

2023

# 十大前沿科技报告

QbitAI Annual Frontier  
Technology Report

# 序言

有个论断是这么说的：

两千年前的古人穿越到一千年前以后，需要适应的东西可能不多。而一千前的古人穿越到一百年前，变化也不至于大到难以适应。但如果一百年前的古人穿越到现在，甚至只是有人“冬眠”个10年、20年，面对的现实变化，堪称天翻地覆。

虽然有段子调侃的成分，但也基本是现状的写实。而变化的核心驱动因素，就是日新月异的科技——它已经并且还在持续带来颠覆式变革。

这是我们这代人的幸运所在，毫无疑问，我们身处一个科技大爆炸、创新密集发生的周期。

于是也是基于对周期的感知，从2021年开始，量子位智库，以一个前沿科技“瞭望所”的姿态，开启了对年度前沿科技趋势的总结、梳理和展望，希望能让更多关注科技变革的人提供参考。

区别于其他智库和研究机构，在量子位智库成立之前，我们通过量子位的方式，站在AI及前沿科技的信息聚合中心，已超过5年，这让我们可以更加近距离观测前沿科技在产学研交汇地带正在发生的一切，也让我们对前沿科技进入大众视野的节奏更有把握。

「大规模预训练模型预示通用智能。」

这是量子位智库《2021年十大前沿科技报告》中传送的趋势，2年前，大模型不算“正道”，谈通用人工智能多少有点科幻色彩。但量子位智库，最后还是把这条趋势写进了年度报告，广而告之。

而2年后的今天，大模型带来的范式变革，AGI的可能性和可能性话题，基本已经成为了新共识。

似的还包括新型AI芯片及其架构、固态电池提升储能上限、量子计算机软硬件协同、AR打开XR新纪元……

我们相信，这些前沿科技或技术创新，最初可能只是一个点，但很快就会变成一条线、一个面，开启一个赛道、一个产业。

这也是量子位智库《十大前沿科技报告》系列想要传递的核心价值，希望能让更多人关注到前沿科技正在发生的变革、更早参与前沿科技创新，然后加入其中实现更大的变革。

当前，前沿科技进入成熟周期的时间正在不断缩短，前沿科技变成经济增长新动能的速度也正在越来越快。

既然已经生逢其时——

If not now, when?

If not us, who?

量子位智库总裁 李根

# 特别鸣谢

科技合作伙伴



\*以上排名无先后顺序



# 目录

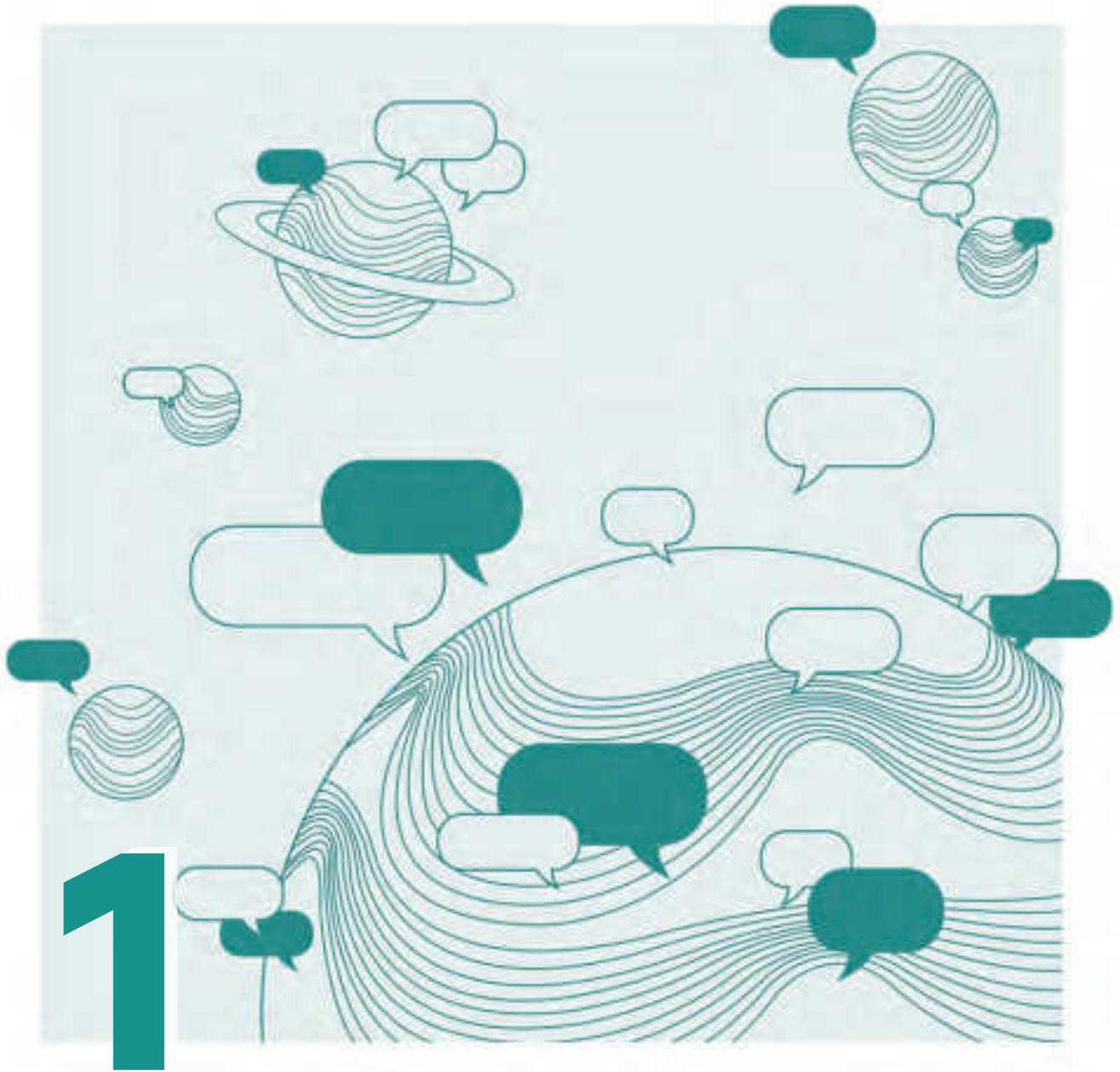
## 序言

- 1 / 智能体热潮：  
人机交互新范式已被大模型打开 P01
- 2 / 3D生成进入涌现期：  
新算法新模型爆发，质量效率可控性日新月异 P08
- 3 / 分割模型大一统：  
计算机视觉即将迎来「GPT时刻」 P16
- 4 / 具身智能带来新想象：  
AGI终极场景下的全新终端 P22
- 5 / 端到端自动驾驶成共识：  
BEV+Transformer重构技术路线 P29
- 6 / 空间计算定义明确：  
消费级产品问世，XR全栈链路打通 P35
- 7 / mRNA打开新象限：  
提供精准医疗新解法，开启生物医药新篇章 P43
- 8 / 脑机接口试验新阶段：  
产品可靠性突破，AI提升数据解码能力 P49
- 9 / 可回收火箭进入「中国轨道」：  
工程化难题被突破，商业航天迎来模式闭环 P55
- 10 / 可控核聚变里程碑：  
点火成功，打开商用想象空间 P63

其他提名前沿科技趋势 P69

2024前沿科技投资观点 P79

结语



## 智能体热潮：人机交互新范式已被大模型打开

AlphaGo的胜利深刻启发了人们对智能体潜在影响的认识。7年后，大模型的崛起再次推动了AI智能体的发展，成为其强大的“动力引擎”。2023年3月以来，产业界产生了多个杰出的AI Agent，如在游戏领域表现出色的英伟达Voyager智能体、协助人们完成日常任务的AI助理HyperWrite，以及专注于提供个人情感陪伴的AI助手Pi等，AI Agent的研究取得了突破性的进展。

# 推荐理由

## 大模型研究与智能体研究互惠互利

目前,许多研究利用大模型作为AI Agent (AI智能体)的认知核心,模型的发展为智能体研究提供了质量保证。从智能体的角度来看待大模型,对大模型研究提出了更高的要求,同时也扩大了大模型的应用范围。

### (1) 大模型为智能体提供了突破性的技术方案

过去基于深度学习框架的智能体能够学习技能,但无法真正理解问题和技能。而大模型智能体带来了深度学习新范式,从思维链到思维算法的推理技术和强大的自然语言理解,这些大模型拥有的技术能力,有望让智能体具备强大的学习和迁移能力,从而让创建广泛应用且实用的智能体成为可能。

### (2) 自主智能体实现复杂流程自动化, 流程效率大幅提升

面对完善的自主智能体,当给定一个目标时,它们能自行创建任务、完成任务、创建新任务、重新确定任务列表的优先级、完成新的首要任务,并不断重复这个过程,直到完成目标。准确度要求高,因而更需要调用外部工具辅助减少大模型不确定性的负面影响。

### (3) 模拟智能体更加拟人可信, 能够提供情感和情绪价值

模拟智能体主要分为强调情感情商智能体以及强调交互的智能体,后者往往出现在多智能体环境中,可能涌现出超越设计者规划的场景和能力,大模型生成的不确定性反而成为优势,多样性使其有望成为AIGC重要组成部分。

### (4) 嵌入自主智能体的软件更符合用户的使用习惯

在未来,大模型驱动的智能体极有可能带来交互方式的变革,从过去用户适应软件应用,变成应用软件适应用户个人习惯,为用户的生活提供更加便利的服务。

“

LLM-based Agent是大语言模型推理能力的展现,是一种能力,而不是一个完整的产品形态,这个能力可以应用于toB和toC领域, toB领域对传统软件是一种补充, toC领域如游戏、内容,对于内容分发有生成+推荐的结合机会。

——明势资本

“

从底层驱动力上看,我们认为主要有两类驱动力在推动着AI Agent的发展:

- **技术驱动力:** 围绕AI Agent开发的技术框架、关键组件以及基础设施一直在过去的几个月中不断地演化和发展,并且,这种推动力不仅来自于闭源公司(如OpenAI),也来自于开源生态的努力(如AutoGPT)。随着底层技术的不断完善与成熟,我们已经看到Agent的开发者生态和应用生态正在逐渐壮大。
- **商业驱动力:** 由于AI Agent具备重构现有互联网应用生态的潜力,因此从商业角度考虑,有望构建围绕Agent的新应用生态的公司都有比较强的动力去推动AI Agent的发展。

在这之中,我们认为主要有几类玩家。一是LLM底层技术提供商,典型的代表就是OpenAI,他们不仅开发LLM底层技术,也会做GPTs的应用生态;二是硬件厂商,他们希望能够延续原先移动应用生态的优势,抓住AI应用生态的机会;三是做开发生态和应用生态的玩家和创业公司,他们也希望从应用开发者和用户的角度切入做新的应用生态。

——BV百度风投

“

我们看好与LLM相关的整体技术栈，包括Agent技术。看好该技术的原因在于其重要性。我们认为AGI很可能是提升整体生产力的核心技术，应用广泛性和对生产要素的替代能力具备划时代性特点，甚至可带动其他诸多前沿科技一起往前发展（如带动算力中的Chiplet技术，带动机器人中的具身智能技术，等等）。而Agent将是迈向AGI的重要路径，目前已看到诸多海外研发进展。虽然道阻且长，我们对其抱有很大期待。

——泰合资本

“

随着AI Agent的逐渐普及和成熟，我们相信未来也会形成一个Agent与Agent、人与Agent进行交互、协作与价值交互的群体智能网络。在这样的群体智能网络之中，信息、商品、资金、服务可以高效地以Agent作为载体进行流动。

——BV百度风投

## 技术原理

### 1. AI Agent原理

AI Agent是能够感知环境并做出反应，通过决策和行动改变环境，并通过学习和反思持续迭代的智能体。此前，智能体经历了符号智能体、交互式智能体、基于强化学习的智能体和具备迁移学习和元学习能力的智能体。在大模型爆发后，AI Agent就采用大模型作为智能体的核心组成部分，通过多模态感知和工具调用来扩展其感知和行动的范围。

LLM-based Agent（基于大模型的智能体）集中了符号智能体推理规划的能力，具备了交互式智能体在反馈中学习、与环境互动的能力，同时具备大模型的少量泛化能力，在任务间实现无缝转移，而无需更新参数。

LLM-based Agent正是基于大模型驱动的Agent，可以实现对通用问题的自动化处理。

“

自主智能体（AI Agent/Autonomous AI Agent）长久以来一直是人工智能界研究的焦点。曾经也出现过一些如SOAR的类似系统，但限于当时的技术水平，应用领域非常狭窄。随着大语言模型的智能涌现，基于大语言模型的自主智能体拥有高度的智能水平，能够自主理解、拆解复杂、抽象的任务，也有更强的工具调用和感知外界反馈的能力，能够自我反思从而对问题提出更优解，甚至能够构建群体智能。我们判断基于大语言模型的智能体将最终成为人类良好的助手、同事和伙伴。

——百度研究院

“

基于强化学习的智能体需要建立基于具体场景的世界模型的仿真器，而LLM-based Agent中大模型的强语言理解能力使得与人相关的仿真器建设变得简单。过去没有大模型的情况下，需要穷举人的问答的各种情况，而现在通过大模型的语言理解能力，智能体就可以在和人的互动中纠正错误并继续逐步推理。

——澜码科技

### 2. LLM-based Agent架构

LLM-based Agent（基于大模型的智能体）的架构可以总结为以下四个模块：配置模块、记忆模块、规划模块和行动模块。

## 配置模块

在配置模块,需要给智能体提供待解决问题的背景信息,比如模拟人类时的年龄、性别、职业等基本信息,让智能体明确自身角色,智能体一般通过IT人员,教师和领域专家等特定角色来执行任务。

## 记忆模块

记忆模块主要是传递知识,让智能体拥有长期和短期的记忆能力。智能体记忆从环境中感知到的信息,并利用记录的记忆来促进未来的动作。记忆模块可以帮助智能体积累经验、实现自我进化,并以更加一致、合理、有效的方式完成任务。

## 规划模块

规划模块是智能体能力的核心,一个好的规划决定了智能体能否顺利执行以及解决问题,规划模块首先将复杂任务分解为简单的子任务,然后逐一解决每个子任务,以及不断根据反馈去重新调整策略。规划模块赋予基于大模型的智能体解决复杂任务时需要的思考和规划能力,使智能体更加全面、强大、可靠。

## 行动模块

行动模块的目的在于将智能体的决策转化为具体的结果输出。它直接与环境交互,决定智能体完成任务的有效性。



基于大模型的自主智能体的架构

“

AI Agent是指能够独立思考、自主行动并可以与环境交互的软件程序或机器人等实体。Agent包含三步:PPA,即感知(Perception)--规划(Planning)--行动(Action)。人工智能之父、图灵奖得主马文·明斯基(Marvin Minsky)在1986年出版了一本里程碑式的著作《思维的社会》(The Society of Mind),试图解读人类思维这个复杂的过程。Minsky认为社会中的某些个体经过协商之后可求得问题的解,这些个体就是Agent,而Agent应具有社会交互性和智能性。

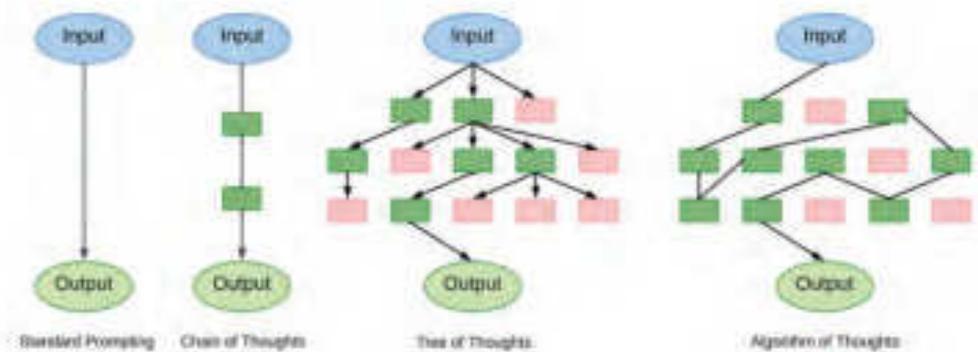
——腾讯研究院

## 规划层面

### 思维算法 (Algorithm of Thoughts)

过去常用思维链(Chain-of-Thought)和思维树(Tree-of-Thoughts)来引导模型任务分解,利用大模型的上下文学习能力模仿类似的任务分解和规划,但这两种技术依赖于大模型的大规模查询,有时单个问题的查询数量可以达到数百个,导致计算效率的下降。

微软联合弗吉尼亚理工大学推出思维算法(Algorithm of Thoughts)具有动态和可变的推理路径,通过维持一条不断发展的思维上下文链条,提高了推理效率并减少了计算成本。这种方法的优势在于它能够灵活地适应不同的问题和情境,并且能够根据需要进行调整和优化。



思维链、思维树、思维算法路线对比

思维算法技术包括四个主要步骤：

- (1) 将复杂问题分解为可理解的子问题，同时考虑它们的相互关系和单独解决的容易程度；
- (2) 以连续和不间断的方式为这些子问题提出连贯的解决方案；
- (3) 直观地评估每个解决方案或子问题的可行性，而不依赖于明确的外部提示；
- (4) 根据上下文示例和算法指南，确定最有希望探索或回溯的路径。

### 3. 基于大模型的多智能体协同技术

多智能体协同是指多个智能体在共享环境中通过相互通信和协作，实现协同行动以达成共同目标的过程。每个智能体都具备一定的自主性和智能性，能够根据环境信息进行感知、决策和执行行动。

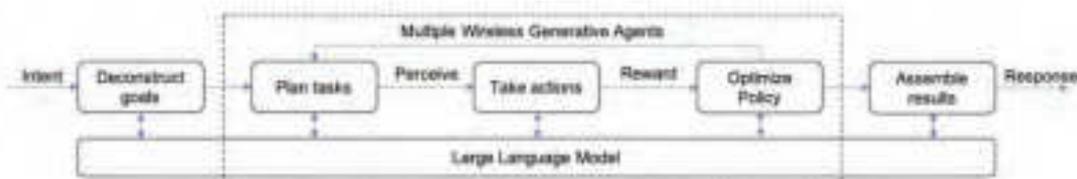


Fig. 1: Close-loop task planning, execution, optimization in wireless generative agents.

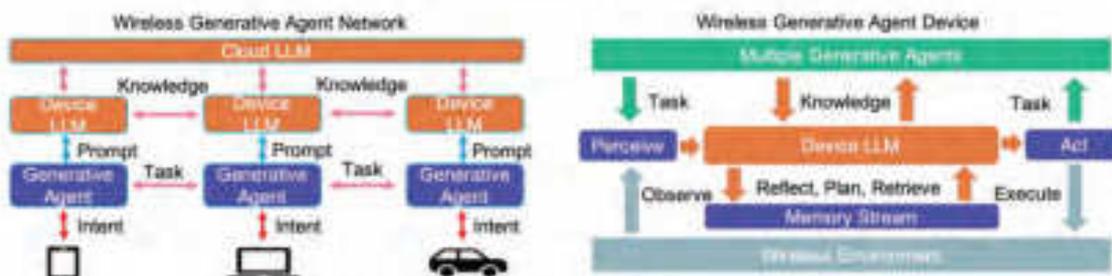


Fig. 2: Wireless generative agent network and device architecture.

无线生成式智能体网络和设备架构图

智能体之间通过网络交换知识、系统规划任务、采取行动并优化策略。其中知识是数据的抽象表示，将其编码为执行特定任务的大模型。来自人类或机器的任务通过不同的终端（手机、电脑、智能车）提供给生成式智能体，这些终端设备通过智能体给大模型创建提示，以完成每个步骤。任务在多个生成智能体之间协同规划，能够最大化地利用不同大模型的知识 and 不同设备的功能。当从其他智能体接收到计划任务时，设备上的大模型可以从云端或其他设

备上的大模型获取特定领域的知识。

无线生成智能体具有观察环境的感知器(传感器)和执行决策的参与者(控制器)。设备上的大模型从观察到的多模态(文本、图像、声音)原始数据中提取语义信息,并将其存储在内存流中,以便将来规划新的任务。相应地,为了执行特定的任务,将检索相关的语义信息以采取行动。在从接收到的高级任务中完成计划的动作后,智能体可以进一步创建低级任务并将其发送给其他智能体以完成目标。

### 多智能体协同技术的主要特征

多智能体协同通过相互之间的交互与合作,使整个系统能够从各个智能体的优势和特长中受益,实现更高效、更智能的决策和行动。多智能体协同主要有以下特征:

#### 通信与信息共享

智能体之间需要进行有效的通信和信息共享,以便相互了解彼此的状态、意图和行动计划。通信可以通过直接传递消息、共享知识库或传感器数据等方式来实现。

#### 决策与合作策略

每个智能体在面临问题时需要进行自主决策,并制定合适的合作策略。决策和合作策略可能涉及到资源分配、任务分工、冲突解决等问题,需要综合考虑整体目标和个体利益。

#### 协同行动与集体效能

多个智能体通过协调动作和行为,实现集体行动的目标。协同行动可以包括任务分工、资源共享、互助、协同学习等方式,以提高整个系统的效能和性能。

“

群体智能的研究在不断推动人工智能的理论技术创新,是目前最通用、最主流的应用,没有之一,其商业价值的想象空间很大。随着大语言模型的成熟,让人看到其商业化的希望。

——跃为资本

## Smallville

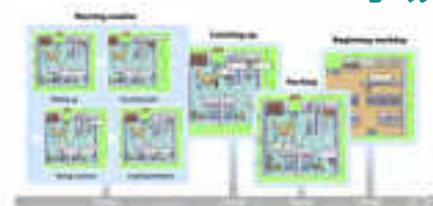
### AI小镇

1

斯坦福大学利用ChatGPT创造了一个AI小镇,将25个智能体作为简单沙盒世界中的角色,让智能体在其中工作、闲聊、结识新朋友等,模仿人类生活。用户可以通过自然语言改变这些智能体及其生活环境的设定,并与之交互。



2



给智能体设定性格、职业以及初始任务关系后,智能体会对一天的生活做一个初始规划,在和环境以及其他智能体的交互过程中,会不断修改自己的计划。右图展示了智能体按照初始计划度过的一上午。

3

智能体会与其他智能体以及环境形成自然交互,交换信息、构建关系以及协同合作等社交行为是由模型自发生出来的,而不是预先编程好的。

——Stanford University 《Generative Agents: Interactive Simulacra of Human Behavior》

## 阿里云魔搭社区——ModelScope-Agent

ModelScope-Agent是一个通用的、可定制的智能体框架，专为开发人员利用开源大模型的功能而设计。通过大模型调用工具魔搭GPT (ModelScopeGPT)，使用者们可以通过一键发送指令调用魔搭社区中的其他人工智能模型，从而实现大大小小的模型共同协作，进而完成复杂的任务。



基于开源的大语言模型作为核心控制器，ModelScope-Agent包含记忆、控制和工具使用模块。它提供了一个用户友好的系统库，具有可定制的引擎设计，支持在多个开源LLM上进行模型训练，同时还能够以统一的方式与模型API和通用API无缝集成。为了使大模型具备工具使用能力，提出了全面框架，涵盖工具使用数据收集、工具检索、工具注册、内存控制、定制模型训练和实际应用的评估。

——《ModelScope-Agent: Building Your Customizable Agent System with Open-source Large Language Models》

## 代表机构

面壁智能



面壁智能是一家人工智能大模型技术创新与应用企业，近期与清华NLP实验室共同开发了XAgent，是由大语言模型驱动的实验性自主智能体，XAgent的设计中创新地引入了一种“双循环机制”，使它能够从“宏观”和“微观”两个视角进行任务规划和执行。XAgent还具备与人类协作和交互能力，可自主与用户进行交互，并向人类发出干预和指导的请求。

实在智能



实在智能是一家人工智能科技公司，致力于用AI技术推动RPA行业向人工智能流程自动化发展，今年8月发布的TARS-RPA-Agent是一个基于“TARS+ISSUT (智能屏幕语义理解)”双模引擎、有“大脑”，更有“眼睛和手脚”的超自动化智能体，是能够自主拆解任务、感知当前环境、执行并且反馈、记忆历史经验的RPA全新模式产品。

联汇科技



联汇科技成立于2003年，以AI多模态大数据处理和分析技术为基础，聚焦在媒体、政务、军民融合和城市信息化四大领域。今年7月，联汇科技发布了基于大模型能力的自主智能体(Auto AI Agent)——OmBot 欧姆智能体，并针对典型场景需求，推出了首批应用。

澜码科技



澜码科技是一家基于大语言模型的Agent平台公司。澜码科技构建了Agent平台“AskXBot”，平台分两层：第一层是专家赋能，专家通过拖、拉、拽以及对话交互的方式定义工作流程，教给机器，从而协助一线员工构建更高效工作的方法论；第二层是员工使用Agent，一线员工可以通过自然语言和Agent沟通并下达指令，让Agent协助完成数据分析、资料调取等工作。



## 3D生成进入涌现期： 新算法新模型爆发，质量效率可控性日新月异

3D生成是一种利用AI+计算机图形学技术，从文字、图像、视频或其他数据中自动创建三维物体模型的方法。它可以用于增强现实、虚拟现实、游戏、电影、艺术等领域，为用户提供更加真实和丰富的视觉体验。它是计算机视觉和图形学的一个重要研究方向，具有广泛的应用价值和创新潜力。由于游戏、影视、XR等产业对3D数据的需求不断增长，目前在有关AIGC的前沿探索中，也出现了不少3D生成相关的研究。

## 推荐理由

### 1. 3D生成技术可以代替传统3D建模管线中大部分流程,为行业降本增效

传统3D建模流程复杂,多为人工手动操作,周期长,成本高。如果3D生成技术成熟,现有的管线中的大部分流程(如创意灵感、草稿生成、模型优化等)将被替代,兼具效率和创新,对行业产生变革性影响。

### 2. 突破性新表达和算法3D Gaussian Splatting (高斯溅射) 出现,在部分数据集上多项性能优于NeRF

今年,在SIGGRAPH2023上被评为最佳论文的《3D Gaussian Splatting for Real-Time Radiance Field Rendering》介绍了一种新的可以实现快速高帧率实时渲染的方法——3D Gaussian Splatting,这一新表达和算法的出现是3D生成领域的一大突破性进展。在实际落地上,目前在3D重建中利用Gaussian Splatting可以解决NeRF与目前渲染引擎难以兼容、单独的开销体系和需要专门的硬件支持几大问题,在速度、质量、可控性上均强于NeRF。但在3D生成的不同场景中两者各有优势。

“

我们看到了NeRF展现出来的充分潜力,并很早就进行了跟进。但NeRF这一表达在实际业务中落地仍然存在一些难点,比如兼容性、开销,与硬件支持。

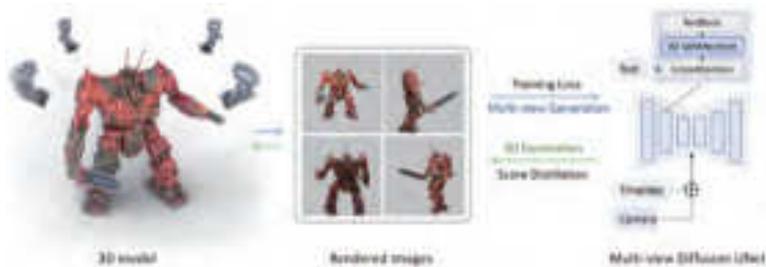
- 在兼容性上,NeRF部分完成了渲染的工作,意味着它需要兼容传统渲染引擎,或者直接替代。而NeRF生成的物件直接在Maya、Blender等常用CG软件中进行直接编辑是很困难的,没有办法很好与该类软件在工作流上融合。而Gaussian Splatting与现有的渲染引擎兼容性很好,海外也出现了大量艺术家基于这一表达进行创作。
- 在开销上,NeRF的特点是“一视同仁”。对于复杂的比如镜面液体,它能以相对传统光线追踪要低的算力去很好地展现出来,但是对一个我们现在已经能够渲染得很好的物体,比如一个漫反射的球,也要用相当的算力去表现出来,这其实是一个优势,也是一个劣势。而Gaussian Splatting在过程中能够根据渲染复杂度,自适应的调整用于表达的高斯数目。
- 在硬件支持上,NeRF在使用时往往需要MLP进行推理,这与当今GPU架构所擅长的并不十分一致,渲染引擎也并没有进行一些匹配硬件底层的实现。而Gaussian Splatting跟现有的GPU与引擎擅长的内容基本上是兼容的。

