华泰证券 HUATAI SECURITIES

AI 2.0：十年之后我们还能做什么

华泰研究

2023 年 6 月 26 日 | 中国内地　　专题研究

科技　　增持（维持）

研究员　黄乐平，PhD
SAC No. S0570521050001　leping.huang@htsc.com
SFC No. AUZ066　+(852) 3658 6000

研究员　谢春生
SAC No. S0570519080006　xiechunsheng@htsc.com
SFC No. BQZ938　+(86) 21 2987 2036

研究员　朱珺
SAC No. S0570520040004　zhujun016731@htsc.com
SFC No. BPX711　+(86) 10 6321 1166

研究员　余熠
SAC No. S0570520090002　yuyi@htsc.com
SFC No. BNC535　+(86) 755 8249 2388

研究员　陈旭东
SAC No. S0570521070004　chenxudong@htsc.com
SFC No. BPH392　+(86) 21 2897 2228

研究员　丁宁
SAC No. S0570522120003　dingning021681@htsc.com
+(86) 21 2897 2228

研究员　张皓怡
SAC No. S0570522020001　zhanghaoyi@htsc.com
+(86) 21 2897 2228

联系人　王兴
SAC No. S0570121070161　wangxing@htsc.com
+(86) 21 3847 6737

联系人　张宇
SAC No. S0570121090024　zhangyu@htsc.com
SFC No. BSF274　+(86) 10 6321 1166

联系人　王心怡
SAC No. S0570121070166　xinyi.wang@htsc.com
SFC No. BTB527　+(86) 21 2897 2228

联系人　权鹤阳
SAC No. S0570122070045　quanheyang@htsc.com
+(86) 21 2897 2228

联系人　于可熠
SAC No. S0570122120079　yukeyi@htsc.com
+(86) 21 2897 2228

AI 大模型展现惊人能力，有望成为下一代通用技术平台

2010 年，麻省理工大学阿齐默鲁教授等提出了科技发展如何影响人类就业的分析框架。当前，随着以 GPT-4 为代表的大语言模型的出现，AI 开始具备文本生成、语言理解、知识问答、逻辑推理等能力，AI 大模型出现通用性强、固定成本高但边际成本递减等平台性技术的显著特征，未来有可能大规模替代当前人类的工作。因此，我们认为现在是重新思考人工智能将如何影响工作和投资的又一重要时刻。我们认为：1）大模型对人类工作的替代将从常规性的脑力劳动开始，金融法律服务、软件外包、传媒等工作可能最先被影响。2）从投资角度，建议按照算力基础设施->硬件载体->大模型平台->应用的顺序寻找受益标的。

投资机会#1：科技巨头在大模型上竞争激化，芯片等送水人最先受益

我们认为，算力需求增长会率先利好芯片、光模块、服务器产业链等送水人。根据 OpenAI 测算，目前算力的增速（翻 1 倍 /3-4 个月）远超过摩尔定律（翻 1 倍 /18-24 个月），未来 AI 应用的逐步丰富，将推动推理芯片等相关市场保持强劲增长。我们预计 2025 年左右当生成式 AI 应用大规模落地后，用于数据中心的推理及训练芯片市场约 900 亿美元，带来先进代工、先进封装、光模块、服务器在内相关产业链机会。

投资机会#2：谁会成为 2.0 时代的硬件载体？

回顾消费电子硬件的发展历程，硬件的升级和应用的发展往往相互促进。年初以来，ChatGPT Plugins 和 Auto GPT 已让我们在应用层面看到了大语言模型的潜力，有望催化智能终端的发展。一方面，大模型将为智能手机、PC 平板、音箱等带来新功能，推动硬件规格升级；另一方面也有望催化 AR/VR、无人驾驶汽车、机器人等新的硬件形态，带动产业链发展。长期来看，或将形成同时拥有 1）超强感知能力，2）通用人工智能（AGI）和 3）灵活的行动能力的硬件产品。虽然离大规模应用仍有距离，但长期来看，建议关注机器人等终端成为 AI2.0 时代的主流载体。

投资机会#3：世界最后需要几家大模型公司？

我们认为，大语言模型是一个资本密集、人才密集和数据密集的产业，是否拥有大语言模型将成为衡量科技企业甚至衡量国家综合实力的重要尺度之一。当前全球大模型呈现“百模大战”的竞争格局，我们认为最终可能会收敛到国内外各数家的寡头竞争格局。在“百模大战”中最终胜出的企业，有望成为可以类比当前百度、阿里、腾讯体量的科技平台型企业。是否能形成“数据-模型-应用”的飞轮是判断大模型企业能否在竞争中胜出的重要标准之一。

投资机会#4：最大机会在行业，看好互联网、办公、金融等行业落地

我们认为，AI 2.0 最大的投资机会在应用，AI 落地的节奏或和行业数字化程度成正比，看好互联网（搜索+广告营销）、办公、金融，教育等行业率先迎来“iPhone 时刻”，关注大模型在：1）电商、金融、医疗等行业替代传统客服；2）文本和图像生成在办公、广告营销、金融、影视游戏等领域成为下一代生产力工具；3）多模态能力在新药开发等科研上的应用前景。

风险提示：宏观经济波动；新技术渗透进度不及预期；中美贸易摩擦加剧；本研报中涉及到未上市公司或未覆盖个股内容，均系对其客观公开信息的整理，并不代表本研究团队对该公司、该股票的推荐或覆盖。

更多免费行业报告 并购家 www.ipoipo.cn

正文目录

AI 2.0: 十年之后我们还能做什么 4
AI 2.0 是下一代通用技术平台 5
投资机会#1：算力需求增长利好芯片送水人 7
投资机会#2：谁会成为 2.0 时代的硬件载体？ 9
投资机会#3：世界最后需要几家大模型公司？ 10
投资机会#4：如何把握大模型应用的投资机会 11
2023 年十大投资机会 14

AI 大模型企业是如何炼成的 16
海外大模型：微软&OpenAI、谷歌、英伟达、Meta、AWS 16
国内大模型：百度、阿里、腾讯、华为、商汤、讯飞 31
相关未上市公司：智谱、MiniMax、第四范式、聆心智能、澜舟科技、云知声、扩博智能 44

算力：AI 应用如何拉动需求变化 46
大模型对训练/推理算力需求的测算 47
计算芯片：算力储备竞赛开启，推理初步具备进口替代能力 49
服务器：AI 服务器市场稳步增长，2021-2026E CAGR 预计达 17% 54
先进封装&材料：高算力需求带动先进封装需求 56
光模块：数据中心市场空间广阔，2023 年有望迎 800G 放量元年 57
服务器产业链相关公司：工业富联、环旭电子 60
计算芯片产业链相关公司：燧原科技、壁仞科技、寒武纪、京微齐力、芯动科技、沐曦集成电路、海光信息、登临科技、中科驭数、摩尔线程、天数智芯、高云半导体、墨芯人工智能、智多晶、云天畅想 60
光通信产业链相关公司：长光华芯、源杰科技、云岭光电、苏州熹联、天孚通信、中际旭创 64
云计算产业链相关公司：道客云、鹏云网络、StreamNative 66

AI 赋能百业：产业变革的奇点 68
AI+企业服务：赋能生产力/沟通/协作工具 68
AI+汽车：AI 大模型加速无人驾驶发展 70
AI+工业：ChatGPT 引领 AI 赋能工业，工业智能化水平或将加速提升 75
AI+金融：坐拥数据富矿，有望成为 AI 大模型率先落地的垂直领域之一 79

AI 赋能传媒：AIGC 应用鹏程万里 83
AI+游戏：降本增效持续验证，AI 新玩法尝试创新 83
AI+电商：多场景应用，降本增效提升体验 87
AI+影视：驱动影视创作升级 92
产业链相关公司：昆仑万维、汤姆猫、蓝色光标、焦点科技、华策影视等 98

AI for Science：人工智能助力科学探索的新范式 100
AI+新药开发：根据功能需求设计/优化蛋白质，加速新药探索速度 101
AI+量子：优化软硬件设计，提升量子技术应用的效率与准确性 103
AI+核聚变：精准调控核聚变控制过程，降低开发周期 105
AI+材料科学：逆向设计是 AI 助力材料研发创新的主要范式 106
AI for Science 产业链相关公司：国仪量子、本源量子、讯飞医疗、BrainCo、脑虎科技 108

附录：AI 大模型算法的前世今生 110
溯源：从经典神经网络到 Transformer 架构 111
发展：从 GPT-1 到 GPT-4，开启大模型新纪元 115
海外落地：BloombergGPT、AutoGPT、SAM 127

附录：AI 大模型需要什么数据？ 132
数据将是未来 AI 大模型竞争的关键要素 132

更多免费行业报告 并购家 www.ipoipo.cn

海外主要大模型数据集.....132
中国主要 AI 大模型数据集.....138
高质量语言数据和图像数据或将耗尽，合成数据有望生成大模型数据.....145
数据产业链投资机会：数据生产环节、数据处理环节.....146
数据服务相关公司: 百度智能云、慧听、标贝、它思、Appen、Telus international、Scale AI.....148

半导体国产化：供需错配推动半导体制造国产化加速.....150
中国半导体供需错配，国产替代主要集中在设计、封测等环节.....150
机会#1 半导体设备：关注先进制程突破以及国产化率提升.....155
机会#2 半导体设备零部件：国产化地位开始受到重视.....161
机会#3 材料：国产基础较好，掩模版、光刻胶国产化率较低.....166
半导体产业链未上市公司：悦芯科技、华封科技、先锋半导体、御微半导体、世禹精密、汉骅半导体、佑伦真空、富乐华半导体、泰科天润、锐杰微科技.....172

汽车智能化：红海竞争的破局点.....175
智能驾驶：高级别辅助驾驶蓄势待发，国产供应商受益.....175
智能座舱：从 4.0 多模态到 5.0 大模型，座舱智能化全面升级.....180
车载芯片相关公司：地平线、黑芝麻智能、芯驰科技、芯擎科技，旗芯微.....186
域控及数据标注相关公司：德赛西威、经纬恒润、纵目科技、魔视智能、宏景智驾、柏川数据.....187
连接器相关公司：电连技术、瑞可达.....188
传感器相关公司：加特兰、森思泰克、行易道、禾赛科技、速腾聚创、图达通、探维科技、灵明光子.....189
AR-HUD 相关公司：泽景科技、未来黑科技.....191

商业航天：大年开启，关注行业机遇.....192
技术端：火箭、卫星制造技术均有突破.....192
需求侧：行业需求加快落地.....203
政策面：利好政策频出，投融资稳健增长.....210
火箭相关公司：星际荣耀、星河动力、蓝箭航天、深蓝航天、中科宇航、天兵科技、凌空天行、东方空间.....214
卫星制造相关公司：银河航天、长光卫星、时空道宇、工大卫星、国博电子、铖昌科技、海格通信、天银机电.....215
地面设备相关公司：震有科技、航天驭星、信科移动.....216
卫星运营服务相关公司：中国星网、中国卫通、航天宏图、中科星图.....217

元宇宙 2023: 苹果这次有什么不同？.....218
Meta Oculus 的失败教训.....218
苹果带来的变化.....221
生成式 AI 可能带来哪些变化.....227
海外公司：英伟达，Epic games，Unity.....230
整机品牌：Nreal、亮亮视野、Rokid、亮风台、大朋 VR.....233
光模组领域：鲲游、耐德佳、至格、JBD、视涯、宏禧科技、赛富乐斯、熙泰、研鼎、蓝特光学.....235

Web3：过去一年发生了什么？.....241
Web3：以所有权为核心的价值网络.....241
风险事件频发，加密合规时代已至.....243
基础设施层：寒冬孕育机遇，Web3 基础设施创新持续.....247
应用层：市场回归冷静，关注 Web3 与真实世界连接.....249
未来展望：Web3 离我们还有多远？.....255
Web3 产业链相关公司：趣链科技、矩阵元.....256
风险提示.....257

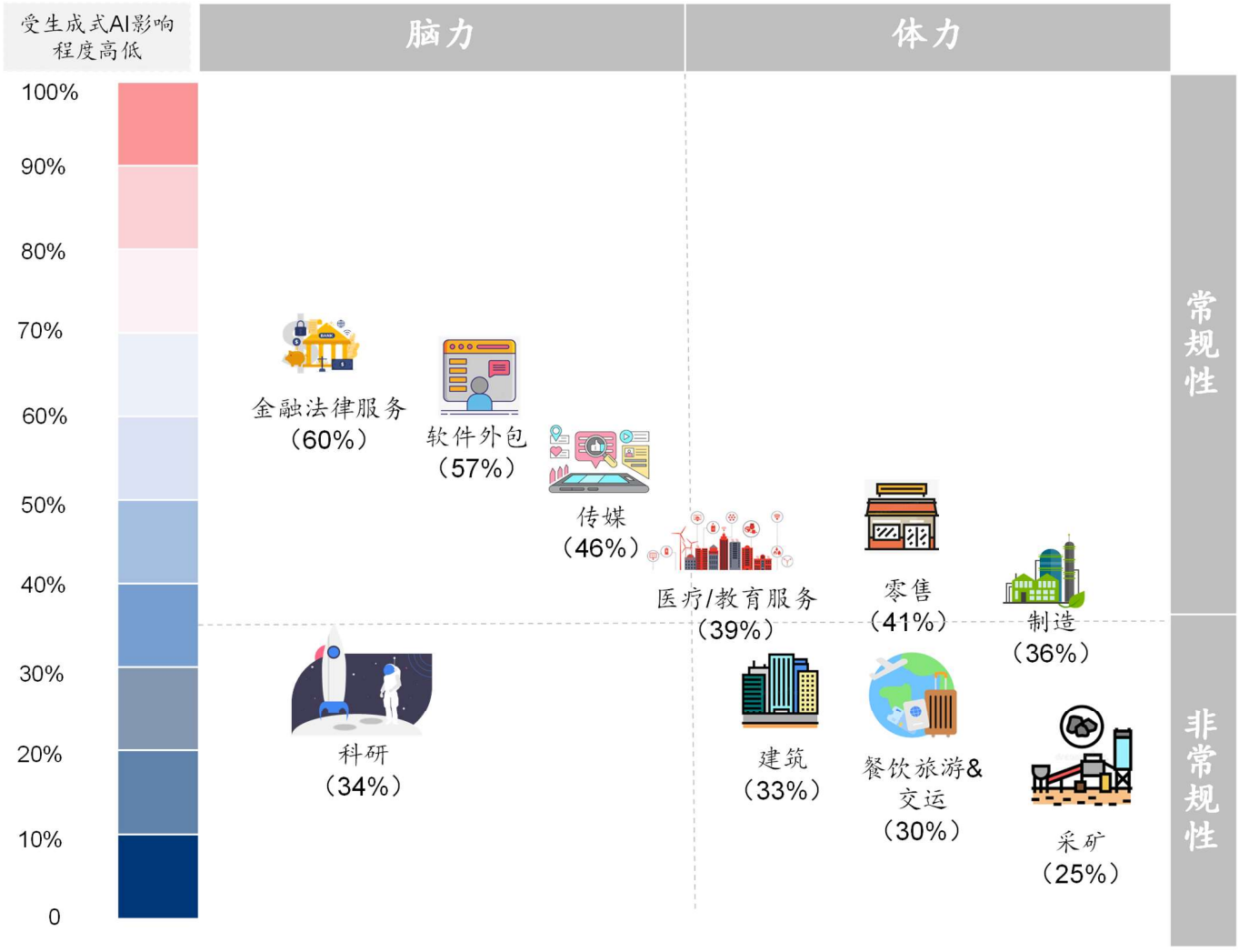
免责
更多免费行业报告 并购家 www.ipoipo.cn

AI 2.0：十年之后我们还能做什么

2010 年，麻省理工大学应用经济学教授达隆·阿齐默鲁（Daron Acemoglu）等人在论文[1]中提出了科技发展如何影响人类就业和收入的分析框架（图 1）。当前，随着以 GPT-4 为代表的新一代 AI 技术的出现，人类正面临重新思考人工智能将如何影响我们今后工作生活的又一重要时刻。

参考阿齐默鲁教授的模型和咨询公司埃森哲近期关于工作受到生成式 AI 影响大小的研究结果[2]，我们把十个主要行业根据：1）工作方式是脑力为主还是体力为主，（2）工作内容是常规性还是非常规性进行划分，归纳了不同性质工作受到生成式 AI 影响的程度。我们认为，随着 AI 能力的不断增强，大语言模型（LLM）将从金融法律服务、软件外包等常规性的脑力工作开始，逐步替代人类工作。

图表1：AI 2.0 时代，生成式 AI 或将逐步代替人类工作



资料来源：Acemoglu and Autor (2010) *Skills, Tasks and Technologies: Implications for Employment and Earnings*，埃森哲，华泰研究

具体结论如下：

1）**常规性的脑力劳动容易被算法替代。**这具体包括金融法律服务、软件外包、传媒等白领工作。这些工作需要收集，分析和解释大量基于语言的数据和信息，且含有较多重复性的知识劳动，与大语言模型的文本理解/生成以及逻辑推理能力高度重合，受到生成式 AI 的影响可能较大；

[1] Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Handbook of labor economics (Vol. 4, pp. 1043-1171).
[2] Accenture(2023). A new era of generative AI for everyone.

