

# 人工智能，无尽的前沿

张亚勤

时光回到 80 年前，1945 年二战刚刚结束，时任美国总统罗斯福在一个深冬的晚上收到了一份提案，名为《科学：无尽的前沿》。这份提案后来成为法案，一举奠定了美国在科学领域的领军地位，催生了众多技术、产品和产业，直接推动了第三次工业革命——我们熟知的无线通信、半导体、互联网、光纤通讯等，都是源于这份科学法案。而今天，我们正迎来一个全新的重大机遇——人工智能，它已然拉开了第四次工业革命的序幕。所以我今天的演讲题目是《人工智能：无尽的前沿》。

## 01 技术大趋势

首先我想谈一下技术大趋势。

### 人类智能的启发

地球上最聪明的物种是什么？是人。经过几十万年的进化，我们的大脑是如此神秘和神奇。我们的大脑不到 3 斤重，功耗只有 20 瓦，但是我们



### 张亚勤

清华大学智能产业研究院（AIR）创始院长。中国工程院外籍院士，美国艺术与科学院院士，和澳大利亚国家工程院院士。曾任百度总裁、微软全球资深副总裁兼微软亚太研发集团主席、微软亚洲研究院院长兼首席科学家等职。

1997 年他 31 岁被授予 IEEE Fellow，成为历史上获得这一荣誉最年轻的科学家。

却如此之聪明。它里面包含了 860 亿个神经元，有百万亿个连接或者说突触，存储量至少有 1 个 Petabyte。我们人类对大脑的理解也是渐进的，现在我们可能对大脑的理解还不超过 10%。很有意思的是，我们有两个最神秘的东西，一个是我们的宇宙，我们对宇宙的了解不到 5%，95% 以上都是暗物质、暗能量；另一个就是我们人类的大脑，我们对它的了解也很少。但这么多年来，随着研究的深入，我们越来越多地知道了它的一些结构、功能。

早年间，保罗·麦克莱恩提出了“三重脑”理论，将大脑分为不同层次：负责呼吸、睡眠、运动等生理功能的物理层次，处理情感的层次，以及负责推理、决策的高级层次。这个理论虽然不够精准，但为我们理解大脑提供了直观的视角。如今我们知道，大脑拥有一百五十多个不同的功能区，860 亿个神经元分布其中，分别负责声音、视觉、语言、运动等不同功能。

人类的记忆功能尤为神奇，包含天生的 DNA 记忆、短期的海马体记忆、长期的皮层记忆，还有显性记忆和隐性记忆。可以说，人类大部分智能都

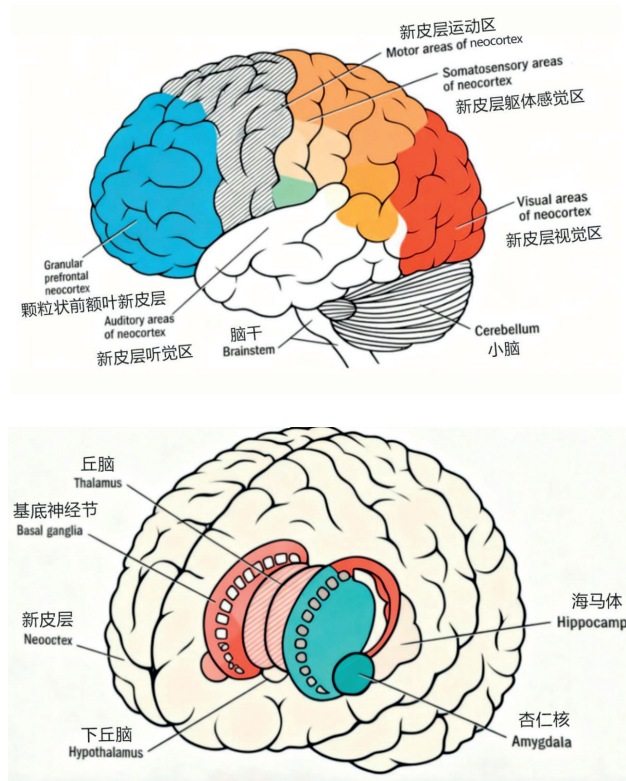
来源于这些不同类型的记忆体。

诺贝尔奖获得者丹尼尔·卡尼曼将人类的思考模式分为两种：系统1是快思考，面对事物时能快速产生直觉、迅速做出决定，无需深入思考；系统2是慢思考，需要经过深度分析和推理，是人类高级智能的体现。这两种系统可以相互转换，当我们对某件事足够熟悉后，慢思考会转化为肌肉记忆和直觉。比如学车初期，我们会刻意关注交通规则、路牌、信号灯和路况，但熟练之后，驾驶就会变成自然、自觉的行为，这就是系统转换的过程。

### 人工智能的发展脉络

什么是人工智能？人工智能，本质上是学习人类智能的过程，多年来我们一直在探寻智能的本质。“人工智能”（Artificial Intelligence）这一概念于1956年正式定义，而它的理论奠基可追溯到更早——图灵率先定义了“计算”与“智能”，并提出了“图灵测试”：若机器能通过多轮对话，让人类无法分辨其是否为人类，就意味着通过了测试。

