

The background is a full-page abstract painting. It features a vertical color gradient from cool blues on the left to warm yellows and reds on the right. The texture is thick and expressive, with visible brushstrokes and splatters. In the lower-left quadrant, there are dark, silhouetted evergreen trees. The overall composition is dynamic and visually striking.

2024 人工智能AI趋势

目录

2023 回顾	3
瞧，这些万亿巨头	5
生成式 AI 十大展望	8
算力“破墙”	11
自动驾驶换道大模型	14
量产人形机器人	17
从元宇宙到“新现实”	20
医疗智能体	22
“通用”基因编辑	25
中美风投再分岔	28
推理的碳足迹	32
结语	34

2023 回顾



准确 对了一半 不准确

新能源

准确性：● ● ●

核心判断：全球化石能源消费将于 2025-2030 年间开始下降。中国碳达峰可能提前至 2025 年。未来 5 年新增可再生能源装机，中国将占一半。

典型事件：IEA 最新预计全球煤炭消费将 2026 年开始下降，中国煤炭需求在 2024 年下降。今年，中国水泥行业提前碳达峰。全球可再生能源新增装机容量刷新历史最高纪录，中国占近 55%。

人形机器人

准确性：● ● ●

核心判断：无论与人一样敏捷的双足机器人，与人自如对话的虚拟人，还是用外骨骼或脑机接口实现人体与机器的合体化，2023 年人形机器人将会加快走向商业化。

典型事件：特斯拉“擎天柱”迅速迭代，小鹏、小米等公司涌入。李飞飞团队试验了 NOIR 脑机接口机器人。福特与亚马逊开始试点人形机器人 Digit 的商业应用。中国出台了全球首部发展人形机器人顶层设计文件。

自动驾驶

准确性：● ● ●

核心判断：2023 年，辅助驾驶将更大规模部署到电动车上，新的应用场景得到不断开拓，中国超越美国成为全球自动驾驶技术最大的试车场。

典型事件：百度近五年来的累计路测里程与累计服务订单超越了 Waymo。中国当前乘用车 L2 及以上智驾渗透率超过了 42%，较去年 30% 大幅提升。换道大模型的特斯拉也准备好让完全版的 FSD 在中国落地。

低轨科技

准确性：● ● ●

核心判断：2023 年，在更多的公司和资本的支持下，火箭发射成本进一步下降，太空是经济发展的下一个前沿。

典型事件：全球全年火箭发射次数首次突破 200 次，主要由 SpaceX 贡献；亚马逊的卫星星座正式升空。中国的朱雀二号成为全球首枚入轨的液氧甲烷火箭。但星舰尚未成功入轨，发射成本大幅下降仍待明年。

合成生物

准确性：● ● ●

核心判断：2023 年，人工智能等底层技术将继续创造与优化酶与底盘等生物元件，更多基于合成生物的碳中和技术与新药研发得以验证。

典型事件：Ginkgo 将与辉瑞共同开发多款 RNA 药物，美国能源部也资助这家行业巨头研究抗菌藻类以减少碳排放。但从研究到量产仍然困难。Ginkgo 市值从 280 亿美元跌至 28 亿美元，另一家合成生物巨头 Amyris 申请破产。

半导体

准确性：● ● ●

核心判断：中国半导体行业在 2023 年需要与时俱进的新一轮产业政策，聚焦制造薄弱环节，打造中国半导体行业的创新生态。

典型事件：历经调整后，大基金二期加速投资步伐，集中于芯片制造环节。大基金三期目标筹集 400 亿美元。华为麒麟 9000S 的突破成为业界继续推进国产替代的定心丸，政策加快创新生态的形成。

元宇宙

准确性：● ● ●

核心判断：2023 年元宇宙的重点，将是设备的突破，即能为用户提供“基本可用”的 VR 和 AR 等设备，并且性价比易于普及。人工智能在 AIGC 的拓展，包括 AI 生成 3D，将成为 2023 年元宇宙的创新亮点。

典型事件：苹果发布了 Vision Pro，但尚未正式销售。它的硬件性能远超当前主流消费级设备，但价格昂贵。受限于 VR 等设备全年出货量不及预期，尽管 AIGC 进展迅速，多数创新并非直接面向元宇宙。甚至 Meta 发布会也很少提及“元宇宙”，它的热度被空间计算取代。

