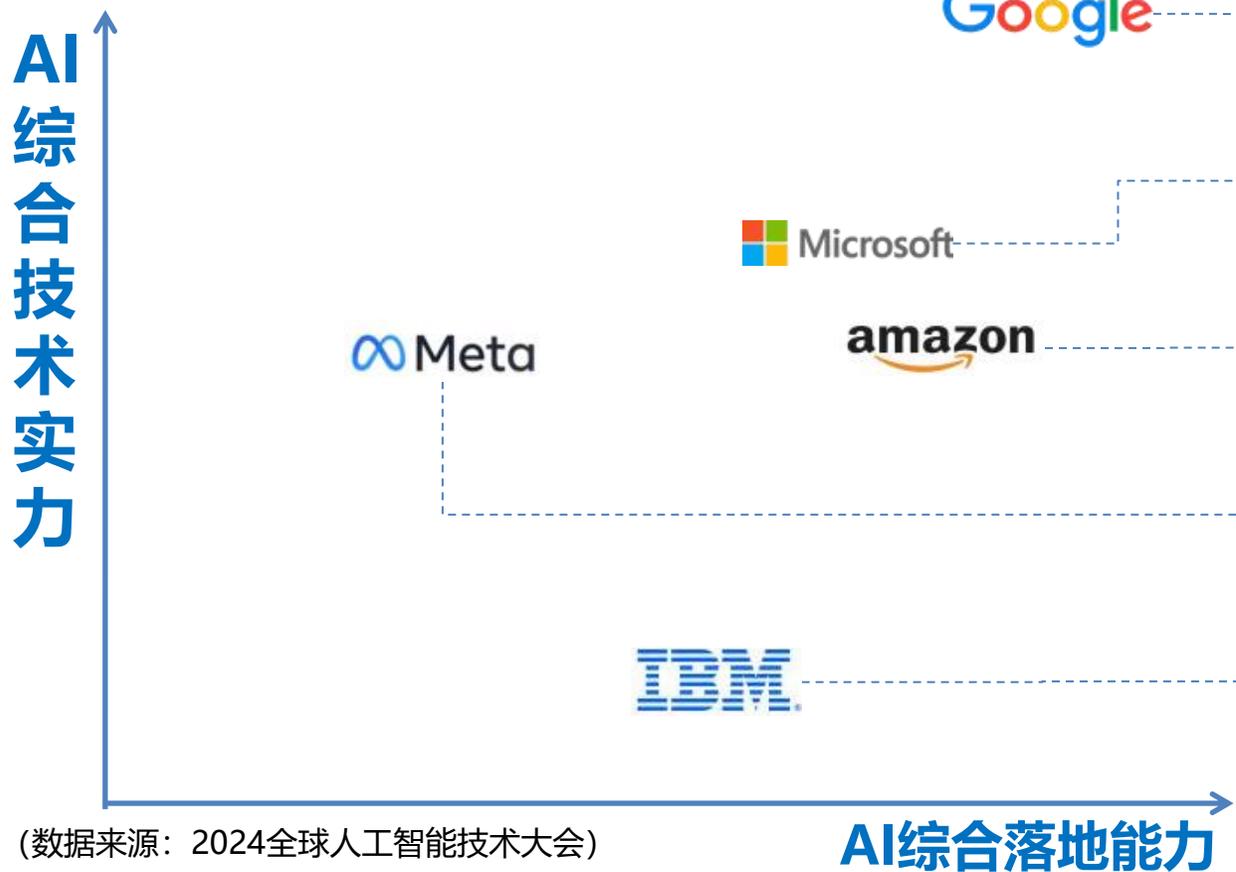
(数据来源：2024全球人工智能技术大会)

注：**AI综合技术实力**包含AI专利、AI论文、AI芯片、机器学习框架等指标；**AI综合落地能力**包括AI开发平台、合作结构数量、AI产业与服务、AI落地表现等指标。

# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：美国

□ 谷歌稳居第一，微软亚马逊紧随其后，Facebook和IBM由于生态与技术深度的欠缺稍落后。



**谷歌** 综合实力居全球首位，以突出的**技术创新**与**业务积累**在技术和商业落地方面均领先其他厂商。

**微软** 综合实力突出，凭借最多的**AI专利储备**及领先的**语音识别**、**计算机视觉**技术实力处于领先地位。

**亚马逊** **应用落地**能力较强，拥有世界第一的**科研投入**与完备的**云生态**。

**Meta** **技术实力**优异，拥有较强的科研实力与技术布局，但在**生态上**稍显不足。

**IBM** **商业落地**突出，在**企业服务**方面的领先地位确立了其商业落地优势，但**技术布局**优势不明显。

(数据来源：2024全球人工智能技术大会)