另外两位重要奠基人常常被忽略：“信息论之父”香农定义了比特和信息量（熵），“控制论之父”维纳定义了负反馈、学习和自适应，这些基础概念



对人工智能的发展起到了至关重要的作用。

这么多年来，人工智能领域出现了很多不同的学派，但整体分一下，主要是两种不同的思路。一种思路认为，可以把大脑的逻辑、规则，还有推理的过程都用符号表示出来，这就是符号学派。基于

这种方式，逻辑体系很漂亮、很简洁，也有明确的因果关系，具备透明性，我们知道机器是怎么推理的，但它唯一的缺点就是不实用，在实际应用中效果不佳。当然还有另外一个学派，叫连接学派。这个学派认为，大脑如



此复杂，智能的实现十分困难，所以要通过大量的数据，通过积累经验、持续学习、不断适应，以及与世界的连接来获取智能，这就是连接主义。最近10-20年主流的深度学习技术，用的就是这样的方法。

人工智能发展史上有几个关键事件值得关注：

2016年 AlphaGo 第一次击败了李世石九段，是以 3:1 的比分击败的。咱们清华的同学柯洁九段不服气，2017年和 AlphaGo 下了三盘棋，最终 0:3 屈居第二。AlphaGo 当然很聪明，它用到了深度学习、强化学习，还有蒙特卡洛搜索，是很了不起的成果，它学习了人类几十万盘棋局。但我觉得更了不起的，是大家不太听说的 AlphaGo Zero。它完全不用学习人类的棋局，它是自己和自己下棋，互相博弈，通过不断地博弈来学习，而且进化速度非常快。AlphaGo Zero 和 AlphaGo 的前一个版本下了 100 场棋，它完胜，比分是 100:0。它不仅可以下围棋，还可以下国际象棋以及别的棋类。所以后来 DeepMind 这个团队说，从此我们和人类下棋了，因为所有的棋类，人类都下不过人工智能了。这是一个特别重要的概念，就是智能体的概念。

基于类似逻辑（算法不同），DeepMind 推出 AlphaFold，解决了人类蛋白质解析预测长达 50 年的难题。原本人类需要 10 亿、众多科研人员耗费十余年才能完成的工作，AlphaFold 仅用一年就全部解决。

2024 年，诺贝尔奖物理学奖、化学奖均授予了人工智能领域的奠基人，其中就包括 DeepMind 创始人德米斯·哈萨比斯——他的团队既创造了 AlphaGo，也打造了 AlphaFold。今年 1 月，我在达沃斯与他有过一场关于新药研发、生物计算及人工智能未来发展的有趣对话。

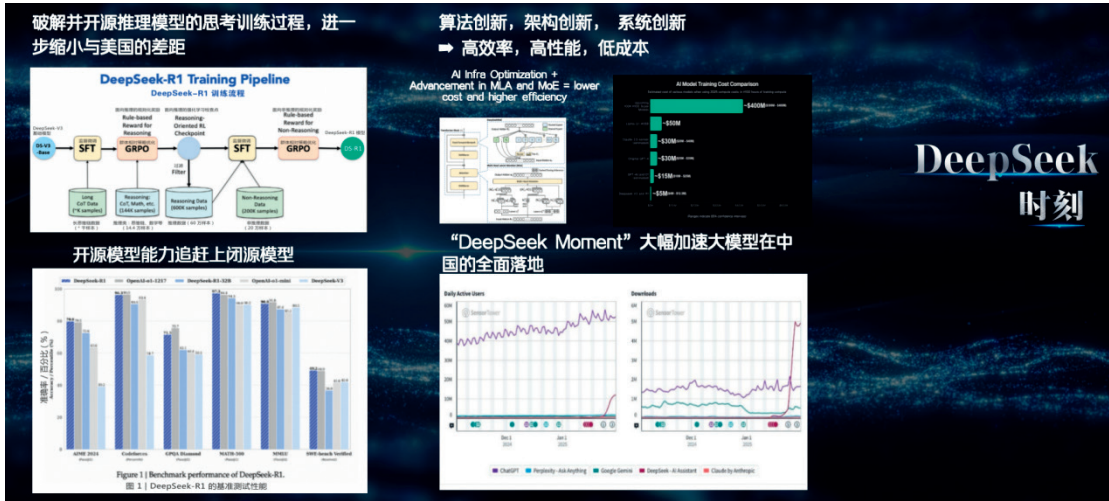
另外一个大的里程碑，是 2022 年，也就是三年前，同样是一个很冷的冬天，OpenAI 的 ChatGPT

出现了，这带来了一个全新的范式。过去的深度学习或者神经网络，主要还是针对具体任务，本质上是一种很聪明的模式识别，比如说语音识别、人脸识别、图像识别，或者字符识别等，算是一种高级的识别技术。但 ChatGPT 出来之后，我们跳跃到了一个新的范式，它不仅可以识别，还可以生成，还可以创造，所以生成式 AI 就这么出现了。