——影眸科技CTO 张启煊

最近,华中科技大学&华为研究团队又继续提出了4D Gaussian Splatting,它实现了实时的动态场景渲染,同时可保持高效的训练和存储效率。

### 3. 新模型MVDream出现,3D视角下的一致性大幅改善

今年10月,字节跳动的研究团队推出了一种全新的多视图扩散模型——MVDream,能够根据给定的文本提示生成几何上一致的多视图图像。



MVDream模型示意图

通过利用在大规模网络数据集上预训练的图像扩散模型以及从3D资源渲染的多视图数据集,得到的多视图扩散模型既能够实现2D扩散的通用性,又能够实现3D数据的一致性。这样的模型可以应用为3D生成的多视图先验,通过分数蒸馏抽样极大地提高了现有的2D提升方法的稳定性,大幅改善了3D一致性问题。

## 4. 多个3D生成优化模型出现,生成质量和效率大大提升

今年OpenAI发布的Shap-E,加州大学圣地亚哥分校发布的One-2-3-45都大大提升了AI生成3D的速度和精确度,其中One-2-3-45能够在仅45秒的时间内从2D图像生成高质量和具备一致性的3D网格。

11月刚发布的升级版One-2-3-45++,通过最初微调用于一致的多视图图像生成的2D扩散模型,随后借助于多视图调节的3D原生扩散模型将这些图像提升到3D,全面提升了三维模型的生成质量,但是速度没有太大损失(45秒到60秒)。

# 技术原理

## 1. 什么是3D生成

3D生成的目标是根据输入数据,生成一个能够表示三维空间中物体形状、纹理、材质等属性的输出表示,通常是指基于文本、图片等借助AI大模型生成3D模型。

3D建模主要有三种方式:3D扫描建模、3D软件建模、AIGC建模,其中的AIGC建模通常指的就是我们所说的3D生成。

## 2. 3D生成主流技术路径

目前3D生成的主流技术路径大致可以分为两种:

### (1) 2D升维

先用Diffusion模型完成text-to-2D,再通过NeRF等完成2D-to-3D,或者直接通过2D素材完成3D建模。

这个路径更为常见,因为相比于3D数据,2D数据更多、更丰富,可生成的3D内容丰富度更高。OpenAI的Point-E、Google的DreamFields和DreamFusion、Nvidia的Magic3D等模型均是此路径。但2D转3D生成速度很慢,生成质量也相对较低,即使单个场景的数据量足够大,目前也很难看到质量有显著提升,或许还需要算法层面的创新。

### (2) 原生3D

指Text-to-3D,该路径直接使用3D数据进行训练。

从训练到微调到推理都基于3D数据,OpenAI的Shap-E、Nvidia的Get3D等模型是基于此路径。Text-to-3D的优势在于生成速度较快,理论上生成的内容质量也较高,但受限于3D数据的数量、质量和多样性,此路径下可以生成的3D场景有限,收集数据的难度更大,成本也更高,可控性相对差些。

## 3. 3D模型的表达方式

3D模型的表达方式可分为显性表达与隐性表达两类。

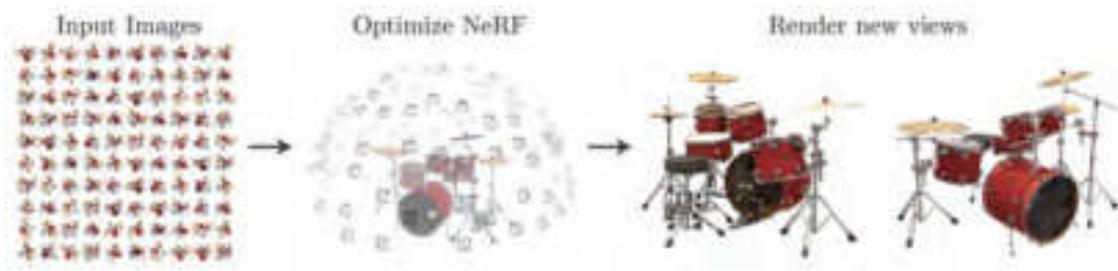
- 显性表达:主要包括体素Voxel,点云Point Cloud,网格Mesh等;
- 隐性表达:是以神经网络参数表达的3D场景,即神经场。主要包括符号距离函数Signed Distance Function(SDF),占用场Occupancy Field,神经辐射场Neural Radiance Field(NeRF)等。

## 4. 新旧3D生成表达与算法对比

### (1) NeRF

作为目前主流3D生成中的最重要的表达与算法之一, NeRF神经辐射场的概念首次被提出于2020年, 由UC Berkeley、Google、UCSD研究机构在论文《NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis》中阐述。

他们提出了一种新颖的神经网络架构, 通过学习一个连续的三维空间中的辐射场来从有限的二维图像中重建高质量的三维场景。



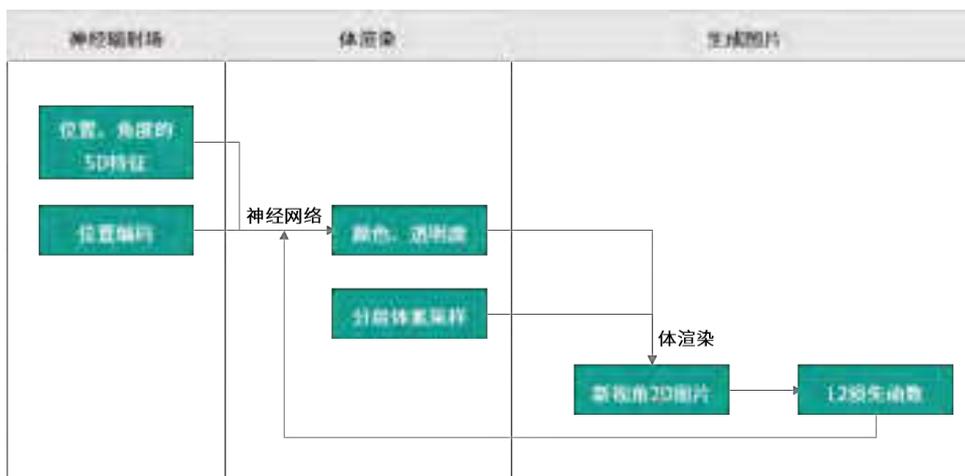
NeRF示意图

2022年, Nvidia发布Instant NeRF, 即将逆向渲染(在几秒钟内将一组静止图像转换为3D数字场景)与NeRF相结合, 生成和渲染速度提高了1000倍以上。

### NeRF整体框架

第一步, 通过神经网络学习场景的辐射场函数来实现对三维信息的编码, 输入为三维空间中的坐标和方向, 输出为对应点的颜色和透明度值。

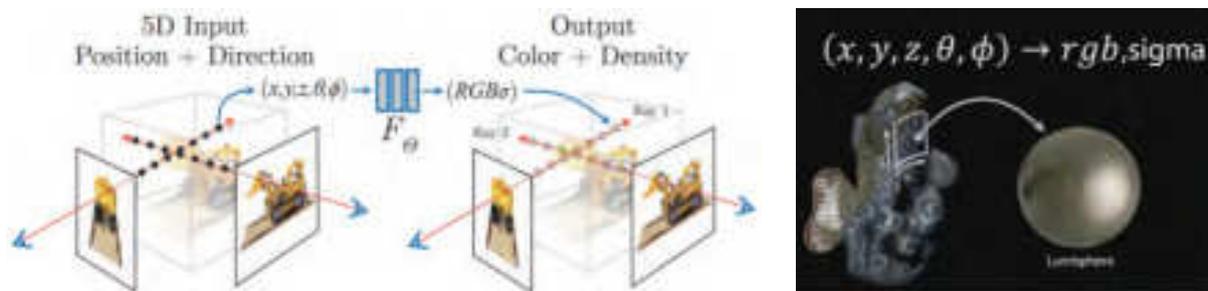
第二步, 体渲染, 首先将场景分成小的体素, 然后对每个体素内的光线进行采样, 通过对所有采样点的辐射场函数进行加权平均, 得到最终的像素颜色值。直观来说, 一个点的透明度越高, 这点在射线下的颜色反应在像素上的权重越小。



NeRF算法整体训练框架图

### 辐射场

在NeRF中, 辐射场被表示为一个神经网络模型, 用于从相机位置和方向计算出每个像素点的颜色和透明度, 从而实现高质量的三维重建和渲染。

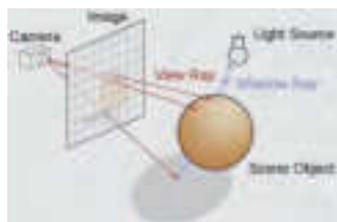


辐射场示意图

辐射场包含三个空间维度 $(x, y, z)$ 和两个方向维度 $(\theta, \phi)$ ，分别表示辐射在空间中的位置和光线的方向。辐射场可以用来描述光线在介质中的传播、反射、折射、散射等现象，以及介质中的吸收、发射等能量转换过程。神经辐射场可以看作是从空间位置和视角方向到颜色 $c=(r, g, b)$ 和透明度 $\sigma$ 的映射。

### 体渲染

在NeRF中，辐射场被表示为一个神经网络模型，用于从相机位置和方向计算出每个像素点的颜色和透明度，从而实现高质量的三维重建和渲染。



渲染示意图

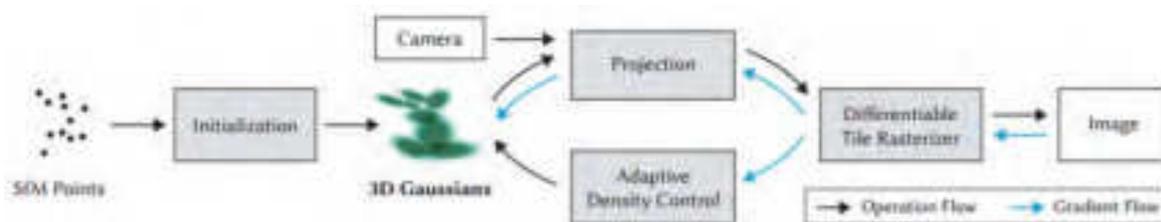
## (2) 3D Gaussian Splatting

SIGGRAPH2023最佳论文《3D Gaussian Splatting for Real-Time Radiance Field Rendering》中介绍了一种新的方法Gaussian Splatting，可以实现高质量的实时场景渲染。

作为3D生成领域具有突破性的一种新的表达和算法，3D Gaussian Splatting利用3D高斯模型来表示场景，通过优化和密度控制实现对场景的准确表达，并使用快速的可见性感渲染算法来加速训练和实时渲染。实验证明该方法在多个数据集上能够达到最先进的视觉质量和实时渲染效果。

### 3D Gaussian Splatting步骤

Gaussian Splatting的输入是一组经过SfM校准后的静态场景的图像以及SfM输出的稀疏点云。该方法整体如下图所示：首先从SfM得到的稀疏点云构建三维高斯函数，在训练过程中通过可微的快速渲染器对3D高斯函数的属性进行优化，并交替进行自适应密度控制。



Gaussian Splatting步骤示意图

### 3D Gaussian

3D Gaussian也叫三维高斯分布或三维正态分布，是统计学中一种特殊的多维正态分布，通常用来建模具有连续性随机性的现象，如图像处理、统计建模、机器学习等。

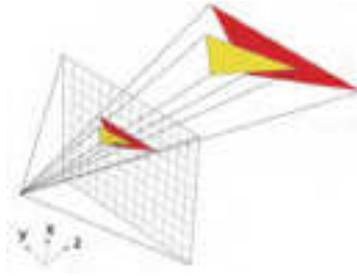
3D Gaussian能够涵盖空间中任意形状的椭球,包括平移、旋转。3D Gaussian在三维空间中定义了一个概率分布,表示如下:

$$G(x) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

因此只要确定了 $\Sigma$ 就可以确定一个椭球的形状。论文中使用的是各向异性 $\Sigma$ ,即在不同方向上具有不同方差值。各向异性协方差更有利于优化,因为它能够适应不同方向的变化,具有更精确的建模、更好的参数优化和更紧凑的表示。

### 光栅化

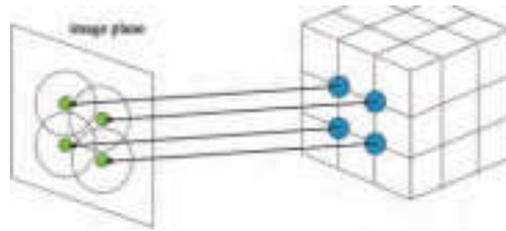
光栅化 (Rasterization) 是采用屏幕空间几何图形、片段着色器和该着色器的输入并将几何图形实际绘制到低级二维(2D)显示设备的操作,是实现计算机屏幕上图形显示和渲染的关键步骤。



光栅化示意图

### Splatting

Splatting是一种用于光栅化3D对象的技术。这些3D对象被映射到投影平面后得到的2D图形称为splat,类似于一个点、圆、矩形或其他形状,就像雪球打在墙上留下的印记,能量从中心向外扩散并减弱。这个过程可以在GPU上并行处理,因为每个Splat之间是独立的。



Splatting示意图

Splatting可以将三维空间中的点投影到二维图像平面上,这些投影的数据点以某种方式在图像上产生视觉效果,从而呈现在最终的渲染图像中。

## 5. 3D生成商业化落地关键

### 生成质量

3D生成的内容质量包括3D模型的精细度、模型的准确性、渲染的分辨率、色彩与光影的准确性、渲染对材质的表达等多个方面。

相比于传统3D建模,由于3D生成模型研究还较为早期,目前利用AI进行3D生成的质量仍然存在一定差距,不能完全满足大规模商业化需求。

### 生成效率

与生成2D图像不同,3D生成由于维度的增加,生成过程中所需计算量呈指数级上升。由于庞大的计算量,目前3D生成的效率较低,多数模型的生成效率难以满足实际应用中的要求。且模型在训练与推理过程中需要占用大量存储空间,因此目前多数3D生成成本较高。

### 可控性

对3D内容的可控性主要包括是否可以生成符合要求的3D内容、是否可以根

据要求对生成的模型与渲染效果进行修改。目前3D生成模型在修改方面表现不足。要满足修改要求的方法有两种,一种是继续加强模型对自然语言指令与图像指令的理解能力。但由于技术突破的不确定性,使模型可以与传统3D建模工作进行衔接的方法则更有可行性,这要求模型可以将3D建模工作与渲染工作分离,且生成的3D模型必须为网格数据mesh。

## 手机生成3D场景

Luma AI 是一家 3D 内容解决方案公司。通过Luma, 用户可以使用手机拍摄少量照片来生成和渲染出 3D 内容。

Luma 产品的主要功能有: 拍摄捕获 3D 图像和视频、图片转 3D、视频转 3D、文本转 3D 等, 用户可以通过 iPhone、网页以及 API 等载体或方式实现用 NeRF、GaussianSplatting等技术构建 3D 模型的功能。



网页版Luma通过自行上传照片、视频来进行三维重建, 网页版目前视频和图片 (ZIP压缩包) 体积最大限制 5GB。

文字转3D模型功能目前依然未全量开放, 只需输入文字描述, 即可生成对应的3D模型。

转换时间在30分钟左右。

# LUMA AI

## 一种文本指导的渐进式3D生成框架

### Dream Face框架



## 影眸科技

### ChatAvatar

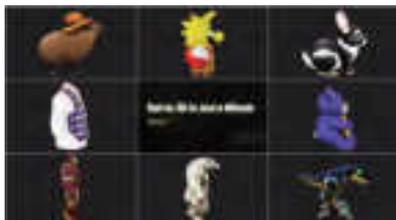
由影眸科技开发的一项新技术, 通过文本/照片/原画生成超逼真的3D面部资产和PBR纹理, 实现了文本到头像的技术。

ChatAvatar利用专有的扩散模型和Production-Ready Facial Assets数据集, 生成符合行业标准并可以直接在Unity等流行软件中使用的CG友好资产。

## 一分钟生成精细3D模型

全球领先的3D资产创作工具,覆盖了文本生成3D模型、图片生成3D模型和AI生成3D贴图,能有效简化复杂的模型制作流程并提升效率,为创作者提供更快、更高效的3D创作体验。

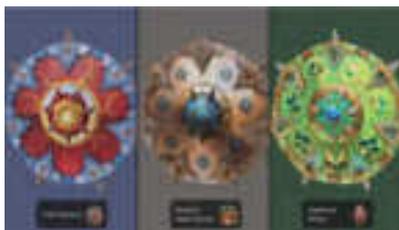
# Meshy



Text to 3D

Text to 3D新版本的特性:

1. 生成的物体品类更加丰富;
2. 避免了多张脸的问题,可以生成角色和动物等;
3. 生成的物体几何质量更高,更加完整。



Text to Texture

Text to Texture 新版本特性:

1. 更好的风格控制(支持写实、2.5D卡通、日漫、卡通线稿、写实手绘、东方水墨等7种风格);
2. 更高的贴图分辨率(可达4K)。

## 代表机构

影眸科技



孵化于上海科技大学,通过与国内顶尖人工智能、计算机视觉实验室合作,积极推动尖端实验室科研成果的民用化、商业化,探索前沿人工智能、计算机视觉技术在大众娱乐市场的推广应用。推出了穹顶光场毛孔级扫描服务与3天打造超写实数字人流程,以及HYPERHUMAN - ChatAvatar 3D角色生成平台。

Meshy



全球领先的3D资产创作工具,覆盖了文本生成3D模型、图片生成3D模型和AI生成3D贴图,能有效简化复杂的模型制作流程并提升效率,为创作者提供更快、更高效的3D创作体验。

凡拓数创



公司以“AI+3D”为技术发展方向,围绕3D可视化技术与数字多媒体集成技术等核心技术,加强对数字孪生技术及自研FT-E数字孪生渲染引擎的建设,紧抓数字创意产品的研发,目前公司在工业制造、市政等领域均有相关应用。

贻深数字

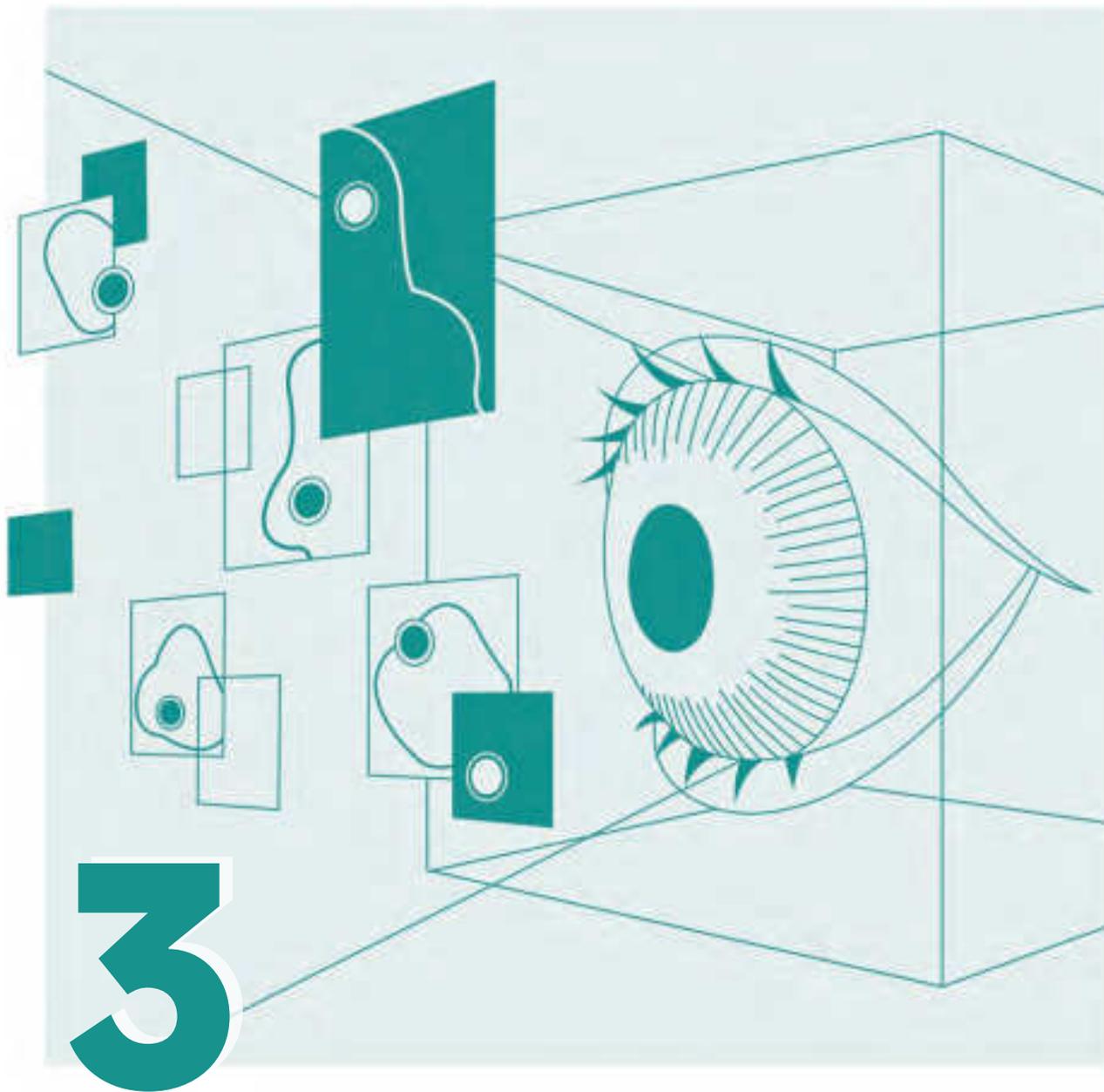


由上海科技大学智能视觉中心的博士生团队联合创建的公司,旨在利用以NeRF为代表的3D AI技术实现通用3D大模型。

生数科技



一家多模态生成式大模型与应用产品开发商。其最新提出的文生3D新算法ProlificDreamer,在无需任何3D数据的前提下能够生成超高质量的3D内容。今年9月正式上线了3D资产创建工具VoxCraft。



## 分割模型大一统： 计算机视觉即将迎来「GPT时刻」

分割一切模型 (SAM) 是Meta在今年推出的AI模型, 10月份相关论文获得ICCV23 的Best Paper Honorable Mention。受SAM影响, CV领域和相关研究在2023年“再度火热”。分割一切的技术在图形标注领域取得了突破性的进展, 这一跨时代的趋势对未来的效用值得期待。

# 推荐理由

## 1. 突破性的分割能力

在SAM出现之前,我们所能看到的其他图像分割模型,都是专有模型。如医学领域分割核磁图像、CT影像的AI模型,但相关模型的良好性能在其他领域无法体现。

SAM的最大贡献,即可以快速分割没见过的图像,将之前零散的图像分割模型统一。在深度学习领域,这类能力被称为零样本迁移,这也正是GPT4备受关注和CV从业者认为计算机视觉领域进入GPT3时代的原因。

“

Reading MetaAI's Segment-Anything, and I believe today is one of the 「GPT-3 moments」 in computer vision.

—NVIDIA Jim Fan

## 2. 迄今为止最大的分割数据集

SA-1B的效应不仅仅局限于创造了分割一切模型,基于该数据的模型可以轻松在其他领域进行零样本泛化。

“

我们构建了迄今为止(到目前为止)最大的分割数据集,在1100万张授权且尊重隐私的图像上拥有超过10亿个任务。

—Meta.AI

## 3. 其他分割模型的涌现

4月SAM发布后,CV领域再次火热,很短时间内涌现出许多基于SAM的二创及实践——

SAM发布后,来自IDEA研究院的刘世隆发布了零样本视觉应用Grounded-SAM,Grounded-SAM把SAM和BLIP、Stable Diffusion集成在一起,将图片「分割」、「检测」和「生成」三种能力合一;

15日,基于SAM的多目标跟踪模型VISAM发布,通过结合SAM的强大分割能力和MOT end-to-end跟踪算法的能力,能够在视频中跟踪多个目标的位置和运动轨迹,并利用SAM还原出每个目标的精细结构和轮廓;

随后浙江大学ReLER实验室进一步解锁了SAM的视频分割能力,开源项目SAM-Track (Segment-and-track anything,分割并跟踪一切);

威斯康星大学麦迪逊分校、微软和香港科技大学发布了一篇名为《一次性分割一切(Segment Everything Everywhere All at Once)》的论文。SEEM能够根据用户给出的各种模态的输入(包括文本、图像、涂鸦等),一次性分割图像或视频中的所有内容,并识别出物体类别;

新加坡国立大学Anything-3D借助SAM,能使分割后的任意物体转化为3D模型,达到“动起来”的效果;

5月初,香港科技大学、南洋理工等开源SAD (Segment Any RGBD,根据几何信息分割图像)的机器学习模型;

6月, 苏黎世理工同样联合港科大, 发布HQ-SAM (Segment Anything in High Quality)。保留SAM预训练模型权重的前提下, 引入最小的额外参数及计算, 设计出负责预测高质量掩码的输出令牌;

中科院开源了FastSAM模型, 能够以50倍速度达到与SAM相近的效果, 并实现25FPS的实时推理;

港中文贾佳亚团队提出的LISA大模型, 使得SAM可以顺畅地理解人类的复杂指令;

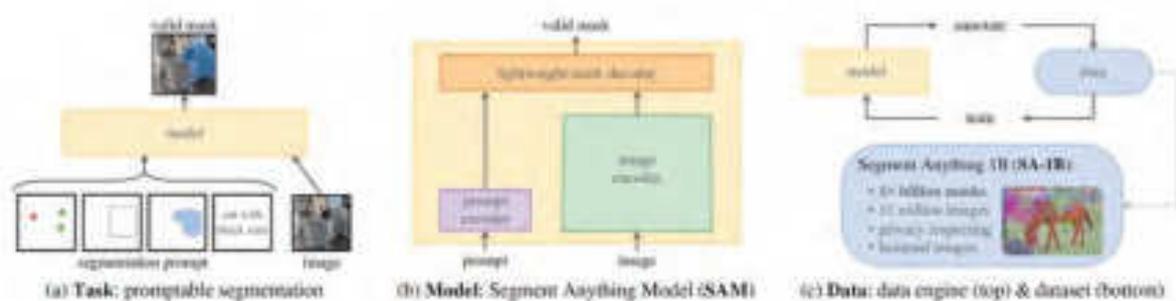
同时SAM还在其他领域拥有广泛的实践, SAM-Med3D是专门针对3D医学图像改进的分割一切模型。

## 技术原理

“

我们的目标是通过引入三个相互关联的组件来建立分割的基础模型: 快速分割任务、为数据注释提供动力并通过提示工程将零样本转移到一系列任务的分割模型 (SAM), 以及用于收集SA-1B的数据引擎, 我们的数据集超过10亿个掩码。

—Meta.AI Research FAIR



SAM框架

Segment Anything (SA) 总的来说是一个包含图像分割新任务、模型和数据集的项目, 这三项内容也正是我们可以用来理解整个模型的维度。

### 任务: 诞生于NLP的提示工程与零样本泛化

SAM的研究团队表示, 他们是从NLP模型中获得的灵感。NLP通过预训练的方式训练基础模型, 同时通过提示工程解决之后的任务。

#### (1) Prompt

SAM中的“分割”就相当于NLP中的Prompt。Prompt可以是一组前景/背景点、一个粗略的框或掩码、自由形式的文本, 或是能指示如何对图像进行分割的任何信息。