人工智能

准确性：●●●

核心判断：2023 年值得期待的，不仅仅是 GPT-4，人工智能大模型将会生成视频、3D 建模，多模态的组合，也将会应用于科学研究，包括蛋白质结构预测分析、新材料、新能源。

典型事件：2023 是大模型军备竞赛的一年，GPT-4、Llama2、Gemini 等相继发布，中美控制了世界上 80% 的大模型。Runway 与 Pika 等初创企业正在颠覆视频行业。人工智能辅助天气预报等研究成果占领了顶级科学期刊封面，预测出 200 多万种晶体结构。

下南洋

准确性：●●●

核心判断：2023 年，中国风险资本与华人创新者将在阵痛中寻找机会。跨境电商与金融支付等数字技术在东南亚扩散；先进制造将产能优势复制到更靠近市场的地区。

典型事件：今年，投资者与创业者热衷于谈论中东主权财富基金。字节跳动收入超过腾讯，优势在于全球扩张；Temu 与 Shein 合计美国用户逼近亚马逊。比亚迪等新能源企业继续在欧洲投产，供应链向周边及海外延伸。中国创业者和企业家出海，走自己的全球化之路。

中美创投

准确性：●●●

核心判断：国家、资本与技术，正在中国形成新的创新生态。2023 年，随着走出疫情，以及对平台经济建立常态化管理，消费和互联网的投资有可能再度活跃。

典型事件：今年，中国“以投促引”进一步下沉，中西部省份新成立基金数量逆势增长，政府引导基金目标规模提升至 13 万亿，硬科技企业上市退出全球最为活跃。但在全球范围内，消费与互联网企业融资都大幅缩水。

人口国运

准确性：●●●

核心判断：在长期推动经济增长的人口、资本与全要素生产率中，中国将更加依靠全要素生产率的提升，更加依靠青年人才参与到创业与创新活动中。

典型事件：年底，官方解读中央工作会议，强调“新质生产力以全要素生产率提升为核心标志”，围绕“技术革命性突破、生产要素创新性配置、产业深度转型升级”展开。

瞧，这些万亿巨头

苹果，微软，字母表（谷歌），亚马逊，英伟达，Meta，特斯拉（马斯克），这些公司之间有什么共同点？

是的，它们都是科技巨头（Big Tech）。它们的市值目前都在万亿美元以上，或者曾经达到过。其中的5家，每年研发支出达到了250亿美元以上。

科技巨头曾经是指那些通过互联网建立起强大平台经济的企业，拥有十亿用户级别的软件及应用。它们往往赢家通吃（Winner takes all）。

科技巨头的概念也在随着技术演变。它们实现了软硬件一体，除了软件，它们也设计和制造终端设备，包括PC、手机、可穿戴、AR/VR、传感器、机器人、智驾汽车，它们也正在渗入到制造业的流程中。

他们还建立起了强大的云计算和企业服务能力，包括芯片、超级计算机和超大规模的数据中心，为第四次技术革命提供最重要的基础设施。

它们已经开始全面竞争一种全新的能力，生成式AI。在深度学习领域的竞争，从2012年视觉计算取得突破时就已经开始，因为大模型的流行而加剧，但只有在ChatGPT发布之后，过去的一年，生成式AI大模型成为科技巨头之间“军备竞赛”的焦点，迅速成为巨头技术栈上标配的一层。它正在成为所有巨头的业务基础模型。

科技巨头的核心能力，越来越区别于其他非科技企业和非巨头企业。它们能把技术和应用端到端地垂直整合到一起，不断扩张业务范围，形成一种科技巨头所独有的、以计算和智能为核心的创新能力。它们都是全球化的企业。

这样看来，华为也是这样一家科技巨头，它也已经建立起了从芯片到硬件应用的全栈技术能力，以计算和智能为核心展开业务范围。它在2022年的研发投入达到了240亿美元。但华为是其中几家巨头的挑战者。