更多免费行业报告 并购家 www.ipoipo.cn

2）**非常规性的脑力活动较难被算法替代。**相较白领工作，新药开发等科研工作需要更高层次的认知和创造力，并且工作内容多变，难以被算法替代，受到生成式 AI 的影响较白领更小；

3）**常规性的体力劳动容易被自动化技术替代。**在零售、制造等行业中，虽然涉及的语言相关工作不多，受生成式 AI 的影响不大，但工作中有较多重复且任务标准化程度高的内容，容易被机器人等自动化技术所替代；

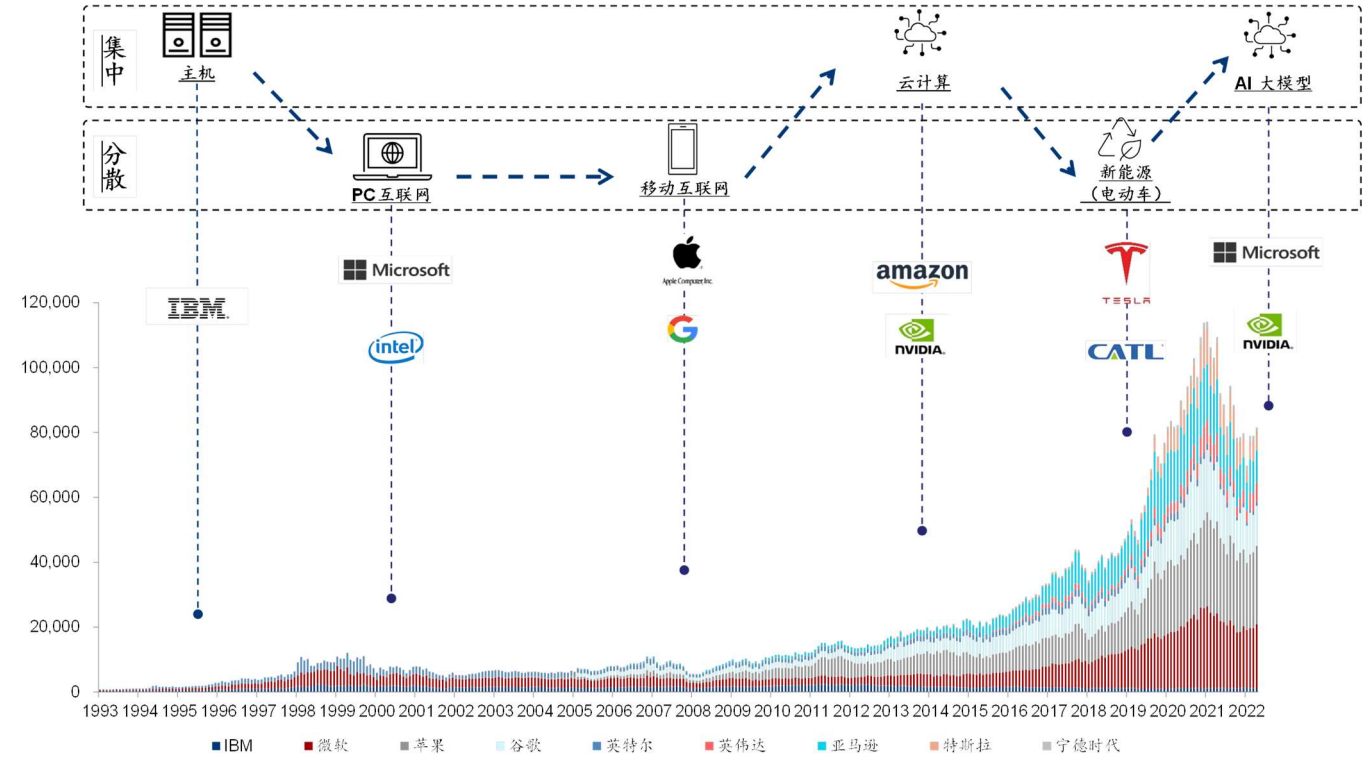
4）**非常规性的体力劳动受影响较小。**建筑、餐饮旅游及交运、采矿等行业虽然也包含较多的体力劳动，但难以被自动化的长尾场景较多，目前看来，受到生成式 AI 和机器人技术的影响都较小。

AI 2.0 是下一代通用技术平台

我们认为以 ChatGPT 为代表的大语言模型的出现，使人工智能技术的发展进入了 2.0 时代。AI 大模型已经具备 1) 通用性强，2）固定成本高但边际成本递减等平台性技术的显著特征，有望成为继 PC，移动互联网，云计算，电动车之后，下一个支撑科技创新的通用技术平台。

如下图所示，每一代通用技术平台的出现，都会伴随几家平台性公司的出现。这具体包括，1）主机时代的 IBM，2）PC 时代的微软和 Intel（Wintel），3）智能手机时代的苹果和谷歌，4）云计算时代的亚马逊和英伟达，5) 智能电动车时代的特斯拉和宁德时代。当我们进入 AI 大模型时代，我们认为，以微软为代表的大模型厂商和以英伟达为代表的算力提供方有望受益于 AI2.0 的崛起，实现平台扩张。

图表2： AI 2.0 是下一代通用技术平台

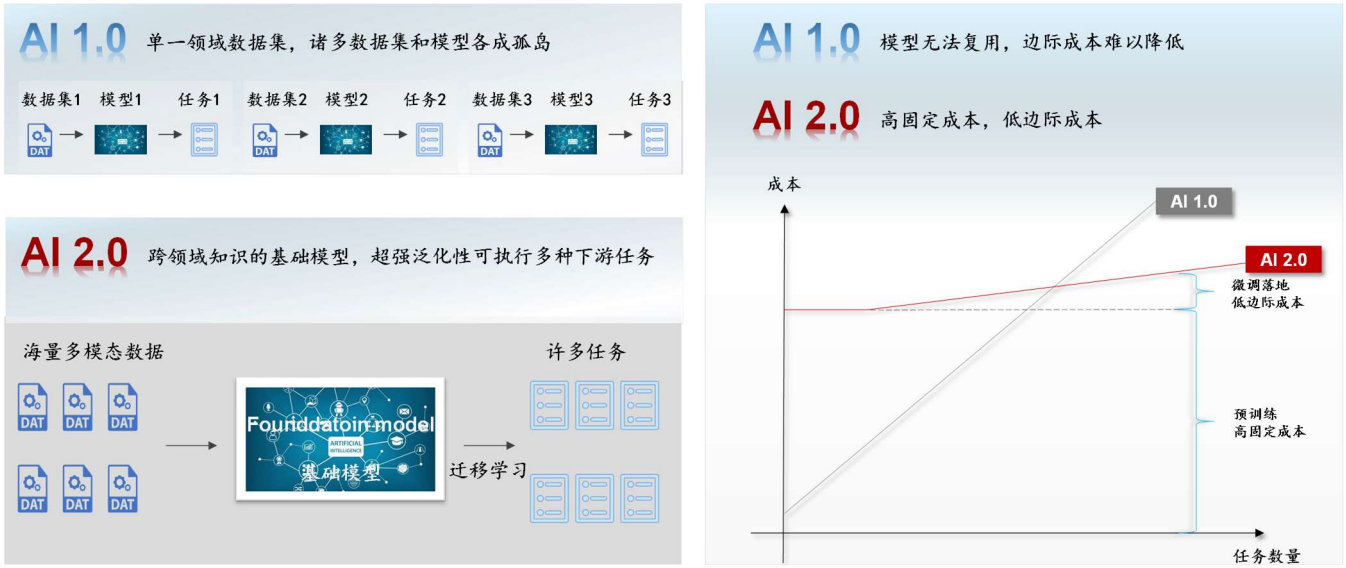


资料来源：，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

过去，计算机视觉等 AI 模型需要针对每个特定任务进行数据采集、模型训练等一系列工作，且训练出来的模型无法用到其他场景，这导致边际成本难以降低、难以大规模落地等问题。大模型的出现改变了 AI 模型的训练和使用的范式。在大模型的新范式下，首先训练出具备跨领域知识的基础模型（Foundation Model），然后通过微调等方式适配和执行下游各领域的任务，能够降低模型落地的边际成本。大模型的新范式带来了成本结构变迁：大模型出现之前的 AI 1.0 时代，模型训练成本和模型数量等比例增加，边际成本较高；AI 2.0 时代，训练基础模型的固定成本高，但微调以适应下游应用的成本极低，预训练模型+微调的新范式边际成本低，具有明显的平台效应，为 AI 公司探索新商业模式提供机会。

图表3：AI 步入 2.0 时代，边际成本大幅度下降



资料来源：《人工智能发展报告 2011-2020》（清华大学，2021），华泰研究

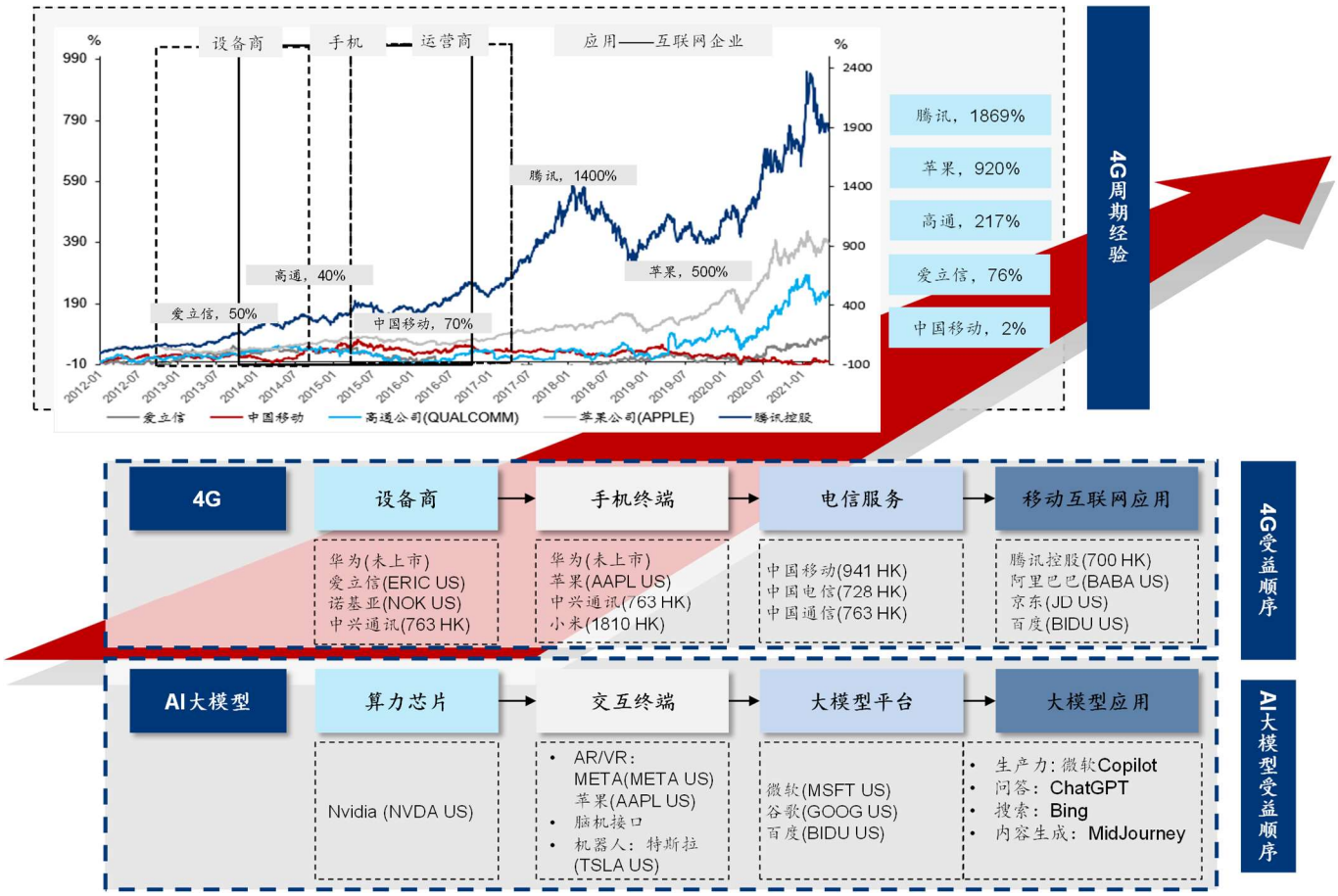
我们相信 AI 2.0 是未来十年科技行业最重要的投资机会，建议投资人参考 4G 的投资逻辑，按照基础设施->终端->平台->应用 顺序寻找投资标的。

复盘 4G 产业发展历程，我们看到在受益板块中电信设备商先行，手机次之、电信服务及移动互联网应用紧跟其后。第一阶段，爱立信等通信网络设备商享受最初的资本开支红利，但由于华为出现导致的格局改变以及设备投入存在周期性，股价波动回到原点；其次，中国三大运营商在 4G 时代发挥着至关重要的作用，但由于来自监管的提速降费压力和业务范围的制约，运营商盈利增速缓慢；4G 时代应用端的主要附加价值被互联网占据，社交、游戏、电商和视频等领域出现了一批杀手级应用，以腾讯为代表的平台型企业也在互联网浪潮中实现市值飞跃。

展望 AI 大模型时代，我们认为：
1)算力基础设施率先受益：AI 大模型推动算力需求快速增长，基础设施的业绩成长确定性最高，主要产业链环节包括计算芯片、服务器、数据中心等；
2)交互终端次之，当前智能手机是大模型应用的主要载体。未来看好大模型能力提升推动 AR/VR，无人驾驶车，机器人等成为下一代硬件载体的潜力。
3)大模型平台是当下产业发展趋势中最核心的一环，看好 MaaS 成为 AI 大模型时代新的商业模式，关注各国对 AI 大模型企业监管政策对行业发展的影响；
4)AI 2.0 最大的投资机会在应用。看好大模型在搜索，电商零售，办公，金融、医药等行业率先落地。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表4：从 4G 周期经验看 AI 大模型受益顺序