生成式 AI 有三个重要的元素，统一表征（Tokenization）、规模定律（Scaling Law）和涌现效应（Emergence）。最重要的，我认为是统一表征。ChatGPT 是怎么做的呢？有点像人类的神经元，我们人类有 860 亿个神经元，不管分布在什么地方，每个神经元的结构都是完全一样的，视觉的、听觉的、运动的、记忆相关的，都是同样的结构。生成式 AI 的统一表征也是类似的道理，不管什么样的信号进来，都把它变成 Token，核心任务就是预测下一个 Token，生成下一个 Token。它可以生成文字、图像、视频，现在大家已经用得很多了。同时它还可以生成新的数据、代码、数学方程式、工具——它不仅能生成工具，还能使用工具；它还可以生成新的蛋白质、分子、材料、药物。当大语言模型的参数量突破百亿级别，便会触发规模定律，出现涌现效应。也就是说，模型的性能并非线性增长，而是随着规模扩大发生跃迁，从而涌现出未经编程的、令人惊喜的新能力。

另外一个重要的里程碑，来自中国的 DeepSeek。这个也是出现在一个寒冷的冬天，就是今年的 1 月份。DeepSeek 出现之前，中国有上百个大模型，这些大模型基本上都在学习 OpenAI，学习 ChatGPT，整个技术路径、算法的架构体系，基本上都是在模仿学习。

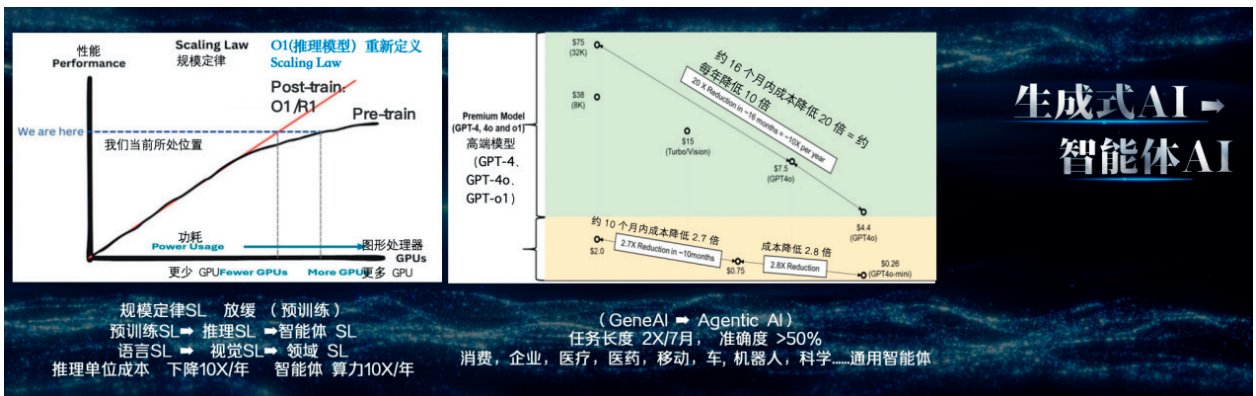
DeepSeek 出现之前，我曾经说过，我们和美国在大模型领域的差距，大概是两到三年。DeepSeek 是一家小小的创业公司，它的工程团队

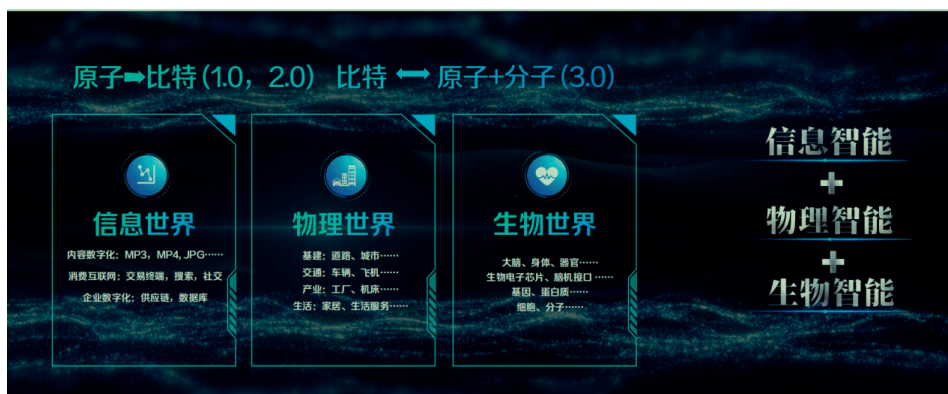


离清华可能就 5-10 分钟的路程，团队里很多都是清华的学生。DeepSeek 所做的，是一条新的路径，在算法、技术、系统架构上都有创新，它用 1% 的算力就可以达到和美国前沿大模型相似的能力。DeepSeek 出来之后，咱们和美国的差距可能就缩短到 2-3 个月，也就是一个版本的差距，从之前的 2-3 年缩短到 2-3 个月，而且在有些应用方面，我们可能做得更好。另外它采用开源模式，它的开源模型很快就被很多买不起大模型的国家、地区所使用，这使得整个模型的落地和应用变得越来越快。所以我把它叫做“DeepSeek 时刻”，一开始有“ChatGPT 时刻”，现在这是“DeepSeek 时刻”，是属于中国的时刻。