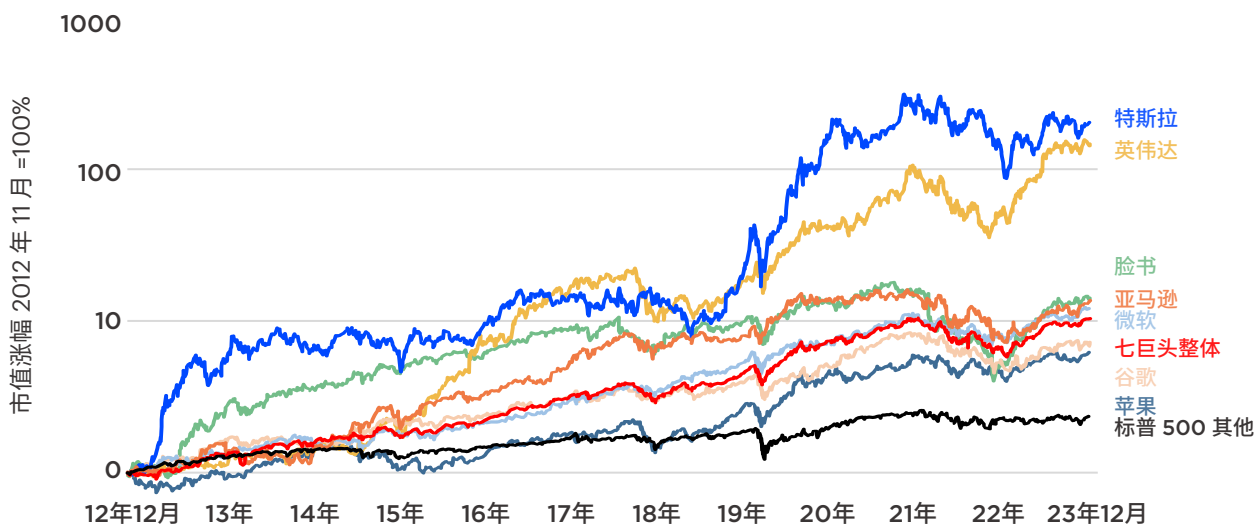
巨头公司的研发投资覆盖了基础技术研究和具体产品的开发。除此之外，它们每年还把研发和投资的20%左右，投入企业内部IT系统、效率软件、先进的技术平台、以及数据中心等基础设施，并持续地提升员工技能，不断把那些复杂和重复的业务和流程进行自动化和简化。

巨头企业保持内部技术的先进性，是其创新和竞争力的一个重要来源。巨头用最先进技术围绕用户数据建立起了正向的反馈机制，形成了飞轮效应。

人工智能正在加快这一飞轮的运转，放大它们在各自领域的核心能力，也在生成新的能力。生成式AI更像是这些科技巨头为自己发明的新工具，一种放大器和加速器。

AI最早的受益者是科技巨头自己。英伟达从2012年起就开始为深度学习提供GPU，营造CUDA生态，10年之内把自己变成了科技巨头。巨头最早受益于AI的，都是其核心业务，如云计算和广告，目前是企业服务和生产力软件，接下来还有硬件和消费智能产品，以及新的“赢家通吃”的领域。

深度学习的黄金十年



来源：Wind，

说明：市值涨跌幅，对数轴。标普500其他指不包含七巨头的标普500成分股的总市值，此外，特斯拉等部分企业部分时段不属于SP500成分股。七巨头整体指七巨头总市值的整体变化。以2012年11月19日为基准100%。

巨头的无限游戏

我们正处于当年个人电脑开始的同样时期。1980年代初，信息技术革命发轫于英特尔发明的CPU。乔布斯创办了苹果电脑，比尔盖茨创办了微软，近50年后，它们跨越了PC、互联网、移动、云计算，直到人工智能，至今是世界上市值最大的两家公司。