SAM中受提示的分割任务即是在给定任何提示的情况下生成「可用」的分割掩码。这一要求类似于期望用模糊的表达让语言模型生成答案。这里的「可用」意味着即使提示不明确或涉及多个对象, 也能输出至少一个合理的掩码。

比如在Meta所给出的示例中, 每列展示出SAM从一个不明确的点提示 (绿色圆点) 生成的三个有效掩码。



SAM模糊生成掩码

## (2) 预训练

基于Prompt的分割任务提出了一种自然的预训练算法。该算法对每个训练样本模拟一系列Prompt（点、框、掩码），并将模型的预测与真实情况比较。



Segment Anything Model (SAM): Meta AI 的新AI模型, 只需单击一下即可「剪切」任何图像中的任何对象。

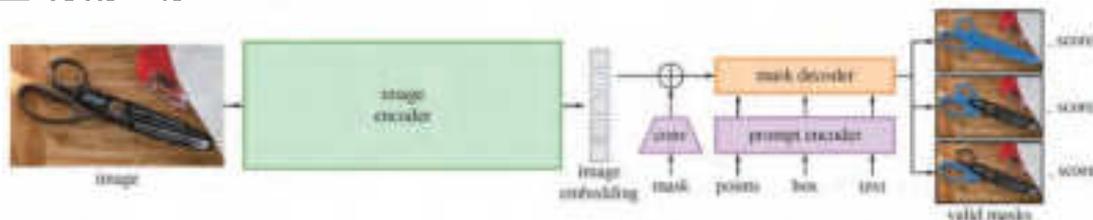
SAM是一种快速分割系统, 可以对不熟悉的对象和图像进行零样本泛化, 无需额外的训练。

—Meta.AI

## (3) 零样本

预训练任务模型能够在推理时对任何提示做出适当响应, 实际的下游任务中, 可以设计适当的提示来解决。如果存在一个能识别猫的检测器, 识别出的猫的框架可以输出为SAM的提示, 来分割猫的不同姿势。

## 模型: 分割一切



SAM的结构

SAM本身建立在Vision Transformer (ViT) 模型的基础上, 同时对实时性能进行了特定的权衡, 具体包括了图像编码器、提示编码器和快速掩码解码器三个组件。

### (1) 图像编码器 (Image encoder) 计算图像嵌入

利用MAE (Masked Autoencoders) 预训练的ViT, 最低限度适应高分辨率的输入, 该编码器在提示编码器之前, 对每张图像只运行一次。

### (2) 提示编码器 (Prompt encoder) 嵌入提示

### (3) 掩码解码器 (Mask decoder) 预测分割掩码

将以上两个信息源组合在一个轻量级掩码解码器中, 来预测分割掩码。

比较重要的点在于, 为了解决输出模糊性问题, 预测输出多个掩码, 在训练过程中, 只回传最小的损失函数。当输入多个提示时, 生成的掩码会比较接近, 此时模型会只预测一个掩码 (当单个提示时不用)。

后图显示了SAM的分割成果, 示例图像来自SAM的数据集SA-1B。这些掩码由SAM自动标注, 每张图像中大约有100个掩码, 模型会根据掩码的数量进行分组。

## 数据:拥有10亿任务的数据集SA-1B

Meta做了一项非常具有奠基性质的工作。由于互联网上的分割掩码并不丰富,Meta构建了一个数据引擎来收集包含11亿掩码的数据集SA-1B。

数据引擎的功能实现有三个阶段:

### (1) 模型辅助的手动注释

以人力对基于浏览器的交互式分割工具进行掩码的标注,注释者可以命名及描述细化后的任何图像。

SAM在前期一直使用公共的分割数据集进行训练,直到收集足够的数据及符号。随着模型的改进,每个掩码的平均注释时间从34秒减少到14秒。14秒比COCO的掩码注释快6.5倍,仅比使用极值点的边界框标记慢一些。

### (2) 混合自动预测和模型辅助

为了提高模型分割任何东西的能力,Meta将注意力集中在不太突出的对象上。具体方法是先向注释器展示了预先收集了掩码的图像,并要求它们注释任何额外的未注释对象。

### (3) 全自动

最后,SAM可以在没有注释器输入的情况下生成掩码。首先是已经收集了大量的人工及半自动掩码来改进模型;其实是开发出了模糊感知模型,使得SAM可以预测有效的掩码。

简单来讲,Segment Anything Data Engine完全采用了大语言模型的训练过程和方式,使得SAM到目前的阶段已经可以分割一切了。



SAM模糊生成掩码

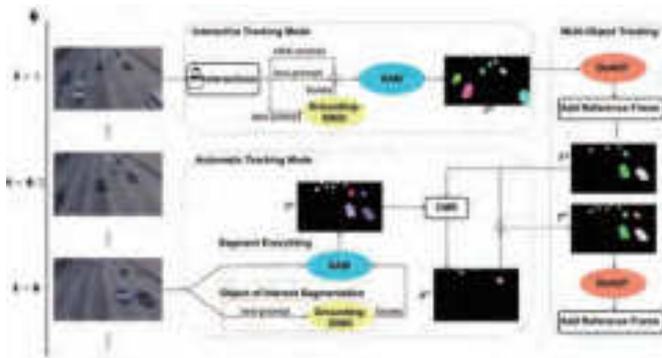


# 浙江大学

## Segment And Track Anything

SAM在视频图像分割中,不会考虑到「帧间的时间连贯性」,因此浙江大学ReLER实验室将DeAOT与SAM结合,提出了适用于视频分割的模型SAM-Track。

- DeAOT: 基于AOT的多对象跟踪模型,可以给定首帧物体标注,并对视频其余帧中物品追踪分割;
- Grounding-DINO: 模型使用的另一种框架,使其可以基于自然语言交互。



—The Pipeline of SAM-Track

SAM-Track的两种跟踪模式:

- (1) 交互式跟踪  
仅在视频的第1帧中获得注释
- (2) 自动跟踪  
第1帧后每隔第n帧调动一次

## 代表机构



- 《Segment Anything Model》



- 《Segment and Track Anything》



- 《Segment Everything Everywhere All at Once》



- Anything-3D



- 《Segment Any RGBD》



- 《Segment Anything in High Quality》



- FastSAM



- LISA



## 具身智能带来新想象： AGI终极场景下的全新终端

今年是人形机器人和大模型同步进发的一年，具身智能 (Embodied AI) 已成为AI研究新热点。自从2021年马斯克在特斯拉AI Day上宣布要开发人形机器人，全世界范围内涌现出各种突破性的仿生机器人产品。在国内：大型科技公司、制造业企业跨界研发仿生机器人；CyberOne、XR-4等全尺寸人形双足机器人产品横空出世；高校团队成立诸多仿生机器人初创企业。大模型的影响之下，具身智能将迈入新阶段。

# 推荐理由

## 1. 关键性技术取得突破性进展

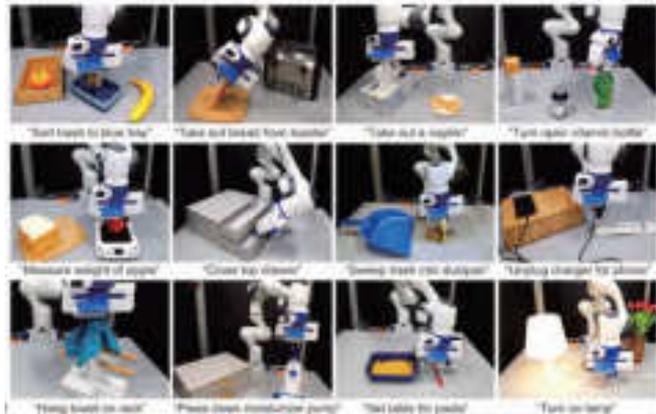
量子位智库首先观察到,在2023年,具身智能领域几项关键性技术都取得了突破性进展。

### (1) 强化学习在具身智能领域的新突破

强化学习是目前具身智能研究的主流思路。2021年,来自UC伯克利的Pieter Abbeel就以强化学习与机器人相结合的工作获得了ACM计算奖,其开创的信任区域策略优化、广义优势估计等方法,被ACM称赞为“为下一代机器人技术奠基”。

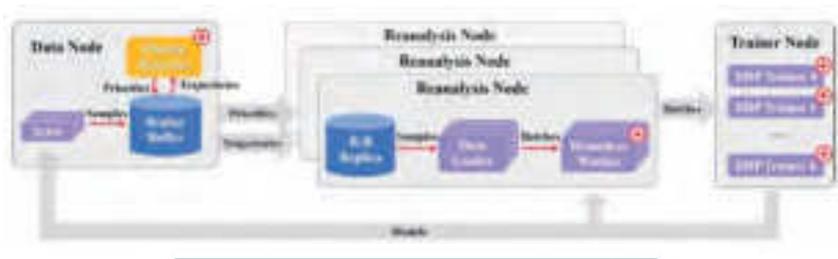
关于这项工作,不可不提来自斯坦福大学的李飞飞教授。她和团队开发了两个训练框架RoboTurk和SURREAL,前者解决机器人学习中的数据问题,后者解决通过数据进行强化学习训练和结果评估的问题。

2023年6月,李飞飞团队发布了具身智能新成果VoxPoser。VoxPoser使用LLM+VLM (Vision-Language Models),能从3D空间分析出目标和环境障碍,帮助机器人实现行动规划,让真实世界中的机器人在未经「培训」的情况下,直接执行任务。



VoxPoser

清华大学梅毅轩等人在2023年提出的SpeedyZero( SpeedyZero: Mastering Atari With Limited Data And Time)算法,就是一项提升现实世界样本效率的工作。基于EfficientZero (一种采样高效的强化学习算法, Mastering Atari Games with Limited Data)的新算法,其工作效率是DQN(Deep Q-Network)算法的500倍。



SpeedyZero架构

### (2) 大模型接入机器人

同样在2023年6月,腾讯Robotics X机器人实验室公布了智能体研究的最新进展。通过将预训练AI模型和强化学习技术应用到机器人控制领域,腾讯机器狗Max大幅提升了灵活性和自主决策能力。

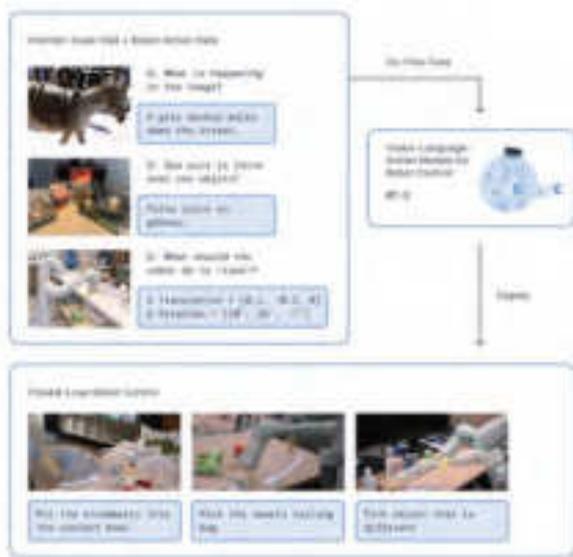
“

具身智能有望让机器人从「自动化的泥沼」中走出来,真正成为人类的伙伴。在大模型的帮助下,机器人能够自主感知环境、学习技能、不断迭代决策质量,在一些特定问题中已经展现出不错的解决问题的能力。这极大提升了机器人创造价值的潜力和范围,也让更复杂的机器人硬件形态可以被接受(收益高于成本)。

——源码资本执行董事 陈润泽

Google DeepMind在今年发布了新的视觉语言模型Robotic Transformer 2(RT-2)。RT-2可以自动从互联网及机器人数据中学习,并将学习成果转化为机器人控制的通用指令,以实现一系列操作动作。RT-2还可通过思维链(Chain-of-Thought)进行多端语义推理,例如推荐哪种类型的饮料最适合疲倦的人。

通过互联网及机器人数据预训练的VLM模型,可以接受机器人摄像头传达的图像,并预测机器人即将执行的动作。



RT-2的训练和架构

“

RT-2是一个基于Transformer的模型,根据网络上的文本和图像进行训练,可以直接输出机器人动作。就像语言模型通过网络文本进行训练来学习一般思想和概念一样,RT-2也可以从网络数据中传输知识来告知机器人行为。

换句话说,RT-2可以与机器人说话。

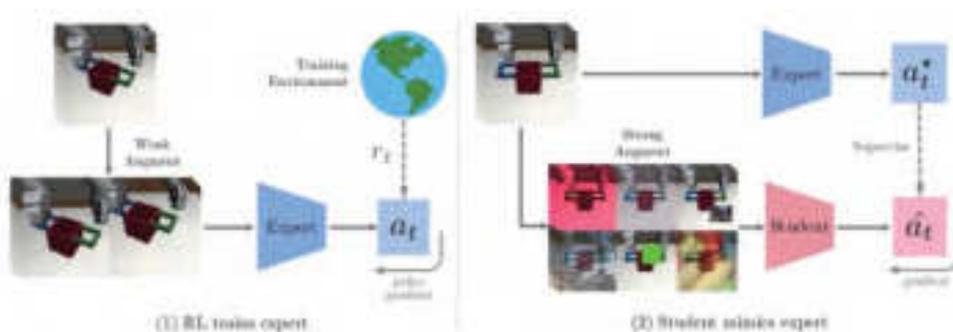
—Google DeepMind科学家 Vincent Vanhoucke

感知,是具身智能的重要部分。它将计算机视觉与机器人相结合,教会机器人通过“看”去分辨和操纵周围的事物,这也是仿生类机器人区别于传统机器人的重要方向。

### (3) 零样本泛化

李飞飞团队的一项最新研究——SECANT模型,正在攻克零样本泛化问题。SECANT模型是一种可以适应新测试环境的自专家克隆方法。

这个方法在两个阶段利用图像增广,分离鲁棒性表征和策略优化,从而实现从未见过的视觉环境的强大零样本泛化性能。



SECANT

## 2. 创业团队涌现,投资活动频繁

人形双足机器人未必是具身智能的唯一形态,但具身智能拥有和人相似的躯干和大脑,最符合人类习惯和情感。

我们发现，许多工业机器人、协作机器人和其他仿生机器人团队在2023年开始研发人形（双足）机器人。并且，具身智能未来发展的可能性受到学界和创投界的支持。

在此期间，足式机器人公司逐际动力完成了总金额近2亿元的天使轮和Pre-A轮融资，融资将用于促进机器人与AI结合，同时帮助其公司研发人形双足机器人。

“

我们认为，未来3~5年，在一些执行技能相对简单的领域，具身智能会逐渐进入到商业化阶段，并拉动供应链向生产低成本的通用机器人大幅迈进。

——源码资本执行董事 陈润泽

吉林大学任露泉院士团队创立的人形机器人公司月泉仿生，宣布完成千万级人民币的天使轮融资。该公司的科研团队正在主攻高功率密度微型电机、仿生人工肌肉作动器、柔性力传感器和人工韧带等人形机器人核心部件。

由清华大学交叉信息学院助理教授陈建宇创立的团队星动纪元，也是一家专注于人形双足仿生机器人的创业公司。

“

我们对具身智能商业落地的评估聚焦在四个环节，即安全性、价格、作业泛化能力和稳定性。相比C端落地，我们认为在B端，因有明确的成本上限，更直接的销售逻辑，对机器人表征，如声音、重量等有更高的忍耐。B端现阶段可从弱人机共存场景、高度结构化场景、简单重复场景切入，其在简单场景、简单任务的环节更容易落地。

——蓝驰创投

量子位智库在今年发布的仿生机器人相关报告中提到，2023年前后十年是具身智能及仿生机器人飞速发展的十年，相关产品首先会在服务业、娱乐业、工业和航天业中有所表现，并进入最复杂的非结构化环境——家庭场景。

作为被我们信任的，继电话、个人电脑、智能手机之后的下一代终端，具身智能将创造出独特的需求。



中国仿生机器人产业  
全景报告

“

未来的机器人不仅具有人工智能的能力，还有与周边环境进行互动、适应和感知的能力。其商业价值可以体现在居家养老、智能化工厂、社区服务等场景，具有很高的成长潜力。

——西安电子科技大学教授 王皓

## 技术原理

1950年，艾伦·图灵在论文《Computing Machinery and Intelligence》的结尾提到：我们也许希望机器能够做到人类的智力能做到的任何事。从哪一个方向开始最好呢？许多人认为从抽象的举措，比如下棋开始好；还有人认为，为机器提供金钱能买到的最好的传感器，然后教它理解和说话，可以模仿对孩子的正确教学会更好。

“

We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English.

——艾伦·图灵《Computing Machinery and Interligence》

我们可以认为这是对“具身智能”最初的阐释。

通用人工智能(Artificial General Intelligence, AGI)一直被认为是人工智能的终极梦想。从去年末ChatGPT的爆发开始,图灵的第一个设想以从未有过的速度飞跃式发展。大模型不完美的当下,已经有更具前瞻的想法,尝试为充满智慧的AGI寻找一个合适的「身体」。为了做到人能够做到的事,帮助人完成生活中会遇到的非固定任务,和人类有着相似样貌的人形机器人在今年受到了大家广泛的关注。对具身智能的关注已经突破了对人形外表的追求,转向与AI的融合。

具身智能(Embodied Artificial Intelligence, EAI)被认为是智能与物理身体相结合的系统。上海交通大学教授卢策吾和北京大学助理教授王鹤提到,具身智能可通过智能体与环境的交互,来获取信息、理解并行动。

以人类的身体机能为参照类比具身智能,理想中的具身智能体应当具备感知、理解、阐述、创造的四大能力。我们使用各类传感器来帮助具身智能识别、判断物体、环境;通过大模型和TTS(语音合成技术),具身智能可真正与人类交流;利用执行器,它可以替代或帮助人类完成工作。

“

我们对具身智能商业落地的评估聚焦在四个环节,即安全性、价格、作业泛化能力和稳定性。相比C端落地,我们认为在B端,因有明确的成本上限,更直接的销售逻辑,对机器人表征,如声音、重量等有更高的忍耐。B端现阶段可从弱人机共存场景、高度结构化场景、简单重复场景切入,其在简单场景、简单任务的环节更容易落地。

——蓝驰创投

以人类的身体机能为参照类比具身智能,理想中的具身智能体应当具备感知、理解、阐述、创造的四大能力。我们使用各类传感器来帮助具身智能识别、判断物体、环境;通过大模型和TTS(语音合成技术),具身智能可真正与人类交流;利用执行器,它可以替代或帮助人类完成工作。

具身智能的一条主要研究思路,就是在虚拟物理世界设计和开发具身智能算法,并将其迁移到真实世界,这被称为Sim2Real(simulation to reality)。在模拟环境中习得的最优策略往往不能在现实环境中使用(reality gap),因此尝试利用Sim2Real来解决这一问题。

目前已经有许多从事具身智能研究的虚拟仿真平台,如AI Habitat等。正如人工智能存在数据集、基础任务等,具身智能的研究同样需要物体及场景数据集、基本任务、学习框架。

“

未来的机器人不仅具有人工智能的能力,还有与周边环境进行互动、适应和感知的能力。其商业价值可以体现在居家养老、智能化工厂、社区服务等等场景,具有很高的成长潜力。

——西安电子科技大学教授 王皓

具身智能的另一条研究思路是在真实世界中采集具身交互数据和学习,但存在采集数据少,采集成本高昂等问题;同时,机器人在真实世界进行实践操作存在一定的危险性,因此暂未规模化。

在发展更加先进的仿真环境和Sim2Real迁移技术,以及在提高具身智能体学习效率、性能、泛化性方面,具身智能依然存在巨大挑战。



Learning Dexterous In-Hand Manipulation  
——OpenAI在sim2real领域的一项开创性工作

- (1) 未来具身仿真环境需要突破更多物体类型的物理交互和传感模拟,这是学习和训练具身智能体的重要平台。
- (2) 仿真到真实之间在外观、物理和控制等层面存在差异,如何更好地解决reality gap仍是重要研究方向。
- (3) 目前在具身智能领域,大模型还远未达到在非具身智能领域的性能表现,包括零样本泛化、基于提示工程的多任务学习、下游任务微调等。

“

具身智能要实现的决策与环境复杂,具有很高的挑战性。数据是限制其发展的主要原因,目前具身智能尚未找到很好的下游落地场景,行业数据较少,或者数据壁垒比较高,在传统数据上得到的模型并没有很强的泛化能力和环境交互能力。其次,目前的建模方式还在初始阶段,智能体还不能很好地利用和理解时序、3D、物理等各种模式,以及解决行走、灵巧操作等控制问题。目前具身智能机器人已初现市场应用的迫切需求。

——上海人工智能实验室

## 「智慧就是适应变化的能力」穹知系统

穹知系统为非夕机器人AI团队研发的通用人工智能系统,是云-边-端协同一体的面向全行业的通用智能机器人“大脑”系统,能够基于机器人视觉、力觉、触觉等多信息感知进行AI算法研究和应用产品开发。

### 云-边-端协同的智能平台框架



# FLEXIV

非夕科技

### 核心工业AI解决方案



2D物体位姿识别



振动检测



通用识别

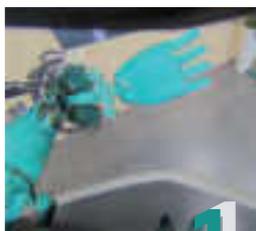


3D物体位姿识别

## 可纠正错误的Optimus



TESLA



最新发布资料片显示，Optimus已经可以自行矫正手和脚的位置，仅靠视觉和关节位置编码器，它就能在空间中精确定位四肢的位置。

可以利用纯视觉分辨物体的颜色、形状、位置，并下达指令按颜色将物体区分开来，受干扰时可以矫正。



Optimus还展示了自主纠正行为。和去年相比手部操作更加精细，视觉内对机器人自体的扫描边缘更加清晰了。



## 代表机构

非夕科技  FLEXIV

通用智能机器人公司，专注于研发、生产集工业级力控、计算机视觉和人工智能技术于一体的自适应机器人产品。非夕科技首席科学家卢策吾，上海交通大学教授，曾在斯坦福大学进行博士后研究。研究兴趣主要集中在计算机视觉、深度学习、深度强化学习和机器人视觉领域。

逐际动力  LIMX DYNAMICS

通用足式机器人公司，专注于运动智能 (Motion Intelligence) 与足式机器人 (Legged Robot) 的研发和制造，产品包括人形双足、四 (轮) 足机器人及相关软硬件解决方案。采用 Model-based 和 Learning-based 的混合方式，聚焦人形双足机器人在地形感知与复杂地形适应能力的突破。

星动纪元  Robot Era

清华大学交叉信息研究院孵化，研发具身智能以及通用机器人技术和产品的新兴科技公司。聚焦于通用人工智能前沿应用，致力于研发适应宽领域、多情景、高智能的通用人形机器人。



## 端到端自动驾驶成共识： BEV+Transformer重构技术路线

随着生成式AI的发展，以ChatGPT为代表的大语言模型的泛化能力有了极大提升，端到端自动驾驶技术也因此受到广泛关注。端到端自动驾驶技术以全部模块神经网络化为特征，对规则的依赖度低，具备智能涌现能力和跨场景应用潜力。

在学术界，今年的CVPR最佳论文颁给了端到端自动驾驶，这几乎代表着一种共识：端到端自动驾驶是行业的未来。在产业界，以特斯拉FSD为代表的自动驾驶系统，表明神经网络算法和AI大模型的加持已经成为自动驾驶技术落地的新路径。

## 推荐理由

### 1. BEV+Transformer 提供高质量感知数据, 推动端到端架构落地加速

驾驶系统在BEV的视角下统一完成感知和预测任务。BEV+transformer形成的高质量3D感知结果, 为后续端到端架构中的大模型计算提供了优质的原始数据。端到端架构的感知决策一体化, 由数据采集原始数据到直接输出驾驶命令, 因此输入的原始数据对最后驾驶命令的生成与执行起到至关重要的作用。相比于直接使用传统CNN卷积网络处理视觉系统采集的2D图像, BEV+Transformer能够直接生成的高质量3D数据, 包含了更多有用的信息, 能够生成更加合理的驾驶命令。

### 2. 更加直接的端到端架构提高驾驶系统整体性能

传统模块化架构的流水线处理机制, 如果前级模型输出的结果有误差, 就会影响下一级模型的输出, 导致整套系统的系统出现偏差。而由于端到端自动驾驶系统更加直接和整合, 输入传感器信号后可以直接输出车控信号, 降低出现误差的概率, 提高了整体系统性能上限。

### 3. 简化系统架构, 降低中间环节标注成本

相比过往的模块化智能驾驶, 端到端自动驾驶以一个模型完成多个模型组合需要实现的分类、追踪、定位等环境感知和规划决策任务, 减少系统中的中间处理步骤, 简化了整个自动驾驶系统的架构。研发人员只需要对一个模型进行整体训练和迭代优化即可实现性能提升, 可以集中资源, 实现功能聚焦, 提高可靠性并降低维护成本。

### 4. 在未知场景下表现出优秀的泛化能力

在未来, 大模型驱动的智能体极有可能带来交互方式的变革, 从过去用户适应软件应用, 变成应用软件适应用户个人习惯, 为用户的生活提供更加便利的服务。

“

端到端自动驾驶系统中, 最终的规划轨迹或控制信号是通过神经网络直接输出。这可以带来两个好处: 一是通过数据驱动代替规则, 减少人工维护规则的需要, 使系统更易于维护; 二是由于模块之间可以通过非结构化的影视特征进行信息交流, 传递的信息量上限比目前人工定义的协议更丰富, 使得整个系统的表现的体验上限可能会更高。

——地平线

“

百度认为大模型将走向多模态, 进一步重构物理世界, 自动驾驶就是大模型重构物理世界的一个案例。大模型将让全无人自动驾驶真正落地。目前, 百度自动驾驶出行服务平台萝卜快跑累计提供服务超过400万次, 是全球最大的自动驾驶出行服务商。

——2023百度世界大会

## 技术原理

### 1. 端到端自动驾驶的原理

端到端自动驾驶利用海量数据进行训练, 将训练好的神经网络部署在一个接受输入数据的模块里面, 就可以直接输出该任务的最终结果。端到端模型将感知、规划和控制一体化整合在一个模型中。

传统的自动驾驶系统采用模块化部署策略, 其中每个功能, 如感知、预测和规划, 都是单独开发并集成到车载车辆中。

目前已经量产的自动驾驶车型, 大多数采用了模块化架构。「模块化」顾名思义, 就是把自动驾驶拆分成一个个典型任务, 然后将这些任务交由专门的AI模型或简单的规则设定来处理。比如说感知、预测、规划等等。

模块化架构的特点是复杂的规则设计，在解决大量驾驶情况时规则设计往往无效。而人工制定的规则除了在使用体验上无法令用户适应，也难以覆盖所有复杂的场景案例。

另外，在一个传统的自动驾驶系统里，包含多个模型，每个模型都要专门进行训练、优化、迭代。随着模型的不断进化，参数量提高，所需的研发人员也水涨船高，研发投入极高。而端到端自动驾驶通过一个模型实现多个模型的功能，研发人员只需要针对一个模型进行训练、调优，在缩减人员成本的同时，也能够提升效率。