### 从生成式 AI 到智能体 AI

2025 年，人工智能领域迎来了又一重要转变——从生成式 AI 迈向智能体 AI。在此之前，我们遵循“规模定律”：数据越多、算力越强，模型效果越好，达到一定阶段后会出现量子跃迁和涌现效应。但 2025 年我们发现，预训练阶段的规模效应正在放缓，数据资源逐渐趋于饱和，继续增加算力的边际收益不断减少。与之相对，后训练（Post-train）阶段的重要性日益凸显。这就像人类的成长：预训练如同上学阶段，从本科到硕士、博士，通过学习积累知识变得聪明；而后训练则像是工作后的实践，在具体场景中不断学习、进化，这也是智能体 AI 的核心来源。





难了。可泛化性是我们人类的特点，但是我们人类的泛化能力也是有一些界限的。比如说我们清华的同学可能理科成绩很好，但文科可能相对差一点；我有一个特别好的朋友，他特别聪明，

什么是智能体？人类作为高智能物种，能够设定任务和目标、规划实现路径、不断试错反馈，凭借强大的记忆完成任务。比如同学们想学人工智能，会想学人工智能我要上哪个老师的课，哪个老师讲得最好，我要找什么参考书，我要怎么准备考试，做什么练习题，会把这个目标分解，然后找到最佳的实现路径，这是我们人类的核心特点。而AI智能体，就是在学习人类这种高级智能，具备三大关键能力：

首先它要自主学习（Autonomous），不是自动学习（Automatic），这两个差别很大。自主学习是没有固定的规则，在不断地探索中学习；而自动化往往是有预设规则的，是按照定义好的规则来运行的。所以智能体的第一个关键特征是自主的（Autonomous）。

第二点是可进化的，就是通过不断迭代可以变得更好，而且进化之后，下一次再做类似的事情，可以把原来学到的知识用上。这是我们人类和别的物种很大的区别——人类的智能是可以叠加的。而和我们最近的物种黑猩猩，它们的智能一代和一代之间是没有本质区别的，所以可进化性十分重要。

第三点就是泛化能力，能够举一反三。比如说我学会怎么去订票，这些功能、技能可以用到别的地方，它可以帮你去报销，可以帮你去购物，我也希望它以后可以帮你开车，当然开车这个任务就很

工作也做得很棒，但学开车拿驾驶执照花了15年还没拿到，最后拿到执照之后马上就撞车了。所以有些地方，人类的泛化能力也会受限，但我们还是希望人工智能能够具备举一反三的泛化能力。

这些智能的实现，离不开最基本的东西，其实就是数据，数据的本质就是数字化，我们的技术底座就是数字化。首先是信息世界的数字化，然后是物理世界的数字化、生物世界的数字化。过去40年，我们最重要的工作就是数字化。早期从1985年开始，我们进行内容数字化、文档数字化，我们的语音、图像、视频、文本、PPT等，这些都是在那个时候变成数字化内容的。后来加上HTML等技术，取得了一个重大成果，就是互联网，先是PC互联网，后来是移动互联网。然后我们又对企业进行数字化，或者说对企业进行信息化，这里面有ERP系统、CRM系统、数据库等，还有企业各种各样的业务流程。这一阶段催生了两个最大的成果，一个是数据库，另外一个云计算。那么现在，我们的物理世界正在被数字化，汽车、公路、交通灯、城市在数字化，我们的电网在数字化，我们的家庭在数字化，我们的车间在数字化，我们的工厂在数字化……整个物理世界都在经历数字化的变革。同时，我们生物世界的蛋白质、大脑、细胞、基因等等也都在被数字化。

MIT媒体实验室主任曾在数字化1.0开启时提出，

我们正在从“原子”走向“比特”。“比特”是香农所定义的数字世界的基本单位。而如今，我们正从比特回归原子、走向分子——新一代智能是信息智能、物理智能和生物智能的融合，是比特、原子和分子的融合，也是碳基生命和硅基世界的融合。

## 清华大学智能产业研究院 (AIR) 的实践

五年前，2020年12月，我来到清华大学创立了智能产业研究院 (AI Industry Research, 简称AIR)。AIR的“I”有三重含义：国际化(International)、人工智能(AI)、产业(Industry)。我们的使命清晰明确：用人工智能创新赋能产业，推动社会进步；目标是打造面向第四次工业革命的国际化、智能化、产业化研究机构。

要实现这一目标，核心是培养未来技术领军人才。我们采用“学术+产业”双引擎模式，大部分老师既具备深厚的学术造诣，又拥有丰富的产业经验。目前研究院已有20多位老师、100多位博士后和博士生、400多位实习生，已成为全球人工智能领域最活跃、最具贡献的机构之一。

我们的研究方向与今天所讲的三元智能高度契合，聚焦信息智能、物理智能和生物智能，开展机器人、无人车、边缘智能、大健康、生物制药等领域研究，合作对象也以产业界企业为主。

## 02 智能体 AI 的多元应用

下面我将结合AIR老师们的研究成果，从信息智能、物理智能、生物智能三个维度，为大家介绍智能体的具体应用。

### 信息智能体：从数学解题到科学研究

智能体的核心难点之一是实现自主、可进化、



可泛化的能力，让其能在手机、PC、眼镜、手表、电视等不同设备上运行，可应用于购物、旅游、企业供应链管理等多个场景。更重要的是，我们希望智能体能够完成更高级的任务，比如解数学题、发明方程式、提出新问题。