同样，2023年，AI开始真正大规模走向消费者。AI时代真正开启，人们称之为苹果时刻、人机交互的范式转移时刻，寻找这个时代的苹果和微软。

但是，科技巨头似乎从源头就控制着这一切。

通用人工智能最具颠覆性的两家初创企业，DeepMind和OpenAI，前者被谷歌收购，后者技术被微软买断和控制。

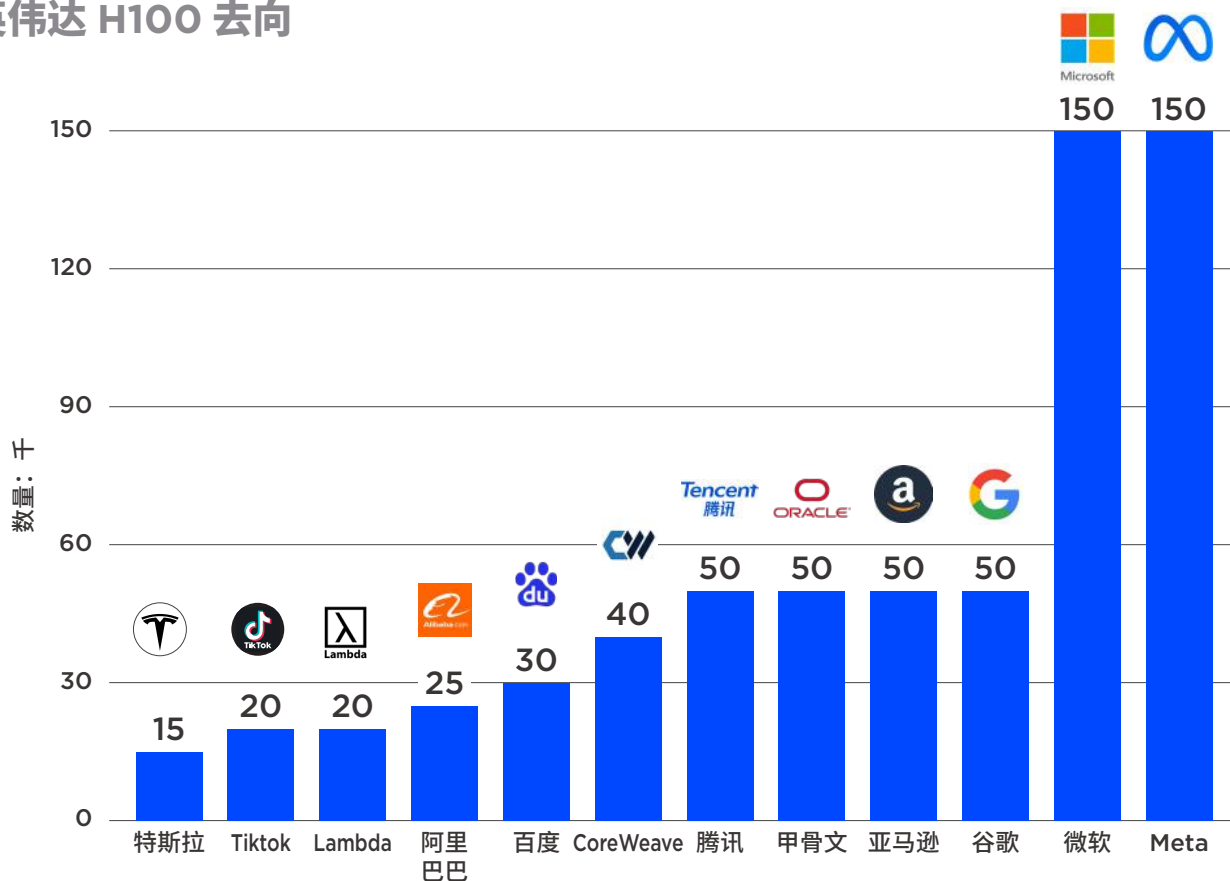
生成式人工智能，从训练最先进的大模型，如GPT-4、Gemini、Claude2，到向数亿用户部署和应用，多数都要依靠微软、亚马逊和谷歌这三家的云计算。英伟达成为新晋的科技巨头，除了用于AI加速的GPU芯片，它也在构建其人工智能的基础设施，还成为过去一年投资人工智能初创企业最多的公司。

它们控制了数据。用来训练生成式人工智能的自然语言数据、视觉数据、代码符号数据、知识图谱，包括合成数据，海量地来自并存储在这些巨头的业务、平台和基础设施中。更重要的是，大数据创造了对自动化和人工智能的需求。

生成式人工智能重新定义了大数据。由于人类已经开始掌握以大型语言模型(LLM)技术以产生智能，网络上的公开数据、企业数据、个人数据、自然界中的数据，以及人工合成的数据，都可以成为训练智能的原料。科技巨头本身拥有巨大的数据库，它们还在获得更多和更好的数据。它们建立起了联盟，为用于人工智能训练的数据建立标准。它们可以合成数据，成为新的模型的训练的来源。目前许多人工智能的研究和模型训练，开始用GPT-4等先进大模型生成或者标注的数据来训练。除了公开数据，OpenAI还开始与各机构展开私有数据的合作。

它们掌握了算力。巨头们已经在全球各地建立了数据中心，拥有最先进的AI加速算力；或者已经囤积的AI芯片，已经超过了世界上许多中等国家所拥有的数量。它们除了用来实现自身业务的AI化之外，还去进行科学探索：AI用于创新药研发、医疗服务、芯片设计、材料发现、能源转型和应对气候变化。