注：**AI综合技术实力**包含AI专利、AI论文、AI芯片、机器学习框架等指标；**AI综合落地能力**包括AI开发平台、合作结构数量、AI产业与服务、AI落地表现等指标。

# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：中国

□ 百度一马当先，携手阿里、华为、腾讯等共同推动中国AI产业发展。



注：AI综合技术实力包含AI专利、AI论文、AI芯片、机器学习框架等指标；AI综合落地能力包括AI开发平台、合作结构数量、AI产业与服务、AI落地表现等指标。

# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：中国其他企业

### VALSE2024 赞助商



### PRCV2023 赞助商



# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：深度学习框架

□ TensorFlow和PyTorch框架仍然是业界使用最为广泛的两个深度学习框架

✓ TensorFlow在工业界拥有完备的解决方案和用户基础

✓ PyTorch得益于精简灵活的接口设计，可以快速设计和调试模型，在学术界好评如潮

□ 近年，中国百度飞桨、旷视天元、清华计图、华为昇思也逐渐成熟



# (一) 中美对比 (Comparison between China and U.S.)

## 7. 中美主要大型人工智能厂商：整体对比

中美人工智能厂商实力仍存在一定差距，美国厂商普遍在技术储备、布局与应用层面实力较强，中国厂商在应用层面实力突出。

维度	微软	谷歌	Facebook	亚马逊	IBM	百度	阿里	腾讯	华为
技术储备									
技术布局									
技术生态									
AI应用									

注：阴影面积大小代表相应维度的相对实力，不代表绝对评分。（数据来源：2024全球人工智能技术大会）

## (二) 人工智能产业前沿趋势



### 趋势一 AI底层核心要素算力提升、数据处理方式优化

AI芯片从通用芯片向专用芯片发展，数据处理方式由人机协作向全面机器化演变

### 趋势二 AI模型从单模态到多模态、预训练大模型趋势明显

国内外多模态大模型井喷出现 (GPT-4o、Gemini、文心一言、盘古等)

### 趋势三 AI边缘应用将进一步拓展，大小模型新思路成型

AI能力将从云端数据中心移动到边缘位置，大模型和小模型协同共存解决实际问题

### 趋势四 AI在部分数据量庞大的应用场景渗透加快

未来AI会在交通、安防、医疗、制造、金融等数据量庞大的领域加速渗透

### 趋势五 AI厂商的安全意识与自主能力大幅提升

未来中国AI厂商将能为政府和企业提供更完备的数据安全服务，实现自主可控

## (三) 美国军事应用 (Military applications in U.S.)

### 1. 致命性自动武器、远程精准打击

2019年9月14日，沙特国家石油公司阿美石油公司的布盖格炼油厂和胡赖斯油田两处石油设施遭到无人机和导弹袭击，导致沙特每天大约570万桶石油原产量暂停供应，这占到了沙特产能的50%，也相当于全球5%的石油日消耗量。



### 2. AI辅助装备 IVAS

2021年，美国陆军与微软签署了一项价值219亿美元的协议，这份合同是12万个基于HoloLens头戴式虚拟现实和增强现实设备。集成视觉增强系统的头戴式设备，增强了态势感知能力，可在各种情况下进行信息共享和决策。