目前，Gemini、ChatGPT等最新大模型在数学奥林匹克竞赛中已能击败人类冠军。而我们的目标不止于此——我们希望让AI证明尚未被人类证明的数学定理。

AIR李鹏老师团队与清华大学丘成桐先生的数学研究院合作，研发了数学智能体AIM。它能够分解任务，完成定理证明。例如在材料科学、分子动力学领域的重要难题“均匀化问题”的证明中，AIM形成了17页的证明文档，其中绿色部分完全由机器生成，橙色部分主要由机器证明后经人工校正，蓝色部分为人工完成，黑色部分是问题描述。这是人机协作的绝佳范例，据数学老师反馈，证明中最难的部分由AI完成。

虽然目前AIM证明的问题仍有一定难度限制，但我相信未来五年，人工智能能够独立证明更难的数学问题——比如千禧年提出的7个最难数学问题（目前已有2个被解决，剩余5个包括计算机领域的NP完备性问题、哥德巴赫猜想、黎曼猜想等）。我与丘成桐先生就此“打赌”，我认为五年内AI能完成其中至少一个难题的证明。无论具体时间如

### 愿景：从数学解题走向数学研究

**基模型**

- 数据以计算题和竞赛型题目为主
- 与数学研究需求存在明显差异

**自动形式化证明**

- 自动形式化困难、证明搜索空间大
- 当前仅能解决特定或简单问题

**AIM**

- 在自然语言空间完成数学证明
- 已呈现解决数学研究问题潜力

**量子算法问题**  
Quantum Algorithm Problem

**吸收边界问题**  
Absorbing Boundary Condition

**高对比度极限**  
High Contrast Limit

**均匀化问题**  
Homogenization Problem

**AIM完成了带有详细分析推导过程的证明**

**AIM基本完成此问题的证明过程和细节**

**AIM完成了主要结论的证明过程，并且证明了新的结论**

**AIM给出了合理的证明思路和部分结论的证明，对数学工作者很有启发意义**

**数学智能体**  
AI Mathematician  
AIM

**AIM workflow AIM workflow**

Y. Liu\*, Y. Huang\*, Y. Wang\*, Peng Li, Yang Liu. AI Mathematician: Towards Fully Automated Frontier Mathematical Research. arXiv:2505.22451. \* 同等贡献

**自主证明 → 人机协同证明**

成功解决均匀化问题，形成17页证明，AIM在证明中最具挑战部分做出非平凡贡献

**人机协同解决均匀化问题**

- 绿色：完全由AI生成+人工辅助验证
- 蓝色：AIM输出具有较高完成度证明后人工校正
- 白色：人工完成
- 黑色：问题描述等

完成叠衣服任务，还能将相关技能迁移到做家务等其他场景，完全通过自主学习适应环境。

无人驾驶是物理智能体的另一重要应用，也是我多年来持续关注的课题。此前在百度担任总裁期间，我们启动了“阿波罗”项目，其落地应用“萝卜快跑”就是一款专门用于驾驶的机器人。无人驾驶的难度极高，需要车辆精准感知复杂交通环境、规划路径、做出实时

何，核心意义在于AI已具备证明难题、提出新问题、生成新方程式的潜力。

安全的决策，集成了人工智能的各个核心技术，因此被称为“人工智能集大成者”。

## 物理智能体：从机器人到无人驾驶

我对阿波罗团队提出了三个核心要求：安全、安全、再安全。这里的“安全”定义为：机器驾驶的事故率比人类驾驶低10倍以上，才能实现商业化上路。经过近10年的研发，无人驾驶L4级（完全

ChatGPT本质上是语言模型，而物理世界的智能体需要具备视觉（Vision）、语言（Language）、行动（Action）能力，构建“世界模型”。AIR的曹婷老师团队研发的系统，就实现了物理世界机器人智能体的核心功能——通过感知、推理、进化、行动和奖励机制，生成决策和动作，指挥机器人完成任务。

### 无人驾驶的技术挑战

系统完整性：感知、规划、决策、执行全面高任务复杂度：针对高复杂性场景的高可靠性需求

**最具挑战的有边界的AI垂直领域问题**  
世界模型+边缘+具身智能

**无人驾驶L4智能体**

Real-Sim-Real

物理世界

数字世界

詹仙园老师团队研发的X-VLA系统，尝试解决智能体的泛化问题。传统机器人学会一项技能后，难以迁移到其他机器人或不同场景。而X-VLA系统仅需9亿个参数，就能部署到不同机械臂和机器人上，实现技能的跨设备、跨场景迁移。比如机械臂学会叠衣服后，更换不同机械臂、调整桌子高度，仍能

### 无人化新阶段：开放复杂道路商业化运营

武汉萝卜快跑 Apollo Go 全球最大的无人驾驶城市  
3000平方公里，1700万人，1500辆车（2025）

200+企业	2亿+公里	1500+
Apollo联盟 理事长：张亚勤	自动驾驶里程	自动驾驶车辆
10+落地城市	10X人类驾驶安全性	2500万+载人订单量
约3000平方公里的服务区	7×24小时服务承诺	