英伟达 H100 去向



来源: Omdia,
说明: 预估 2023 年 H100 主要买家购买的 H100 芯片数量。

巨头所拥有的强大的算力，实际上是人工智能所引领的第四次工业革命的基础设施。微软以后每年将在数据中心投入 500 亿美元，包括自行研发的芯片的支出，这已经相当于一个科技大国的 AI 基础设施的投入。亚马逊和谷歌，都在为数据中心更高的计算效率研发芯片。苹果研发的手机和个人电脑的芯片，已经超过了专业的芯片设计公司。

美国及许多国家的政府、大学、研究机构，会日益依赖这些巨头的算力基础设施。发展人工智能成为许多国家优先考虑的事项，这些巨头的云中心已经遍布世界各地，它们可以在这些数据中心的基础上，轻松地与当地政府合作。

拥有数据和算力，加上资本的力量，这些巨头可以吸引世界上最好的人才。它们建立起了最先进的研究部门，吸引了世界上最优秀的图灵奖获得者和理工科博士，也掌握了最好的算法——Transformer 论文就出自谷歌，而谷歌一直是高质量 AI 论文的来源地。Meta 也建立了超级算力集群，研发出最流行的开源机器学习库 PyTorch，不仅支持自己在社交媒体上的推荐算法，支持它建立起最大的线上广告系统，而且推出的大模型 Llama 引领了开源大模型潮流。

这些巨头公司还有一个重要特点，都是从初创企业成长起来的，除了苹果公司之外，其创始人依然在管理企业，或者对企业的方向与战略发挥着影响力，其 CEO 依然能让日益庞大的企业保持敏捷。企业体内部活跃着技术基因，工程师思维主导了企业文化。

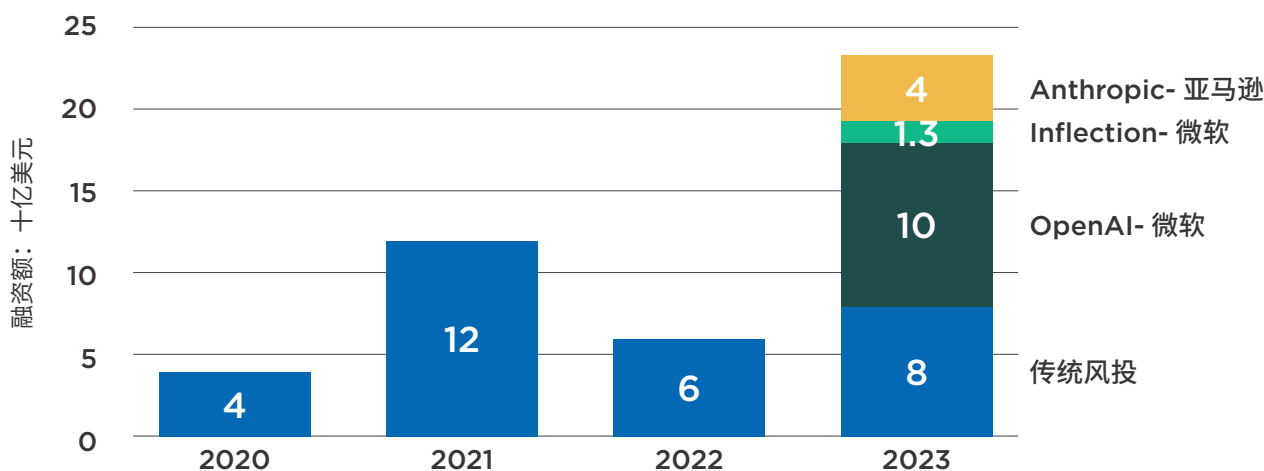
巨头不仅砸下巨资进行前沿技术的和新产品研发，打造先进的企业 IT 系统和技术平台，它们还投资、收购初创企业。2023 年，对生成式 AI 的投资，微软、谷歌、亚马逊、英伟达等几家大型科技巨头对大模型初创企业的投资，金额上远远超过了独立的风险投资机构。