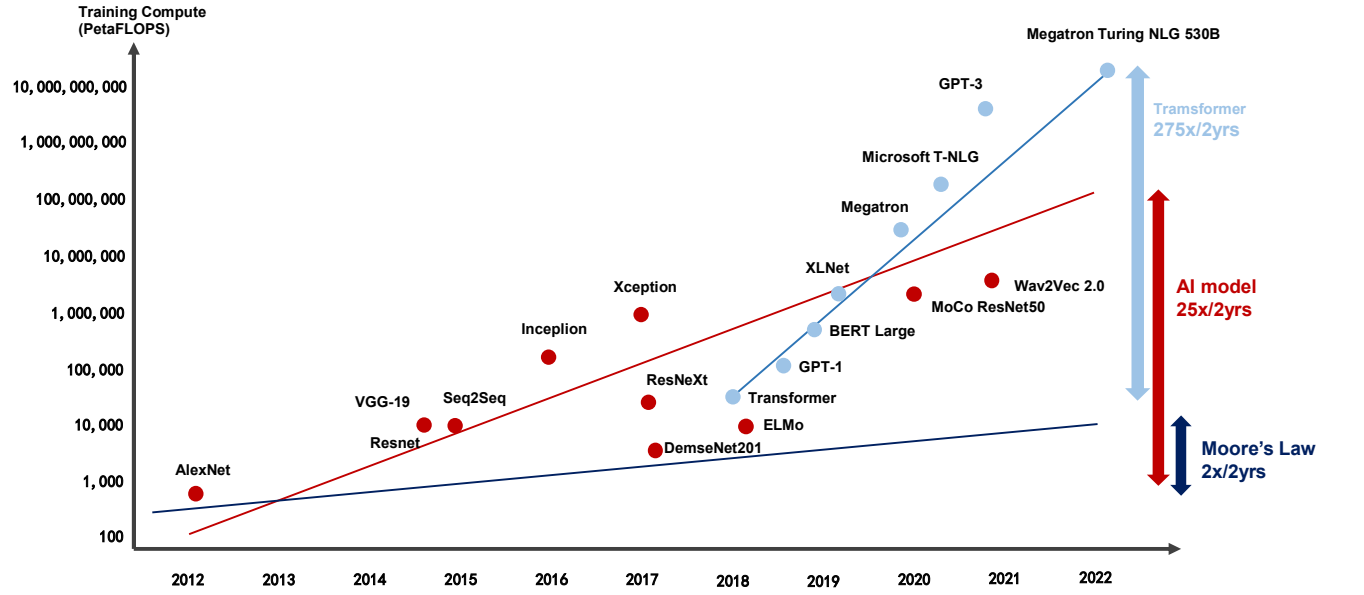
资料来源：Bloomberg，华泰研究

投资机会#1：算力需求增长利好芯片送水人

我们认为，算力需求增长会率先利好算力芯片，光模块，服务器产业链等送水人。根据 OpenAI 测算，目前算力的增速（翻 1 倍 /3-4 个月）远超过摩尔定律（翻 1 倍 /18-24 个月），未来 AI 应用的逐步丰富，将推动推理芯片等相关市场保持强劲增长。我们预计 2025 年左右当生成式 AI 应用大规模落地后，数据中心用推理及训练芯片市场约 900 亿美元，带动先进代工、先进封装、光模块、服务器在内相关产业链。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表5：大模型对半导体需求增长快于摩尔定律



资料来源：NVIDIA Ian Buck statement (2018)，，华泰研究

算力芯片：全球 900 亿美金市场，训练英伟达一家独大，推理百花齐放。我们认为是否拥有大模型将成为科技巨头科技平台企业的重要分水岭。我们预计全球有约 30 家科技巨头和 300 家 AI 大模型初创企业在进行算力相关投资。到 2025 年，这对应大约每年 300 亿美金训练芯片和 600 亿美金的推理芯片市场。其中，英伟达凭借 CUDA 软件框架构建的强大行业壁垒，在训练芯片市场一家独大，推理芯片方面，海外客户除了采用英伟达、AMD 等的通用 GPU 以外，也会采用 TPU 等 ASIC 以提高性价比。国内客户出于供应链安全考虑，逐步提升国产芯片的适用占比。海思，寒武纪，燧原等企业都有较好发展。

服务器/PCB/先进封装：AI 服务器增速超传统服务器，高算力需求带动先进封装需求。从 GPU 到 AI 工厂，我们认为 AI 计算会为整个服务器产业链带来增长。相较于全球传统服务器需求持续低迷，根据 IDC 预测 AI 服务器 21-26 年复合增速有望达到 17%。我们看到从单颗 GPU 芯片 H100，通过 NVLINK Switch 形成一颗巨型 GPU，然后通过 Quantum InfiniBand 技术，搭建有上百张 GPU 的 DGX 服务器，最后把多台 DGX 联通形成一台 AI 超级计算机，整个系统的搭建过程会带动包括服务器整机、PCB、光纤光缆、电源在内的整个服务器产业链的性能升级。在先进封装领域，产业从 2.5D 封装正走向 3D 封装，根据 Prismark 预测 ABF 载板作为先进封装关键材料有望在 21-26 年实现复合增速 11.5%，而相关测试设备有望长期受益于 Chiplet 和国产替代趋势。

光模块：AI 大模型推动 800G 光模块迎放量元年。以 ChatGPT 为代表的 AI 大模型，对数据中心内外的数据流量都提出了新要求，将带来光模块行业“量”和“质”的双重提升。以 ChatGPT 为代表的 AI 大模型训练中，由于需要海量数据及跨机器协作，产生了大量的数据中心内部通信需求，无阻塞的胖树结构成为当下最主流的 AI 训练网络架构，带来两方面主要变化：1）量方面，AI 网络架构带来数据中心内部交换机、服务器数量的增加，光模块作为数据中心内外部连接的核心部件，用量将明显提升；2）质方面，AI 对于高速率、大带宽的网络需求将推动光模块向 800G 加速升级，根据 Lightcounting 预测，2023 将成为 800G 光模块放量元年，2028 年 800G 出货量有望达 998 万只，对应 2023-2028 年复合增长率为 72%。在此产业趋势下，我们认为能够提供高速光模块的龙头企业以及上游光器件、光芯片厂商有望迎发展机遇。

免责
更多免费行业报告 并购家 www.ipoipo.cn

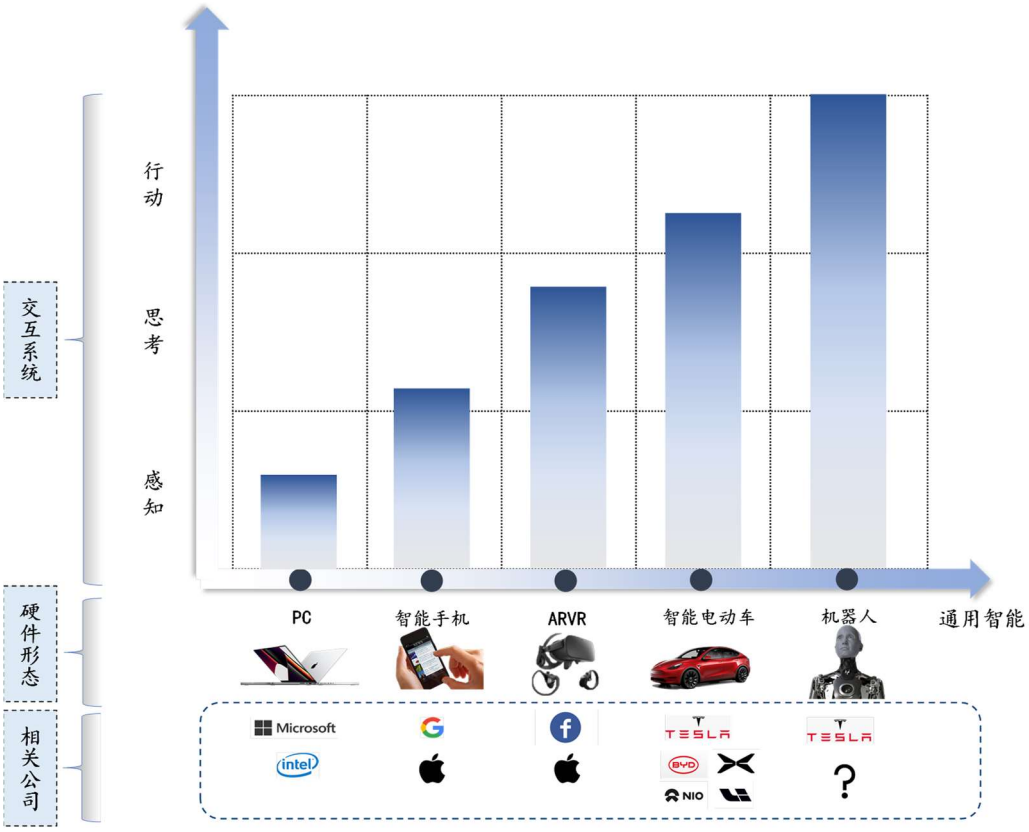
投资机会#2：谁会成为 2.0 时代的硬件载体？

我们认为大语言模型能力的不断增强，在为智能手机等现有硬件产品带来 AI 聊天等新的功能的同时，也会催生机器人，ARVR，无人驾驶车等新的硬件形态。从消费电子硬件变迁历程来看，硬件的升级和应用的发展是一个互相促进的关系。一个新的硬件形态（例如 2008 年的 iPhone, 2023 年的苹果 Vison Pro）会吸引开发者开发新的应用，另一方面，新的应用（2013 年左右的手游）会推动硬件规格的升级。

我们认为一个应用的能力可以分为感知、思考和行动三个维度。过去 15 年智能手机和电动车的发展，使手机，电动车等产品已经具备了较强的“感知”能力。过去几年 AI 云计算的发展，使硬件初步拥有了“思考”的能力。

展望未来，我们认为大语言模型会大幅提高硬件产品的“思考”能力，机器人技术的发展会提高硬件产品的“行动”能力。一个同时拥有（1）超强感知能力，（2）通用人工智能（AGI），和(3) 灵活的行动能力的硬件产品，可能就是现在大家畅想中的人形机器人。虽然很难预判什么时候这个产品能够实现，但在最终实现人形机器人的过程中，我们相信会催熟 ARVR，无人驾驶车等很多新的硬件形态，带动产业发展。

图表6：大模型发展推动硬件载体的变迁



资料来源：华泰研究

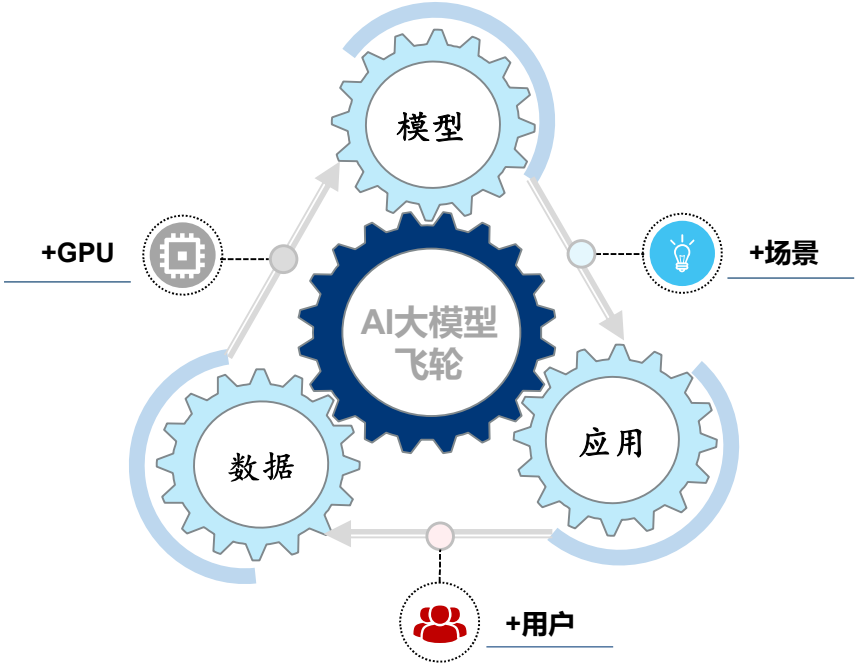
免责
更多免费行业报告 并购家 www.ipoipo.cn

投资机会#3：世界最后需要几家大模型公司？

我们认为是否拥有大模型将成为科技巨头科技平台企业的重要分水岭。目前 AI 行业呈现出“百模大战”的格局，我们认为未来基础大模型可能呈现国内外各有数个赢家的**寡头竞争**格局。类比云计算领域，国外有 AWS（亚马逊）、Azure（微软）和 GCP（谷歌）三巨头，国内有 BAT 和华为，前期基础设施的巨大投入造成了云计算的寡头市场格局。与云计算类似，基础大模型需要耗费大量算力和数据标注成本，并且随着规模的增加，其训练费用将指数级增长。在高壁垒和地缘政治等原因下，未来基础大模型可能在国内外各自形成寡头竞争格局。

AI 大模型是一个资本密集，人才密集和数据密集的产业，如何形成“数据-模型-应用”的飞轮，是大模型企业成功的关键。具体而言，大模型的诞生以算力基础设施为根基，以算法为骨干，以高质量数据为血肉；大模型结合具体应用场景的落地产生应用；用户使用应用的过程中产生数据，数据又得以反哺模型。由此，从大模型的训练到应用落地产生闭环，我们认为“数据-模型-应用”的飞轮能否运转是大模型企业可持续发展和迭代的关键。

图表7：形成“数据-模型-应用”的飞轮，是大模型企业成功的关键

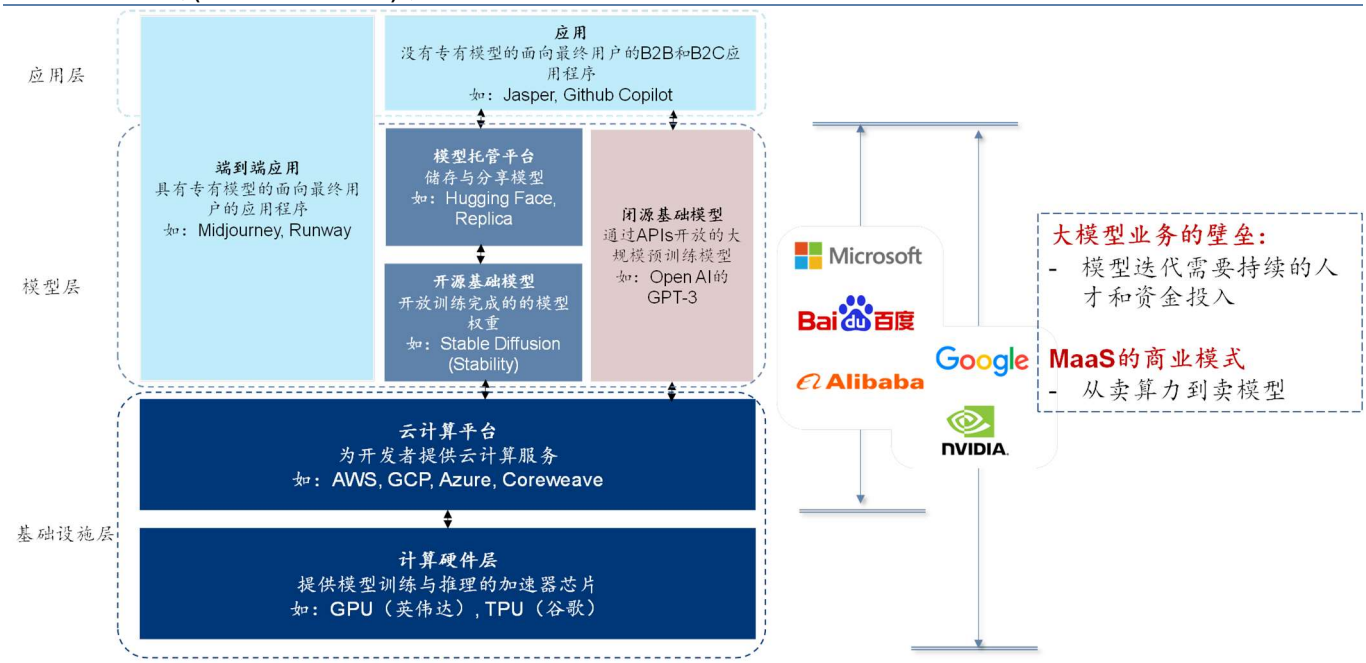


资料来源：华泰研究

MaaS 可能成为 AI 大模型时代新的商业模式。过去云计算架构主要分为 IaaS、PaaS、SaaS 三层，经过多年发展，出现了像微软，MongoDB，Adobe 等一批代表性企业。未来，我们认为，MaaS (Model as a Service）可能成为 AI 大模型时代一种新的商业模式。如上所述，大模型具有固定成本高，通用性强等特点，MaaS 企业依托已经训练好的大模型，向 2B 客户提供包括 API 调用、模型训练、模型微调、模型部署服务等各类服务，数据，算法，算力将是 MaaS 企业的主要竞争壁垒。目前，国内，阿里，百度，商汤等都已经宣布提供类似服务。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表8： 大模型服务(Model as a Service)有望成为 AI 时代新业态