## 2.感知层的变化

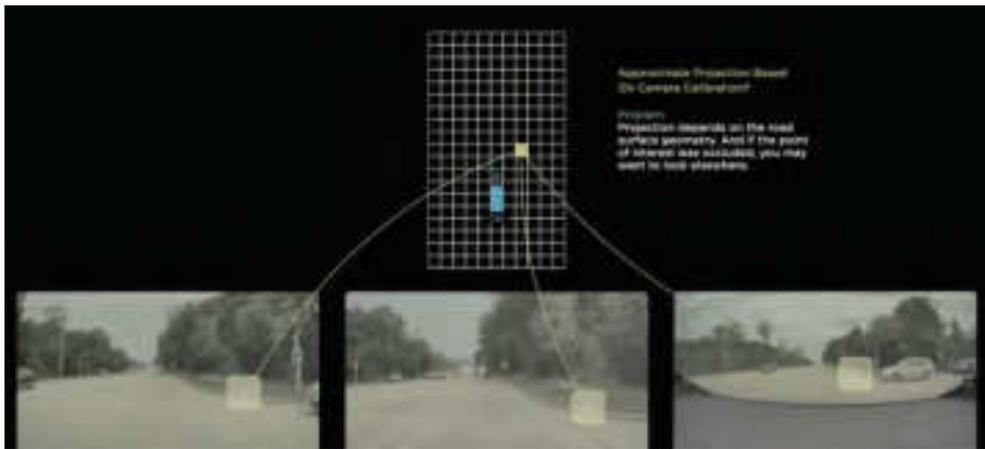
自动驾驶的一个整体趋势：「重感知，轻地图」

在感知层面，端到端自动驾驶一定是基于3D感知，其中用到的一项关键技术为BEV。

### (1) BEV

BEV的全称是基于多视角摄像头的鸟瞰图感知 (Bird's-eye-view Perception)，是指从多个摄像头获取图像信息，并将其融合成一个全面的、统一的视图，从而对车辆周围的环境进行准确的感知和判断。

基于BEV，延伸出了很多细分技术，都是为了能够实现端到端自动驾驶方案量产落地。



特斯拉BEV视图

#### 占用网络 (Occupancy Networks)

特斯拉在2022年AI Day上发布Occupancy Networks(占用网络)，实现BEV从2D到3D的转化。具体来讲，基于占用网络能够实现在线3D重建。首先，将世界划分为一系列网格单元，然后定义哪个单元被占用，哪个单元是空闲的。通过预测3D空间中的占据概率来获得一种简单的三维空间表示。

占用网络主要从以下方面提升车辆对障碍物的感知：

- (1) 通过动静态物体统一的障碍物感知方式，能够大幅减少corner case；
- (2) 摆脱检测框的约束，对不规则障碍物的感知能力增强。对障碍物感知的增强，能够提升自动驾驶过程中的安全性；
- (3) 在自动驾驶实现的传感器层面，占用网络理论上可以替代补盲冗余作用的激光雷达，使系统硬件成本大幅下降。

#### 4D标注 (4D-Label)

4D-Label技术主要是在3D空间的基础上，加入时间维度的标注过程。3D标注只能捕捉物体的静态信息，而4D标注可以记录物体在时间轴上的运动和变化。一般通过BEV视角观察物体，再在3D空间中对车辆、行人、交通标志进行4D标注，可以准确标注。

物体的位置、大小、形状以及运动轨迹,精准完成物体检测和路径规划等任务。

传统的人工标注方式需要大量时间和人力资源,且易出现主观误差,而4D标注的技术路线主要通过4D重建实现点云级别或object级别的重建,通过人工标注积累原始数据。随着数据积累到一定程度,可以训练云端大模型替代人工标注,能够提升80%的标注效率。

## (2) BEV给车企带来的变化

- ① 自动驾驶从堆砌配置变为缩减配置BEV可以实时生成语义地图,进而替代高精度地图,甚至还可以去掉超声波雷达。
- ② 需要芯片厂商更好地适配,与车企形成深度合作一套标准的BEV算法,现在的很多芯片都不支持,大家在用各种各样的算力去替换它们,这就需要芯片厂商和主机厂深度合作。

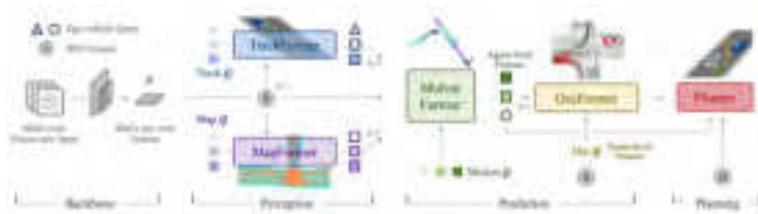
## 3.端到端自动驾驶面临的挑战

端到端自动驾驶采用的transformer架构,与生成式AI的大语言模型面临类似的困境——可解释性差。可解释性差意味着端到端模型在遇到难以处理的场景时,开发人员难以精准快速确定问题所在,也就导致目前阶段,端到端自动驾驶模型的通用泛化性较差。

解释性、泛化性差,会让自动驾驶产品在实际落地过程中出现安全隐患,而安全问题是自动驾驶领域首当其冲需要克服的难点。目前在这个问题上,尚未出现好的解决方案,只能通过模型的不断迭代来提高准确性。

此外,端到端自动驾驶需要学习的是连续驾驶行为,而非以帧为单位的图片。因此需要用大量标注驾驶行为的视频进行训练。当前这类视频数量有限,且主要在仿真环境中进行训练和验证,难以实现闭环验证。这也影响了端到端自动驾驶的量产落地。

### —Planning-oriented Autonomous Driving



CVPR 2023

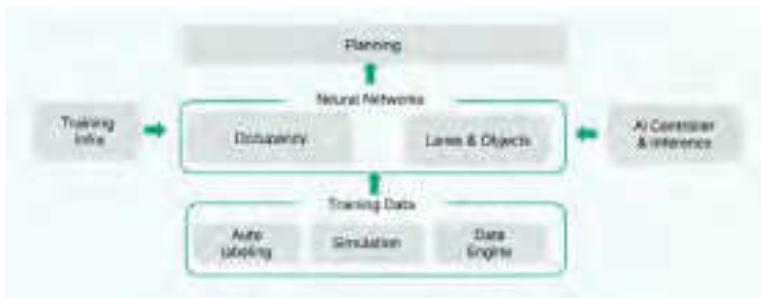
最佳论文

论文中提出“感知决策一体化”的自动驾驶通用大模型UniAD, UniAD包含四个基于transformer解码器的感知和预测模块,以及最后的规划器。

多视角图像通过BEVFormer进行特征提取,获得BEV特征B;TrackFormer中,通过track queries的可学习嵌入,从B中查询物体的信息来进行检测和跟踪。MapFormer中,使用map queries从B中查询道路元素(例如车道和分隔线)的信息,并执行地图的全景分割。MotionFormer捕获B+物体+地图之间的交互信息(作为k和v),并预测每个物体的未来轨迹,由于场景中每个物体的行动可能会对其他物体产生重大影响,该模块为所有考虑的物体做出联合预测。OccFormer使用BEV特征B作为查询,物体motion信息作为k和v,并且预测物体未来occupancy的情况最后,Planner使用来自MotionFormer的表达性自车查询来预测规划结果,并通过避开OccFormer预测的占用区域来避免碰撞。

## 特斯拉 FSD V12系统

FSD V12系统都是通过AI实现,99%的决策都交给神经网络给出,通过将高质量的视频数据来训练AI模型,视觉输入,控制输出。面对从未见过的建筑、道路标志,特斯拉也能够顺利通过。



特斯拉FSD系统架构

特斯拉在数据、算法、算力层面打造了一套包含感知、规控、执行在内的全链路自动驾驶软硬件架构FSD。特斯拉端到端方案将神经网络作为核心,以BEV+Transformer+Occupancy为主流路径,去除了各模块基于规则的代码,将传感器收集到的信息直接输入神经网络,经过处理后能够直接输出自动驾驶指令,使得整个系统端到端可导。

# TESLA

## 代表机构

国内车企使用BEV,目前主要用于城市辅助驾驶(NOA)的快速落地。BEV+Transformer已成为自动驾驶算法主流趋势,目前包括比亚迪、蔚小理、智己等车企,以及华为、百度 Apollo、毫末智行、地平线、轻舟智航、觉非科技、商汤科技等自动驾驶企业均已布局 BEV+Transformer。

**地平线**  地平线  
Horizon Robotics

地平线成立于2015年,是行业领先的高效能智能驾驶计算方案提供商。地平线于2021年7月推出专为高阶智能驾驶打造的车载智能芯片——征程5,单芯片算力达128 TOPS,支持如 BEV 等领先智能驾驶算法模型的应用部署。地平线自动驾驶通用算法框架UniAD,建立了以全局任务为目标的端到端网络框架,为自动驾驶技术与产业的发展提供了新的方法与思路。

**理想**  理想

辅助驾驶系统ADMax3.0,从技术上解决对物理世界的还原和对交通参与者的行为预测,使智能驾驶车辆能够像人一样感知、规划、决策,实现静态BEV道路结构还原、动态BEV交通参与者还原、Occupancy道路占用的物理世界还原、预测模型与实时规划等功能,并使用NeRF技术增强Occupancy网络算法使之实现更高的精度和细节。

## 小鹏



XNet系列是小鹏实现量产的BEV感知技术架构,也是XNGP辅助驾驶系统的核心技术模块。XNet通过将多个摄像头采集的数据,进行多帧时序前融合,输出BEV视角下的动态目标物的4D信息(如车辆速度、运动预测等)和静态目标物的3D信息(如车道线位置等),利用神经网络实现端到端数据驱动算法迭代,大大提升车辆的智能辅助驾驶能力,尤其是面对城市复杂场景时的环境感知、动态物体运动轨迹预判、决策、执行等能力。

## 小马智行



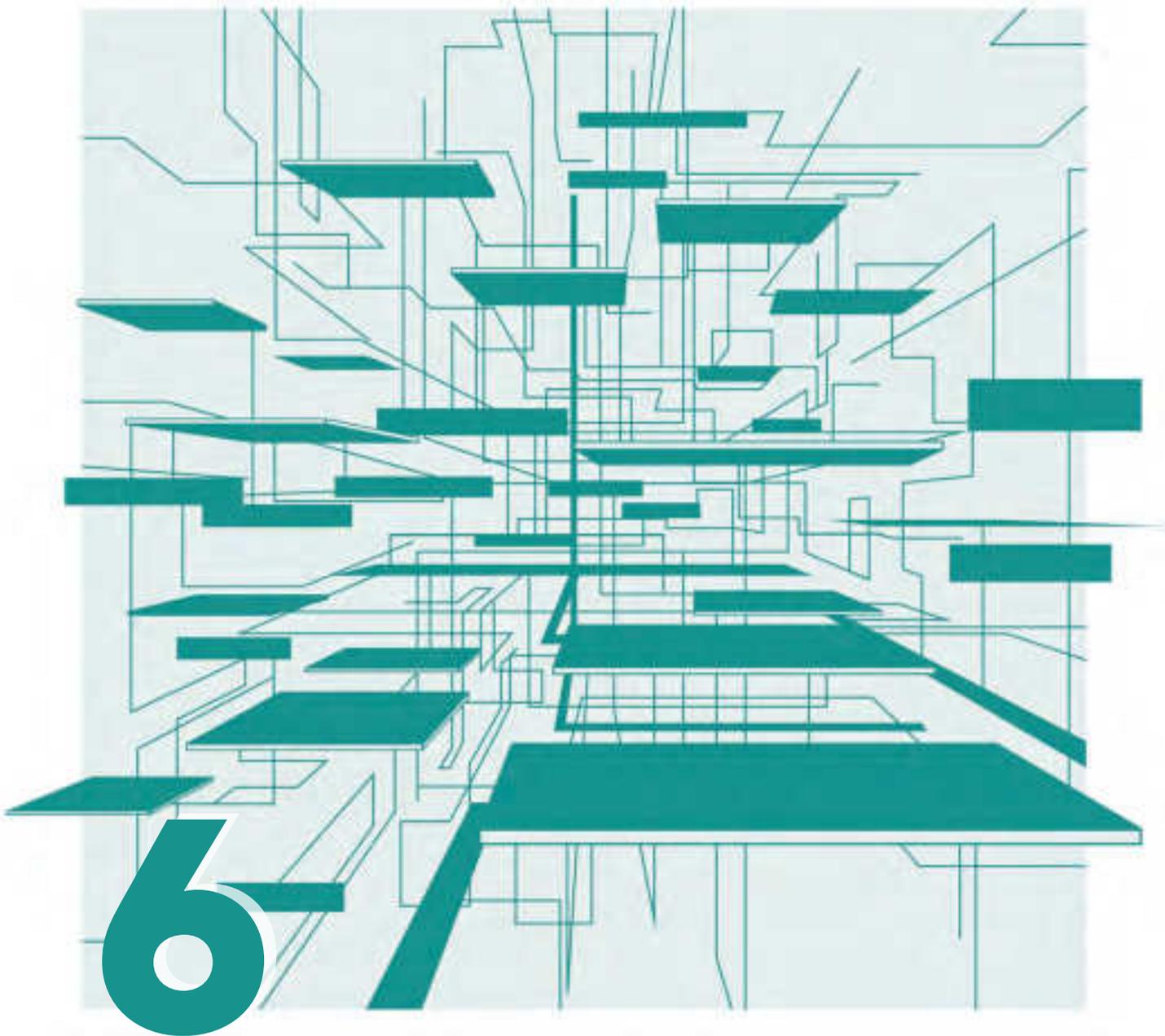
自研BEV感知算法,大模型识别各类型障碍物、车道线及可通行区域等信息,最大限度降低算力需求,同时在无高精地图的情况下,仅用导航地图也可实现高速与城市NOA功能。

在高速场景中,小马智行利用传感器,用鱼眼相机参与行车BEV感知模型,减少对传感器数量的依赖——使用6个摄像头(4个鱼眼相机及前后向各1个长距相机)与1个前向毫米波雷达。

## 毫末智行



毫末智行的DriveGPT“雪湖·海若”,基于BEV视觉方案重塑智驾技术路线。DriveGPT输入感知融合后的文本序列,输出的是自动驾驶场景文本序列,即将自动驾驶场景Token化,形成“Drive Language”,最终完成自车的决策规控、障碍物预测以及决策逻辑链的输出等任务。具体实现过程是,DriveGPT雪湖·海若在预训练阶段通过引入量产驾驶数据,训练初始模型,再通过引入驾驶接管Clips数据完成反馈模型(Reward Model)的训练,然后再通过强化学习的方式,使用反馈模型去不断优化迭代初始模型,形成对自动驾驶认知决策模型的持续优化。此外,DriveGPT视觉方案重新配置感知硬件,毫末在BEV框架中引入车端鱼眼相机。



## 空间计算定义明确： 消费级产品问世，XR全栈链路打通

今年6月，苹果在WWDC2023开发者大会上正式推出了其首款混合现实(MR)头显Apple Vision Pro，苹果CEO库克发表「时代宣言」：“如同Mac将我们带入个人计算时代，iPhone将我们带入移动计算时代，Apple Vision Pro将带我们进入空间计算时代”。

作为元宇宙的基础技术之一，空间计算能实现数字世界和现实世界的无缝对接。随着国内外空间计算硬软件技术取得突破，以及代表性消费级空间计算产品问世，将进一步推动各行业发展，或将开辟人类生活新方式。

## 推荐理由

### 1.空间计算作为下一代计算方式,将彻底变革传统信息呈现和交互范式

传统互联网的本质是2D平面计算,各类应用给用户展示的文字、图片、视频和直播等内容,均是将原有信息丰富度和空间均进行压缩后的2D内容。人与人、人与内容、人与企业、人与数字世界的互动都被局限在平面上,一定程度限制了交互的更多可能。

空间计算则让计算升维,使内容变得更立体,几乎可以将信息100%复原,保真度极强。语音、视觉、手势等方式相较于传统交互模式(鼠标、键盘、触摸屏等)更加自然灵活,使人们能以更趋于本能的方式在不同场景和业务完成人机互动,是底层交互模式的创新变革。

由此在空间计算时代里,人们可以更加沉浸、自由地探索数字世界,信息丰富度空前扩增,用户学习感知理解的成本大大降低,人机交互的上限与边界被更进一步拓展。



空间计算与传统计算

### 2.五大领域基础技术基本成熟,AIGC发展推动数字内容生产

空间计算业内核心关注的五大领域:感知、理解、协同、展现、内容生产创作,技术和产业供应链均基本达到商业化要求,成为今年空间计算迅速发展和落地的重要基础。

在之前硬软件技术基本成熟的情况下,今年AIGC的爆发和快速发展也开始解决数字内容快速生产创作的问题,实现了空间计算领域的闭环。

感知层面,传感器已经非常成熟,随着手机和半导体技术的普及,很多传感器可以复用;理解层面,用AI进行手势识别、空间识别、语音识别等都已成熟,基本达到“万物识别”;协同层面,低延时高带宽支持多人参与协同的平台,从PC到移动互联网时代一直都在做;展现层面,近几年大家比较关注的XR,更多关注光学技术的发展,空间计算领域更多关注画面的展示方式、内容展现和组织方式上;内容生产创作上,原来的内容创作更多强调3D引擎和工具,现在大家达成共识未来一定是AIGC的天下。

——Rokid CEO 祝铭明Misa

### 3.硬软件技术实现突破,代表性消费级空间计算产品Vision Pro将「空间计算」概念迅速普及

2023年6月苹果在WWDC2023开发者大会上正式推出了第一款空间计算设备Apple Vision Pro,这款产品的发布真正将“空间计算”这一概念推向大众,并直接瞄准To C市场。

#### (1) 硬件技术突破

- ① 硬件配置层面:采用M2+R1双芯片驱动处理器,这是苹果的独家优势,其他主流XR芯片在短期内难以追赶;
- ② 交互方式方面:首次设计三维互动界面,摒弃了手柄控制模式,采用了更先进的眼动追踪和虹膜识别等技术,可以实现眼球追踪、语音和手势等多种交互方式融合,逐步接近自然用户界面,开创了XR交互新范式;
- ③ 光学方面:采用了3p pancake光学方案,相比单片式方案更轻薄,同时支持IPD调光和屈光镜片定制,可以最大程度地适应用户的眼睛。

## (2) 生态系统建设突破

其中苹果专门研发的空间操作系统VisionOS, 实现了苹果软硬件生态资源无缝对接, 是构建生态、实现商业化的关键。

VisionOS是基于macOS、iOS和iPadOS数十年工程创新之上的成果, 苹果又赋予了它许多新能力, 来支持空间计算的延迟要求。如全新实时执行引擎, 可满足高性能任务需求; 如动态注释点渲染管线, 可确保每一帧画面都能在用户注视位置呈现最优图像质量; 以及现有应用架构, 也为原生支持空间体验做好了扩展。

“

苹果发布Apple Vision Pro后, 带来了巨大想象空间, 通过动作、语音、眼手配合, 空间环境、具体的物件进行输入、操作, 重新定义了XR交互方式。虽然第一代价格昂贵, 量产也存在问题, 但相信苹果会在持续发展中优化迭代, 实现产品普及。在技术上, 苹果加入了虹膜识别技术、数字人生成技术等交互互动技术, 均是在实际产品使用场景中很重要的功能。场景上, 苹果提出了空间计算概念, 虽说不会瞬间改变生产力, 但可以想象到会带来人类与数字世界交互范式的变革。伴随着LLM的成熟, XR设备交互会更加容易, 会成为连接数字世界和物理世界的载体, 因此上层的应用生态会在未来两年迎来一波新的机会。

——真格基金合伙人 尹乐

## 4. 空间计算VST/OST两条路线齐头并进, 国内已有OST代表选手

在今年8月26日2023 Rokid Jungle发布会上, Rokid对外发布了消费级OST (Optical See Through) 个人空间计算平台Rokid AR Studio, 成为国内代表性消费级OST空间计算产品。

## 5. 将进一步推动XR、数字孪生和元宇宙的增长潜力

代表性消费级空间计算产品商业化落地后, 将带动XR开发者生态、内容生态以及供应链逐渐完善, 应用开发将迎来革命性的创新时刻。XR设备也向“空间计算平台”定义转变, 终端销量天花板有望逐步打开, 将进一步推动XR、数字孪生和元宇宙等领域加速增长。

# 技术原理

## 1. 空间计算是什么

空间计算是元宇宙的基础技术之一, 最初是指对地图及其他地理位置数据进行计算和分析以实现定位与测量的技术。随着XR、虚拟人、数字孪生等技术领域的发展, 微观空间的计算需求也在逐渐增加, 现在提到空间计算, 多指微观空间的计算。

空间计算实质上是一种通过空间数据和算法对空间信息进行处理和分析的技术, 使得虚拟世界和现实世界实现无缝混合, 两个世界之间可以相互感知和交互。

早在2003年麻省理工学院Simon Greenwold的硕士论文中就有相关的描述, “空间计算是人类与机器的交互, 机器可以处理并操纵真实对象及空间的相关信息。它是我们实现让机器成为我们工作和娱乐的合作伙伴的重要手段。”

## 2. 空间计算核心技术

包括三维重建、空间感知、用户感知、空间数据管理等。从计算载体维度考量, 空间计算可以分为端计算、云计算以及由5G作为重要媒介的云边端协同计算。

### 三维重建

是在计算机环境下利用虚拟现实来表达可观世界的关键技术。三维重建主要是利用相机采集到的图片数据、深度图数据或两种数据的融合重建出物体和场景的3D模型。目前三维重建技术具体应用场景包括3D打印、机器人、空间场景识别、无人驾驶、虚拟现实、SLAM等。

### 空间感知

指的是获取人或物在空间中的位置、速度、方向等状态的能力,主要通过一系列传感器之间的数据校准来进行确认。汽车、智能手机、机器人、XR头显等智能终端在复杂空间中的定位、以及和位置相关的服务,都需要应用空间感知技术。

### 用户感知

用户感知首先要对用户表达的信息内容进行处理,即系统或设备要捕捉、分析和理解用户的形象、状态和行为表达等信息,然后在此基础上对用户进行认知引导和交互驱动。

### 空间数据管理

空间数据主要由三维模型、点云、图片、网格、位置点、向量、表面模型、路径等视觉图像数据转换而来,具有海量多模的特点。出于数据资产管理和未来大量多维分析计算能力的需求,对空间数据进行采集、存储、处理、生产、调用全生命周期的高效管理尤为重要。

实时可交互的空间计算平台,需要“感知-理解-重建-编辑-共享”的全链路能力。

空间计算平台发展趋势:从空间感知(多源传感器融合+计算机视觉技术结合,对周围环境和物体的距离、形状、大小等空间特性做出快速及准确的感知)——空间理解(将空间感知结果进行深度学习处理)——空间重建(实时构建与真实世界1比1,机器可读的空间网格,任何虚拟物体都可以合理放置在真实空间里与人们进行交互)——空间编辑(通过云技术与工具链,开发者可以实时在真实空间里放置、编辑和删除虚拟物体)——空间共享(通过不同终端(如智能汽车、智能眼镜)可以实时接入同一空间,和该空间内的同一物体进行互动,从而产生链接和网络效应)。

——联想创投

## 3.主要应用场景

无论是B端还是C端市场,空间计算在各行业领域都有着丰富的应用场景和发展潜力。

在行业应用侧,空间计算可应用于智能终端,扩展生产工具能力,实现“降本增效”。

在庞大的消费者市场,空间计算目前最大的载体是XR.XR设备可以深入到每个消费者的日常生活中,丰富娱乐、办公、游戏体验。

目前空间计算在工业、文旅、会展、商业、游戏、协同办公、影视媒体等场景中已有多个落地应用案例,以下为几个重要领域。

#### 工业领域

空间计算结合5G与XR,可以将设备3D模型、设备状态、工序流程等信息与机器、产品进行关联,提升制造商捕捉和分析数据的能力,并在运维过程中为工程师呈现更准确的可视化实时信息,提高巡检效率,降低工序风险,提升人力资源利用率。

#### 文旅领域

空间计算技术可以帮助实现数字孪生城市、智慧旅游等多种业务模式。例如,利用AR、VR等技术,可以为游客提供更加生动、丰富、互动的文旅体验,增强游客的参与感和记忆点。

#### 社交娱乐领域

空间计算技术在娱乐和文化领域也有广泛应用,可为用户提供多维内容和感官体验。例如,通过XR技术,用户在电影、游戏、音乐等场景中可获得身临其境的体验,未来也将融入日常社交、办公等场景中。



空间计算或者 XR 产业永远是解决三个问题。第一是展现力是否足够丰富和真实,包括延迟、分辨率、渲染模型的真实程度;第二是设备本身的便携性或舒适度的问题,不管是挂在身上、戴在头上、还是随身携带,或者全息成像,也许未来也会有新的模式,但都要关注这一问题;第三是整个生态是否成熟,什么样的应用或内容形态会成为手机之后的新形态,带来革命性的变革,这个就拭目以待了。

—Rokid CEO 祝铭明Misa



现在的XR-空间计算技术发展还很早期,但有了它的技术支撑,人们除了吃喝拉撒、生老病死之外所有的需求都可以在元宇宙中得到满足,包括社交、商业、娱乐等,所以未来的商业价值巨大。

—跃为资本人工智能组负责人

## 4.未来XR的两大方向

### (1)VST

即视频透视,英文全称为Video See-Through,简称VST。利用摄像头等传感器,捕捉真实世界的影像,然后投射在屏幕上,用户看到的一切都是虚拟设备想让他看到的样子,真实世界与虚拟世界都是数字信号,可以同步修改。代表产品有Apple Vision Pro、Meta Quest2/Pro、Varjo XR、Pico等,更准确地说属于MR,兼有AR和VR的体验。普通品牌细节上有区别,例如Apple Vision Pro能通过小旋钮,让用户体验在VR和AR间调整,看电影时,更偏向于VR的沉浸式效果。

### (2)VST优势

#### 虚拟与真实物体的遮挡更容易控制

- OST的一个基本问题是,由于光学合成器同时允许来自虚拟和真实世界的光源,虚拟物体通常不会对真实物体进行完全遮挡,这使得虚拟物体看起来像幽灵一样呈半透明状态。
- 相比之下,VST在融合真实图像和虚拟图像方面会更容易控制。由于都是以数字图像的形式提供信息,VST合成器可以逐像素地采用真实、虚拟或两者混合来模拟透明程度的方式,从而产生比OST更令人信服的遮挡效果。

#### 视场(FOV)更大

- OST受光学元件以及显示器等核心元件的限制,加上轻量化的需求,主流的OST方案的FOV在30-70°左右。
- VST使用的显示器尺寸较大,而且可以将显示器和光学器件完全设置在眼睛前方,因此能够获得比较大的FOV,目前主流在90-120°之间。

#### 虚拟和真实世界的延迟更容易匹配

- OST提供了几乎实时的真实世界视图,但虚拟世界的视图或多或少总存在一定延迟,这种时间的不匹配可能会导致配准等问题,影响用户体验。
- VST可以通过延迟真实世界的视图,来匹配虚拟视图的延迟。换句话说,VST提供了一种减少或避免真实和虚拟图像时间不匹配问题的方法。

## 配准信息更加丰富

- 对于OST,系统关于用户头部位置信息的唯一来源是头部跟踪器。
- VST则提供了另一种信息来源:真实场景的数字图像,这意味着VST可以通过检测图像中的特征,并使用这些特征来增加配准的准确性。

亮度匹配控制  
更加方便

- OST由于允许用户直接看到真实世界,受到真实环境亮度的影响较大。如果真实环境太亮,现实世界会冲掉虚拟图像;如果真实环境太暗,则虚拟图像将会冲淡现实世界。
- VST因为相机本身的动态响应有限,亮度控制就会相对更加容易。VST真实和虚拟图像都是由监视器生成,并且都经过了一定的动态范围压缩,因此能够更加方便地进行亮度匹配控制。

### (3) OST

即光学透视,英文全称为Optical See-Through,简称OST。它就像一层玻璃,同时通过摄像头定位和芯片计算,虚拟物品和现实可以产生交互,以真实世界为主,可以让用户对现实世界的感受更加出彩。

代表产品有Microsoft HoloLens系列、Magic leap系列、Rokid Max Pro、Xreal X等,属于严格意义上的AR。

### (4) OST优势

## 亮度范围更大

- OST看外界真实环境,由方案透过率和外界光线亮度决定,可达到6600尼特甚至更高,是目前VST需求的10倍以上。
- VST方案由于真实环境是通过相机拍摄后在屏幕上显示出来,因此看到的外界环境亮度也受限于屏幕亮度,目前在100-600尼特之间。