它们投资的战略性也越来越强。巨头投资大模型初创公司，其中相当大的金额就是算力信用的投入。例如微软对 OpenAI 投资 130 亿美元，其中很大一部分是 Azure 云计算；亚马逊以 40 亿美元投资大模型初创公司 Anthropic，其中多数是 AWS 的算力信用。而巨头们自研的 AI 芯片，也将用于这些大模型的训练和推理功能。谷歌则更早建立起这样一个共生链条。硅谷人称“云洗钱”。但目前食物链的顶端仍然是英伟达。

巨头们不仅控制了最强大的闭源大模型，而且控制了最流行的开源大模型，如 Meta 推出的 Llama，微软研发的 Phi 系列的小型开源模型等。Google DeepMind 还能结合最强的强化学习模型进行科学发现。

这些企业还拥有全球化的优势，它们的用户、业务、供应链和数据中心遍布世界各地。美国巨头与中国巨头的国际竞争主要在东南亚，AI 的兴起，让两国巨头在社交、视频、电商、云计算的竞争愈发激烈。

巨头主导生成式 AI 风险融资



来源：Pitchbook，
说明：全球范围内，截至 2023 第三季度。

生成式 AI 十大展望

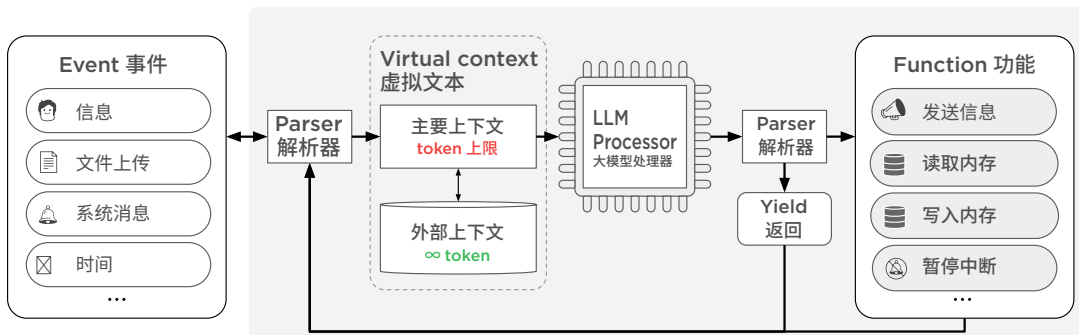
2024 年，优化训练和部署大模型仍然非常重要，大模型的生态加速形成，应用开始在一些领域大规模展开，主要表现在如下十个领域：

1. 智能体作为任务助理进入更多应用场景和业务流程

智能体能有一定的主动性，能帮助完成任务，而不仅仅是问答。在感知环境后，通过其大脑（大型语言模型），调动其他的程序、应用、知识，甚至自己编程，规划和执行更复杂的任务。有了智能体，许多人可以用经验和专业知识，通过自然语言而不同于编程代码去写软件。

2. 操作系统集成下一代大模型，成为下一代操作系统

大型语言模型日益操作系统化，AI 芯片为它设计，PC 和手机的操作系统为它升级，AI 应用成为它的下游，上下文管理类似于操作系统的内存。微软将推出 Windows12 操作系统，在 PC 上与下一代大模型和 Copilot 深度集成。在移动设备上部署的模型，也与 iOS 与安卓操作系统紧密结合，实现 AI 功能和建立 AI 应用商店。



来源：MemGPT: Towards LLMs as Operating Systems

3. 生成式 AI 制作的影视剧大量出现，冲击影视行业发生剧变

图像和视频是生成式 AI 迭代最快的领域之一，GPT-4V 等多模态大模型的推出，基于扩散模型的 Dall-E 3、Midjourney 和 Stable Diffusion 的功能不断增强，LCM-LoRA 等技术达到了实时生成图像和视频的效果，对影视、音乐、游戏等内容娱乐行业的影响是颠覆性的。这方面的应用也是巨头目前还染指不多的领域。2024 年将大量出现由生成式 AI 产生的影视剧，创作者、用户以及角色之间将会出现崭新的交互方式。

4. 人形机器人开始量产，自学习与环境互动能力进一步强化

在已有的机器人技术之上，多模态和具身智能的大模型，不断展示出惊艳的效果。大型语言模型的推理和规划能力，与视觉模型结合，可以通过获取周围环境数据，学习人类用手脚完成任务。2024 年人形机器人开始量产，开始在工作场景中进化迭代人类的灵活性。