资料来源： A16Z 官网，华泰研究

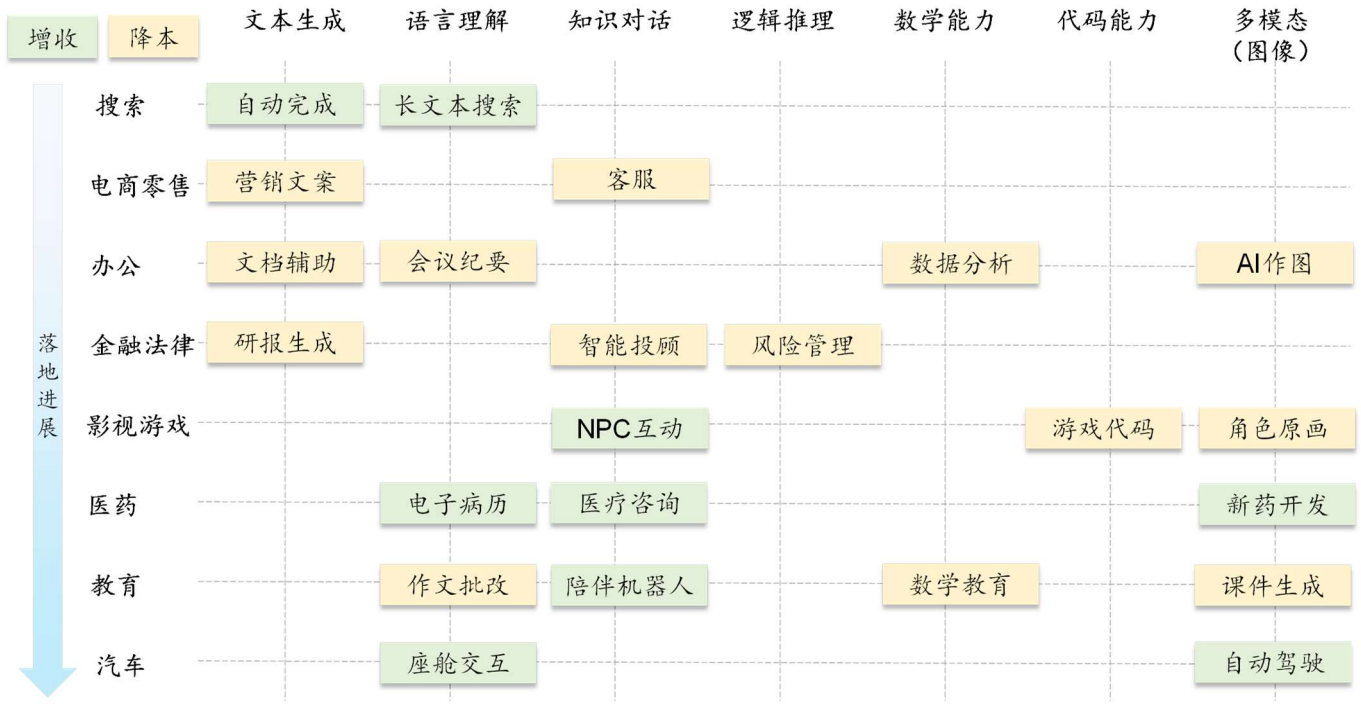
监管政策是规范 AI 大模型健康发展的基础，关注国内政策落地进展。大模型由于涉及到大量个人和企业数据，生成内容中存在虚假信息等风险，是当前全球各国监管关注的重点。23 年 6 月，欧洲议会在全球率先通过《人工智能法案》，对 AI 模型提出了版权披露、保障隐私与非歧视等基本权利、风险分级监管等要求。23 年 4 月，国家网信办发布发布《生成式人工智能服务管理办法（征求意见稿）》，明确要求，符合大模型生成的内容需要符合"社会主义核心价值观"，并要求人工智能企业在提供服务前，向国家网信部门申报安全评估；并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更，注销备案等手续。关注监管政策落地对行业发展的影响。

投资机会#4：如何把握大模型应用的投资机会

我们认为，长期来看 AI 2.0 最大的投资机会在应用。根据科大讯飞董事长刘庆峰的观点，通用人工智能应当具备**文本生成、语言理解，知识问答、逻辑推理、数学、代码、多模态**等七大维度的能力。通过对 AI 在搜索、电商零售、办公、金融法律、影视游戏，医药、教育、汽车等行业应用前景的分析，我们认为，AI 应用的落地节奏或与行业数字化程度成正比，我们看到 AI 大模型在互联网（搜索+广告营销）、办公、金融等领域率先迎来“iPhone 时刻”。其中最值得关注的应用包括：1）知识对话在电商、金融、医疗等行业替代传统客服；2）文本和图像生成在办公、广告营销、金融、影视游戏等领域成为下一代生产力工具；3）多模态能力拓宽人类探讨世界的范围，特别是在新药开发上的应用前景。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表9：大语言模型如何赋能行业



资料来源：OpenBMB，华泰研究

AI+搜索：输入输出模态多样化，有望集成全新商业模式，打开行业天花板。传统的搜索行为往往需要经历长时间的判断过程和多轮的互动，而 AI 技术的加入，将从底层结构上改变搜索形态，以 GPT 为代表的大模型加持赋能搜索行业：1）创造性内容生产与输出，AI 大模型可以利用大量实时文本数据学习语言的规律，并根据用户指示创造性地完成特定需求下各种类型的文本输出，如故事、诗歌、摘要、代码等。2）拓展搜索的多模态能力，利用 AI 技术处理视频、图片、语音等多种形式的信息，使搜索不再受限于信息的形态。AI+搜索的诞生便吸引了全球的注意力，内置 AI 聊天功能的 New Bing 搜索引擎在推出一个月后日活跃用户已突破 1 亿，自 New Bing 推出以来，聊天功能使用次数累计超过 4500 万次。根据 Industry growth insight，2021 年全球搜索市场规模为 109.0 亿美元，预计 2023 年达到 123.6 亿美元。同时 AI+搜索也催生出未来新的商业模式，例如微软目前内部正在探讨在生成回复的引用链接中放置订阅制和弹出式广告的可行性，有望进一步打开新的市场空间。

AI+电商零售：定制化内容生成；个性化营销推广。目前广告/营销行业存在单位推广成本高、转化率低的痛点，AI 能赋能行业生产力的提升，主要体现在：1) 创新广告形式，提升信息传递效率，形成强流量、高互动，提高广告营收；2) 提高营销内容生产效率、推进个性化营销；3) 有望推动营销服务商商业模式革新、毛利率提升。通过以上赋能，行业增长空间有望提升，盈利能力。目前海内外 AI+广告/营销应用百花齐放，例如，国外 Adspert 利用 AI 将广告全自动化，优化亚马逊广告竞价投放，增加销售额，提升投产比；国内腾讯广告大模型赋能广告制作环节，同时助力提升广告主 15%GMV。

AI+办公：生产力的又一次跃升；重塑内容生产、沟通、协作模式。生成式 AI 当前在办公领域的应用主要包括改善生产力、沟通、协作工具等应用方向。以 Microsoft 365 Copilot 为例，Copilot 通过把 GPT-4 提供的内容生成功能，与存储在 Graph 数据库中的企业数据，以及 Word、Powerpoint、Excel、Outlook、Teams 等办公工具相结合，提供包括内容创作、数据分析、辅助决策等一系列新功能。生成式 AI 将传统的点式交互升级为自然语言的交互方式，能够实现更灵活的功能调度，进一步降低办公场景的沟通协作成本，提升工作效率。此外，生成式 AI 能够挖掘办公场景中的数据价值，提供知识管理等功能，进一步提升产品的价值空间。

免责
更多免费行业报告 并购家 www.ipoipo.cn

AI+金融：广泛应用于银行、保险、资管、投顾等环节，助力金融机构降本增效。我们认为垂直行业的高价值量数据对于 AI 大模型的训练和垂直领域应用至关重要，金融行业数字化程度领先，拥有数据富矿，有望成为 AI 大模型率先落地的垂直领域之一。应用端，我们看到生成式和理解式大模型在银行、保险、资管、投研、投顾等多个细分领域正在落地或拥有潜在落地场景，帮助金融机构降本增效。过去，理解式大模型主要用于信息识别与挖掘，例如风险识别、客户识别等等。生成式 AI 的主要落地场景包括应用于各金融子行业的智能客服以及内容生成，例如投研领域的研报生成、投顾领域的投资建议生成、保险/资管领域的合同模板生成等等。

AI+游戏：AIGC 带来玩法体验端的创新升级，以及生产营销端的降本提效。1)玩法体验端:目前看 AIGC 有望在 NPC 智能交互及玩法设计创新方面加速应用，AINPC 的不断成熟将改变游戏玩家与角色的交互方式，进一步提升游戏沉浸感；同时，AIGC 带来的 UGC 等玩法创新也有望带来全新的游戏类型，进一步提升游戏内容的丰富度及自由度。2)生产营销端:以 ChatGPT 等大语言模型为基础的文本类 AI 生成工具、StableDiffusion 及 Midiourney 等 AI 图像生成工具、Gen2 等 AI 视频生成工具等已在游戏行业各环节有所应用，同时英伟达、微软、Epic Games、Unity 等基于原有产品或服务纷纷推出 AIGC 功能。随工具化 AIGC 产品和服务快速迭代，行业“卖铲人”将加速 AIGC 对行业赋能，带来进一步的降本提效空间。

AI+教育：虚拟个性化辅导教学，提升教学质量同时降本增效。传统教育具有人工成本高、内容同质化以及资源分配不均匀的弊病。基于多模态模型在文本、图像等领域取得重大突破，生成式 AI 技术有望协助构建智慧教育新生态。近期国内外多家公司发布 AI+教育产品，例如多邻国在“Max 订阅方案”中推出两项基于 ChatGPT-4 设计的教辅功能、网易有道发布 AI 口语老师以及科大讯飞发布“1+N”星火认知大模型等。随着大模型的迭代升级，我们看好生成式 AI 应用于个性化辅导、AI 虚拟教师等领域，助力打造全新的低成本、个性化自适应教育模式，推动优质教育资源的规模化应用。同时建议关注 AI+教育带来的智慧校园、智能终端等硬件设备，以及相关 IT 基础设施、云服务等软件设施的投资机会。

AI+医疗：拓宽人类对自然界的认知半径，探索科学边界。医疗健康领域是人工智能率先落地的行业之一，也已经催生了 Nuance、IBM Watson 等一批全球知名企业。随着基于大模型的生成式 AI 的出现，我们看到 AI+医疗有望迎来一波新的发展机遇。我们看好生成式 AI：1）应用于实时病例生成、医疗影响分析等领域，从而发挥提高问诊效率、降低数据分析的人工成本等作用；2）赋能药物发现源头创新，探索自动根据功能需求设计/优化蛋白质、给定抗原等目标蛋白生成抗体等，加速药物发现流程。

AI+汽车：降低数据搜集、处理成本，优化驾驶体验，高级别智能驾驶落地可期。无人驾驶看上去很美，但一直很难落地。无人驾驶是最早被提出的人工智能应用场景之一，谷歌、苹果、特斯拉、百度等海内外科技巨头从 2016 开始就积极布局，但直到现在仍很难实现大规模商用落地。我们认为，1）多维度数据的获取和标注成本高，2）对小概率事件的决策准确度和人类还存在较大差距，3）事故时法律权责归属不明确，是制约其发展的部分原因。我们认为，1）以 ChatGPT 和 SAM 为代表的 AI 大模型的引入，自动标注、虚拟仿真等技术的引入大幅降低模型中数据的采集和处理成本；2）感知、预测、执行全流程算法迭代，体验能力的提升有助于从行业整体层面加速高级别智能驾驶的量产落地。看好 AI 大模型赋能之下，降低数据标注成本，提升感知预测执行全流程能力。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表10：AI 产业链相关公司

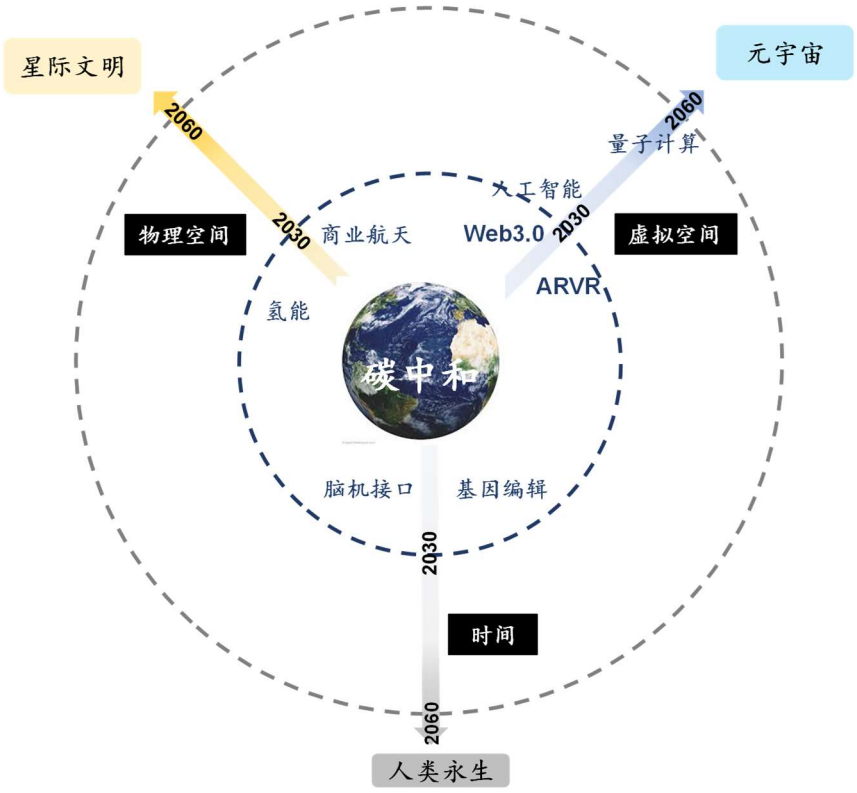
		海外/中国台湾公司	中国大陆公司
算力	芯片	英伟达、AMD、台积电	寒武纪、海光信息、芯原股份
	光模块/光芯片	Coherent	中际旭创、天孚通信、长光华芯、源杰科技
	先进封装	味之素、Ibiden、Shinko、ASMPT、Shibaura	长电科技、通富微电
	服务器	惠普、戴尔、广达、纬颖、英业达	工业富联、联想、浪潮信息
硬件	人形机器人	特斯拉	小米集团、优必选、绿的谐波、拓普集团、三花智控
	AR/VR	苹果、Meta	字节跳动
模型		微软、谷歌、英伟达、Meta、亚马逊	百度、阿里巴巴、腾讯、商汤、科大讯飞
应用	搜索	谷歌、微软	百度、三六零、昆仑万维
	电商零售	亚马逊、Shopify	阿里巴巴、京东、焦点科技
	办公	微软、谷歌、Adobe、Salesforce	金山办公、福昕软件、泛微网络、致远互联、万兴科技
	金融法律	Bloomberg	同花顺
	医药	Schrodinger、Exscientia、Atomwise	英矽智能、晶泰科技
	教育	DUOLINGO	科大讯飞、视源股份、传智教育、好未来、有道
	汽车	特斯拉	小鹏汽车、蔚来汽车、理想汽车