**在中国打造全球最大的无人驾驶平台和运营**

无人、无安全员)的核心技术挑战已基本克服,长尾问题不断优化。

目前,“萝卜快跑”已累计行驶2亿公里,在国内外十几个城市落地运营,安全程度超过人类驾驶10倍,未发生过一次恶性事故。仅在武汉,3000平方公里的区域内、1700万人口的城市中,1500辆萝卜快跑车辆可实现7×24小时服务,我们已在中国打造了全球最大的无人驾驶平台和运营体系。

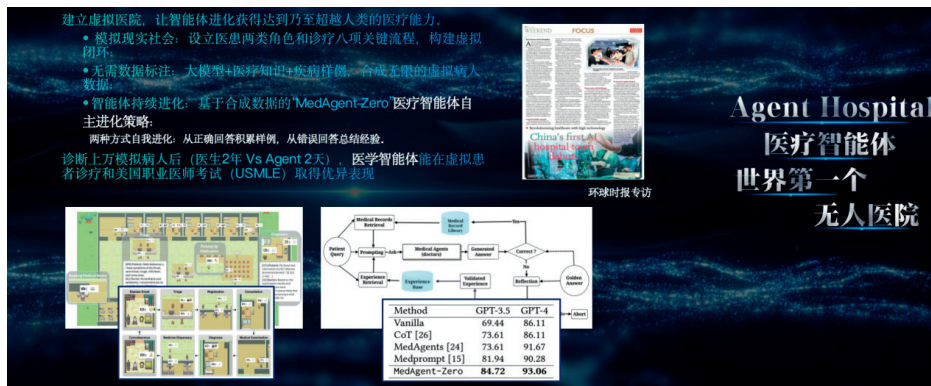
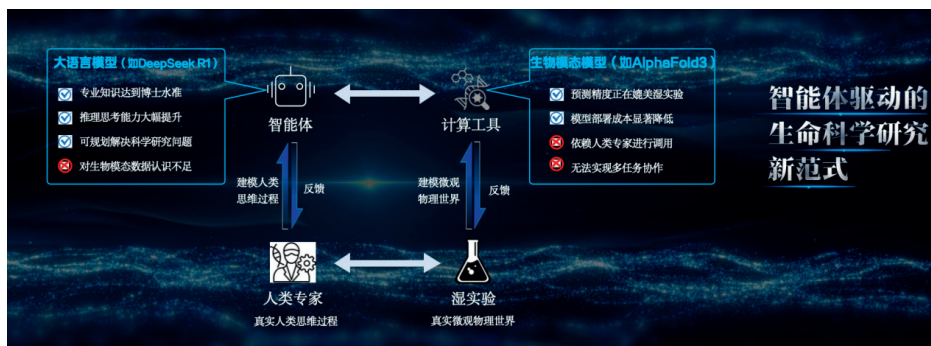
全球范围内,谷歌的Waymo、特斯拉,以及国内的Pony(小马智行)、WeRide(文远知行)、地平线等企业在无人驾驶领域也取得了显著进展,整个产业正

从技术研发走向商业化落地。我预计到2030年,10%的新车将具备L4级无人驾驶功能,迎来无人驾驶的“DeepSeek时刻”。

### 生物智能体:从新药研发到智能医疗

人工智能在生物智能领域的应用,首先体现在新药研发的加速上。德米斯·哈萨比斯在达沃斯对话中提到,未来十年左右人类所有疾病都可能被治愈,这一观点也许过于乐观,但人工智能确实能大幅缩短新药研发周期。

AIR的兰艳艳老师团队研发了新药筛选新技术,

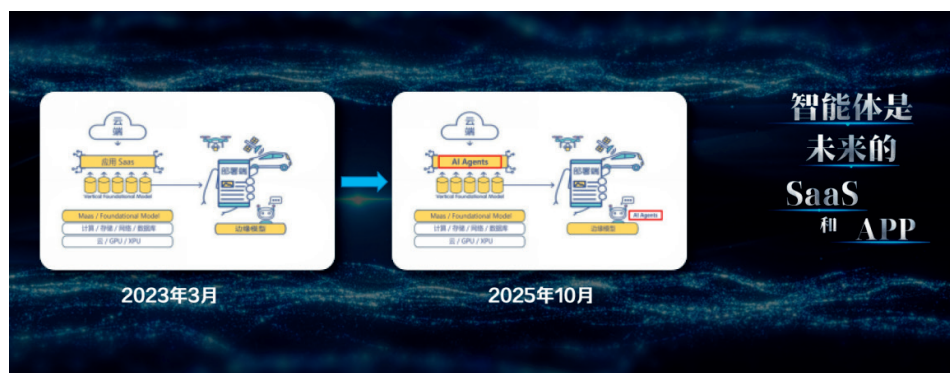


通过 AlphaFold 解码 2 万多个蛋白质结构,找到“口袋靶点”,再与几十亿、上百亿个蛋白质进行对接。目前仅有不到 10% 的蛋白质可用于制药,大量蛋白质分子结构的应用价值尚未被发掘,而该技术通过人工智能算法实现了快速对接,相关成果即将在《科学》杂志发表。