7. 数据来源的深度和广度进一步开拓，进一步规范，更多合成数据与自然数据结合用于大模型训练

数据决定了泛化的边界。自然语言数据，以及直接从现实世界事件或对象中收集得到的数据，已经无法满足下一代大模型的训练的胃口。在专业领域和垂直场景，非公开的数据将会发挥更大的价值。大模型训练、自动驾驶、机器人、图像生成、模拟仿真等，都在大量使用合成数据的同时生成新的数据。越来越多的数据标注也由 AI 来完成。但是，只使用合成数据可能会造成数据多样性不足和自循环训练的问题。2024 年将会看到 AI 企业寻求合法获取更多非公开数据，以及使用更多的混合数据。

8. 苹果真正入局，力争复现 AI “iPhone 时刻”

2023 年被称为大模型之年，苹果表面上在作壁上观，但实际上在芯片及硬件方面的研发一直在加大力度，只做不说。2024 年苹果将把 Vision Pro 推向市场；PC 和手机加载大模型，苹果是其中最重要的玩家；为了建立 AI 应用生态，操作系统封闭的苹果拥抱开源模型。苹果被广泛期待能给消费市场带来更好的 AI 产品体验。

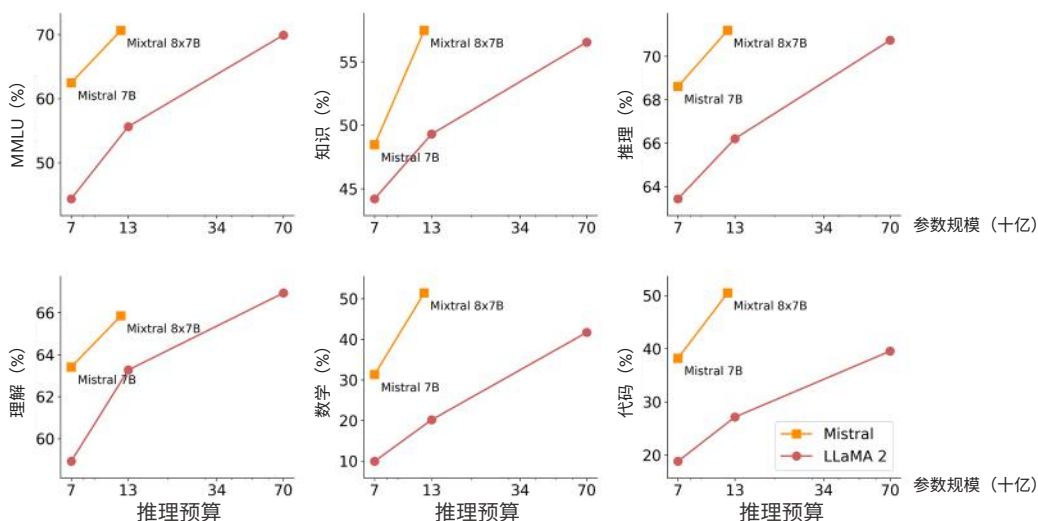
9. 一些开源模型及 AI 应用，因为无法建立起商业模式将面临生存危机

绝大多数初创企业的开源模型，目前还无法在提供推理服务、授权、训练和部署模型方面建立起用户基础；消费类模型 + 应用的初创企业，在激烈的竞争中多数将遭淘汰；纯应用类的初创企业，许多将遭到巨头碾压或者很快在更新的开源技术迅速推广中出局。快速获取用户并且在反馈中建立起数据飞轮的企业将赢得生存。而能结合起应用场景、行业深度和垂直数据来源的企业，将能保护好自己。

10. 小模型结合软硬件应用，新物种涌现

2023 年是大模型之年，2024 年也将是“小”模型之年。更多几十亿到上百亿参数的小模型，通过模型架构、算法、训练和精调的创新，以及结合外部检索，性能可以叫板百亿参数大模型，甚至追平 GPT-3.5（1750 亿参数）。开源模型许多来自中国、欧洲、韩国、甚至中东等地，以更快的速度推广到各行各业。小模型尤其适于下载到设备上，在许多功能上可以替代从云上提供的大模型服务。小模型 + 终端设备是 2024 年的重要看点。

Mistral 7B 和 Mixtral 8x7B，碾压 Llama2



来源: Mistral, <https://mistral.ai/news/mixtral-of-experts>

说明: Mistral 模型参数规模分别为 70 亿与 120 亿, LLaMA 2 模型参数规模分别为 70 亿, 130 亿与 700 亿。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/57716614000006030>