## 真实世界分辨率更高

- OST分辨率由人的视网膜决定,约为单眼24k+,是目前VST的10倍左右。
- VST受限于显示屏幕的分辨率,目前约单眼2k-4k。

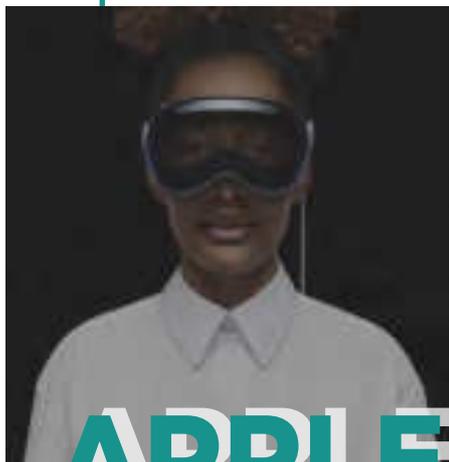
## 延迟更低

- OST能够直接看见真实外界环境,完全没有延迟。
- VST因对外界真实环境的采集、显示需要软硬件处理,故存在延迟。

## 更低的眩晕感

- OST透过光学模组看见的真实世界存在无数个焦平面(可以随意看远景或近景),可以防止辐辏冲突和眩晕的产生。
- VST因光学模组硬件显示,目前只有1个焦平面,容易产生眩晕感。

## 全球首款空间计算产品「Apple Vision Pro」



2023年6月6日, 苹果在2023年全球开发者大会 (WWDC23) 发布了MR产品Apple Vision Pro, 计划将于2024年初在美国首发, 售价为3499美元, 搭载了苹果公司首款3D摄像头, 用户可用其观看和拍摄3D视频。



硬件端

采用M2+R1双芯片驱动处理器, 计算性能强大, 其中M2负责运行VisionOS以及计算机视觉算法, R1主要负责传输和处理传感器信号, 可以实现在12毫秒内将图像传输至显示器。

搭载了全球首个空间操作系统VisionOS, 其底层核心主要基于iOS、macOS和iPadOS等, 集成多应用3D引擎、实时子系统、空间音频、注视点渲染、空间计算架构等子系统。

软件端



应用端

苹果Vision Pro宣传短片中, 设计师将3D模型放在桌面上后, 可以360度去观察。在办公和居家场景中, Apple Vision Pro将显著提升用户的生产力, 用户可以获取常用的APP、打开全新的多任务处理方式等, 还有诸多使用场景。

## 国内首个轻薄型空间计算平台「Rokid AR Studio」

在今年8月26日2023 Rokid Jungle发布会上, Rokid最惊艳的是对外发布了消费级OST (Optical See Through) 个人空间计算平台Rokid AR Studio。Rokid AR Studio延续Rokid分体式设计思路, 由AR空间计算眼镜Rokid Max Pro和AR空间计算主机 Rokid Station Pro组成。



### AR空间计算眼镜Rokid Max Pro

硬件设计方面:

机身重76g, 具备「以手势交互为核心的空间计算能力」;

可实现SLAM、微手势交互、第一视角分享、视觉定位VPS能力。

视觉方面:

全新BirdBath模组, FOV达到50度;

索尼半导体科技Micro OLED屏幕;

屏幕分辨率为单眼1920×1200, 支持90Hz刷新率。



### AR空间计算主机 Rokid Station Pro

CPU:高通XR2+Gen1, 12G运存+128G存储

操作系统:YodaOS-Master

视频解码: 4K 60fps 解码

摄像头:索尼IMX586 4800万像素后

置摄像头, 视场角115度

传感器:内置9轴IMU

传输:NFC、WIFI 6、蓝牙5.1

## 代表机构

Rokid

**Rokid**  
AR时代引领者

致力于AR眼镜等软硬件产品研发,以YodaOS操作系统为载体的生态构建。首提AR主机概念,推出RokidStation系列,产品强调更AR原生的体验,强调计算性能,不只是提供巨幕体验,开拓AR办公场景。

XREAL

**xreal**

主要聚焦于混合现实技术研发,AR终端眼镜生产,主打消费市场。现已推出两款面向C端的AR眼镜产品Light、Air。产品强调轻薄、美观,关注用户佩戴体验。在AR效果呈现上,主打“大屏体验”,将2D内容引入到3D空间。先后进入中美日韩英西德等多国市场。

雷鸟创新

**雷鸟**

专注于AR领域的新锐品牌,具备关键光学器件MicroLED全彩光引擎自研、生产和高效整机量产交付的能力。今年10月雷鸟创新举办其首场线下发布会,正式发布了双目全彩波导一体式真AR眼镜——雷鸟X2。

影目科技INMO

**INMO**  
影目科技

以消费级AR眼镜这一产品终端为核心,拥有从软硬件及系统应用研发到测试、生产、市场销售具备完整产业链、完备上下游供应链的智能硬件公司。今年4月全面上市全彩无线一体式AR眼镜INMO Air2。

凡拓数创

**frontopus**

国内第一家3D引擎上市公司。20年专注研发3D可视化技术,为客户提供数字展示一体化服务,包括3D数字内容制作、VR/AR展示等。

超图软件

**SuperMap**  
超图软件

国内三维空间数据模型龙头。旗下SuperMap软件是亚洲最大、全球第二大GIS软件,是数字中国、数字政府、企业数字化、数字孪生、元宇宙、智慧城市的重要技术底座。

利亚德

**leyard**

旗下虚拟动点拥有光学动捕、惯性动捕和基于AI无标记点追踪等一系列基于空间计算的机器视觉动作捕捉技术。



## mRNA打开新象限： 提供精准医疗新解法，开启生物医药新篇章

今年mRNA领域取得了令人瞩目的成就。5月，百度美国研究院与斯微生物、俄勒冈州立大学等团队合作，发表了一篇关于mRNA序列优化算法的论文。这是中国科技企业首次在《Nature》正刊上发表论文。10月，卡塔琳·卡里科 (Katalin Karikó) 和德鲁·魏斯曼 (Drew Weissman) 因在核苷碱基修饰方面的发现，加速了针对COVID-19的有效mRNA疫苗的开发，获得了诺贝尔生理学或医学奖。由于mRNA技术的不稳定性及人体免疫等多原因，无法应用于新药研发。直到脂质纳米颗粒递送技术、碱基修饰技术的突破，解决了mRNA疫苗应用的难题，实现了快速量产。

# 推荐理由

## 1.mRNA疗法优势明显

### 质粒DNA介导疗法

质粒DNA介导的疗法中,由于编码后的质粒DNA尺寸过大,显著增加了系统递送和表达的难度,降低了进入细胞核后的转录效率导致治疗过程的延迟。此外,基于质粒DNA的表达蛋白会长时间存在细胞中,可能导致更高的脱靶效应和免疫应答等问题。

### Cas9蛋白疗法

Cas蛋白注入血液后易被蛋白酶降解;RNP分子(由Cas9蛋白和sgRNA组成)较大,穿透细胞膜进入细胞质的难度较大,而且在生产过程中需保证Cas蛋白的纯度和活性,保证其发挥基因剪切作用。

### mRNA疗法

mRNA的规模较小,更利于传递,而且mRNA在体内的半衰期较短,并且不会进入细胞核改变基因组,也就避免基因突变的风险。因此,一旦mRNA疗法出现副作用,通过剂量调整等方式可以控制治疗效果,可能出现的有害副作用就会逆转。

## 2.mRNA应用于预防疫苗、治疗疫苗、治疗药物等领域

mRNA可以作为一种极具潜力的通用技术平台。目前mRNA可应用于传染病预防、肿瘤免疫治疗、蛋白替代、CAR-T、基因编辑等,应用场景总体可分为三大类:预防疫苗、治疗疫苗、治疗药物。其中预防疫苗领域的布局最丰富,其次是治疗药物领域。

mRNA疫苗的优势主要表现为:

- ① **安全性优势:** mRNA仅在细胞质中表达,避免整合基因组的风险,表达效率高、免疫原性高、半衰期短,体内存在天然代谢降解机制。
- ② **靶点数量优势:** 很多难成药的蛋白或胞内蛋白均可由mRNA编码表达。可分泌至胞外、靶向受体或循环系统,靶向性选择更加丰富。
- ③ **疗效优势:** 各种修饰使mRNA更加稳定,可高度平移。通过将mRNA配制成载体分子,可以实现高效的体内递送,使其在细胞质中快速表达。
- ④ **制造工艺优势:** mRNA采用DNA模板通过体外转录过程合成,制造工艺简单,可高效率、低成本、大规模生产。
- ⑤ **研发周期优势:** mRNA疫苗的研发仅需在成熟技术平台上更换抗原序列,因此在应对病毒变异时,优势突出,4-6周可实现更新换代。

## 3.mRNA疫苗商业落地迅速,国产供应链逐步完善

mRNA疫苗的广泛关注给mRNA研发带来了巨大的资本投入和时代红利,mRNA的产业链迎来历史性机遇。

(1) mRNA疫苗商业落地迅速。在疫苗开发方面,石药集团、康希诺生物、沃森生物、复星医药等上市公司都在mRNA技术上进行了布局。其中,石药集团和康希诺生物在mRNA新冠疫苗的研发和生产方面取得了重要进展,并在国内已经获批上市。

(2) 国内的关键原材料及耗材供应完善。在酶、尿苷原料，RNA帽类似物等原料方面，兆维科技，瀚海新酶，诺唯赞等国内mRNA上游公司能够提供mRNA合成所需的完整试剂，且上游公司不断进行新技术探索和专利布局。

(3) LNP生产厂商逐步掌握产业先进技术。目前国内已具备提供工业化生产用的LNPs递送系统辅料，包括 SM-102，M-DTDA-2000 在内多种阳离子脂质及 PEG 化的阳离子脂质。

## 技术原理

mRNA技术将编码疾病特异性抗原的mRNA引入人体内，利用蛋白质合成机制产生抗原，从而触发免疫应答，达到预防或治疗疾病的效果。mRNA技术具有产品质量标准、可重复、安全性高和相对便宜的优势，因此能被快速用于不同疾病的治疗中，mRNA优势十分明显，但也有两大难点需要突破：

① **mRNA的分子设计与化学修饰**。由于核酸酶的存在，RNA在细胞外液及细胞质内均不稳定，且可刺激细胞表面受体激发免疫系统反应，mRNA修饰可以在一定程度上增加mRNA的稳定性，降低其体内免疫原性。此外，mRNA的调控区及编码区的序列设计也极大地影响了其在细胞内的稳定性和翻译效率。因此，科学家们需要通过不断设计mRNA的分子结构及优化序列组合，以达到稳定mRNA结构，提高蛋白质翻译效率的目的。

② **mRNA在体内的有效递送**。mRNA属带负电的生物大分子，很难透过同样带负电的细胞膜进入细胞发挥疗效，此外，它们易被核酸酶降解，难以完整地到达靶部位。因此mRNA需借助相应的递送系统，才能将RNA药物传递到治疗部位。能否克服胞外的降解及从溶酶体逃逸是mRNA在细胞内发挥治疗作用的关键所在。

近年来，在mRNA分子设计和递送技术层面均有突破性进展：

### 1. 分子设计层面

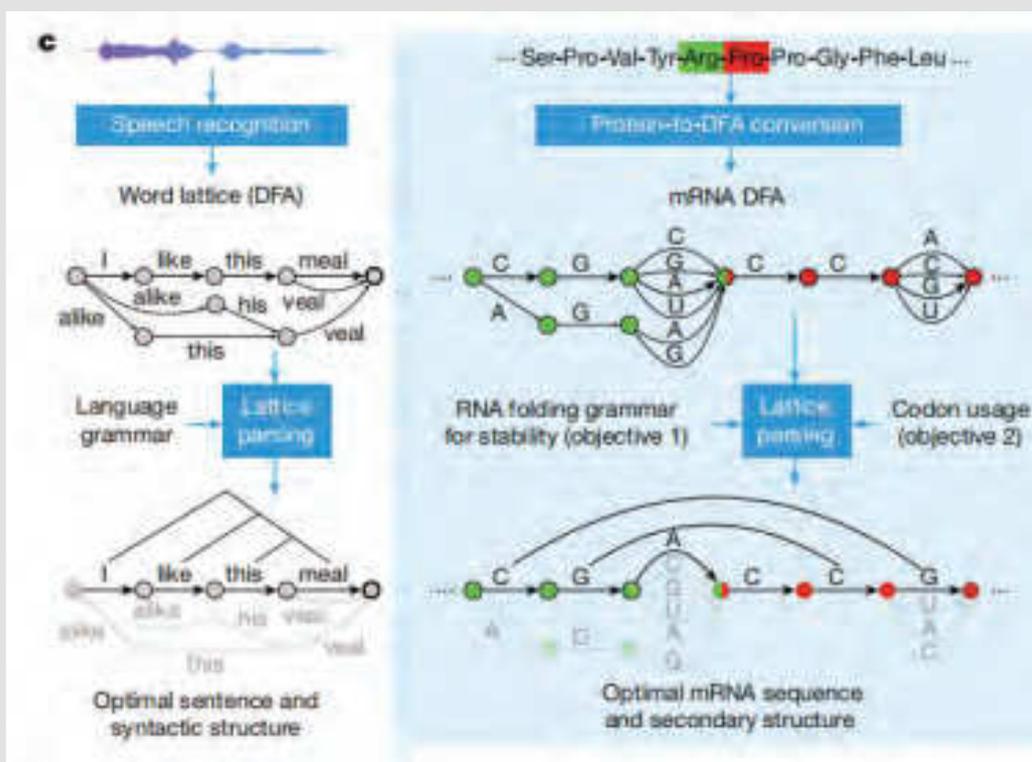
过往的疫苗抗原设计中，是在已知某个核心抗原的氨基酸序列，然后利用CodonOptimizer这样的密码子优化工具，根据不同表达平台的密码子适应性，获得编码序列，只要表达量能够经过克隆筛选和工艺优化满足工业生产需求就算完成。但mRNA技术疗法是将蛋白生产过程搬到机体细胞内，所以对于编码序列的筛选和设计就显得极为重要。

### 线性设计 LinearDesign

LinearDesign主要解决的是在众多mRNA编码序列中的搜索难题，通过动态规划算法将指数级的搜索空间降低至多项式级，这其中应用到了网格解析技术。

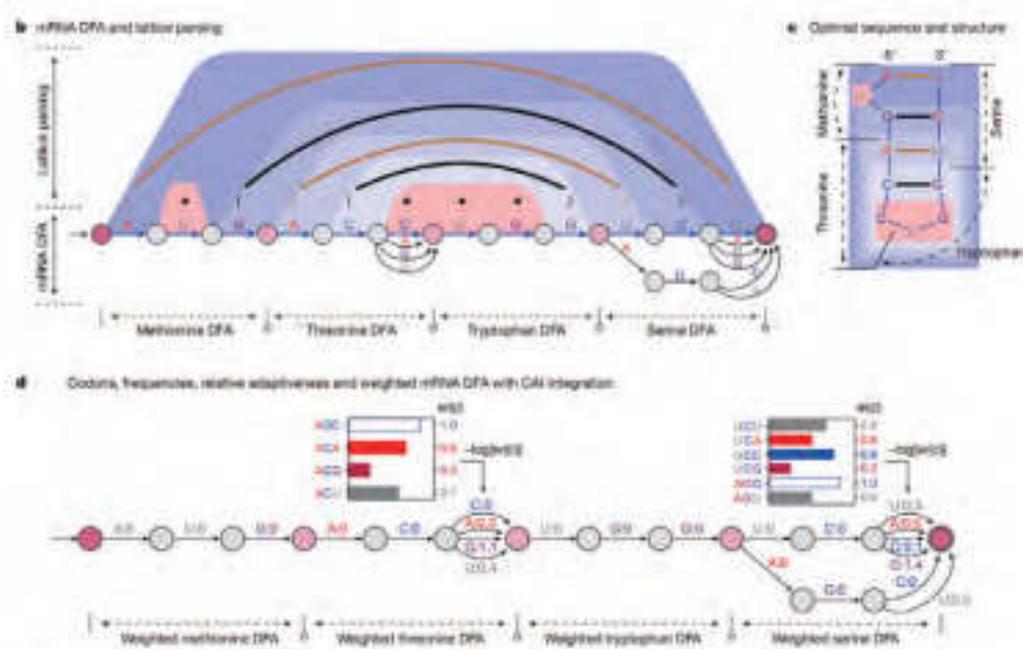
### 网格解析 (Lattice Parsing) 技术

mRNA 序列设计问题与自然语言处理领域解析句子文本有相似之处。mRNA序列设计问题的输入是一个蛋白质序列，而这个序列对应了海量 mRNA 候选序列。将这些候选序列转换成一个网络图，也就是mRNA DFA，根据 RNA 折叠语法规则和密码子使用偏好，在mRNA网络图上做网格解析，就能找到最优的mRNA序列。



用网络解析技术将mRNA转换成网络图 (mRNA DFA)

用**确定有限状态自动机 (DFA)** 来表达氨基酸和蛋白质，这样不同位置上密码子的选择就可以抽象为计算理论中常用的DFA图，将各类氨基酸以及终止密码子抽象为DFA图。在此基础上，将氨基酸的DFA串联起来，即可得到一段蛋白质序列的DFA图。再通过取DFA与SCFG的交集，从多个mRNA序列中找到二级结构最稳定的序列，后图中“Methionine Threonine Tryptophan Serine”最稳定的序列是“AUGACCUGGAGC”。



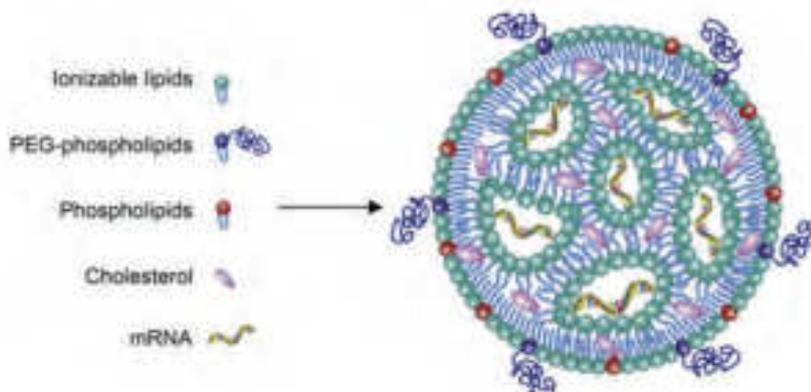
根据蛋白质的DFA图找出最优序列

以新冠病毒刺突蛋白举例,要设计mRNA新冠疫苗,需要在 $10^{632}$ 次方个序列中确定满足条件的序列。过往的设计方式多以自然界存在的新冠刺突蛋白对应的mRNA段为基础,经过稍微改动后作为备选序列,但这种设计方式得到的mRNA序列其二级结构一般不够稳定。

而LinearDesign算法可以在保证翻译的蛋白质不变的情况下,能够迅速设计出二级结构更稳定的mRNA序列。基于LinearDesign的mRNA序列在稳定性、蛋白表达和免疫原性等方面明显优于传统算法设计的基准序列。计算机模拟表明,使用这种新算法仅需11分钟就能设计出最优的mRNA序列,能够改善候选mRNA疫苗的稳定性和蛋白表达。

## 2.递送技术层面

目前临床上最先进的mRNA递送系统主要是脂质纳米颗粒(LNP)——由可电离脂质(Ionizable lipids)、磷脂(Phospholipids)、聚乙二醇化脂质(PEG-phospholipids)和胆固醇(Cholesterol)组成。



用脂质纳米颗粒(LNP)封装mRNA疫苗示意图

## 脂质纳米颗粒 (LNP) 递送技术

脂质纳米颗粒递送技术是一种常用的载体技术,能够将mRNA序列包装在一层脂质纳米粒子中,通过特定的化学反应组装成球形结构,这种结构可以保护mRNA序列不被降解,增加被人体细胞摄取的效率。

今年6月,宾夕法尼亚大学Drew Weissman 教授和Michael J. Mitchell 教授团队化学合成了一种新的LNP组分——佐剂类脂质,增强了mRNA-LNP疫苗的佐剂性(佐剂:能够增强疫苗免疫应答的分子)。实验结果表明,用佐剂类脂质部分替代LNP的可电离类脂质组分,增强了mRNA的递送效率,显著提高了mRNA-LNP疫苗在小鼠体内的先天免疫力,并具有良好的耐受性。

## FixVac —— mRNA癌症疫苗平台

FixVac (固定疫苗) 候选平台由mRNA编码的非突变肿瘤抗原的固定组合构成。mRNA是采用专有的RNA脂质体递送配方配制而成,旨在提高mRNA在体内的稳定性,并靶向抗原呈递树突状细胞(DCs)——免疫系统的训练营。通过增强树突状细胞向免疫系统呈递这些特异性肿瘤抗原,FixVac 候选药物能够针对表达一种或多种肿瘤抗原的癌细胞触发强烈而精确的先天和适应性免疫反应。通过基于mRNA的癌症疫苗方法,能够为癌症患者提供一种高度肿瘤类型特异性的现成治疗方法,特别是处于晚期、难以治疗的疾病。

虽然癌症病人的肿瘤构成各不相同,但这些肿瘤可以共享抗原,这些抗原在特定癌症类型中是一致的,并且经常在许多相同类型的癌症患者中表达,但在体内健康细胞中未发现。

所以,癌症治疗的难题在于免疫系统无法识别癌细胞为恶性细胞,无法触发免疫反应。FixVac平台的目标是通过向免疫系统提供正确抗原,来激活识别癌症特异性抗原的免疫细胞。

# BIONTECH

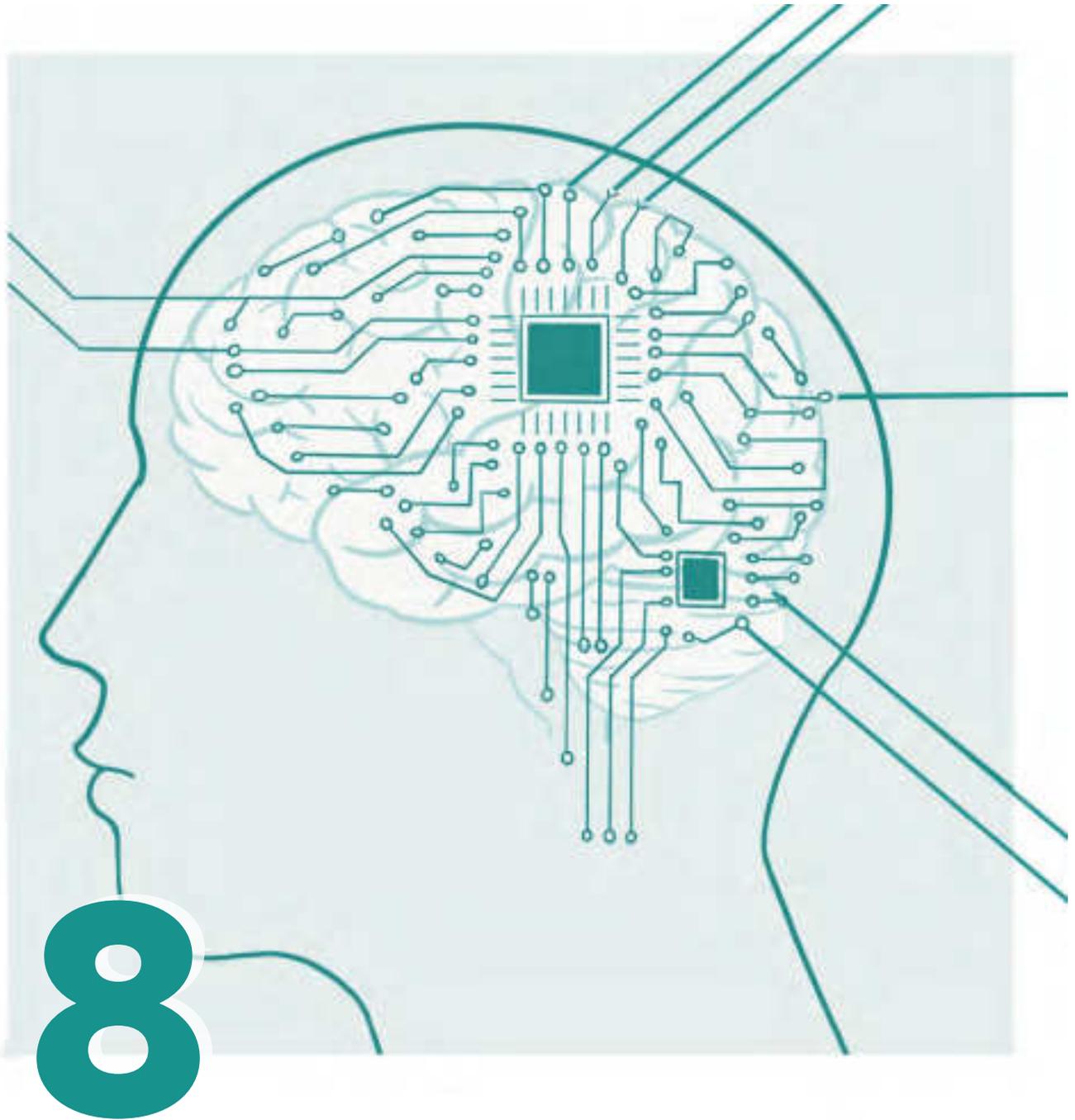


## 代表机构

深势科技

DP Technology  
深势科技

2023年7月深势科技推出的Uni-RNA 大模型大大提升了 mRNA 疫苗设计、RNA 结构预测、ASO 开发、SiRNA 疗法创新、靶向 RNA 小分子开发、Aptamer 研发等几乎所有RNA相关研发场景的模型能力。



## 脑机接口试验新阶段： 产品可靠性突破，AI提升数据解码能力

自脑机接口 (BCI) 概念提出至今, 已经有50年历史。当前脑机接口技术已经进入多个医疗场景, 并且进入人体临床试验阶段。

脑机接口的最终目标是发挥人类大脑的全部优势, 并且使大脑能够直接与外接装备进行高效互动。随着科学技术的不断发展, 脑机接口将从单向脑机接口 (即产生、获取和解析脑信号), 到双向人机交互, 最后实现人机共融、协同决策的高智能状态。

# 推荐理由

## 1. 植入式设备安全性能达标, 进入人体临床试验阶段

今年5月, 脑机接口公司Neuralink获得FDA批准进行人体临床实验。据悉, Neuralink计划在2024年进行11台植入手术, 2030年这个数字预计增长到2万台以上。在迈出了满足安全性要求这一步后, 脑机接口下一步需要通过人体临床试验证明其稳定性和有效性。

## 2. AI加持下解码速度提升, 科研推进加速

今年科研方面的突破主要是解码速度提升, 更接近正常人的文字或语音输出速度。《Nature》连续发布两篇论文报告新的脑机接口装置。科学家利用人工智能技术, 让两名瘫痪患者能够靠“意念”重新开口说话, 并在准确性和速度方面都实现突破: 两组研究中, 脑机接口设备解码语音的速度分别为每分钟62个单词和78个单词, 且错误率分别降至23.8%和25.5%。

“

数据驱动的脑科学研究是近年来脑科学研究的主要方向之一。通过大量的动物和人类实验, 科学家们积累了海量、多模态的实验数据, 例如脑电、基因、脑影像等等, 因此如何更好地分析这些数据对于脑科学的研究至关重要。大模型技术对于脑科学实验数据的分析和理解很有帮助。例如, 微软亚洲研究院在今年提出的脑电数据大模型和基因表达预测大模型都证明了这一点。未来我们期待能够在多种大脑数据上构建能力更强的基础大模型。