资料来源：华泰研究

2023 年十大投资机会

长期来看，我们认为全球科技行业的发展会沿着元宇宙，人类永生、和星际文明三个维度演进。其中，1） 追求永生是推动人类进步的动力之一，未来脑机接口、纳米机器人等前沿科技有望继续延长人类寿命；2）元宇宙是互联网的下一站，也是人类数字化迁移的下一步；3）移民火星的意义在于为人类文明留下“备份”，随着航天技术突破，星际旅行与火星移民终将实现。

图表11：科技板块投资机会



资料来源：华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

过去一年，在这三个维度上，除了 AI 大模型以外，我们还看到科技行业发生了以下变化。

1) 商业航天：近一年来，我们观察到中国商业航天产业链上下游发生明显变化。在技术端，国内民营液体火箭公司积极探索发动机技术与可回收技术突破，国内卫星制造在国家统筹规划、民企合力组建下步入规模化发展。在需求端，我国多项星座计划有序开展部署，中国星网预计明年上半年开始发射；“卫星直连”成为行业热点，手机、汽车、模组厂商积极开展尝试；工信部提出加快 6G 研发，为卫星互联网提供重要机遇。在政策面，国内多省市利好政策频出，助力商业航天产业成长。建议关注我国商业航天产业链发展机遇。

2) 元宇宙：受 Meta 缩小在硬件上的投入以及币价下跌等影响，元宇宙产业链从 2022 年下半年开始出现回调。6 月，苹果正式发布第一代头显产品 Vision Pro，在微显示技术（硅基 OLED)和人机交互技术（眼动和手动追踪）上和 Oculus 等现有产品比有了显著的提升。后续关注苹果如何发挥其在开发者和用户生态上的优势，以及生成式 AI 等最新的技术发展，激活元宇宙生态，值得关注。

3) Web3：回顾 Web3 过去一年，加密寒冬中市场风险事件频发，加速合规框架建立成为美国、欧洲、新加坡、香港等各国家/地区监管强共识。同时，行业创新持续，基础设施层：1）以太坊完成合并驱动生态进入新阶段，市场关注 Layer2 扩容、流动性质押等赛道。2）央行数字货币在过去一年取得了积极进展，或将成为区块链技术逐步迈入主流的铺垫。应用层：1）金融领域，流动性质押衍生品、现实世界资产等新赛道兴起，印证普通人进入门槛降低，Web3 正逐步完善与真实世界的连接；2）非金融领域，NFT 成交量高位跌落后，市场投机情绪减弱，关注 Web3 和音乐、社交、游戏等具体场景结合。

4) 半导体国产化：我们认为半导体设备和材料是中国半导体行业国产化的关键一环，美国、日本进一步收紧出口限制，中国半导体当前面临国内需求远大于国内供应链产能的供需错配现象。2022 年中国计算、手机、汽车、通信等系统厂商半导体消费额超过 1068 亿美元，约占全球半导体消费额的 25%，而中国代工收入仅为全球 10%。在半导体设备（全球市占率 5%）、材料（全球市占率 9%）等环节存在严重“卡脖子”问题。在国产化推动下，我们看好：（1）设备：刻蚀、薄膜沉积、量检测等；（2）零部件：射频电源、光学镜头、金属结构件等，（3）材料：光刻胶、抛光液/垫等投资机会。

5) 汽车智能化：在汽车电动化渗透率不断提高的背景下，汽车智能化较大可能成为下一个红海竞争的破局点，具体表现在：智能驾驶领域：领航辅助驾驶（NOA）等高级别辅助驾驶功能模型成熟度不断提高，市场领先者的测试版逐渐落地。我们认为今年或是城市 NOA 普及的元年并且成本有望进一步下降，国产供应商有望受益。智能座舱领域：现阶段智能座舱已经处于全面向 4.0 阶段发展的时期，硬件的装载是实现软件层交互的基础，座舱硬件已率先迎来大规模上车机遇；其次在智能座舱大模型生态发展下，我们认为拥有大模型和机器视觉技术布局的软件层企业同样有望受益。

免责
更多免费行业报告 并购家 www.ipoipo.cn

AI 大模型企业是如何炼成的

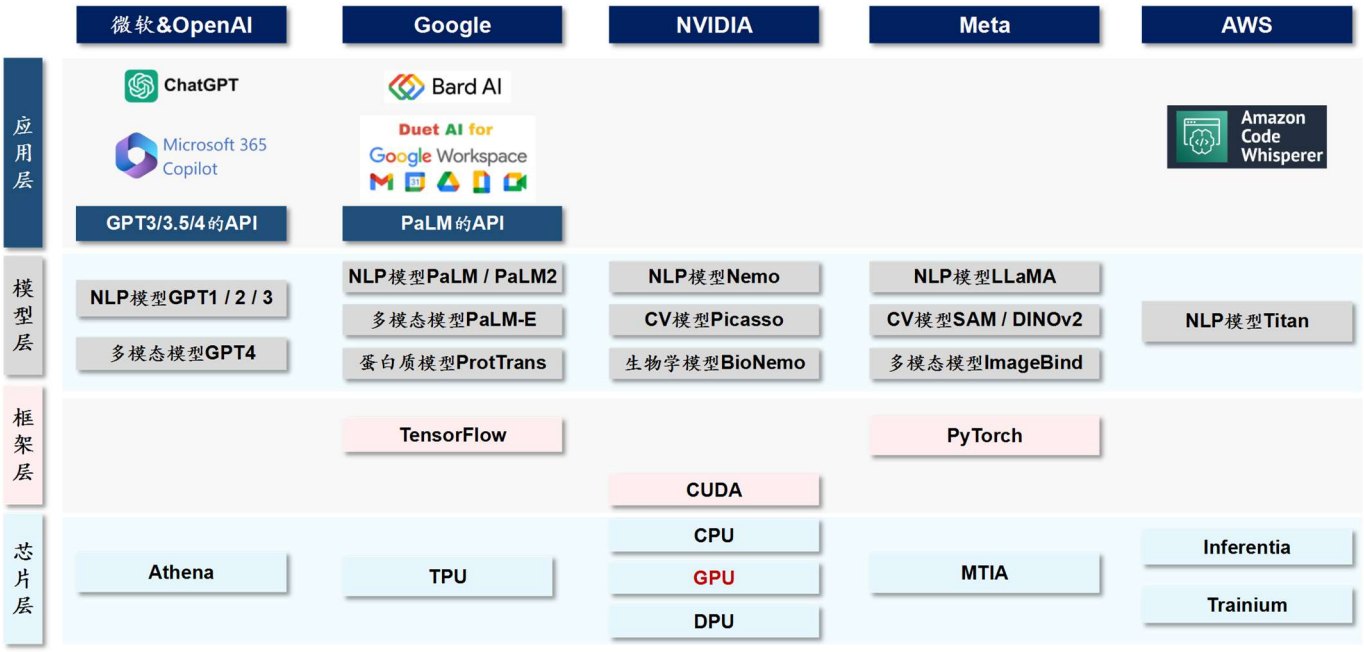
我们认为 AI 大模型是一个资本密集、人才密集和数据密集的产业，如何形成“数据-模型-应用”的飞轮，是大模型企业成功的关键。我们看到海外企业中，微软&OpenAI、谷歌已经逐步形成 AI 大模型的飞轮。当前，国内百度、阿里、商汤、华为等积极加入，行业呈现“百模大战”的竞争格局，能否形成飞轮是最后能否胜出的关键。

海外大模型：微软&OpenAI、谷歌、英伟达、Meta、AWS

微软&OpenAI 领先，谷歌追赶，Meta 防御性开源，英伟达转型算力云服务

微软和 OpenAI 是目前大模型技术水平、产品化落地最为前沿的领军者，其对颠覆式创新的持续投入是当前领先的深层原因。谷歌技术储备丰厚，自有业务生态广阔并且是 AI 落地的潜在场景，但管理上未形成合力，目前正在产品化、生态化加速追赶。英伟达是 AI 芯片领军者，CUDA 框架构筑了其它芯片公司难以逾越的护城河，目前正在从硬件向算力云服务、MaaS 等商业模式转型。Meta 在产品化上进展缓慢，选择模型开源的防御性策略，以应对 OpenAI、谷歌等竞争对手的强势闭源模型。AWS 作为领先的云服务厂商，超算技术布局领先，但是在 AI 大模型竞争上的应对稍显迟缓。

图表12：海外科技巨头 AI 布局



资料来源：公司官网，华泰研究

OpenAI：全球领先的 AI 初创企业

发展历程：从非营利开端到向营利性全面转型

OpenAI 是美国一家人工智能研究实验室，由非营利组织 OpenAI 和其营利组织子公司 OpenAI LP 所组成，公司致力于构建安全的通用人工智能（AGI）以造福人类。

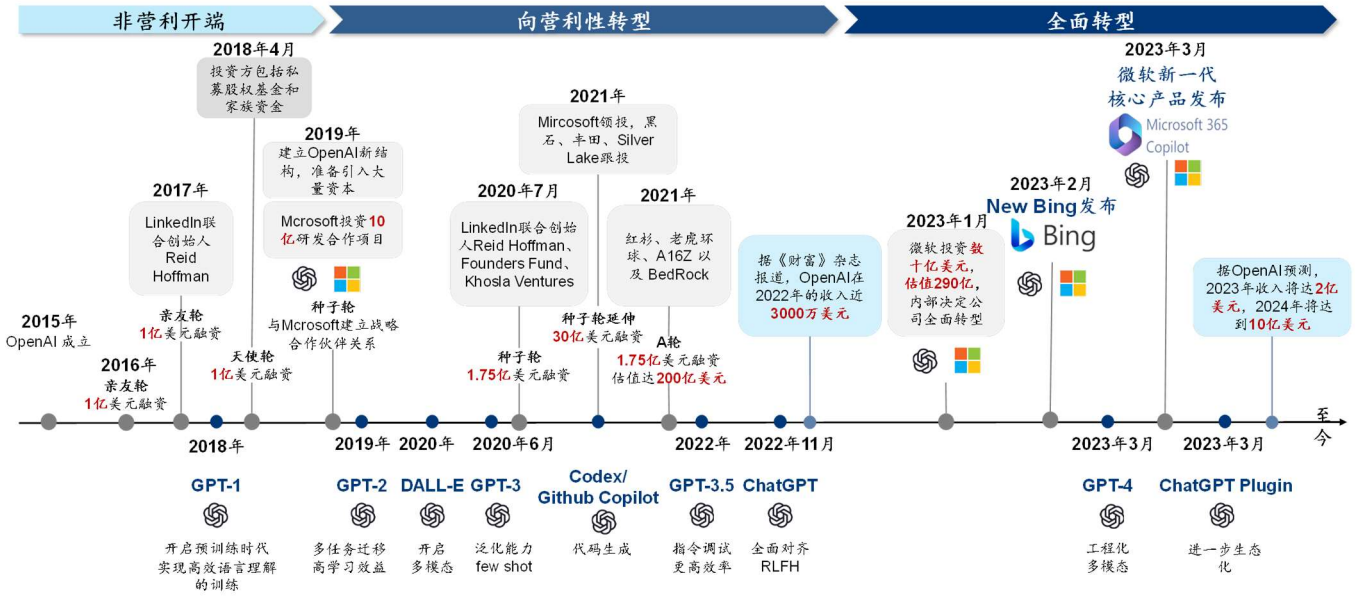
非营利开端：硅谷领军人物云集，创建非营利组织以促进 AI 发展。非营利性的 AI 项目 OpenAI 于 2015 年宣布正式启动，由许多硅谷领军人物共同创建，例如硅谷创业孵化器 Y Combinator CEO 的 Sam Altman、Google Brain 的 Ilya Sutskever、时任互联网支付处理平台 Stripe CTO 的 Greg Brockman 以及特斯拉的 CEO Elon Musk 等，许多创始人都曾是被誉为“深度学习教父”Geoffrey Hinton 教授的学生。公司把生成式预训练模型（Generative Pre-trained Transformer，GPT）确定为主要研究方向，先后推出并开源预训练 NLP 模型 GPT-1，以及采用迁移学习技术、能实现多个 NLP 任务的 GPT-2。

免责
更多免费行业报告 并购家 www.ipoipo.cn

向营利性转型：与微软深度绑定，推出掀起生成式 AI 浪潮的 ChatGPT。2018 年，由于 Elon Musk 担任 CEO 的 Tesla 等公司也在开发 AI 技术，存在利益冲突和人才争夺等矛盾，Elon Musk 辞去 OpenAI 董事会席位。为支撑大模型训练的高算力和资金需求，同年，有限营利公司 OpenAI LP 成立。2019 年，OpenAI LP 接受微软 10 亿美元投资，与其达成独家合作伙伴关系。在强大算力和充足资金的助力下，OpenAI 沿着 GPT 路线持续发力，2020 年推出拥有小样本泛化能力的 GPT-3，2022 年推出加入指示学习（Instruction Learning）和人类反馈的强化学习（Reinforcement Learning from Human Feedback，RLHF）的 InstructGPT，并于 2022 年发布产品化的 ChatGPT，掀起了一股席卷全球的生成式 AI 浪潮。

全面转型：加快商业化步伐，构建生态圈。2023 年 1 月，微软宣布与 OpenAI 的长期合作伙伴关系进入第三阶段，将继续向 OpenAI 投资数十亿美元，并加速产品与技术的整合。2023 年 3 月，OpenAI 发布工程化的多模态 GPT-4，并与各个领域的软件开展合作；同月，OpenAI 发布 ChatGPT Plugins（ChatGPT 插件集），将 GPT 大模型能力与面向用户的第三方应用程序互联，应用空间想象力广阔。