聂再清老师团队打造的新药研发智能体,能够根据研发需求分解任务,自动查找资料、分析蛋白质结构和功能,生成初步研发图谱,极大提升了新药研发的效率,为科研人员提供了重要支持。

人工智能在医疗领域的另一突破,是 AIR 刘

洋老师团队打造的全球首个无人智能体医院——清华大学人工智能医院（今年4月成立）。这是一个虚拟医院，医生、病人、护士等角色均由智能体担任，涵盖不同科室，形成完整诊疗闭环。



智能体之间通过协作、博弈不断进化，无需人工标注数据，仅需两天时间就能完成相当于两年的病例诊断学习，诊断准确率超过传统医院。

需要强调的是，AI 智能体医生并非要替代人类医生，而是作为医生的助手，提升诊断效率和准确性。目前，该系统已在清华校医院、长庚医院等十几家医疗机构开展测试，由真实医生和病人参与验证。

### 03 未来技术发展和产业格局

#### 人工智能时代的“操作系统”

下面我想讲一下未来技术的发展趋势，特别是产业格局的变化。我在微软公司工作近16年，期间主持开发了全球最大的嵌入式操作系统 WindowsCE，所以对操作系统有一个特殊的情结。操作系统，它是定义一个时代最重要的技术平台，有了操作系统之后，下面的芯片，上面的应用程序，整个技术生态其实都是围绕着操作系统来部署的。在 PC 时代，我们知道操作系统是 Windows( 视窗操作系统 )，芯片是 X86 架构，上面围绕着这个平台开发了各种各样的应用程序。到了移动互联网时代，我们用的手机操作系统是 iOS 和安卓，在国内我们也用华为的鸿蒙系统。下面的芯片也变了，变成了 ARM 架

构，上面的应用也变了，有各种不同的移动应用，像微信、短视频等各种应用。到了人工智能时代，大模型就是人工智能时代的操作系统。围绕着这个操作系统，下面的芯片架构变成了以 GPU 为主流，NVIDIA（英伟达）为什么全球市值这么高？因为它主要就是做 GPU 的。上面的芯片架构变了，下面的应用生态也变了，这次人工智能时代的技术规模，比移动互联网时代、比 PC 时代要大很多倍，可能会达到一个数量级、两个数量级，甚至更大的规模。

2023年3月，我绘制了人工智能时代的架构图：以前沿基座大模型为操作系统，上层涵盖行业垂直系统、SaaS 应用软件，端侧（手机、PC）则通过大模型蒸馏或压缩后的小模型运行 APP。到 2025 年 10 月，我对这一架构进行了更新，核心变化是将 SaaS 和 APP 替换为智能体——我认为智能体是未来的 SaaS，也是未来的 APP。虽然短期内手机 APP 仍是主流，但智能体功能将逐步融入其中。

#### 通用人工智能 (AGI) 的实现路径

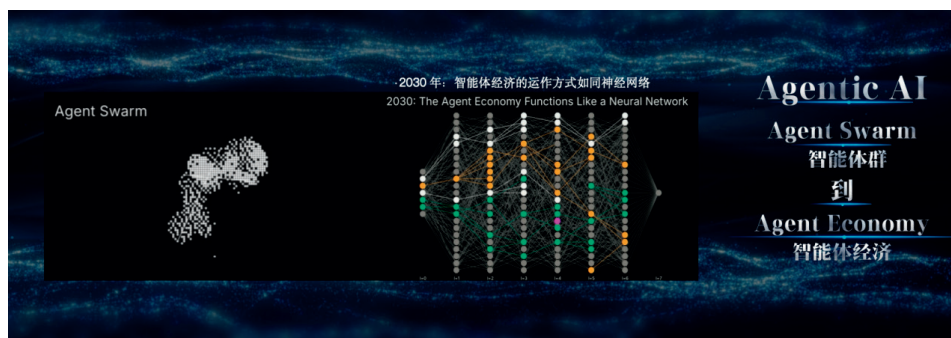
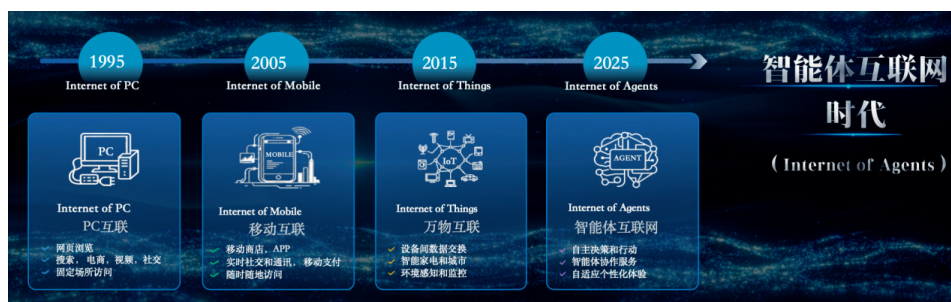
智能体是实现通用人工智能 (AGI) 的必然路径。目前 AGI 的定义尚未统一，我对它的理解是：可进化、可泛化、具备长期记忆，在 99% 的任务上超过 99% 的人类。