## (三) 美国军事应用 (Military applications in U.S.)

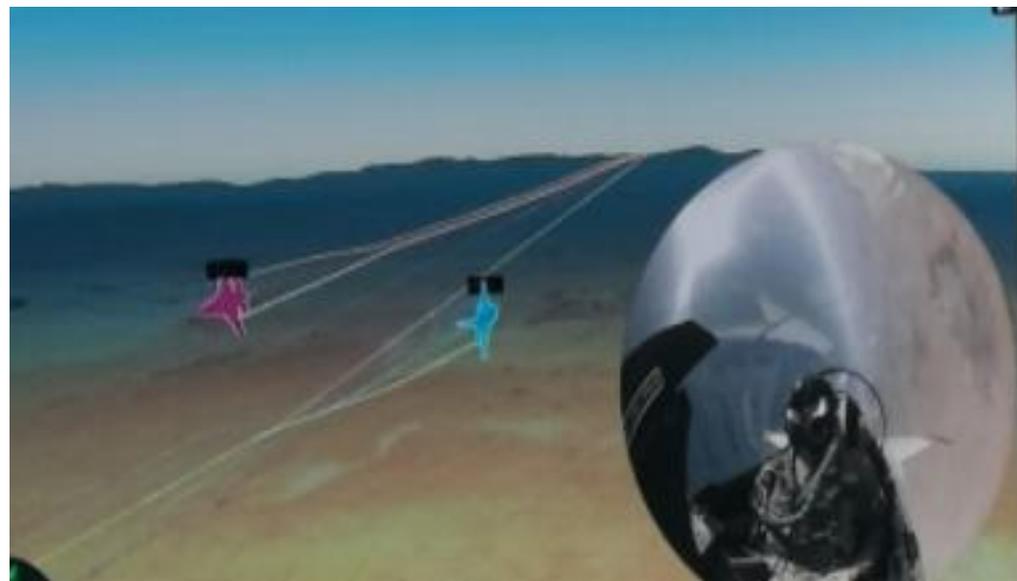
### 3. 商业AI系统的恐怖再利用

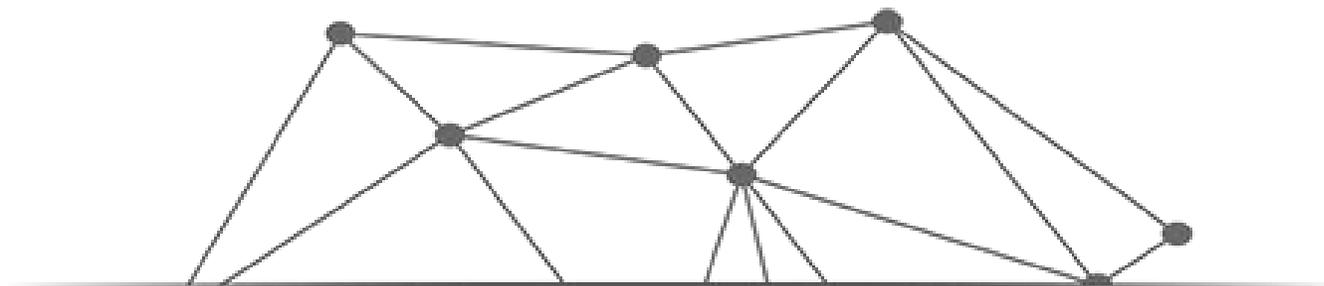
2021年10月，美国陆军年会展出了一种杀人狗，该设备装备了射程达1200米的机枪，它能够在复杂环境奔跑，还能夜视。



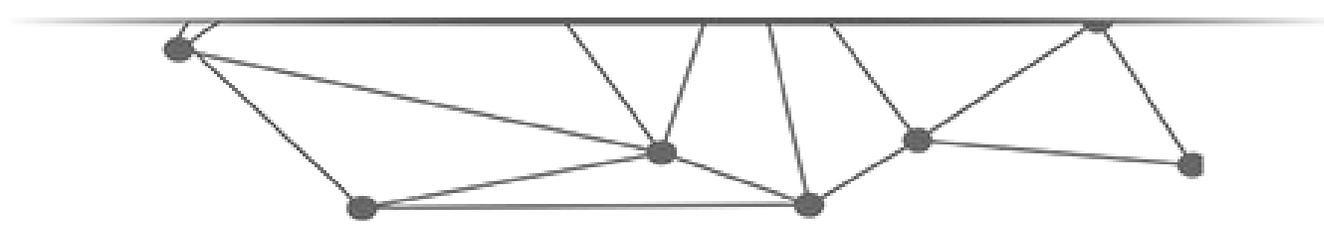
### 4. DARPA “空战演进” 项目

2024年4月，美国国防部宣布“空战演进 (ACE)”项目首次实现**AI智能体**驾驶的X-62A与王牌飞行员驾驶的F-16战机进行“阿尔法狗斗”（视距内空战），该智能体有望应用到**更复杂的多机场景**，为实时、战役级“马赛克战”奠定基础。





# 小 结



## 小结: 第四次工业革命之人工智能时代

### 人工智能时代

#### 第一次工业革命

1769年, 瓦特~蒸汽机  
大大提高了纺织业、冶金业和采矿业的劳动生产力, 人类社会进入 **“机械时代”**

#### 第二次工业革命

1866年, 西门子~发电机  
电力开始成为补充和取代蒸汽动力的新能源, 人类社会进入 **“电气时代”**

#### 第三次工业革命

1946年, 冯·诺依曼~ENIAC  
手工计算速度的20万倍, 加上后来的通信技术, 人类社会进入 **“信息时代”**



国务院关于印发  
新一代人工智能发展规划的通知  
国发 [2017] 35号

人类现代社会发展已经迈过机械时代、电气时代、信息时代, 正在经历着向**人工智能时代**发展和变迁的过程。。

人工智能带来了全球产业发展的重大机遇, 也带来了诸多安全风险与全球性挑战。需要国际社会共同应对, **促进人工智能技术造福人类!**

读万卷书 行万里路 只为最好的修炼



QQ: 14777591 (宇宙骑士)

Email: [ouxinyu@alumni.hust.edu.cn](mailto:ouxinyu@alumni.hust.edu.cn)

Website: <http://ouxinyu.cn>

Tel: 18687840023