图表13： OpenAI 的发展历程



资料来源：腾讯科技公众号，OpenAI 官网，华泰研究

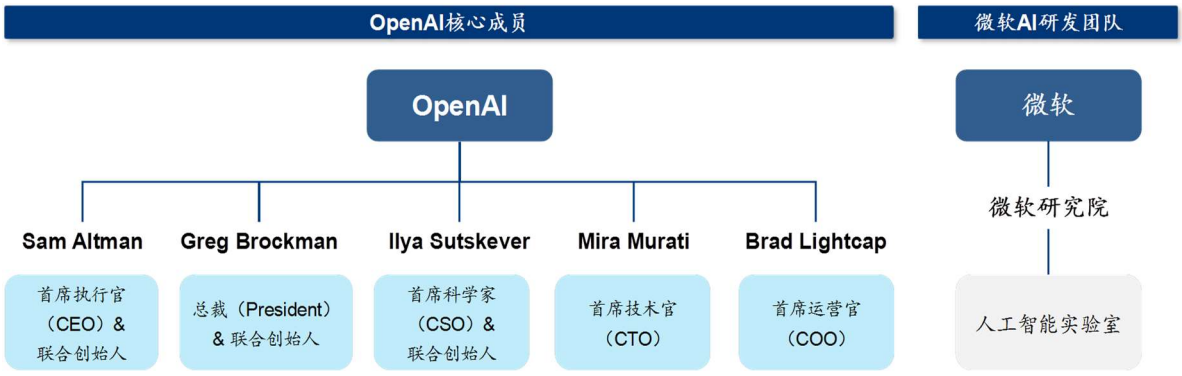
团队：年轻、背景豪华且高度聚焦技术

OpenAI 有着一支高人才密度、高效率的“特种兵”创始人团队。首席执行官兼联合创始人 Sam Altman，曾任硅谷创业孵化器 Y Combinator 的 CEO，于 2015 年带领创建了 OpenAI，致力于构建安全且人类级别 AI。总裁兼联合创始人 Greg Brockman，曾任互联网支付处理平台 Stripe 的 CTO，在 OpenAI 曾用九个月时间从“零”学起成为一名 1 顶 10 的 AI 工程师，是技术产品化的第一推手。首席科学家兼联合创始人 Ilya Sutskever，曾就职于 Google Brain，参与了深度学习框架 TensorFlow、生成式对抗网络（GAN）和序列到序列（Seq2Seq）模型的研究，在 OpenAI 领导了 GPT-1/2/3、DALLE 以及 ChatGPT 等模型的研发。

微软的 AI 研究实验室（Microsoft Research AI）是推进 AI 研究的主力团队。实验室于 2017 年成立，隶属于微软研究院，凝聚了来自感知、学习、推理和自然语言处理等多个 AI 研究子领域的顶尖科学家，旨在将不同学科结合起来以推进 AGI 技术研发。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表14：OpenAI 和微软的核心成员或 AI 团队介绍



资料来源：MicroSoft Research 官网，华泰研究

AMiner 与智谱研究发布的报告显示，团队规模虽不足百人（共 87 人），但有着以下显著特征：1）平均年龄 32 岁，“90 后”是主力军；2）学术能力和业界经验均突出，其中有 10 人来自谷歌；3）技术人员近九成，高度聚焦技术研发；4）LLM 领域的经验丰富，1/4 团队成员曾参与 Codex 项目。

算力：强大的算力支撑 GPT 释放大模型潜能

2020 年，微软为 OpenAI 专属打造拥有超过 28.5 万张 CPU、1 万张 GPU 的 AI 超级计算机，通过 Azure 云平台给予其算力支持，并保证该部分算力不会被 Azure 其他服务占用，为 OpenAI 的大模型训练配置强大且灵活的计算资源和基础设施。此外，芯片龙头企业 NVIDIA 也与 OpenAI 保持紧密合作，供应最先进的 GPU 以支持 GPT 系列模型的加速计算，例如 2021 年 OpenAI 将 NVIDIA 全球首款 DGX AI 超级计算机用以训练 GPT-3.5。

微软正在自研 AI 芯片，首要目标便是为 OpenAI 提供算力。2023 年 4 月，据 The Information 报道，微软正在秘密研发自己的 AI 芯片，代号雅典娜（Athena）。雅典娜芯片由台积电代工，采用 5nm 先进制程，首个目标便是为 OpenAI 提供算力引擎，以替代昂贵的英伟达 A100/H100，节省成本。微软从 2019 年就开始研发这款芯片，目前已在测试阶段。

模型：坚持 GPT 技术路径，持续探索生成式 AI 潜力

OpenAI 从 18 年起一直沿着 GPT 路线持续发力，通过 GPT1/2/3/3.5/4 等模型不断地进行技术探索。在模型架构上，GPT 是侧重生成的单向模型。模型的基础架构是 Transformer 的解码器，解码器的第一个自注意力层加入了掩蔽机制，使得句子的未来信息被隐藏，由于其只能通过学习当前和历史的文本信息，来对下一个字进行预测，因此属于单向的生成式模型。生成式模型相比理解式模型通常具有更高的灵活性和更强的泛化能力，在应用时更具通用性。GPT 系列模型结构秉承了不断堆叠 Transformer 的思想，通过不断提升训练语料的规模、网络的参数数量来完成 GPT 系列的迭代更新。

图表15：GPT 系列的迭代更新

	Transformer层数	训练语料量	参数量	模型性能
GPT-1	12层	5GB	1.17亿	生成的文本存在不流畅和不合理
GPT-2	48层	40GB	15亿	可以生成更长、更自然、更连贯的文本；具备对未知词的生成能力
GPT-3	96层	1350GB	1750亿	可以在没有接受任何指导的情况下完成一些任务；支持多语种生成和对话生成等其他任务

资料来源：斯坦福 Ecosystem Graphs，陈巍谈芯公众号，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

InstructGPT 极大地提升了通用 AI 系统与人类意图的对齐能力。InstructGPT 基于 GPT-3 的架构，引入了指示学习（Instruction Learning）来使得 AI 系统更遵循人类指令，减少产生有害或错误内容的可能性；同时，引入人类反馈的强化学习（Reinforcement Learning with Human Feedback，RLHF）训练机制，通过强化学习范式让生成式 AI 产出的内容更符合人类意图。

多模态能力和预测扩展工具让 GPT-4 与万物互联。2023 年 3 月，OpenAI 发布 GPT-4，模型支持文本和图像输入的多模态能力，应用空间想象力大，有望重塑从浏览器到文档智能等的软件交互。同时，OpenAI 推出具备预测扩展性的深度学习堆栈，用极小算力成本就能可靠预测 GPT-4 在下游垂直领域应用的性能，让小成本广泛试用成为可能。在选择合适的大模型后，模型微调让通用基础模型在细分领域的表现进一步提升。

应用：以产品为导向形成数据飞轮，逐步构建生态圈

OpenAI 始终以产品为导向，积极推出产品以迅速获取用户，从而得到用户反馈数据来训练出更好的模型。2020 年，OpenAI 在发布 GPT-3 后快速开放商用 API 服务，让下游客户通过 API 调用模型能力。在 ChatGPT 的训练过程中，则融入用户数据反馈：1）采用 GPT-3 API 获取的真实用户请求微调基础模型；2）引入 RLHF 训练机制，采用人工标注的数据，通过人类的反馈进行强化学习，针对性地进行模型优化。这使得 ChatGPT 在多个自然语言任务如文书写作、代码生成与修改、多轮对话等中展现出远超 GPT-3 的优秀能力，并有效减少有害和不真实输出。

技术与产品的整合进一步推动数据飞轮的运转。当前，微软将把 OpenAI 的技术与旗下各应用生态的产品整合，涵盖企业级云计算、办公、底层代码生成和娱乐交互等各个应用层面。

图表16：微软产品与 OpenAI 模型的整合

资料来源：微软科技公众号，大数据应用公众号，机器之心公众号，华泰研究

谷歌：LLM 领域的奠基者

发展历程：从 AI 技术研发的领军者到产业化进程的推动者

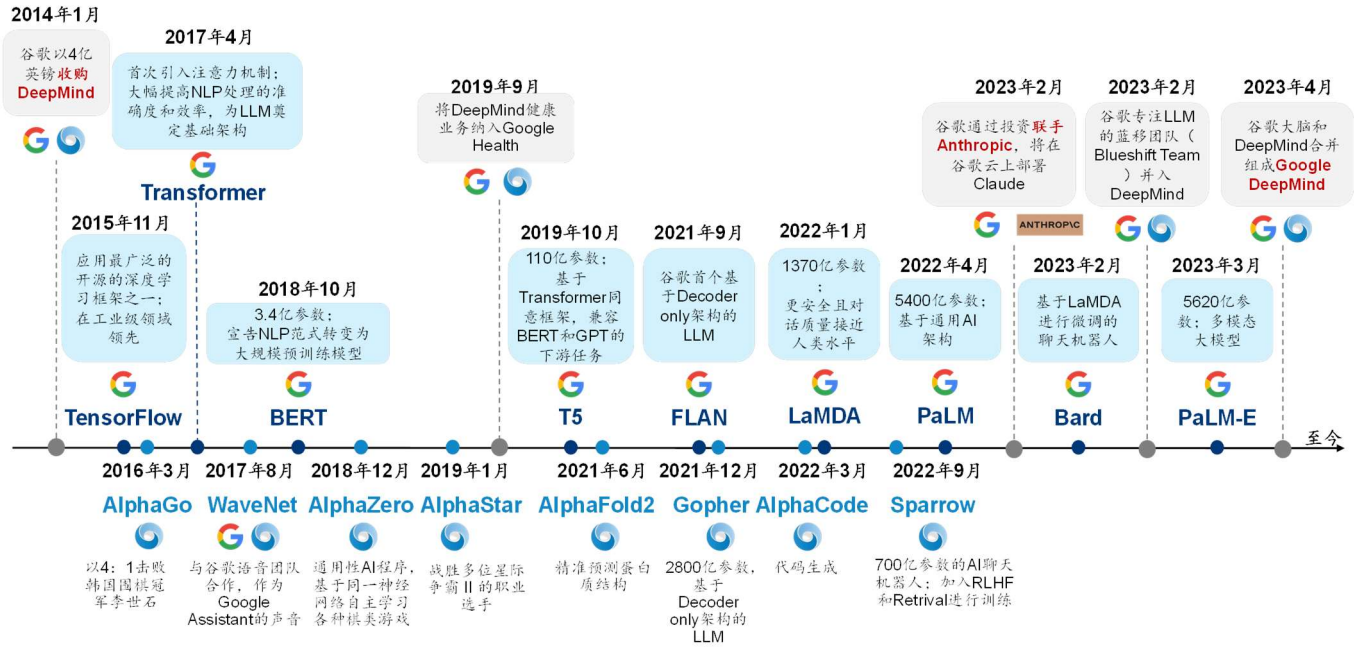
谷歌联手 AI 初创企业 DeepMind，率先展开 AI 领域的技术布局。2014 年，谷歌收购 AI 初创公司 DeepMind。DeepMind 持续探索模拟、生物科研、游戏操作、对话聊天等领域的高性能算法，代表产品包括打败世界围棋冠军的 AlphaGo、精准预测蛋白质结构的 AlphaFold、战胜电竞职业玩家的 AlphaStar、智能聊天机器人 Sparrow 等。

在 LLM 领域，谷歌是新技术研发的领军者和奠基者。2015 年，谷歌开源了深度学习框架 TensorFlow，其具备灵活、可扩展且易于部署等优势，成为应用最广泛的深度学习框架之一，尤其在工业应用处于领先地位。2017 年，谷歌发布 NLP 领域的里程碑——Transformer，它首次引入注意力机制，使得模型能更好地处理长文本和复杂的语言结构，大幅提高 NLP 任务的准确度和效率，为后来所有大语言模型奠定基础架构。2018 年，谷歌发布 3.4 亿参数的大规模预训练模型 BERT，在理解复杂的语言结构和语义中表现优秀，在多项 NLP 任务上的效果都远远超过了传统的浅层架构，宣告 NLP 范式转变为大规模预训练模型的方法。

免责
更多免费行业报告 并购家 www.ipoipo.cn

谷歌加快生成式 AI 产业布局。OpenAI 于 2022 年底推出的 ChatGPT 掀起了席卷全球的生成式 AI 浪潮，使得 GPT 能力在各个行业和领域不断拓展，并对谷歌 AI 领先地位构成挑战。作为应对，谷歌于 2023 年 2 月宣布通过投资联手 AI 初创公司 Anthropic，将在谷歌云上部署 ChatGPT 的有力竞品 Claude，并于同月推出自研的基于 1270 亿参数大模型 LaMDA 的聊天机器人 Bard。

图表17：谷歌 AI 领域的发展历程

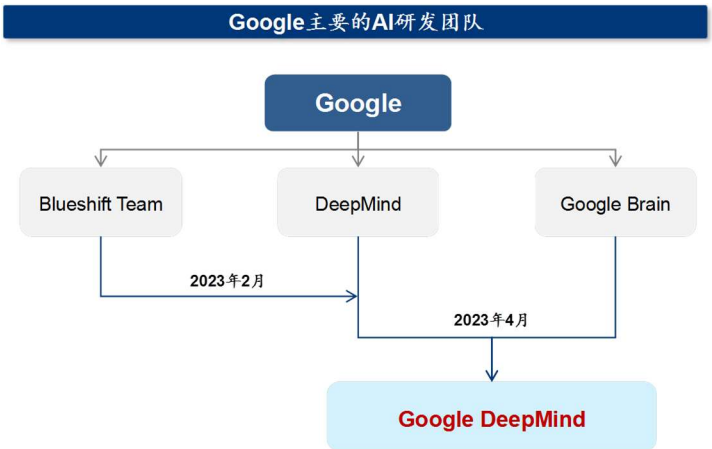


资料来源：DeepMind 官网，Google 官网，斯坦福 Ecosystem Graphs，OpenBMB 网站，华泰研究

团队：Google DeepMind 汇集谷歌 AI 领域人才

经过两轮人才整合，当前谷歌主要 AI 研发团队合并为 Google DeepMind。2023 年 2 月，谷歌旗下专注 LLM 领域的“蓝移团队”（Blueshift Team）宣布加入 DeepMind，以共同提升 LLM 技术的能力。4 月，谷歌宣布正式将谷歌“大脑”和 DeepMind 两大团队合并，组成“Google DeepMind”部门。Google DeepMind 部门首席执行官将由 DeepMind 联合创始人兼首席执行官 Demis Hassabis 担任；原谷歌大脑团队负责人 Jeff Dean 转任谷歌首席科学家，领导谷歌 AI 相关的最关键和战略性的技术项目，其中首要项目就是系列多模态 AI 模型。AI 领域的世界级人才与计算能力、基础设施和资源的汇集，将显著加快谷歌在 AI 方面探索的进展。

图表18：Google 主要的 AI 研发团队



资料来源：Google 官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

算力：TPU 系列芯片和新一代超级计算机支撑 AI 大模型训练和创新

TPU（Tensor Processing Unit）是谷歌专门为高效计算和加速神经网络训练和推理过程而设计的专用芯片（ASIC）。2016 年，谷歌发布第一代 TPU，成为 AlphaGo 背后的算力。与 GPU 相比，TPU 采用低精度计算，在几乎不影响深度学习处理效果的前提下，大幅降低了功耗、加快运算速度。最新一代的 TPU v4 发布于 2021 年，在相似规模的系统训练中，TPU v4 比 A100 计算速度快 1.15-1.67 倍、功耗低 1.33-1.93 倍。TPU v4 卓越的性能、可伸缩性和灵活性使其成为支持 LaMDA、MUM、PaLM 等大规模语言模型的主力产品。

谷歌推出新一代超级计算机 A3 Virtual Machines，为大型语言模型创新提供动力。谷歌于 2023 年 5 月的 I/O 大会发布 A3 超级计算机，其采用第 4 代英特尔至强可扩展处理器、2TB DDR5-4800 内存以及 8 张英伟达 H100。此外，A3 是首个使用谷歌定制设计的 200 Gbps IPU 的 GPU 实例，GPU 间的数据传输绕过 CPU 主机，通过与其他虚拟机网络和数据流量分离的接口传输，网络带宽相比 A2 增加 10 倍，具有低延迟和高带宽稳定性。A3 超级计算机专门设计用于训练和服务对计算资源要求极高的 AI 大模型，这些模型将推动当今生成式 AI 和大型语言模型的创新进展。