要实现 AGI，还需要解决一系列关键问题，比如构建符合物理定律的世界模型、理解因果关系、

优化记忆系统等。当前 AI 的记忆较为粗糙浮浅，而人类记忆是智能的核心复杂部分。

如果按照这样的定义，我认为我们在 15-20 年内会达到 AGI 的水平，并且能够通过“新图灵测试”。图灵测试，一开始只是文本对话层面的测试，现在它已经延伸到各个领域了。首先在信息领域，在内容生成方面，我觉得 5 年之内就可以达到 AGI 的水平，大家看最新的 Sora 视频生成模型，它生成的视频已经和人类制作的差不多了，甚至在某些方面比人类做得还好。在 10

年之内，在物理智能方面可以实现 AGI，无人车从技术方面已经基本过关了，人形机器人还需要更长的时间。我们现在看到各种人形机器人的表演都很好，也有很多相关的研究，包括人形机器人的灵巧手、人脸肌肉控制等技术，但要做到真正和人类相似，至少还需要 10 年的时间。不过我很看好这个产业，我认为 2030 年，也就是十年以后，机器人的数量会超过人类的数量，这会是一个巨大的产业。但是现在，人形机器人还处于科研阶段，还没有到完全量产的地步。那么更重要的是生物智能领域，比如脑机接口、生物体与 AI 的融合、生命体的数字化等，这个领域要实现 AGI，还需要差不多 20 年的时间。



另外一个角度看我们的产业发展，是从互联网的发展脉络来看。1995 年开启了 PC 互联网时代，2005 年开启了移动互联网时代，2015 年开启了物联网时代，也就是万物互联的时代。现在我认为，我们进入了一个新的时代，就是智能体互联网时代，也就是 Internet of Agents。这里面有一个特别有意思的概念，叫做 Agent Swarm（智能体群）。

红杉资本在 2025 年的演讲中提出了“Agent Swarm”（智能体群）这个概念：未来人类交往将通过智能体实现，智能体之间将形成群体智能，通过协作、博弈、纠错不断进化，如同蜂群、动物群体一般，最终进化为类似人类大脑神经网络的结构，催生“智能体经济”。



这种智能体经济将彻底改变经济形态、人类组织架构和企业运作模式：企业的核心资产将变为芯片、数据中心、数据和 AI 模型；团队组建不再局限于招聘人类员工，智能体将成为重要组成部分；就业形态、分配模式等深层次社会问题也将随之重构。

## 人工智能的风险与治理

还有一件最重要的事，就是人工智能智能体在带来巨大机遇和强大能力的同时，也伴随着不可忽视的风险。

这个风险有几个层面：首先是信息智能领域的风险，我们已经看到了，它可以生成虚假信息，它可以进行深度伪造，有的时候它还会产生幻觉，还可以用来欺骗他人，另外还有版权归属的问题。我们现在使用的互联网信息，从上个月开始，已经有 50% 以上的信息是人工智能所产生的。这个时候我们怎么防范这些风险？比如现在有很多不实的信息，这些不实信息又被用来训练新的大模型，然后生成更多的不实信息，形成恶性循环。我们怎么解决这些问题？需要从技术、政策、法规方面共同努力来解决这些问题，不过这个领域的风险我认为还是可控的。

到了物理世界，我们把大模型、智能体和无人车、机器人、无人机，包括军事系统连接起来，智

能体之间的协作和博弈，如果出现失控，如果被恶意滥用，那么造成的风险就会更大。到了生物智能领域，假如我们的大脑和 AI 连接在一起了，碳基生命和硅基世界通过芯片或者外挂的传感器连接在一起了，到了那个时候，尽管它可以给我们带来巨大的好处，我们可以想象到，如果一旦出现失控、被滥用，风险就会特别大。所以这个领域需要我们人类最聪明的人

去研究这些问题、解决这些问题、面对这些挑战。这里面有做科学研究的、技术开发的、产品设计的，同时也需要政府的政策法规专家一起，共同打造有效的治理框架，而且这个治理框架需要是全球范围的。但我自己是有信心的，人类进化了这么多年，我们有一个特殊的能力，就是可以发明高级的工具，同时我们也可以管理好高级的工具。

当前，人工智能正从鉴别式 AI 走向生成式 AI，并逐步迈向智能体 AI。新一轮人工智能是信息智能、物理智能和生物智能的融合，是原子、比特和分子的融合，是碳基生命和硅基世界的融合。在这一进程中，我们拥有天文级海量数据、指数级运算能力，更重要的是人与机器将协同进化，催生巨大的产业机遇——达沃斯 AI 理事会预测，到 2030 年，人工智能带来的新机遇将创造 20 万亿美元的经济价值，超过当前许多国家的 GDP 总量。同时，我们也面临着隐私保护、安全保障、就业转型、社会公平、风险治理等一系列社会挑战，人工智能将重构全球社会、经济版图。

80 年前，《科学：无尽的前沿》法案推动了第三次工业革命，而在前三次工业革命中，中国始终是旁观者或跟随者。而人工智能带来了新的无尽的前沿，正在开启第四次工业革命，这一次，我坚信凭借强大的国力、众多的人才和有利的政策，中国必将成为领军者！