另一方面, 大脑数据不同于自然语言和图像, 模态更加丰富、数据不统一并且信噪比低, 直接采用现有的自然语言大模型的技术并不能很好的解决问题。为了训练更强的脑电数据大模型和基因表达预测大模型, 我们在数据建模、算法设计、训练优化等多个方面进行了创新, 我们相信这些创新成果不仅适用于脑科学相关的大模型训练, 也对训练垂直领域的大模型具有普遍的借鉴意义。

—微软亚洲研究院

## 3. 意念控制机器人——脑机与机器人融合加速

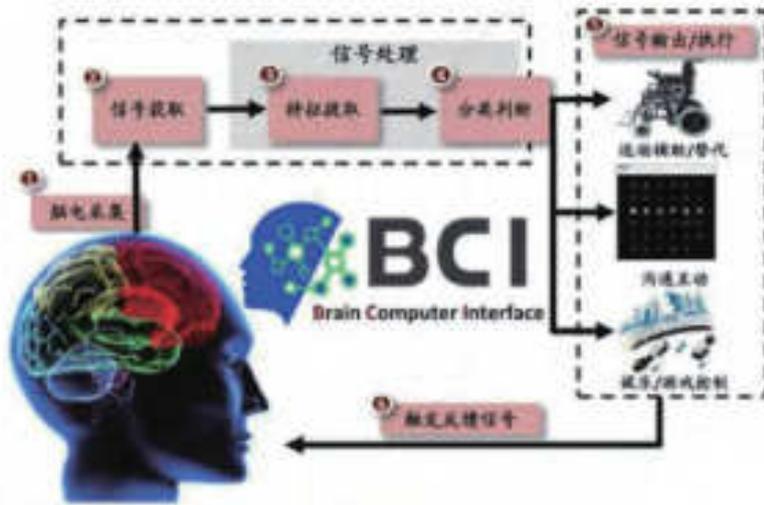
「具身智能」是今年最热门的前沿方向之一。而脑机接口技术, 能够将人类脑电波中的信号转换为机器人可以执行的技能集, 使人类通过大脑信号指挥机器人执行日常活动。同时, 机器人能够在与人类的协作过程中学习人类的意图目标。通过人类协作来构建智能机器人系统的方法, 可为残障人士提供辅助系统。

# 技术原理

## 技术定义

脑机接口是在脑与外部设备之间创建的信息接口, 通过检测(读)或调控(写)神经网络活动, 实现脑-机之间的直接信息通讯。

脑机接口系统工作的三大核心流程包括信号采集、信号处理、以及信号的输出和执行。与之对应的技术包括神经信号采集技术、刺激技术、范式开发、算法能力、外设技术和系统集成等。

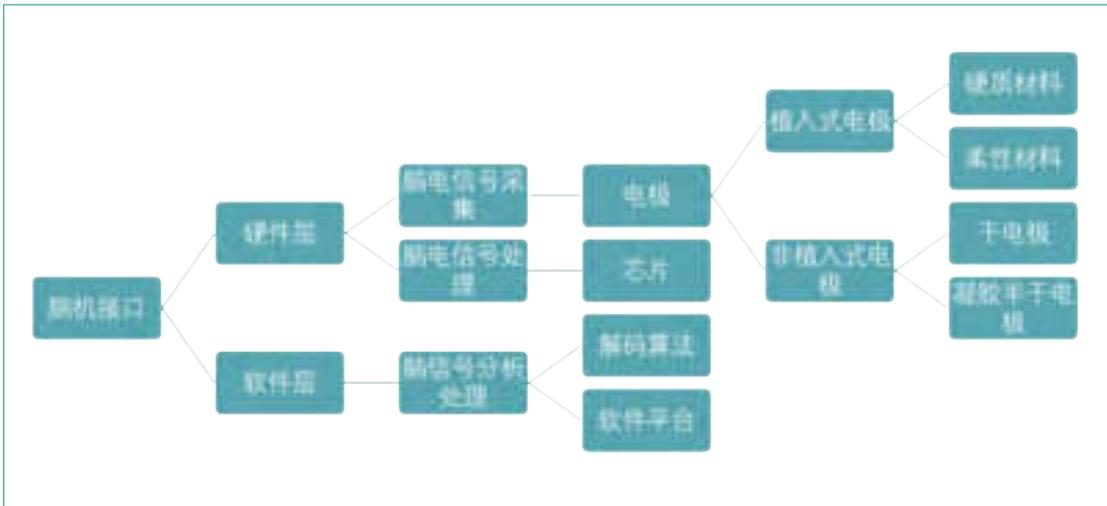


脑机接口原理图

## 技术分类

从技术路径上讲，脑机接口分为植入式和非植入式两种。植入式设备能够获得更高质量的神经信号但存在安全风险，主要应用在医疗健康领域，如神经替代、神经调控相关技术和产品。非植入式设备能探测到的脑信号精度不高但无安全风险，主要应用在更广泛的生活消费领域。

脑机接口技术整体上分为硬件层和软件层，其中电极、芯片和解码算法是业内普遍认为的脑机接口领域最关键的技术分支。



脑机接口流程图

### • 硬件

脑电信号采集主要依靠脑部电极，电极决定了所采集脑电信号的空间分辨率和质量。目前柔性电极是多数植入式脑机接口公司选择的电极材料，其中包括「皮层刺入」和「皮层表面」两种植入方式。Neuralink选择的是「皮层刺入」，国内微灵医疗选择了「皮层表面」，脑虎科技则同时涵盖两种方式。

芯片方面，目前主要有通用芯片和专用芯片两种方案。在专用芯片里中，使用ASIC芯片的有Neuralink、复旦大学等；使用FPGA芯片的有天津大学。今年，我国成功研发出一款具有完全自主知识产权的新一代8通道脑电采集国产芯片，可用于非植入式脑机接口。

### • 软件

解码算法方面，通过机器学习和深度学习的加持，研究人员能够开发出更高精度的大脑信号解码算法，以实现更准确和可靠的脑机接口系统。目前脑机接口的性能提升主要依靠人工智能加持的神经信息解码技术实现。

“

在未来,我们需要关注如何让科学技术走向原型,再从原型到产品,最后从产品实现商业价值。无疑,脑机接口技术将应用于临床重大脑疾病诊治和脑功能的探索,可以视为是“人-机-物”三元共融万物感知的核心技术,是解决渐冻症、高位截瘫、癫痫等重大脑疾病的重要手段,未来进一步结合人工智能技术,通过人脑与AI的融合来增强大脑功能,实现全脑AI交互,释放人脑的全部潜能。

——脑虎科技

“

微软亚洲研究院与脑科学相关的研究主要集中在两大方向:

(1) 利用AI更好地理解大脑,帮助提升大脑健康

在临床和科研上,依靠人力对于脑电信号进行分析具有很大的局限性。微软亚洲研究院训练了首个跨数据集的脑电预测模型,可以针对不同场景的脑电数据进行统一分析。此外,微软亚洲研究院还开发一个AI Neurologist系统,能模拟医生写脑电图报告的过程,自动生成脑电分析报告,并可以辅助临床和科研场景下的脑电信号分析工作。

(2) 通过理解大脑的结构和活动,设计脑启发的AI

在脑启发的神经网络方面,受大脑神经信号传输模式的启发,微软亚洲研究院设计了一种新的神经网络 CircuitNet,采用统一的结构灵活实现大脑中前向、后向、抑制、促进等各种信号传输机制。通过在函数逼近、时序分析、图像识别、强化学习这四种任务上与多层感知器、CNN、RNN、Transformer等经典神经网络的对比,研究员们发现CircuitNet可以用更少的参数实现相同甚至更好的效果。

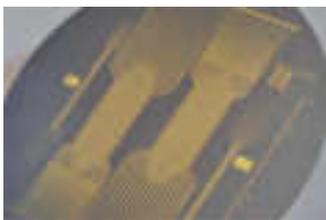
——微软亚洲研究院

## 医疗级全植入式无线脑机接口全栈技术

### 高密度柔性微型皮层电极阵列

技术创新:神经信号传感纳米增强技术

在提升聚合物传感器对神经信号感应性能的同时保证结构更稳固,使用寿命更长。



### 脑机接口系统专用集成电路

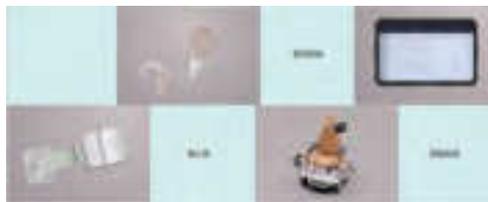
技术创新:神经信号采集刺激一体闭环芯片

在多通道同步采集神经电信号的同时,进行精确的、可编程的高功率神经电刺激。



全植入  
运动脑机接口系统

未来技术:新型光纳米神经遥控技术  
分为植入体、数字终端、智能轮椅机器人三部分



# 微灵医疗

## 一款简洁的「一体化的设备」

从2024年开始, Neuralink将开始为人类植入脑机接口。  
2024年计划进行11台植入手术, 而到2030年, 这个数字要增长2000倍, 计划为超过22000人植入脑机接口。



植入物N1



生物相容性外壳



电池



芯片



柔性电极



通过机器人手术实现完全自动化植入N1设备; 可根据患者的大脑结构和需要来定制植入方案

# NEURALINK

## 柔性脑机接口技术方案

脑虎科技是国内首家通过柔性脑机接口临床伦理审批的公司。目前, 脑虎科技在植入式脑机接口领域已经完成了核心技术突破以及关键器件制造。

### 科研产品



CereCube 高通量神经信号采集系统



NeuroInterface集成式BCI



MindExplorer软件系统



智能手机器人



SilkTrove & SurfTrove 柔性电极



皮层电极

# 脑虎科技

### 应用案例



犬类脑机接口案例

非人类灵长类柔性脑机接口案例

## 代表机构

**脑虎科技**  脑虎  
NeuroXess

NeuroXess脑虎科技成立于2021年10月,是一家通过柔性脑机接口技术来保护及探索大脑的生命科技公司。目前,脑虎科技在植入式脑机接口领域已经完成了核心技术突破以及关键器件制造。

**微灵医疗**  微灵医疗  
WE-LINKING

微灵医疗成立于2019年4月,主要从事医疗级全植入式无线脑机接口全栈技术研发,是深圳脑机智能融合技术初创企业。微灵医疗发端于脑科学与植入式脑机接口的深耕研究,奠基于医疗级脑机接口系列关键技术从零到一的突破,定位于脑机接口的医疗产品研发、应用示范和普及推广。

**臻泰智能**  臻泰智能  
ZhenTec

臻泰智能是一家脑机接口医疗产业化研发商,依托西安交通大学孵化,致力于各类脑控交互、VR/AR及医疗康复机器人系统的研发、生产及销售。



## 可回收火箭进入「中国轨道」 工业化难题被突破，商业航天迎来模式闭环

近8年来,美国太空探索技术(SpaceX)公司的“猎鹰9号”火箭凭借着超高的运载效率和可回收技术,在世界运载领域独树一帜。随着航天产业的发展,研制低成本、高可靠、使用方便灵活的可重复使用运载火箭是下一阶段航天技术发展的重要方向之一。当前阶段,运载火箭的回收复用已是世界航天的发展趋势,中国众多商业航天企业也正在研制各自的可回收复用型号。

# 推荐理由

## 1. 可回收相比于一次性运载火箭能大幅降低成本

航天技术的发展和规模一直受到发射成本的制约。从成本结构来说，硬件设备价值在单个火箭中占比较高。火箭中占总重90%以上的都是推进剂，但推进剂其实只占成本的1-2%。一二级火箭均呈现硬件成本占比高、推进剂占比低的特点。若在实际可回收的条件下，可以保留火箭发动机和制导系统，火箭可重复使用的次数足够多，那么硬件成本分摊到单次发射的成本将大幅下降。

## 2. 星座大规模部署的需求推动火箭向可重复使用发展

由于火箭相关配套零件的生产速度往往是缓慢的，可重复使用的火箭相比于单次发射的火箭在产能有限的条件下可以实现规模更大周期更短的发射，能在一定程度上提高整体运力规模，同时提升了发射频次，从而能更快实现星座（用于同一任务的大规模卫星组网）的大规模部署。

2023年7月9日，中国在酒泉卫星发射中心使用长征二号丙运载火箭，成功将卫星互联网技术试验卫星发射升空。开启“中国要发射近1.3万颗”卫星的布局，将实现“中国星链”。



星座组网的核心需求是把足够数量的卫星发射到指定轨道当中，拆解这个规模庞大的任务，一方面是提升运力，另一方面是提升发射频次，在这两个基础都能保证的情况下，要尽可能的降低发射成本，而可回收火箭正是降低成本的利器。

——蓝箭航天

## 3. 国内首次液体火箭全尺寸一子级的重复使用飞行试验验证成功，动力系统的推力调节技术实现突破

2023年11月2日，星际荣耀自主研发的液氧甲烷可重复使用验证火箭双曲线二号（代号SQX-2Y）在我国酒泉卫星发射中心点火升空，约1分钟后，火箭在目标着陆点平稳精确着陆，火箭状态安全恢复，飞行试验任务取得圆满成功。

通过此次飞行试验，验证了星际荣耀重复使用液体火箭一子级总体方案的正确性，连续变推力液氧甲烷动力系统方案的正确性，火箭着陆段垂直软着陆导航、制导与控制方案的正确性，液氧甲烷火箭加注、发射、回收及后处理流程的正确性与协调性，验证了火箭着陆装置缓冲及支撑功能和性能，获取了液氧甲烷火箭重复使用关键技术核心试验数据。这是国内首次成功验证液体火箭一子级可重复使用方案，是中国在火箭可回收工程化上的一大突破。

## 4. 液氧甲烷发动机运载火箭的工程化突破

2023年7月12日，“朱雀二号”成为全球首款发射入轨的液氧甲烷运载火箭，蓝箭航天国内首次突破液氧甲烷中型液体运载火箭关键技术，填补国内液氧甲烷火箭型谱空白。目前国际上公认的新一代可重复使用火箭发动机技术方向就是液氧甲烷发动机，液氧甲烷以环保、高可靠、低成本、不易积碳结焦、可重复使用等特点，成为可回收火箭的理想推进剂。

“

未来有大批量的市场需求,对于火箭批量化、重复使用,单次的运载能力有更高的要求,液氧甲烷火箭是较为可靠的选择。当下需要先摸清液氧甲烷火箭的特性,工程的原理,下一阶段通过液氧甲烷发动机,建立更大的火箭——直径十米级别,质量到两百吨的火箭。因为这个阶段才能到人类航天的下一个阶段,通过火箭实现航天的航空化,洲际火箭才可能实现。未来五年中国所有的商业火箭全部会选择液氧甲烷火箭,从当下和历史存证都能看到。

——蓝箭航天

“

液氧甲烷火箭技术实际上已经初步验证,包括蓝箭航天在今年实现全球首飞液氧甲烷火箭入轨(整箭级别产品验证),同时这项技术有相当明确的下游场景,甚至头部客户已经开始进行方案验证、招投标、采购。系统级技术创新+工业化体系打造=性价比大幅优化。我们认为2024年开始,中国的商业航天或将进入技术商业化的上行周期。

——泰合资本合伙人 梅林

## 5. 全球最先进可回收火箭「星舰」本年内进行两次试飞

2012年,SpaceX创始人兼CEO埃隆·马斯克提出了星舰的早期概念,当时它被称为“火星殖民运输系统”(Mars Colonial Transporter)。

星舰被称为“人类历史上体积最大、推力最强运载火箭”,其意义主要可以有两方面:可复用性与运力。

之前的猎鹰9号,则仅能回收一子级,而星舰目标是两级运载火箭皆可回收并重复使用。若这一目标实现,星舰将取代SpaceX的其他所有运载火箭和航天器(猎鹰9号、猎鹰重型火箭与龙飞船),接下卫星发射与载人运输的重担,甚至在未来飞向月球与火星。

- 2023年4月20日,星舰首次与超重型火箭合体试飞,由于出现两级火箭未按计划分离等问题,自毁系统将星舰烧毁在墨西哥湾上空。此次试飞积累了大量宝贵的数据,SpaceX对星舰系统和发射平台进行了1000多处改进,包括采取“热分离”方式实现两级火箭的分离以提升运载性能、给平台加装喷淋系统等。
- 11月18日,星舰重型运载火箭时隔7个月后再次发射。底部高70米的第一级“超级重型”助推器,与顶部高50米的第二级“星舰”飞船成功分离,虽然最后因为没有进入预定轨道而触发了自毁系统试飞失败,但相比第一次发射解决了推进器和火箭的分离等关键问题,取得了巨大进步。SpaceX下一步最主要的,应该就是要解决火箭和推进器的回收问题。

# 技术原理

## 1. 什么是火箭的可回收

火箭回收技术,是指火箭发射后回收,并重复使用的技术。一次性使用的运载火箭发射成本很高,使用火箭回收技术能够降低发射成本。

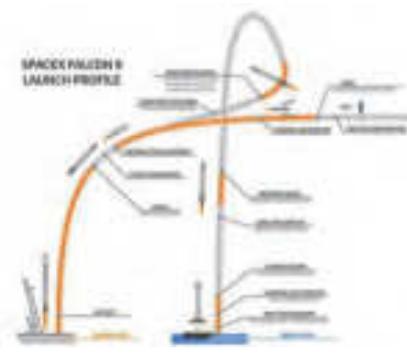
## 2. 三大主流回收方式

### (1) 垂直起降回收

通过重启一级火箭发动机以减速并调整成和地面垂直的姿势回到指定地点的火箭回收方式。重启发动机需预留推进剂，火箭运载能力损失约30-50%。

SpaceX的猎鹰9号火箭正是目前最具代表性的垂直起降可回收火箭。相比于伞降回收，垂直起降回收着陆精度高、着陆冲击小，能实现包括一级发动机等核心部件在内的箭体整体无损伤回收。

相比于带翼水平回收，垂直起降回收对地面场地及保障要求低、对运载效率影响小、技术难度较低。因此，垂直起降回收因商业应用价值最高而成为香饽饽，是当前火箭回收的最主流方式。



“猎鹰9号”火箭一级回收飞行剖面

“

垂直回收一方面继承了传统运载火箭的结构，另一方面它是动力型元素，返回过程是通过发动机工作降低速度。这种方式消耗的是推进剂，对火箭的结构和外形设计难度降低，对发动机的要求提高，而发动机反喷功能的实现难度并不是特别大，所以综合权衡下，垂直回收成为技术门槛低，效率非常高的一种技术路线，也得到了目前全世界的认可。

前些年欧洲还在坚持传统的基于空气动力学的方案，而现在他们也开始采用这种跟猎鹰9号一样的模式。日本和俄国同样，印度现在虽然没有很官方的计划出来，但他们目前的概念性设计也是垂直回收方案。国内也是一样的，从运载火箭的角度来说都采用了垂直回收的方案。

——星际荣耀

### (2) 伞降回收

是一级火箭完成级间段分离后使用降落伞进行回收的方式，具有技术成熟度高、运载能力损失小（约10%）、成功率高等优势，尤其能减少对地面人员设施的危害。但伞降回收对着陆地形要求高且难以控制着陆点，火箭落地后发动机也随之报废，与真正意义上的“回收利用”差距很大。

采用伞降回收的代表火箭如联合发射联盟公司（ULA）的“火神”火箭，该火箭一级发动机结束工作完成级间分离后，使用降落伞进行减速，由直升机在空中实现回收。

### (3) 带翼水平回收

是一级火箭装备机翼、起落架等结构利用空气动力滑翔降落的回收方式。该方式对火箭整体结构设计要求高，机翼结构重量等因素致使火箭运载能力损失约40%。

采用带翼飞回回收的典型代表是俄罗斯提出的贝加尔号有翼助推器方案，贝加尔号为第一级带翼飞回式助推器，可飞回发射场以自动方式像飞机一样着陆，可重复使用100次。

## 3. 可回收火箭关键技术

与传统的运载火箭相比，可回收火箭主要有以下关键技术需要攻克。

### (1) 可重复使用总体设计技术

全流程不仅涉及发射上升段,还涉及到回收返回段。从趋势上看,总体设计还需考虑全寿命周期的系统融合设计,采用智能技术,提高火箭的智能化水平,有利于重复使用运载火箭结构效率的提升。



可重复使用总体设计技术

### (2) 可重复使用火箭发动机技术

返回时发动机多次起动需克服失重、超重、高动压等复杂飞行环境带来的各种内、外部干扰,发动机**多次起动及推力精确调节**技术研究是实现火箭回收的基础。

发动机点火,火箭起飞,发动机关机,火箭一级和二级分离;火箭一级在返回过程中再启动,先做一个减速;高速进入大气层后利用大气减速;在着陆时发动机再点火,继续进行减速和控制。在其中发动机要经历 2~3 次启动。

回收落地关键时刻,需要发动机快速变推力关机,以减小落地瞬间发动机后效冲量对箭体的冲击和扰动。猎鹰9完美实现可靠回收的关键技术之一就是采用了面关机针栓式喷注器技术。

### (3) 可重复使用控制制导关键技术

垂直起降运载器再入返回过程中经历多个飞行段,全程高精度导航系统的设计需要综合考虑导航精度、观测条件、安装位置及大气和发动机喷气尾焰扰动等使用环境限制,可采用天文导航、视觉辅助、激光雷达及卫星导航地面差分站相结合。

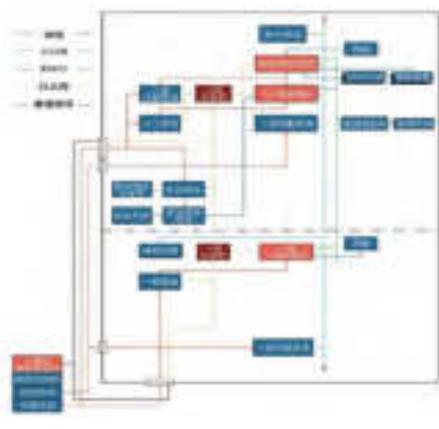
### (4) 结构轻质化技术

运载能力取决于结构效率,轻质化的结构设计是提升结构效率的重要途径。结构轻质化的前提条件是火箭飞行和返回过程中力/热环境的精细化辨识,针对重复使用,需兼顾结构材料的强度、疲劳、损伤容限。结构的能力提升则取决于轻质化比强度高材料的应用、先进制造工艺的应用。在材料结构应用方面,二级结构、常温贮箱采用碳纤维复合材料结构。

特殊工艺和复杂结构则可采用3D打印技术,减少零件和工艺非必要结构,例如“人族”火箭的结构和发动机使用专门的3D打印合金制造,与传统火箭相比,部件数量减少了99%。

### (5) 一体化综合电子技术

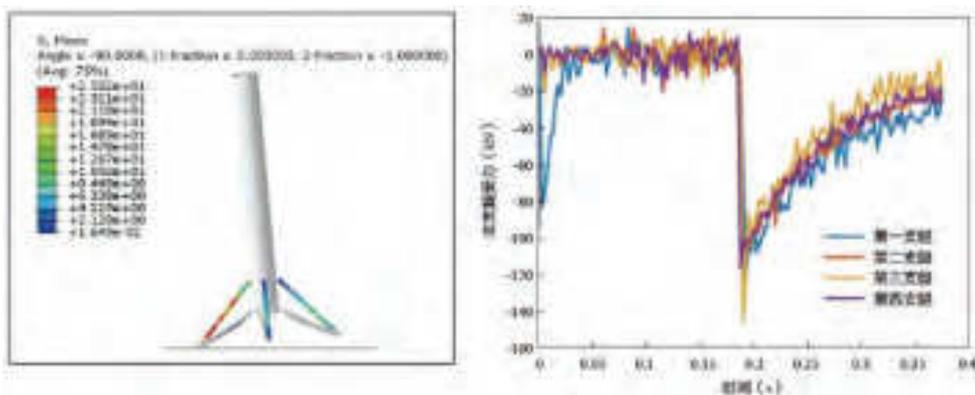
传统的航天电子系统按控制系统、测量系统、推进剂利用系统、总体网系统等开展独立系统研制,存在计算、供配电、通信等资源重复,电子设备功能单一,CPU计算资源利用率低等问题。因此,需要统筹考虑飞行控制、测量、测发控等功能需求,突破传统的设计约束,解决各功能存在交叉重叠等现象。



一体化综合电子技术架构图

### (6) 缓冲着陆技术

火箭着陆过程中会受到冲击载荷,其内部缓冲器是着陆支腿的关键组成部分,用于吸收火箭着陆期间的动能。其次,支腿具有足够的刚度和强度,保证碰撞过程中不被损坏,并经过简单检测维修后能够重复使用。着陆支腿一般由着陆腿、缓冲装置、展开锁定机构及足垫等组成。火箭着陆支撑机构的作用是:能够在火箭起飞上升阶段收拢、压紧,返回软着陆前展开、锁定,能够缓冲着陆瞬间的冲击力,碰撞时能够调整适应火箭姿态,以防侧翻,着陆完成后承担火箭的重量。



某型商业运载火箭缓冲结构回收仿真

“

从火箭的使用角度来说,回收后对火箭的处理技术,包括怎么把推进剂卸出,怎么把高压气卸出,发动机回温等一系列回收后的处理过程,在实际使用中是非常影响回收火箭应用的一个技术,也是非常关键的需要验证的一个流程。

目前我们的火箭已经在着陆区进行回收处理后进行检测,为我们后面入轨火箭一子级复用效率的提升,或者说复用技术在入轨火箭上的应用进行了一次重要验证。后续我们也会在高空实验经历过更恶劣的工况之后,进行更加综合性的检测,得到一个更全面、更完整的验证,为后续发射提供基础。

—星际荣耀

## 4. 可回收火箭现状

放眼国内外研制进展,已经真正意义上成功实现垂直回收全比例火箭的仅有美国的太空探索技术公司SpaceX和蓝色起源Blue Origin,这也直接说明了这种火箭回收方式的“闯关”难度级别之高。

国内民营方面,星际荣耀、蓝箭航天、深蓝航天、星河动力等一系列公司都有开发和试验可回收重复使用火箭的计划,进展也都相对顺利。它们基本上都是采用“变推力可重复启动发动机+栅格舵+脚架”的模式实现火箭自主返回着陆,这对发动机性能提出了极高的要求,为此这些公司也都在积极自主研发发动机。

“

我们现在研制的可重复使用火箭双曲线三号预计2025年首飞，2026年投入商业化发射。从技术验证角度，一般来说如果连续两次成功就可以进入大规模商业化发射。

——星际荣耀

“

最早2025，最晚2026年，我们会争取实现运载火箭一级箭体的可重复使用，我们坚信我们会成为国内最早的中大型火箭一级箭体可重复使用的企业。

从明年开始我们会向市场开始商业化的交付，从3发到6发再到12发，实现每年翻一番的状态，并且在2026年可以率先取得盈利。这是中国商业航天完整的从0到1的过程，到2026年我们花了11年的时间。对标SpaceX从0到1的过程，我们基本处于同一水平。

——蓝箭航天

## 国内首次液体火箭一子级回收试验成功

北京星际荣耀空间科技股份有限公司成立于2016年10月，是中国首家具备自主研制运载火箭能力并成功完成高精度入轨发射的民营企业。近日，公司完成了国内首次开展的液体火箭全尺寸一子级的垂直起降（VTVL）与重复使用飞行试验，实现了中国商业航天在可重复使用运载火箭技术上的重大突破。

# 星际荣耀

### 双曲线系列运载火箭产品型谱



四级小型固体运载火箭，主要用于近地轨道和太阳同步轨道卫星的发射任务。2019年7月25日在酒泉卫星发射中心成功完成高精度入轨发射，实现了中国民营商业运载火箭成功入轨零的突破。

双曲线一号小型固体运载火箭 (SQX-1)



SQX-2Y火箭是用于验证液氧甲烷重复使用运载火箭一子级垂直回收和重复使用的一型验证火箭，火箭采用3.35m全尺寸箭体直径，全长约17m，装配星际荣耀自主研发的焦点一号深度变推力液氧甲烷发动机（代号JD-1），高精度圆满完成飞行试验任务。