框架：领先的自研深度学习平台 TensorFlow

2015 年，谷歌开源了深度学习框架 TensorFlow，其是当前 AI 领域的主流开发工具之一。TensorFlow 采用计算图来表示神经网络计算过程和数据流动，从而实现高效的并行数据处理和运算，尤其在大规模分布式训练方面具有优势。TensorFlow 还提供了广泛的工具和 API，可用于各种类型的深度学习应用，例如计算机视觉、自然语言处理和语音识别等领域。此外，TensorFlow 具有很好的可移植性，能够在多种硬件设备和操作系统上运行。目前，TensorFlow 生态系统完备，为研发人员提供了从模型构建到部署和运营管理的全套工具链，是深度学习领域应用最广泛的框架之一，特别在工业级领域应用处于领先地位。

模型：各技术路线的模型储备丰富，多模态大模型实现人机交互领域突破

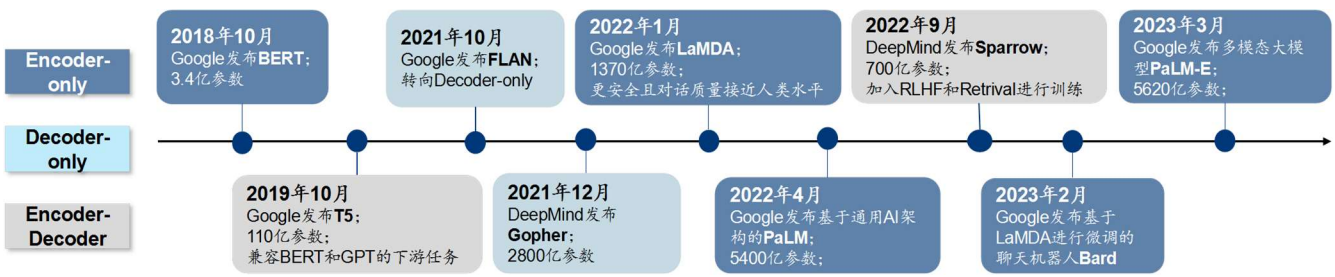
谷歌在 AI 各技术路线上积累了丰富的模型储备。从基础架构的角度，其技术路线主要可以分为三个方向：

1）基于 Transformer 纯编码器（Encoder-only）：与单向生成的 GPT 模型不同，纯编码器模型是侧重理解的双向模型，通过并行计算和多头注意力层来学习单词之间的相关性，实现上下文理解。主要模型包括语言模型 BERT、LaMDA、PaLM、基于 LaMDA 的对话模型 Bard 以及多模态视觉语言模型 PaLM-E。

2）基于 Transformer 纯解码器（Decoder-only）：纯解码器是单向生成模型，主要用于句子生成。主要模型包括谷歌的语言模型 FLAN、DeepMind 的语言模型 Gopher。

3）基于 Transformer 统一架构（Encoder-Decoder）：统一模型框架能兼容以上两类模型的理解和生成的下游任务。主要模型包括语言模型 T5、对话模型 Sparrow。

图表19：谷歌在 AI 各技术路线发布的主要模型



资料来源：Google 官网，斯坦福 Ecosystem Graphs，华泰研究

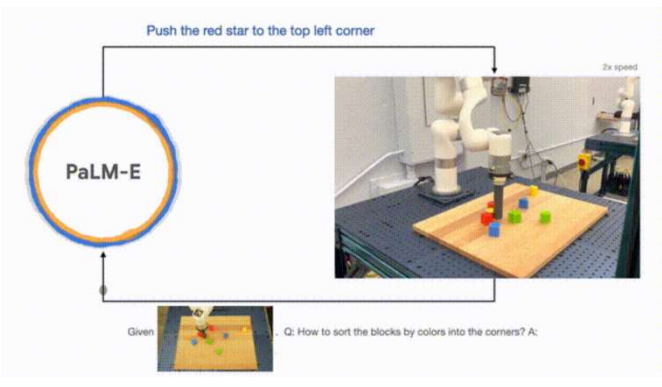
免责
更多免费行业报告 并购家 www.ipoipo.cn

多模态视觉语言大模型 PaLM-E 实现人机交互领域的重大飞跃。2023 年 3 月，谷歌发布 5620 亿参数的多模态视觉语言模型 PaLM-E，是 PaLM-540B 语言模型与 ViT-22B 视觉模型的结合。与传统的语言模型相比，PaLM-E 具有更高的效率和灵活性，可以根据输入的复杂度和任务的需求动态地激活不同的模块。同时，PaLM-E 可以通过文本、图像和视频等多种类型的输入来理解和控制机器人等实体化系统，从而实现具象化的能力。PaLM-E 的发布代表着人机交互领域的一次重大飞跃，在工业、商业、办公等领域的机器人应用上有着较大的想象力。

图表20：PaLM-E 机器人执行“从抽屉里拿出薯片”的多步骤指令

资料来源：GitHub PaLM-E，华泰研究

图表21：PaLM-E 机器人规划“将积木按颜色分类”的长周期任务



资料来源：GitHub PaLM-E，华泰研究

最新的大语言模型 PaLM 2 在多语言能力、代码能力和逻辑能力上得到显著提升。谷歌于 2023 年 5 月的 I/O 大会发布 PaLM 2，模型基于 2022 年 10 月提出的 AI 新架构 Pathways，并基于 TPU v4 和 JAX 框架训练。在语言能力上，PaLM 2 的训练数据超过 100 种语言，能理解、翻译和生成更准确和多样化的文本；同时，PaLM 2 在 20 种编程语言上进行训练以提升代码能力。在逻辑能力上，PaLM 2 在推理、数学任务测试中显著优于 PaLM，并与 GPT-4 相当。此外，谷歌在 PaLM 2 的基础上还训练了用于编程和调试的专业模型 Codey，安全知识模型 Sec-PaLM 以及医学知识模型 Med-PaLM 2。

模型压缩技术将加速 AI 技术与智能终端的融合。PaLM2 主要包含 4 种规格（Gecko、Otter、Bison 和 Unicorn），其中体积最小的 Gecko 可以在手机上运行，每秒可处理 20 个 Token，大约相当于 16 或 17 个单词。模型压缩技术或将成为实现 AI 大模型在边/端部署的核心技术，加速 AI 技术与智能终端融合。

应用：积极布局生成式 AI 应用以迎接挑战

谷歌在过去更注重发表论文，未能及时将成果产品化。谷歌在各种路线上发布了很多模型，论文成果丰富，但均未能走出象牙塔落地；同时，旗下 DeepMind 研发的基础语言模型 Gopher、Chinchilla 和对话模型 Sparrow 也尚未产品化。

积极布局生成式 AI 领域应用以迎接挑战。谷歌于 2023 年 2 月发布基于 1370 亿参数大模型 LaMDA 的 AI 聊天机器人 Bard。谷歌在 4 月发布了 Bard 的更新版本，升级了数学与逻辑能力。2023 年 5 月 I/O 大会，谷歌展示了 Bard 支持的文字、代码、图像生成与理解能力，并重点强调了 Bard 与谷歌应用及外部其他应用的协同能力。此外，谷歌通过投资联手 AI 初创公司 Anthropic，押注 ChatGPT 的有力竞品 Claude。

面向开发者，谷歌开放 PaLM 的 API，并发布工具 MakerSuite，帮助开发者们快速构建生成式 AI 应用。PaLM API 将为开发者提供面向对轮交互而优化的模型，如内容生成与对话；也能为开发者提供摘要、分类等多种任务的通用模型。工具 MakerSuite 则能帮助开发者快速、轻松地构建和调整自定义模型，在浏览器中就能快速测试和迭代调整后的模型。

免责
更多免费行业报告 并购家 www.ipoipo.cn

PaLM 2 赋能超 25 种功能和产品的升级，包括办公套件、聊天机器人以及企业级 AI 平台。在 PaLM 2 的加持下，Workspace 推出的 AIGC 工具包 Duet AI，其中包含 Google Docs 和 Gmail 中的智能写作工具、Google Slides 中的文本生成图像、演讲备注以及 Google Meet 的会议摘要生成等功能。升级版 Bard 不仅将支持图文多模态，还将接入网络以实时网页搜索；同时，大量 Bard+tools 的组合将推出，Bard 有望深度融合到 Google 所提供的原有工具中。ToB 应用方面，Vertex AI 集合了谷歌研发的基础模型，用户可按需调用模型并对模型的参数权重进行微调；此外，Vertex AI 还能够根据内部数据建立企业版本的搜索引擎。

英伟达：AI 时代的芯片领路者

发展历程：英伟达的二十年辉煌史

英伟达成立于 1993 年，是全球最大的 GPU 供应商，也是 AI 时代的芯片领路者。上市初期公司主要关注 PC 图形业务，与 PC 出货量具备高度相关性，主要以 OEM 的销售模式。后续随着智能手机、平板电脑等消费电子的应用出现，公司敏锐捕捉到终端需求的变化，将业务重心向高端游戏显卡市场过渡。按照过往经验，英伟达基本每 1-2 年更新一次游戏 GPU 架构，凭借强大的性能和生态优势迅速提升在全球独立 GPU 市场的市占率，近几年均保持在 70%以上的份额。进入 2018 年数字货币价格大幅下跌，数据中心开始承接公司新增长点， 2020 年该业务营收增速超过 120%。此外，英伟达持续以 CUDA 为核心构建护城河，结合 CPU+GPU+DPU 三大硬件，形成统一的生态闭环，且在各细分领域形成全套解决方案。

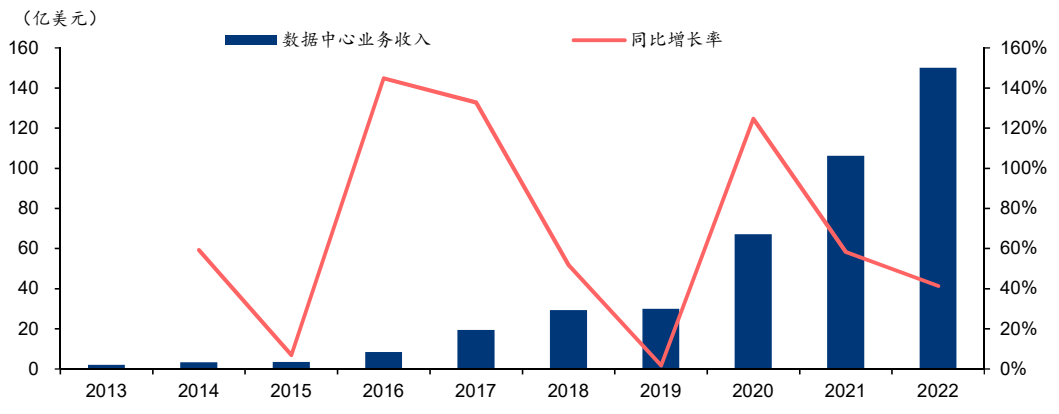
第一阶段（1992-2015）：从图形芯片到游戏显卡，确定主流赛道形成稳定现金流

持续推动业务升级，战略核心转移至高端游戏卡领域。在 GPU 推出初期，以 OEM 形式销售显卡是公司重要的收入来源之一，市场终端应用以 PC 为主，笔记本及平板的普及度尚低，英伟达的产品与 PC 出货存在较高程度的绑定。随着半导体行业发展，终端应用趋于多元化，PC 市场热度从 2012 年起开始出现明显下降。此外，随着集成显卡性能的提升，独立桌面显卡出货情况也开始出现衰退。在此市场环境下，英伟达将战略核心转移至高端游戏卡领域。在 2010-2015 年在全球 PC 市场逐步倒退的情况下，英伟达游戏显卡出货量 5 年实现 9%的年复合增长，销售均价/收入分别呈现 11%/21%的年复合增长。2019 年公司游戏业务已经为其贡献了过半营收，而 OEM/IP 业务收入占比下降至仅 6%。

第二阶段（2016-2021）：构建通用计算生态，数据中心业务实现爆发式增长

推出革命性架构 CUDA，进入发展高速期。2006 年英伟达研发出了能够让 GPU 计算变得通用化的 CUDA（Compute Unified Device Architecture）技术，运用 GPU 并行处理能力来应对复杂的计算，次年公司就在图形市场取得飞跃式增长，同时 Tesla GPU 的推出让英伟达成功进入通用计算时代。经过二十多年的发展，公司已形成包括：面向 PC 和笔记本的 GeForce 系列，面向工作站的 Quadro 系列，面向超大型计算的 Tesla 系列和面对移动处理器的 Tegra 系列。

图表22：Nvidia 数据中心业务收入及同比增长率



资料来源：NVIDIA 官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

第三阶段（2022-）：AI 加持下，游戏、数据中心、自动驾驶三驾马车助力公司迈向千亿美金市场

英伟达经过前期的路径选择已经逐步清晰了三条主要赛道，同时在人工智能飞速迭代的情况下，从元宇宙平台搭建到协助内容创作到 AI 时代的“台积电”定位，精准把握每次技术变革：

1）游戏方面：英伟达背靠强大的 GPU 实力不断在游戏行业进行创新，推出 DLSS、光线追踪等技术提高游戏体验，并进一步推动云端游戏业务发展。23 年由于 PC 端出货减少短期内影响了游戏业务增长，但据彭博一致预期，未来 5 年游戏业务 CAGR 仍有望达到 10%。

2）数据中心方面：除了训练/推理芯片性能更新，英伟达更侧重于 AI 超级计算机的打造和对整个服务器产业链的赋能。其在 GTC2023 推出的 NVIDIA DGX 超级计算机，将成为大语言模型实现突破背后的引擎。

3）自动驾驶方面，英伟达在硬件上推出了自动驾驶汽车的平台 DRIVE Hyperion、算力达每秒 254 TOPS 的系统级芯片 DRIVE Orin SoC、新一代集中式车载计算平台 DRIVE Thor。22 年秋季 GTC 大会公布的中央计算引擎 Drive Thor 芯片由 Grace CPU、Hopper GPU 和下一代 GPU 组成,可提供 2000TOPS 算力，预计将在 25 年搭载于极氪。软件方面则推出了操作系统 DRIVE OS、AI 辅助驾驶平台 DRIVE Chauffeur、可实现对话式 AI 等智能服务的 DRIVE Concierge 等。过去 5 年汽车业务 CAGR 为 10%，仅占 FY23 收入的 3%，但未来会随着产品逐渐落地而稳步提升。黄仁勋在 22 年 9 月的 Computex 大会上预计未来 6 年，将能够创造 110 亿美元营收价值。

团队：灵魂人物引导产业变革，数次收购完善业务版图

灵魂人物带领数次决策，英伟达引导产业变革。英伟达由黄仁勋与 Chris Malachowsky、Curtis Priem 于 1993 年 4 月共同创立，创业初期仅是数十家 GPU 制造商中的一员。但凭借黄仁勋非凡的见识和魄力，从 1999 年大规模裁员后发布世界首款 GPU 拿下大客户微软，到 2006 年推出 CUDA 技术，到 2012 年宣布计算卡与游戏卡分离成为与 AMD 竞争转折点，英伟达引领了显卡行业的每一次重大变革，成为占据独立显卡 80%以上份额的巨头。

英伟达不断通过收购完善业务版图。英伟达早期的竞争对手 3dfx 在 2000 年被英伟达收购，3dfx 是 20 世纪 90 年代显卡芯片的市场领导者，主营业务包括显卡以及 3D 芯片，在 2000 年时因为在一场与 NVIDIA 的官司中败诉而被 NVIDIA 仅仅以 7000 万美元和 100 万美元的股票收购，并在 2002 年年底破产。2019 年，英伟达宣布成功收购以色列芯片制造厂商 Mellanox，进一步加强对数据中心的布局，加速 HPC 和 AI 芯片的发展。另一具有代表性的案例是最后由于监管问题以失败告终的 2020 英伟达收购 ARM 事件，如若达成或将打通 ARM CPU 的庞大生态系统，由此也可以看出英伟达利用收购不断扩大商业版图，营造完整生态的动机。