双曲线二号可重复使用技术验证火箭 (SQX-2Y)



该型火箭是目前已知国内规划的直径最大、运载能力最强、每公斤载荷发射价格最低的可重复使用液体运载火箭，目前正在开展方案设计和关键技术攻关，计划于2025年首飞。

双曲线三号中大型可重复使用液氧甲烷运载火箭 (SQX-3)



# SpaceX

## 成就SpaceX的「猎鹰9号」

猎鹰9号采用反向垂直起降方案,在箭身第一子级加装了四个折叠的可展开式栅格舵,用于恢复转向和控制四个着陆缓冲器。

- 2015年12月,首次实现了内陆着陆场的发射和回收。
- 2023年,猎鹰9号Block 5,连续174次发射,100%成功率。



## 全球最先进可回收火箭「星舰Starship」



2018年11月至今,围绕正式命名为**星舰(Starship)**的架构不断取得实质性的进展。

本年度星舰进行两次试飞均失败。

这一新的全不锈钢版箭船组合架构中,助推火箭为超重,即超重鹰火箭,技术参数与BFR一样:高63米、直径9米、31台猛禽发动机,新变化是箭体材料由碳纤维复合改成不锈钢合金。

上面级即飞船级为星舰,技术参数也与BFR相同:高55米、直径9米、7台猛禽发动机,关键改进包括:增加驱动鳍和隔热层、加压空间扩大至超过1000立方米、材料也改为不锈钢合金等。

## 代表机构

### 星际荣耀 星際榮耀

是中国首家具备自主研发运载火箭能力并成功完成高精度入轨发射的民营企业。近日,公司完成了国内首次开展的液体火箭全尺寸一子级的垂直起降(VTVL)与重复使用飞行试验。

### 蓝箭航天 LANDSPACE 蓝箭航天

致力于构建以中大型液氧甲烷运载火箭为中心的“研发、制造、试验、发射”全产业链条,打造航天领域的科技综合体,为全球市场提供高性价比、高可靠性的航天运输服务。

### 中科宇航 中科宇航

是中国研制中大型运载火箭的商业航天火箭企业,致力于发展创新性、经济性运载火箭,具备承担低轨卫星星座建设和空间站货物运输的能力。

### 星河动力 星河动力

致力于为人类低成本、高频次进入空间提供解决方案,也为天地往返、空间运输、太空资源开发提供澎湃动力让太空资源更加轻松地融入我们的生活。其核心产品及服务包括“谷神星”系列小型固体运载火箭发射服务,以及“智神星”系列中型重复使用液体运载火箭发射服务等。

### 深蓝航天 深蓝航天

一家以液体可回收运载火箭研发及提供商业发射服务为主的,集科研、产试验、发射服务为一体的商业航天新企业,主打产品“星云”系列可回收复用运载火箭。



## 可控核聚变里程碑： 点火成功，燎原商用产业链

今年，我国新一代人造太阳“中国环流三号”取得重大科研进展，首次实现100万安培等离子体电流下的高约束模式运行。

近年来，核聚变的底层技术不断突破，尤其是高温超导材料的突破使得产业界开启了核聚变商业化转化的步伐。核聚变作为一种全新的能量来源，因其自身极高的能量密度、清洁等特性，已经成为产学研对“终极能源”的定义。一旦核聚变能够实现从核能到电能的转变，将颠覆人类社会的能源供给形式。此外，因其自身能量特质，使用核聚变作为能量源，可以进一步拓展人类现阶段可探索的边界，包括AGI的实现、深空探索等。

# 推荐理由

## 1. 核聚变本身的优势

核聚变,简单来说,是两个轻原子核在超高温或高压的条件下,远离核外电子的束缚,结合成较重的或质量较大的原子核的过程。由于这个过程会以极高的效率释放巨大的能量,而且污染小,因此可控核聚变一旦实现,人类将彻底解决能源问题,因此被人称之为“终极能源”。

 <p><b>从原料上来看</b></p>	 <p><b>从能量供给角度看</b></p>	 <p><b>从安全性角度看</b></p>	 <p><b>从环保的角度看</b></p>
<p>核聚变是两个轻原子核结合成一个较重的原子核并释放出巨大能量的过程。核聚变反应发生在一种叫作等离子体的物质状态中。</p>	<p>核聚变燃料的能量密度极高,0.6吨铀所蕴含的能量相当于2000000吨煤,换算成石油则是1300000吨,相当于30吨氧化铀所蕴含的能量。</p> <p>核聚变产生的能量非常大——是核裂变反应的四倍,而且聚变反应可以成为未来聚变动力堆的基础。理论上,只要有几克这些反应物,就有可能产生一太(万亿)焦耳的能量。这大约是发达国家的一个人在80年内所需要的能量。</p>	<p>可控核聚变更可靠。核聚变反应的条件极为苛刻,燃料必须达到上亿摄氏度高温,满足足够高的密度条件,但凡有一个条件不满足,都会导致核聚变反应温度、密度的下降,使聚变反应停止。所以,当聚变堆出现事故苗头时,即时停止加料可即时停止核聚变反应;不产生余热,也就杜绝了堆芯熔化和泄漏事故的发生。</p>	<p>采用核能作为动力源,不仅减少了对化石燃料的依赖,还可以降低数据中心的碳足迹。</p>

## 2. 高温超导材料的突破

高温超导磁体相较于同体积下的低温超导产生的磁场更强,且能够突破低温超导材料磁体磁场上限,推动可控核聚变装置的小型化与成本降低。

2020年上海超导带材良率已经达到90%左右且产品性能大幅提升。随着高温超导带材规模化和良率的提升,其价格已经下降到百元/米。高温超导材料的技术成熟与成本降低使得越来越多公司入局可控核聚变领域,加速了可控核聚变的商业化进程。近几年商业公司研发的紧凑型、小型化托卡马克装置均采用高温超导磁体。

## 3. 与AI相辅相成地发展

### • 核聚变为AI提供能量来源

咨询机构Tirias Research曾预测,到了2028年,数据中心功耗将接近4250兆瓦,比2023年增加212倍,数据中心基础设施加上运营成本总额或超760亿美元。AI训练的背后,是对能源的大量消耗。

什么样的能量供给,才能保障日益强大的AI训练能力?OpenAI创始人Sam Altman认为虽然太阳能与储能也可以实现能量供给,但核能才是最好的方式。其本人已经押注了一家核聚变初创公司Helion。

### • AI帮助提升核能系统的设计水平

人工智能已经参与到核电领域的设计、制造、运行和维护等各个环节中。

DeepMind最新研究使用RL算法,通过实验模拟将等离子体形状精度提升65%,并且新任务学习所需的训练时间减少了3倍以上。

## 技术原理

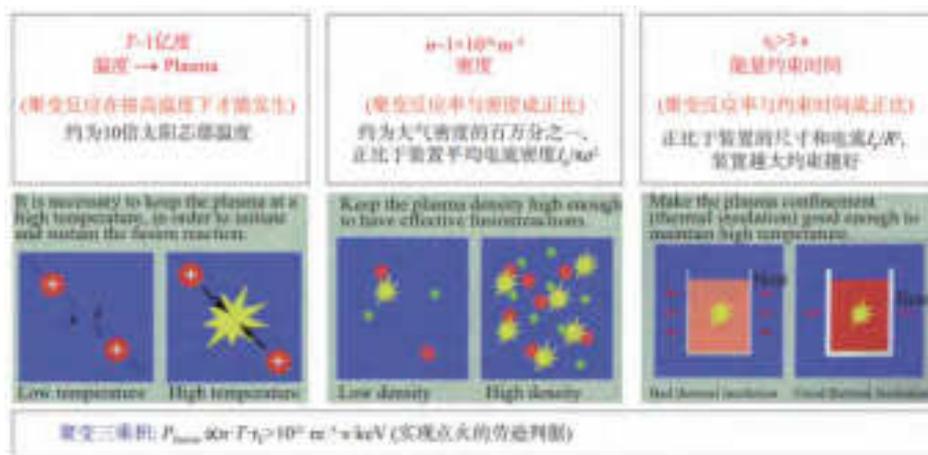
### 核聚变如何工作

核聚变是两个轻原子核结合成一个较重的原子核并释放出巨大能量的过程。核聚变反应发生在一种叫作等离子体的物质状态中。等离子体是一种由正离子和自由移动的电子组成的高温带电气体,具有不同于固体、液体和气体的独特性质。

太阳的巨大引力自然会诱发核聚变,但如果没有这种力,就需要更高的温度才能发生反应。在地球上,我们需要超过1亿摄氏度的温度和强大的压力,以使氘和氚发生聚变,还需要充分的约束,使等离子体和聚变反应保持足够长的时间,以获得净功率增益,即聚变放出的功率大于输入功率。

### 实现核聚变的三个条件

足够高的温度(T)、一定的密度(n)和一定的能量约束时间( $\tau_E$ ),三者的乘积称为聚变三乘积。



#### 1. 极高温

由于原子核带正电,必须在极高温下才能获得足够的能量以克服彼此间的库仑势垒;

原子核靠得更近时,通过量子隧穿效应产生核聚变反应的几率更大。

此外,温度也不能过高。过高温下原子核接近时间将缩短,反应几率反而减小。在地球上实现高效核聚变反应,需要将温度维持在1亿摄氏度以上。

#### 2. 密度

用以衡量等离子体约束区单位体积内粒子的个数。保持足够的密度意味着单位体积内拥有更多的氘氚原子核,能够有效提高原子核间的碰撞效率,以获得足够的核聚变反应率。

#### 3. 能量约束时间

高温等离子体的能量以辐射和热传导的形式逸出,为唯象描述热传导损失功率,将等离子体总能量热传导损失所需时间定义为能量约束时间。

高能量约束时间意味着装置具有良好的隔热性能,能量流失得缓慢,可以进一步提高核聚变反应率。

## 主流技术路线

达到聚变条件后,还要对高温聚变物质进行约束,以延长可控核聚变反应时间。可控核聚变的约束有两种主要途径:磁约束和惯性约束,目前磁约束被认为是最容易率先落地的方式,也是各国研究最多的方式。国际上已知最大的可控核聚变装置ITER和中国的东方超环(EAST)使用的均为磁约束。

### 磁约束

利用强磁场将等离子体约束在装置空间内,以避免高温等离子体与壁材料接触。

#### · 约束装置

当前磁约束核聚变研究的主流主要分为托卡马克和仿星器两类研究装置,前者在工程实现难度上低于后者,也是目前全球磁约束路线的机构选择的主流装置。

#### · 材料

托卡马克高温等离子体真空室中,面向等离子体的部件通常成为第一壁部件。第一壁部件接收来自等离子体的高能粒子的轰击。第一壁材料需要在长时间忍受来自芯部等离子体的强热流和强粒子流轰击。由此可见,磁约束核聚变装置中第一壁材料的选择尤为重要。目前,广泛研究的金属材料包括钨、钨和铍,ITER已经确定使用钨作为偏滤器材料、铍作为第一壁材料。

### 惯性约束

激光聚变是点燃核聚变反应的一种方式,也是磁约束的一种替代方案。激光聚变通过惯性约束,使用高功率激光器加热和压缩含有氢同位素(例如氘和氚)燃料芯块的微小球形胶囊。强烈加热胶囊表面,使燃料发生微内爆,结果燃料芯块表层被烧蚀,并发生爆炸。这个过程所产生的惯性保持燃料长时间受到约束,足以产生聚变反应。

惯性约束核聚变的特点是驱动器和反应器是分离的,因而相对磁约束核聚变来说结构较为简单,不需要庞大的磁路系统,系统的工作条件也相对宽松。然而由于惯性约束聚变的连续性仍有待解决,尚无法应用于能源领域。目前美国国家点火装置(NIF)、中国神光III装置是正在运行的最大的激光聚变驱动器。

## 技术的商业落地:核聚变发电

核聚变发电是证明其商业可行性的关键一步。

核聚变发电利用轻原子核“聚变”所释放的能量。当两个轻原子核融合时,所产生的原子核质量比原来两个原子核质量之和略轻。这一质量差没有消失,而是被转化为能量。这种微小的质量损失转化为巨大的能量,使得追求核聚变能非常值得。

### 核聚变商业化的过程

在可控核聚变领域,提到商业化离不开Q值。所谓Q值即能量输出/输入比值。当Q值大于1时,能够证明核聚变反应可以产生比启动反应过程所提供的能量多得多的能量。

迄今,JET已从24兆瓦加热功率产生16兆瓦聚变功率,实现的最佳增益Q值为0.67。然而,实现商业价值的发电,需要将Q值做到30左右。目前全球最大的核聚变实验堆ITER将目标Q值设到10以上。此外,为了保持反应时长,使核聚变成为稳定的电力来源,需要将Q值维持在5。Q值为5代表临界阈值,高出这个阈值,等离子体开始自加热,以维持核聚变反应。

在ITER证明通过核聚变获得净能量后,证明核聚变能产生净电力是下一阶段任务,这也是核聚变示范场的作用所在。这个阶段的关键挑战是如何使聚变等离子体保持足够长时间的稳定性,以便持续地产生能量。

虽然目前ITER还尚未完全建成,但建设过程中,由于覆盖的产业链足够广,在建设核聚变反应堆的过程中,也能够推动产业链各环节发展。

在整个产业链当中,材料是最关键的一环,尤其是第一壁材料。



## 「脉冲不点火」系统

### 等离子体加速器



一个6×40英尺(1英尺=0.3048米)的杠铃形状的「等离子体加速器」,使用强大磁体将气体混合物加热到能让原子分裂的程度,在设备的两端形成被称为等离子体的超高温物质环。

磁体会以每小时100万英里(约合161万公里)的速度推动这些环相互撞击,并在装置中间进一步压缩它们,从而产生超过1亿摄氏度的温度,引发核聚变反应,其中原子核碰撞,质子和中子结合,释放出各种粒子,并产生能量。

### 第七代原型核聚变发电机“北极星”

Helion的核聚变过程可以直接回收电力。当等离子体继续加热和膨胀时,它自己的磁场会推动设备周围磁体产生的磁场,驱动了带电粒子流,也就是所谓的电流。而这反过来又给一个被称为电容器的能量存储装置充电,该装置为磁体供能发射下一个脉冲。



# HELION



## 能量奇点

### 高温超导强磁场托卡马克装置

#### 高温超导托卡马克理论样机(洪荒70)

洪荒70是能量奇点独立研发建造的全高温超导托卡马克装置，位于上海自贸区临港新片区。

该装置自2022年初开始研发，于2023年8月进入总体安装阶段，计划于2023年底建成。洪荒70装置的全部磁体系统均采用高温超导材料建造。届时，洪荒70将成为全球首台建成运行的全高温超导托卡马克装置，率先在完整装置层面验证高温超导托卡马克技术路线的工程可行性。

#### 25特斯拉高温超导聚变装置磁体(经天磁体)

经天磁体是基于高温超导材料的大孔径强磁场环向场(TF)磁体，高约3米，磁场强度超过25特斯拉，工程电流密度超过150兆安培每平方米。

该研发项目自2022年第四季度开始，目前已完成概念设计、工程设计，进入加工建设阶段，计划于2024年完成建造和测试，届时将成为全球场强最高的聚变装置磁体。

## 代表机构

### 能量奇点



能量奇点2021年成立，是中国首家聚变能源创业公司，致力于研发有商业发电潜力的高温超导强磁场托卡马克装置及其运行控制系统，为未来可控核聚变发电堆提供高性价比、高可靠性的核心组件和服务。

### 星环聚能



星环聚能成立于2021年10月。公司致力于聚变能商业应用及相关技术研发，以建成我国首个商用可控聚变堆为己任，潜心于小型化、商业化、快速迭代的可控聚变能装置，专注成长为集研发、设计、运维于一体的全球顶尖的聚变能科技企业。

### 上海超导



上海超导成立于2011年，是一家以新材料等学科为基础、利用镀膜技术，从事第二代高温超导带材研发、应用及销售的高新技术企业。第二代高温超导带材可以在温度大大高于绝对零度的情况下产生强磁场以约束高温等离子体、使其不规则热运动发生连续碰撞、产生大量的聚变反应并释放出能量，从而实现聚变的可控运行。上海超导的国产超导带材已进入产业化推广阶段，带动了超导电力、超导节能、可控核聚变、高速磁悬浮交通、高场磁体等一系列新兴产业的形成与发展。

# 其他提名前沿科技趋势

## HBM

### 推荐理由

生成式AI爆发后, AI服务器中GPU搭载HBM已经成为主流。HBM通过增加带宽和减少功耗有效解决了“存储墙”和“功耗墙”问题, 是目前冯诺依曼架构下, 能够满足AI高性能计算要求的量产存储方案。现在, HBM已经成为生成式AI下, AI训练芯片的标配, 主流的大模型训练芯片A100、H100等均应用了HBM。除此之外, 其他AI训练芯片如AMD MI300也将采用HBM。

HBM的优势主要体现在3D堆叠封装上, 通过直接在芯片上打孔(TSV)和布线, 将芯片直接进行电气连接——实现高速互联和高带宽的同时也能降低数据传输功耗。

作为一种高性能存储器技术, HBM将成为未来存储厂商的技术演进方向。

### 技术原理

HBM(高带宽存储器)是一种新型CPU/GPU内存芯片, 基于2.5D/3D封装技术将DRAM Die垂直堆叠, 在较小的物理空间里实现高容量、高带宽、低延时与低功耗。DRAM Die之间通过TSV(硅通孔)的方式连接, 逻辑控制单元对DRAM进行控制, GPU和DRAM之间通过u Bump和Interposer(起互联功能的硅片)连通。

目前最先进的HBM为第五代HBM3以及第六代HBM3E, 封装的DRAM Die层数达到12层。HBM主要由SK海力士、三星与美光三家公司供应, 其中SK海力士具有先发优势, 三星、美光公司呈追赶态势。

## 忆阻器

### 推荐理由

今年, 清华团队研制出全球首颗全系统集成、支持高效片上学习的忆阻器存算一体芯片。据悉, 该芯片可执行的任务包括运动控制、图像分类和语音识别。相同任务下其能耗仅为ASIC的1/35, 同时有望实现75倍的能效提升。

在人工智能高算力需求下, 基于忆阻器的存算一体架构能够突破冯诺依曼架构下的“存储墙”和“功耗墙”, 以更低的延迟和更低功耗进行计算。此外, 忆阻器的非易失性也使其在计算密度方面有更大潜力。

### 技术原理

到目前为止, 设计电路时使用的无源元件, 除了电容器、电阻器和电感器, 但还存在第四个基本元件, 称为“忆阻器”。忆阻器是在电压或电流输入下电阻可编程的器件, 阻值与流经的电荷相关且具有非易失性(可记忆流经它的电荷量)。

研究发现, 忆阻器电学特性与脑内神经突触的工作原理相似, 单个忆阻器便可模拟一个突触的基本功能, 与传统使用多个晶体管和电容器相结合的互补金属氧化物半导体(CMOS)来模拟一个突触相比, 减少了很多能耗, 也降低了集成电路的复杂性。

根据材料和物理机制,忆阻器件可分为阻变随机存储器(ReRAM),相变存储器(PCRAM),磁随机存储器(MRAM)和铁电随机存储器(FeRAM)四种。其中,MRAM产品进入量产阶段,FeRAM和PERAM有小规模应用,而ReRAM尚未实现商业落地,目前处于测试验证阶段。

## 卫星互联网

### 推荐理由

今年11月,马斯克宣布星链(Starlink)已实现现金流收支平衡。自2019年以来,星链已经将5000颗卫星送入近地轨道,目前有4200多颗保持活跃状态。今年10月,SpaceX星链官网推出“Starlink Direct to Cell(星链直连手机)”业务,可与现有LTE手机配合使用,无需更改硬件、固件或特殊应用程序,便能实现“无缝访问文本、语音和数据”。国内,银河航天已完成“翼阵合一”卫星的二维展开关键技术攻关,未来,该卫星将用于支持手机直连卫星的宽带通信需求。此外,今年7月23日,银河航天研制的首款柔性太阳翼通信卫星“灵犀03星”顺利升空,主要功能为低轨宽带通信。

手机厂商方面,今年8月,华为Mate 60 Pro发布,成为全球首款支持卫星通话的大众智能手机。

2023年,能够实现手机与卫星直接相连的技术成为新的发展方向 and 竞争焦点,各类厂商加速布局实现“手机直连卫星”。卫星互联网的商业路径已然清晰,在有限的太空资源下,各卫星制造商面临激烈的外部竞争,从而进一步加快卫星互联网的建设速度。未来,手机直连卫星、“天地一体化”建设将成为卫星互联网的竞争高地。

### 技术原理

按照卫星所在轨道高度的不同,可将通信卫星分为低轨(LEO)通信卫星,中轨(MEO)通信卫星和地球同步轨道(GEO)通信卫星。近年来,为了避免中高轨卫星引起的信号衰减和时延,低轨卫星逐渐成为研究重点。低轨卫星具备质量轻、结构简单且成本低等优势,但由于卫星高度低,要实现地球表面的信号全覆盖,需要发射大量卫星组件卫星星座。

### 定义

卫星互联网是指基于大规模卫星星座、以通信卫星为接入手段的互联网宽带服务模式。不同于分别以地面基站和光缆通信为媒介的移动通信和光缆通信仅能向地面终端发射信号,卫星互联网则可以同时向地面和空中终端提供宽带互联网接入服务。

### 趋势

卫星互联网与地面网络融合是6G的关键支柱和技术,星地融合通信网络的形成将成为未来全球网络覆盖的发展趋势。

# 二氧化碳合成糖

## 推荐理由

### 1.解决了传统制糖过程中的原料供应和效率问题

人类对糖的获取主要依赖于植物类生物质资源,这种传统的“二氧化碳-生物质资源-糖”的加工过程,受到原料供应和植物光合作用能量转换效率的限制。而二氧化碳人工合成糖这一技术,可以摆脱自然合成途径,提供一种灵活可控可拓展的糖制造模式。

### 2.二氧化碳合成糖新技术达目前人工制糖路线中碳转化效率最高水平

今年8月16日,中国科学院天津工业生物技术研究所通过化学-酶耦联的人工合成系统,研发出了人工转化二氧化碳从头精准合成糖技术,相关成果在《科学通报》杂志发表。

这项技术比自然光合作用糖合成更为高效、精准,且步骤更为简单,可以精准控制合成不同结构与功能的己糖,其碳转化率高于传统植物光合作用,同时高于化学法制糖以及电化学-生物学耦联的人工制糖方法,是目前人工制糖路线中碳转化效率的最高水平。

“

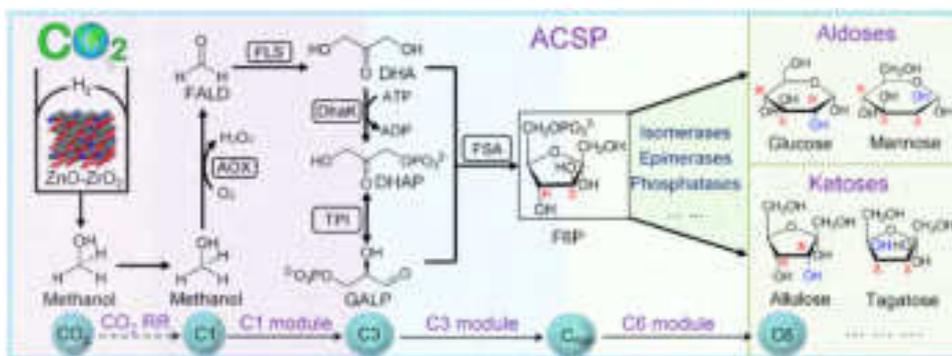
这项基于一碳化合物(二氧化碳或甲醇)精准合成己糖的工作,进一步拓展了一碳生物转化途径,通过醛缩酶及异构酶的挖掘实现了糖产物构型精准控制,一碳生物转化效率也获得显著提高,为人工生物制糖提供了一个有潜力的路线。

——西湖大学教授、德国工程院院士 曾安平

## 技术原理

天津工业生物技术研究所团队创建了一个人工二氧化碳到糖的途径(ACSP),以促进二氧化碳的碳原子对映选择性转化为特定的糖立体异构体。ACSP的路线图可以分为四个步骤:

1. 无手性控制的二氧化碳热化学还原为单碳甲醇;
2. 串联酶促醛醇缩合C1单元,以在磷酸化糖中建立特定的立体化学构型,如果糖-6-磷酸(F6P)与(3S,4R);
3. 酶异/外映体化,将磷酸化糖前体的手性中心改变为确定的对映体,如具有(3R,4R)的6-磷酸异丙糖(A6P);
4. 最终,将生成的中间“糖-磷酸”去磷酸化为立体糖。不可逆的去磷酸化步骤可以从热力学上推动整个反应向最终产物发展。



人工CO<sub>2</sub>到糖途径(ACSP)的设计

# AlphaFold

## 推荐理由

### 1.最新版AlphaFold能够预测蛋白质数据库中所有分子

2020年底,AlphaFold 2 预测蛋白质结构的准确率达到90%,而最新版AlphaFold将准确预测蛋白质结构的能力,推广至准确预测DNA、RNA、小分子配体和生物分子结构。

### 2.AlphaFold的扩展功能有助于加速生物医学突破

AlphaFold可为疾病通路的功能研究、基因组学、生物可再生材料、植物免疫、潜在治疗靶点、药物设计机制提供各种全新的可能性。

“

百度提出了全球首个开源、并提供在线服务、无需MSA输入的蛋白结构预测大模型HelixFold-Single,这一举措打破了AlphaFold2等主流依赖MSA检索模型的速度瓶颈,将蛋白结构预测速度平均提高数百倍,实现了秒级别预测,该工作的发表也为产学研各界带来了使用门槛更低、适用范围更广的蛋白结构预测解决方案。

——百度研究院

“

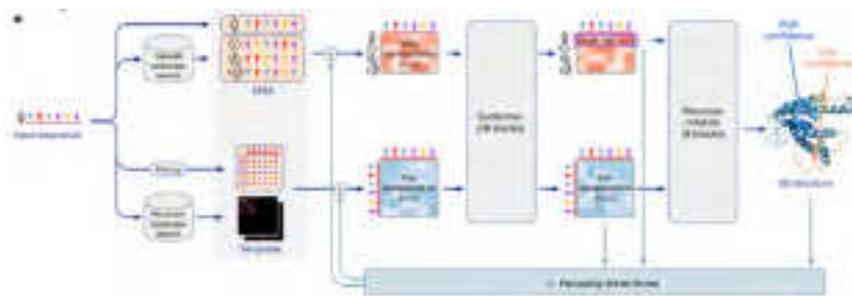
AlphaFold 是由AI驱动的技术,该技术通过预测蛋白质结构来实现蛋白质研究。这个程序通过利用大量蛋白结构数据集、先进的蛋白结构解析技术以及持续发展的AI硬件和软件,可以预测蛋白质相互作用的过程,以及蛋白质与DNA和小分子的相互作用。

AI技术进步可能是一个阶梯式的,以AI制药为例,蛋白质的复杂程度是好几个数量级的,超过人脑可以处理的程度,超过了一个单纯靠实验室筛选的程度。引入AI技术是一个必然的规律,AI处理能力除了从大量的、复杂的数据里面发现一些规律,还有在新层面的尝试,通过训练AI能不能有超越人类的洞察力,拓展人类对于科学认知的边界。

——联想创投

## 技术原理

AlphaFold 2 基于最先进的深度学习算法以及进化中蛋白质结构的守恒,使用一个端到端深度神经网络,该网络被训练为通过利用同源蛋白质和多序列比对的信息从氨基酸序列生成蛋白质结构。



Deepmind 《Highly accurate protein structure prediction with Alpha-fold》

AlphaFold 2包含三个模块:第一个是输入模块,以氨基酸序列作为输入,并生成MSA表示和对表示。第二个是Evoformer模块,从第一个模块获取MSA表示和对表示,并对他们传递到深度学习模块Evoformer。第三个是结构模块,实现了从蛋白质结构的抽象表示到目标蛋白质的三维原子坐标的过渡。

# ControlNet

斯坦福大学在ICCV23上获最佳论文奖的《Adding Conditional Control to Text-to-Image Diffusion Models》提出了一种名叫ControlNet的模型，只需给预训练扩散模型增加一个额外的输入，就能控制它生成的细节。

## 推荐理由

### 1. 增强扩散模型

类似Stable Diffusion的大型扩散模型可以通过ControlNet来增强，丰富其扩散方法。

### 2. 可个人训练

ControlNet可以在个人设备上训练，实现在有大型显存和多个GPU的集群上一样的效果。



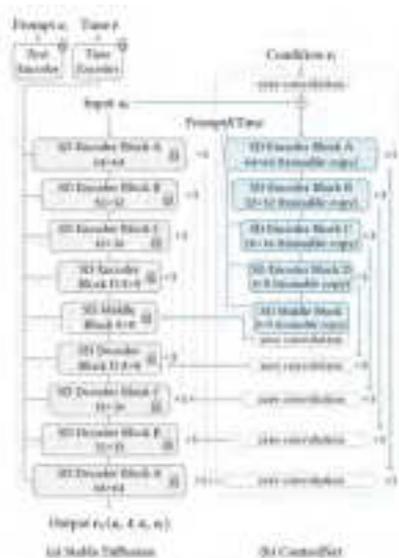
给出提示的稳定扩散

## 技术原理

它的思路和架构如下：

ControlNet先复制一遍扩散模型的权重，得到一个「可训练副本」(trainable copy)。原扩散模型经过几十亿张图片的预训练后，参数是被锁定的，而可训练副本只需要在特定任务的小数据集上训练，就能学会条件控制，在小数据量的情况下也能训练出很好的生成效果。

「锁定模型」和「可训练副本」通过一个 $1 \times 1$ 的卷积层连接，名为「0卷积层」。它的权重和偏置初始化为0，训练时速度非常快，接近微调扩散模型的速度。



Stable Diffusion的U-net架构与编码器部分及中间部分的ControlNet相连

——量子位



无提示下的稳定扩散

# 混合AI

## 推荐理由

### 1. 更高的AI算力需求带来的能耗负担

生成式AI搜索的查询成本是传统搜索方法的10倍，混合AI可以利用现有具备AI能力的数十亿边缘终端进行运算。边缘终端运行生成式AI模型的能耗显著低于云端，还能帮助云服务提供商降低数据中心的能耗。

### 2. 可靠性和可用性更强的处理方式

混合AI架构可以在云服务器和网络连接拥堵时，向边缘终端转移计算负载。

### 3. 安全与隐私保护

终端侧AI查询及个人信息完全保留在终端，不向云端暴露保密信息，可以从根本上保护用户隐私。混合AI架构中的「隐私模式」能够充分利用终端侧AI向聊天机器人输入敏感提示。



《隐私计算产业展望报告》

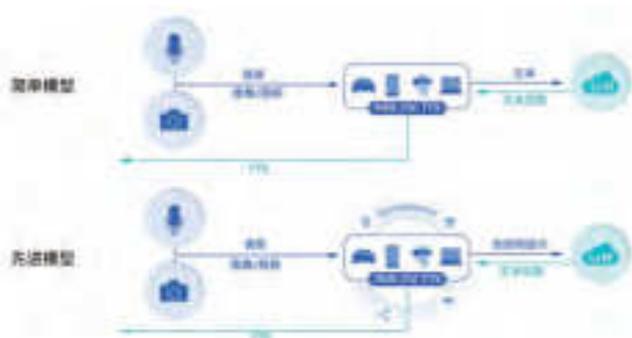
## 技术原理

混合AI指终端和云端协同工作，在适当的场景和时间下分配AI计算的工作负载，以提供更好的体验，并高效利用资源。在一些场景下，计算将主要以终端为中心，在必要时向云端分流任务。

而在以云为中心的场景下，终端将根据自身能力，在可能的情况下从云端分担一些AI工作负载。

终端侧神经网络或基于规则而运行的判决器(arbiter)将决定是否需要使用云端。

用户对处理结果不满意时，再次尝试发起请求就有可能引入一个更好的模型。终端侧AI处理能力会随终端升级和芯片迭代来提升，分流更多云端负载。



在先进模型中，终端侧AI将能承担更多处理，并提供经过改进的个性化提示。借助终端侧学习终端上的个人数据，通过编排器(orchestrator)程序协作，可创建用户个人画像，基于更多情境信息提供更完善的提示。

——高通《混合AI是AI的未来白皮书》

## 基础模型

“

大模型是当前最有前景的方向,有一些新的点值得密切关注,一是多模态的进一步打开,因为多模态是人类世界的本来样貌,AGI的发展趋势一定是朝向多模态。技术将从单一的文本、图像、视频(2D和3D),再到声、光、电,甚至分子、原子等各类模态,而且具备跨模态迁移的特性。通过设计不同模态对应的解码器,可以实现多模态之间的切换。未来理想的框架是:“多模态的对齐和融合+统一的编码器和解码器”,可以更好地解决多模态的理解和生成任务。

——腾讯研究院 前沿科技研究团队

“

AI大模型的企业应用将迎来爆发式增长,大模型将帮助各行各业的企业软件进行智能升级,未来将更加智能化、自动化、通过自然语言与人交流、大幅提高效率。AI生成技术也会迎来爆发,助力媒体、娱乐等行业的内容生成效率大幅提升、成本大幅降低,更加平民化。如何控制所生成内容的价值、质量、安全性等成为重大挑战。

——华为云

“

大语言模型已在理解、生成、逻辑和记忆等人工智能基础能力上取得突破,为发展通用人工智能带来曙光。大语言模型将成为人工智能技术研发和应用的基础设施,支撑产业升级。进一步地,以大语言模型为基础,融合语言、视觉和听觉等多元数据,发展跨模态大模型的脚步将加快。

今年以来,大模型是科技领域的热点。大模型究竟能为消费者、能为千行百业做什么,成为了互联网、科技企业探索的方向。随着大模型技术不断提升,大量AI原生应用将不断涌现,从而开启一个繁荣的AI原生应用生态。未来的AI原生应用将会走向多模态,在信息世界之外重构物理世界,数字技术与实体经济将深度融合,引领世界进入一个AI原生的时代。大模型不但将重构搜索引擎、地图等各类应用,也将催生更多的AI原生内容工具和生产力工具,解决过去解决不了或解决不好的问题。

——百度研究院

“

微软亚洲研究院认为现在需要聚焦人工智能的第一性原理,从根本出发,构建能实现人工智能效率与性能十倍甚至百倍提升,且具备更强涌现能力的基础模型。在勾画人工智能的未来蓝图时,基础模型是人工智能的第一性原理,其中,基础网络架构(Foundation Architecture)和学习范式(Learning Paradigm)是其两大核心基础。

新一代的基础模型应当具备两大特质:强大且高效。其中,“强大”体现在其性能、泛化能力和抵抗幻觉能力等方面的出色表现,“高效”则是指低成本、高效率和低能耗。

——微软亚洲研究院

## AI基因检测

“

基因检测AI的发展得益于超大规模基因数据的整合和大模型技术的进步,有望解决下游场景中数据受限问题,为解决生物学关键问题提供新视角。在基因检测AI的研究上,数据整合、算法开发和模型解释性是关键,对于发展大规模多模态基因数据整合有重要推动作用。商业上,基因检测AI在药物研发、疾病预防与治疗等领域具有巨大价值。例如,基于基因组数据加速新药开发、预测遗传疾病风险和药物不良反应,提供更精准的治疗。在科研应用方面,基因检测AI技术在基于序列的表观遗传预测、跨物种基因表达预测等领域具有重要的应用,将有效推动分子生物学、基因组学等生物学基础研究。

——上海人工智能实验室

## 无线通信技术

“

在人工智能时代,无线技术将不仅仅是通信的工具,还将成为人工智能发展和应用的“数据桥梁”和“第六感官”。IoT设备或智能终端生成的数据,需要强大、稳定的无线通信将其传输到云端的计算资源,才能实现实时的处理和分析;当人们需要更丰富的数据来训练人工智能模型时,无线感知则能够揭示那些视觉、语音、文字之外的现象,为人工智能提供更多元的数据类型,如视野外,穿过障碍物的透视,像呼吸心跳那样的微小移动,以及材料检测等等。

无线感知极大延伸了人类感官,激发了人们对其未来应用的无尽想象。比如将其应用于智慧农业,了解农作物的生长状况,更有效地优化浇水、施肥和农药播撒来提高产量;再比如将其应用于食品安全和质量监测,无接触地检测食品是否出现变质、成分是否符合描述、添加剂是否合适等等。无线感知还能有力推动人工智能的发展和应用。它能为大模型提供远超人类感官范畴的数据类型,帮助人工智能用更多方式去学习和了解真实世界,以及为未来的机器人提供更大的传感器,使之更精准地与环境互动。

— 微软亚洲研究院

## 量子计算

“

下一代计算机有可能在物理原理上完全不同,基于光的特性进行矩阵运算、基于量子物理进行计算。如果能够实现从专用机走向通用机,或者哪怕在特定领域中先实现专用机的规模化量产,都会带来巨大的算力提升和计算变革。

— 泰合资本

## Chiplet技术

“

Chiplet技术指的是使用小型模块化的“Chiplet”来组成更大、更复杂的系统级芯片(SoC)。这些Chiplet可以独立设计和生产,从而在芯片设计和生产中提供更大的灵活性和可扩展性。Chiplet本身是一系列整套的芯片设计和制造的方法论,包含了一系列各种各样的子技术,事实上大多数技术其实都实现了商业化。Chiplet技术在很大程度上降低芯片设计的复杂程度,有效降低研发与设计成本,还可以通过更加更加合适的制程选择大幅降低芯片制造成本,缩短上市周期。

基于芯片行业的发展情况,摩尔定律趋于见顶,现如今先进工艺已经到了2nm、3nm,硅基芯片的晶体管密度已经逐渐接近极限,即使出现了FinFET、GAAFET等结构技术,但依然难以抵抗摩尔定律放缓的趋势。Chiplet技术从成本上、本质上有机会去突破摩尔定律的见顶的趋势;对于中国来说,要解决大算力芯片的制造问题,可以用相对成熟Chiplet技术做一些能够去逼近先进制程的芯片,满足算力需求。

— 联想创投

## Societal AI

“

基于对人工智能的理解和发展趋势的预测,微软亚洲研究院认为建设 Societal AI 应该包含五个方面:价值观对齐、数据及模型安全、正确性或可验证性、模型评测、以及跨学科合作。其中,价值观对齐是一个新兴的领域,但其重要性已经得到了工业界和学术界的广泛认可。在 Societal AI 的研究中,微软亚洲研究院认为人工智能对齐的目标应该从指令上上升至人类的内在价值观,让人工智能可以通过自我判断,来使其行为与人类价值观保持一致。

— 微软亚洲研究院

## 器官移植技术

“

器官移植是临床需求巨大的领域，用户面临的主要问题是器官来源的短缺。目前在器官移植领域有两个方向：一是利用基因编辑等手段改造的大型动物器官，克服器官移植后的排异反应后，猪、牛、羊这些大型动物，就可以为很多器官衰竭的中末期患者提供解决方案；二是通过干细胞诱导分化和体外培养技术培养出新的器官，目前体外的器官培养能培养出一些肺、心，甚至脑、肝脏，但离移植还比较远，更多是在药物发现，还有进行药物功能验证中使用。

整体来看，器官移植是非常新的领域，中美均属于探索阶段，在不断临床试验的过程中发现和解决问题。去年马里兰第一例转基因的猪心移植，延长了患者两个月的生命，移植后一开始患者状态非常好的，但是后来出现了急剧的恶化，在后来的详细分析报告里才发现了之前被忽视的会造成排异反应的基因位点。另外，还有一个重要的问题，一些猪身上特异性的病毒，猪的免疫器官可以压抑住的，但是移植到人体内，人的免疫系统没有针对猪具细胞病毒相关的免疫压制，病毒就会引起功能障碍。不在进行实践当中是很难预料到相应的问题，所以这个领域还是需要更多的研究和积累。一部分分泌型器官的移植可能首先实现突破，比如胰岛，不会像人体全器官这么复杂，国内外已经有一些团队进行体外胰岛的培育，移植到患者体内，从而实现胰岛素依赖型的糖尿病患者功能性治愈，也有做到临床一二期的产品。全心等大器官的移植不是社会资本适合介入的阶段。

——联想创投

## 生命科学工具

“

过去几年，生命科学工具行业在疫情催化下迅速发展。尽管当前创新药行业处于平缓发展的低潮期，但对比海外市场，以及站在更长的历史纬度、时间纬度来看，中国整个新药产业进程，无论是生物药、新型小分子药物或其他药物形式，已经进入第二次浪潮的前夕。而医药/诊断行业与生命科学工具行业是一个相互促进的过程——上游促进下游的发展，同时下游再反哺上游，如同DNA结构一样，是双螺旋式的发展趋势。基于上述因素，再结合全球巨头表现，中国生命科学工具市场方兴未艾。

过去三年，新冠疫情导致全球生命科学供应链都面临着较为紧张的状态，大量产能优先供应于新冠疫苗和核酸/抗原检测。上游供应出现系统性缺货或涨价情况，部分货期甚至超过1年，这客观上为一批国内上游产业链企业获得了头部客户的试用机会，可以快速打磨产品，同时从政府和监管机构角度也意识到卡脖子技术在生物医药生产及研发的重要性，积极扶持本土企业发展，为这一领域的国产化提供了优良的外部环境。

从投资机构的角度，如果我们只看生物制药的上游产业链，即生物生产工艺的解决方案领域，整体市场规模大概从100亿元人民币快速提升到200多亿元。一些核心单品及相关市场规模超过20亿元，这使得很多投资机构开始关注其中的部分细分领域。这个领域开始迎来了系统性的投资机会。

——经纬创投

# 2024 前沿科技投资观点

## 如何判断一项前沿科技的发展空间和合适的发展节点？

技术的演进通常是一个渐进的过程，往往需要数年，甚至数十年的持续研发。在这个过程中，我们需要保持耐心，等待底层技术的成熟拐点的到来。同时，要牢记商业成功是前沿科技持续进步的前提条件。为确保商业成功，关键在于将技术有效地转化为产品。

这一步往往不仅仅是单点的技术突破，更需要考虑产业链上下游的整体成熟。这包括了从研发到生产再到市场推广的全过程，需要形成完善的商业生态系统。在判断发展节点时，需要密切关注市场反馈、用户需求以及竞争态势。这有助于及时调整发展策略，确保技术的应用能够更好地满足市场的实际需求。同时，与行业内的合作伙伴建立紧密的合作关系，共同推动整个产业链的发展。

——BV百度风投

技术可靠可验证、「工业化」进程清晰、应用具备广泛性即市场空间大。

——泰合资本合伙人 梅林

从投资的角度，一项前沿技术的发展空间和发展节点，是判断投资的两个关键。

发展空间也就是科技的应用范围和落地场景，利用技术能切实解决哪些问题，能改造多少行业，能重塑多少产业链，能提升多少的效率。这依赖于这项技术的变革性，以及给社会带来的「熵减」效果。

发展节点在于技术成熟度和市场进入时机，如果一项技术的成熟度和市场的成熟度能够契合上，满足社会与日俱增的需求，那么就迎来发展节点。就如同当下的AGI，当大模型使得智能不断涌现，就使得其具备了渗透到各个领域进行改造的能力，而社会各个行业「蓄势已久」的需求也被刺激起来，无论是发展速度还是迭代速度都非常快。

——九合创投

判断技术发展空间需要考虑技术所对应的应用是否重要、高频，对于发展节点主要需要考虑技术和市场的成熟度。

——华为云 张教授

# 2024 前沿科技投资观点

## 如何判断一项前沿科技的发展空间和合适的发展节点？

需要了解项目的属性和风险，其中就包括技术的差异化门槛、知识产权的完备性、应用场景以及投入产出比。投资前沿科技最关键的是要关注大势，时机很重要。时机由基本要素组成，比如人才储备、关键技术成熟度、政策、社会环境等等。

比如，2015年我们开始看新能源汽车的时候，当时中国已经具备了电子和锂电池的生产能力，且续航里程数已经超过300公里；另外，政府制定政策鼓励大家购买电动汽车，有一些驱动因素。如果2010年就有人喊这是风口，但这些要素不具备的话，就很难形成大趋势。

——蓝驰创投

要看一项科技能否解决过去无法解决或解决不好的问题，能否刺激出新的需求，能否快速找到PMF (Produce Market Fit) 点。

——百度研究院

前沿科技取得重大突破与其最佳商业化的时间节点很多时候并不是重合的，科技突破与市场反馈的不谋而合这件事可遇而不可求。优秀的科学研究不一定是好的，但一定是新的；而优秀的产品设计不一定是新的，但必须是好的。

专注于科学研究容易陷入供给端思维，而忽略了市场对于这项技术的真正需求。当需求特别明确时，我们对某项未必是最新的技术进行工程开发、加工与包装，有可能打造出一个很好的产品；当需求尚不清晰，如果盲目将技术进行商业化，就可能陷入拿着锤子找钉子的困境。

——真格基金合伙人 尹乐

我们会更关注几大领域：信息、能源技、生命的前沿技术，它们交叉融合发生的创新往往有更大的发展空间。从关注的角度，我们可能会注意到这些领域影响重大的TRL2-3成熟度的工作，但从投资角度TRL6-7是关注更多的发展节点。

——明势资本

我们认为了解市场需求和行业前景，以及观察投入资金和资源在某领域的情况，可以粗略评估某前沿科技的可行性和成熟度，这可以用来判断一个科技的发展空间和发展节点。

——西安电子科技大学教授 王皓

# 2024 前沿科技投资观点

## 如何判断一项前沿科技的发展空间和合适的发展节点？

发展空间,主要是看该科技对原有技术链和产业生态可能带来的变革型影响,一旦成熟,能够带来多大的新市场,或者节约巨大的成本。合适的节点,主要看该技术是否可以进入到产业落地尝试的阶段,这个阶段,会出来更多的新产品、新模式、新业态,后续重点观察期可持续性影响。

——腾讯研究院 前沿科技研究团队

一项技术的发展空间不是判断出来的,是观察出来的。一项技术的发展空间,是需求侧问题,而非供给侧问题。承担技术发展速度风险的收益远好过承担需求风险的收益。因此,从创业和投资的角度,都应当踏踏实实回到需求中,面向能“观察”到的需求研发技术。

最合适的节点是踮脚刚刚够着的时候,最好不要是起跳才能够着的时候。比较好的情况是,技术上还没有完全解决,但剩下的主要都是工程性问题,而非科学问题。这样,一家创业公司有窗口期去定义一个新产品,但又不至于被科学问题束缚。

——源码资本执行董事 陈润泽

三四年前,团队认真做这个领域的投资与研究时,挑战也很大。因为前沿科技本身就是一个比较交叉的学科,如果想在这个领域有所建树,要求投资人既要对产品 and 市场有充分理解,又要对商业模式的本质有深度思考,要求非常高,我们也是边摸索、边前进。

——经纬创投

看应用场景是否刚需,是否能真正解决问题。

遵循新技术发展的定律,一项前沿科技从诞生到发展应用起来,要经过超10年的时间,期间人们会经历初期的亢奋,发展中的质疑,信心的打破与重建,再次看到希望这样一个跌宕起伏的过程,像是走到愚昧之巅又跌进绝望之谷,在经历漫长的开悟之坡后,终于到了一个持续平稳的高原。

——跃为资本

# 2024 前沿科技投资观点

## 前沿科技从实验室走向广泛落地的关键要点和里程碑？

以我们关注的一些具有颠覆性的的前沿领域如认知机器人为例，其复杂度是远超想象的，且技术不是单点的，其中多个技术点是成熟度较低，远没有到能发生质变的时刻。这种情况下如果想尝试技术落地的关键点是找到能够的迭代技术能力、高效积累有用数据的应用场景、商业模式甚至组织发展路径。在技术不完全成熟的情况下，找到局部的Product-Market Fit可以看作一个里程碑。

——明势资本

技术已经基本成熟、且可用于刚需高频应用。

——华为云 张教授

我们过去数年支持了大量的AI硬件公司。以终为始来看，科研化、工程化、产品化、商业化、全球化是成功硬件公司的五个必要阶段。

前沿科技创业成功的本质是人才进化，对创始人与核心团队的综合能力有很高的要求。在上述的不同阶段，创始团队的精力分配与关注点不尽相同，需要灵活调整组织结构来保证绝对的执行力。同时，前沿科技商业化的周期通常较长，为了应对资本市场的周期轮动，创始团队需要进一步提升资本使用效率。

——BV百度风投

技术的前沿性、变革性和实用性，是非常关键的要素。不同于在实验室里的科学，打破科技成果转化的「一公里」最核心之处在于「商用的能力」，要有巨大的实用意义。前沿科技要能带来明显的效率提升，提供更好的解决方案，创造真实的价值，并保持稳定的表现，持续降低成本，达到更好的投入产出比，才有可能实现大规模的应用。

——九合创投

有「工业化」目标和能力的团队 + 重点下游的深度支持 + 良好的技术生态 + 可持续的资金支持。

——泰合资本合伙人 梅林

# 2024 前沿科技投资观点

## 前沿科技从实验室走向广泛落地的关键要点和里程碑？

蓝驰创投认为,应该寻找需求与创新实现统一的拐点,当这个拐点到来,硬科技会实现软着陆。

其中的关键要点包括:首先,从更早、更准确地洞悉需求出发。底层技术的成熟只是基础,只有追问深层变化因子,主动和需求做连接,并且随着时间变化持续更新对于需求的洞察,才能达成广泛落地的前提条件;其次,在发展过程中,科技企业要学会更快适应变化,快速感知应用场景、经济政策、竞争格局的变化,迅速迭代。最后,保持长期主义和自律专注。底层技术从研发到真正应用到商业场景、再到被市场广泛接受是一个漫长过程。在这场长跑中,创业者应当始终明确公司的主干道。既要着眼长远,更要专注自律。

——蓝驰创投

在AI原生时代,数据飞轮是AI原生应用成功的充分必要条件。在这个时代,数据越多得到的反馈就越多,得到的反馈越多算法和产品的体验就越好,用户就越多,产生的数据也就越多,这就是数据飞轮。因为深度学习人工智能是数据驱动,因此AI原生应用产生的数据不论从质还是量都超越上一个时代。因此,关键的要点和里程碑是让数据飞轮转起来。

——百度研究院

技术到产业,不仅仅是技术本身,而是技术、商业、政策和社会的统一和精妙平衡。比如商业上是否可能形成可持续的收入,是否可以形成规模化的产品使得成本落入可推广的区间,社会大众对于新技术的接受度,以及相关的治理等。

——腾讯研究院 前沿科技研究团队

首先要验证技术的可行性,这可能需要多次迭代和测试。其次开发技术原型并在试验场景中验证其有效性,以确保技术在实际应用中的可行性。当然这也受商业模式的制约。一个可行的商业模式在一定程度上可以为实验室环境下的科技产品提供从市场角度的需求,这对科技从研发到落地有着重要的可行性的评估。另外一点是最终用户评估并在用户反馈中进行学习,改进和优化产品和服务。

——西安电子科技大学教授 王皓

# 2024 前沿科技投资观点

## 前沿科技从实验室走向广泛落地的关键要点和里程碑？

除少部分领域(部分药物、材料)外,大部分科技从实验室走向落地都会经历一段痛苦的工程化阶段。关键的要点是团队是否能转变思维模式,把视角从追求创新度/差异化转化为满足现实需求,把核心的技术优势与公司财务和运营报表上的具体指标对应起来。

最重要的里程碑:公司找到了需求的基本盘,团队建立起了基本的经营理念和能力。

——源码资本执行董事 陈润泽

是从功能到工程,是从单一技术的突破到现实的可用、好用。

是新技术能够适应旧生态。

人的影响因素很大,需要掌舵者有复合背景,具备体系化、工程化的能力。

——跃为资本

前沿科技创新能否被接受的关键要点:

- 1、相对优势(差异优势有多大,是否可以取代);
- 2、兼容性(与现存价值观、接受者背景、需求的符合程度);
- 3、复杂性(被理解和被使用的难易程度);
- 4、可试验性(被实验的可能);
- 5、可观察性(多大程度上接受者可以看到创新的结果)。

具体而言,一个创新和它所取代的事物相对优势越大,与现存的价值观、潜在接受者过去的经验和需求冲突很小,复杂性不高,可以通过试验进行考察,并且人们可以直接观察结果,那么这个创新就更容易被人们接受。在前沿科技走向落地过程中,不能再单纯追求实验室里对于技术参数的数值提升,而要真正考虑科学技术是否可实现工程化和产品化,并能找到具体且广泛的需求场景应用其中。

关键里程碑是看前沿科技是否跨越了鸿沟,也即是科技使用者人群是否从创新者和早期使用者,跨越到早期大众,也即是是否完成了PMF验证(满足了未被满足的需求;你的定价对你的目标用户来说是可以接受的,并且你对未来的盈利能力没有任何问题;可以规模复制,拥有健康的留存率、市场份额和客户满意度),如果是的话,就成功实现创新的扩散,前沿科技实现广泛落地。

——真格基金合伙人 尹乐

# 2024 前沿科技投资观点

## 前沿科技从实验室走向广泛落地的关键要点和里程碑？

**技术可行性验证：**确保技术在实验室环境中是可行的，并且具有足够的性能和稳定性。这可能涉及进行一系列的实验和测试，以验证技术在实际应用中的可用性。

**市场需求分析：**了解市场对该技术的需求和潜在应用。这包括识别潜在用户、市场规模、竞争对手以及技术在解决实际问题上的独特价值。

**商业模式设计：**开发可行的商业模式，包括收入来源、合作伙伴关系和客户获取策略。这有助于确保技术的商业化过程更加有序和可持续。

**投资和资金筹集：**寻找资金支持，可能包括政府拨款、风险投资或合作伙伴投资。这有助于支持技术的开发、推广和商业化。

**建立生态系统：**与其他相关技术、合作伙伴和利益相关者建立合作关系，构建一个完整的生态系统，促进技术的广泛应用。

**可持续性规划：**制定长期的可持续性规划，包括技术的未来发展方向、升级计划和适应新需求的能力。

以上都属于，创业不易。

——经纬创投

# 结语

# KEYWORD



量子位智库  
往期报告



量子位智库小助手  
请备注姓名来意



量子位的小智库



量子位公众号

在连续三年出品《年度十大前沿科技趋势报告》后，我们发现，前沿科技的发展绝非一蹴而就。理论萌芽、科研成熟、技术工程化、创企出现、产品落地等等，每一个节点都有其瓶颈与必备要素。

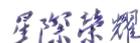
一方面，我们的确进入了科技爆发的历史性时期。众多前沿科技正在交叉融合，从不同角度改写我们的生产生活。另一方面，也有更多要素和资源需要链接与分配，加速前沿科技落地应用的进程。

作为在早期就开始关注AI领域的科技内容平台，量子位将继续发挥自己在学术界、产业界、投资界以及爱好者之间的链接作用，继续观察并推送这次科技浪潮，助力各个前沿领域的进步和成长。

作为量子位旗下科技创新产业链平台，量子位智库致力于提供前沿科技和技术创新领域产学研体系化研究。我们将通过媒体、社群和线下活动，帮助创新者更多展示科研成果，决策者更早掌握创新风向。

量子位智库期待与更多科研人员、初创公司、技术爱好者等共同携手，见证科技创新浪潮。

## 鸣谢 受访机构



\*以上排名无先后顺序

## 编撰团队

量子位智库总裁 李根

量子位智库 郑钰瑶 丁乔 游艳馨 兰壹凡

美术团队 罗玉 高佳乐



QbitAI Annual Frontier  
Technology Report

20  
23

# 十大前沿科技趋势报告