图表23：NVIDIA 收购公司一览

收购公司	年份	被收购公司简介
3dfx	2000	显卡技术先驱
Media Q	2003	无线领域图形和多媒体技术的领导者
Uli Electronics	2005	核心逻辑技术开发商
Hybrid Graphics	2006	手持设备嵌入式 2D 和 3D 图形软件开发商
Portal Player	2007	个人媒体播放器半导体、固件及软件供应商
Mental Images	2008	视觉渲染软件公司
AGEIA	2008	游戏物理技术开发商
Portland Group	2013	加速计算开发者工具提供商
Mellanox	2020	高性能互联技术领域龙头
Excelero	2022	存储方案提供商

资料来源：英伟达官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

商业模式：深度绑定台积电走 OEM 模式，轻装上阵重视研发投入

深度绑定台积电，外包制造环节轻资产专注研发。1998 年，公司与台积电正式建立策略联盟伙伴关系。从一开始，黄仁勋就致力于使英伟达成为一家无晶圆厂的芯片公司，通过将制造芯片费用外包以降低资本支出。OEM 模式虽然令投资者担忧生产外包的风险，但另一面英伟达将腾出的资金用于研发，逐渐建立起了高技术壁垒。自上市以来，英伟达研发投入绝对值基本呈现持续增长，单季度研发费用率平均值超 20%，2022 年公司研发投入达 73 亿美元。公司研发投入基本上全部用于 GPU 产品，长期研发投入与积累也使公司产品技术壁垒及竞争力得以充分提升。

拓展云服务商业模式，将与企业、云厂商实现互惠共赢。在硬件制造的基础上，英伟达近期还推出了 NVIDIA AI Foundations，定位为“超算云服务+模型代工厂”，用于企业构建自定义语言模型和生成式 AI。NVIDIA AI Foundations 包括语言、视觉和生物学模型构建服务，分别名为 Nemo、Picasso 和 BioNemo，使用企业专有数据进行训练，用于处理特定领域的任务。**我们认为超算云服务+模型代工厂的商业模式是英伟达作为芯片送水人在大模型时代的商业模式的自然延伸。**

图表24： NVIDIA AI Foundations 服务与基础模型

资料来源：2023 年 GTC 大会，华泰研究

图表25： NVIDIA Picasso 输入文本，输出图片、视频、3D 模型

资料来源：2023 年 GTC 大会，华泰研究

硬件迭代：训练/推理芯片性能参数持续领先对手

芯片性能优势优于竞争对手，在推理芯片方面将继续扩大优势。英伟达通常 1-2 年左右更新一次架构，2015 年以来经历了 Pascal、Volta、Turing、Ampere 和 Hopper 时期，其制程也从 16nm 快速推进到 7nm 和 4nm，预计下一代 Blackwell 架构 GPU 将延续与台积电合作选用 3nm 工艺。为了加速在混合精度计算、光线追踪领域的发展，英伟达在 Volta 架构引入 Tensor Core 来加速 AI 处理，而在 Turing 架构上加入 RT Core 来实现光线追踪的混合渲染。从性能对比来看，英伟达在 22 年初推出的 H100 在 AI 性能方面速度比 A100 提升 9 倍，在大型语言模型上推理速度高 30 倍，在部分性能上优于 21 年 12 月 AMD 推出的高端 GPU MI250，在训练芯片具备较强的话语权。

而在推理方面，尽管算力要求比训练端更低，市场需求更大也更为细分，英伟达面对的竞争更激烈，但公司的推理侧解决方案更具备通用性、低延时、低功耗。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表26：英伟达与 AMD 产品对比

品牌产品	AMD	AMD	NVIDIA	NVIDIA	NVIDIA	NVIDIA	NVIDIA
产品	INSTINCT MI100	INSTINCT MI250	P100	V100 SXM2	T4	A100 80GB PCIe	H100 PCIe
发布日期	2020	2021	2016	2017	2018	2020	2022
工艺	7nm	6nm	16nm	12nm	12nm	7nm	4nm
半精度(FP16)	184.6 TFLOPS	362.1 TFLOPS	21.2 TFLOPS	125 TFLOPS	65 TFLOPS	312 TFLOPS*	NA
半精度(FP16 Tensor Core)	NA	NA	不支持	不支持	不支持	不支持	1600 TFLOPS*
单精度(FP32)	23.1 TFLOPS	45.3 TFLOPS	10.6 TFLOPS	15.7 TFLOPS	NA	19.5 TFLOPS	48 TFLOPS
单精度(FP32 Tensor Float)	46.1 TFLOPS (AMD为Matix Core)	90.5 TFLOPS (AMD为Matix Core)	不支持	不支持	不支持	156 TFLOPS	800 TFLOPS
双精度(FP64)	11.5 TFLOPS	45.3 TFLOPS	5.3 TFLOPS	7.8 TFLOPS	8.1 TFLOPS	9.7 TFLOPS	24TFLOPS
双精度(FP 64 Tensor Core)	不支持	90.5 TFLOPS (AMD为Matix Core)	不支持	不支持	不支持	19.5 TFLOPS	48 TFLOPS
INT8	184.6 TOPs	362.1 TOPs	NA	NA	130 TOPs	624 TOPs*	NA
INT8 (Tensor Core)	不支持	NA	不支持	不支持	不支持	不支持	3200 TOPs
CUDA兼容	否	否	是	是	是	是	是

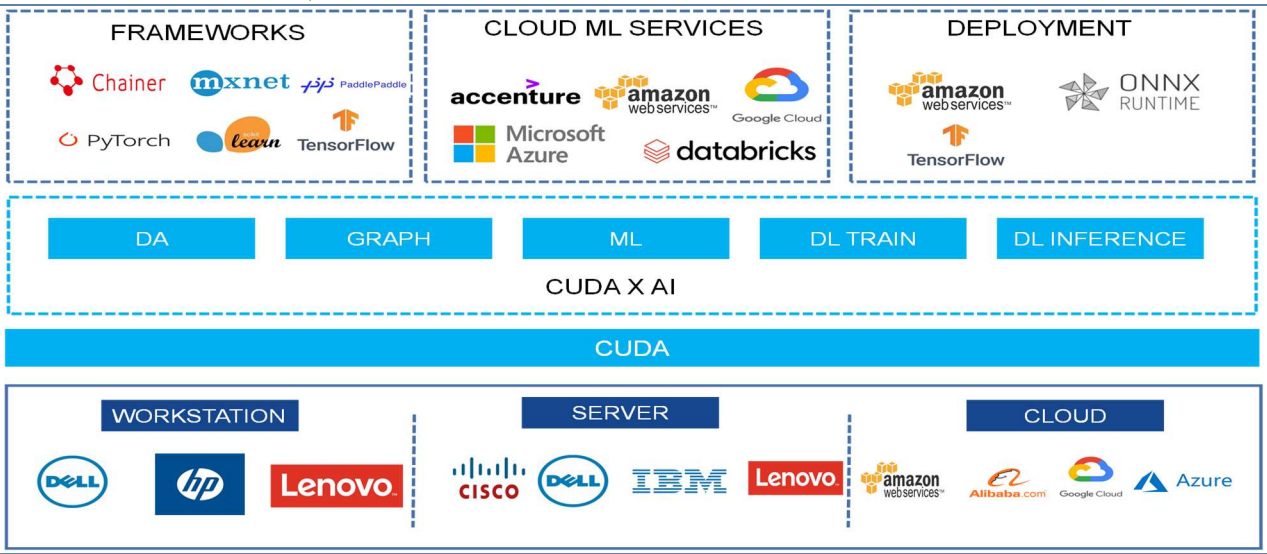
资料来源：英伟达官网，华泰研究

软件：CUDA 构建完整生态，CUDA-X AI 结构拉开差距

CUDA 是一种将 GPU 作为数据并行计算设备的软硬件体系，不需要借助图形学 API，而是采用了比较容易掌握的类 C 语言进行开发，开发人员能够利用熟悉的 C 语言比较平稳地从 CPU 过渡到 GPU 编程。与以往的 GPU 相比，支持 CUDA 的 GPU 在架构上有了显著的改进：1）采用了统一处理架构，可以更加有效地利用过去分布在顶点着色器和像素着色器的计算资源；2）引入了片内共享存储器。两项改进使得 CUDA 架构更加适用于通用计算，加上 2008 年后苹果、AMD、和 IBM 推出的 OpenCL 开源标准，GPGPU 在通用计算领域迅速发展。

英伟达在 CUDA 上的战略眼光及持续坚持，是如今形成软硬集合生态的最核心壁垒：1）易于编程与性能提升：CUDA 包含 CUDA 指令集架构（ISA）以及 GPU 内部的并行计算引擎，采用通用并行计算架构，使 GPU 能够解决复杂计算问题，相较 AMD 采取的通用 OpenCL 平台而言，并行计算能力可提高 10-200 倍不等；2）不断丰富的生态系统：英伟达通过十余年迭代扩充已形成丰富的函数库、API 指令、工具链、应用程序等，对开发者友好程度更高。在 CUDA 基础上英伟达进一步推出 CUDA-X AI 软件加速库，提供对深度学习、机器学习和高性能计算的优化功能，进一步拉开与竞争对手在 AI 时代的差距。

图表27：CUDA 及 CUDA-X AI 生态系统



资料来源：英伟达官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

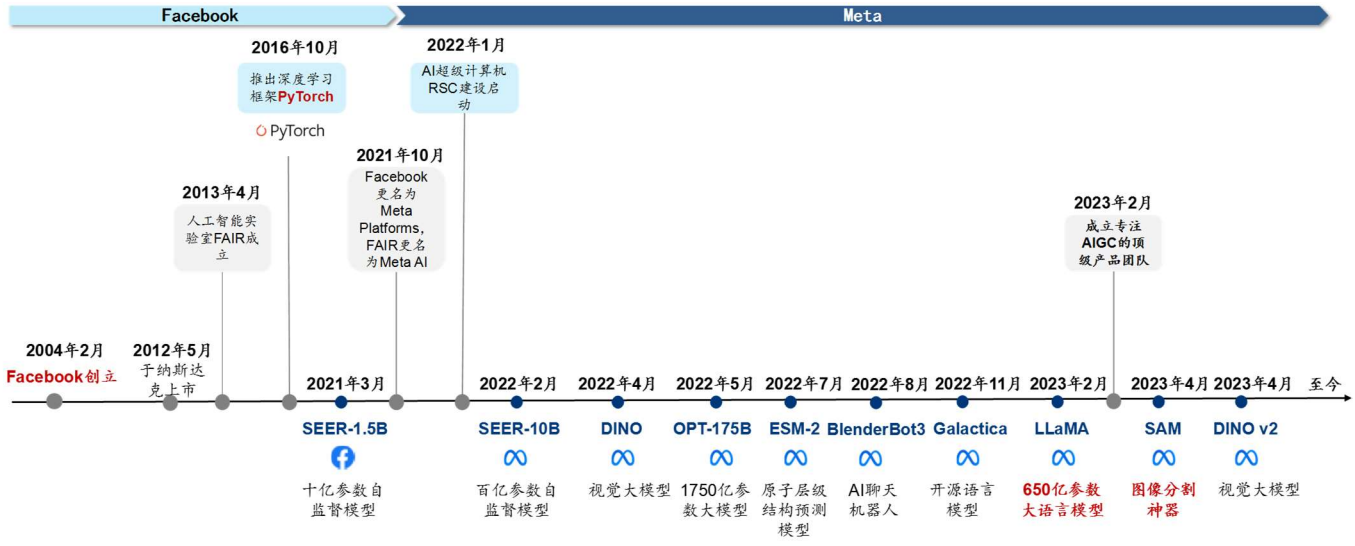
Meta：AI 和元宇宙双轮并驱

发展历程：全球最大社交媒体巨头拥抱元宇宙和 AIGC

Meta 是全球社交网络龙头企业。2004 年 Mark Zuckerberg 创立 Facebook，2021 年公司改名为 Meta Platforms，取意 Metaverse (元宇宙)，体现了公司在虚拟世界中工作和娱乐的愿景。Meta 主要经营社交网络服务、虚拟现实、元宇宙等产品，旗下拥有 Facebook、Instagram 和 WhatsApp 等社交软件。

在大模型领域，Meta 奋起直追开源多个大模型。2017 年，公司开源了深度学习框架 PyTorch，是深度学习领域最常用的框架之一。2023 年 2 月，Meta 推出针对研究社区的大型语言模型 LLaMA。4 月，Meta 发布机器视觉领域首个用于图像分割的通用大模型 SAM 和其训练数据集 Segment Anything 1-Billion，将自然语言处理领域 prompt 范式延展到 CV 领域，为 CV 领域带来革命性技术突破。5 月，Meta 发布大规模多语言语音 MMS 项目，将彻底改变语音技术，支持 1,107 种语言的语音转文字和文字转语音以及超过 4,000 种语言的语言识别。同月，Meta 发布 650 亿参数语言模型 LIMA，无需 RLHF 就能对齐，根据其论文《LIMA: Less Is More for Alignment》(Zhou et al, 2023)，在人类评估结果中甚至可与 GPT-4、Bard 和 DaVinci003 相媲美。

图表28： Meta 发展历程



资料来源：Meta 官网，华泰研究

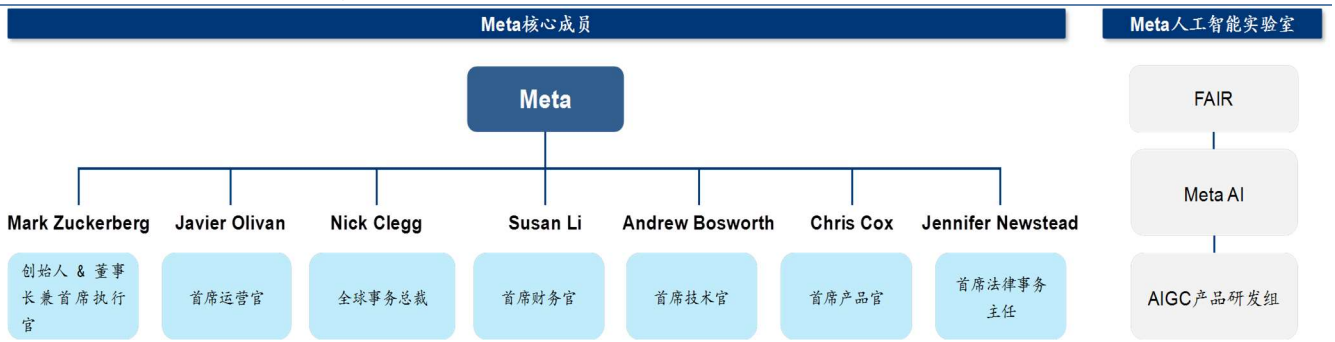
团队：AI 人才汇聚，旨在创建 AIGC 顶级产品团队

Meta 拥有一支经验丰富的核心团队。创始人、董事长兼首席执行官 Mark Zuckerberg，曾就读于哈佛大学，2004 年创立了 Facebook。首席运营官 Javier Olivan，持有斯坦福大学工商管理硕士学位和纳瓦拉大学电气和工业工程硕士学位，在加入公司之前曾担任西门子移动的产品经理。首席技术官 Andrew Bosworth，毕业于哈佛大学，曾任微软 Visio 开发人员，2017 年创建了公司的 AR/VR 组织，领导公司发展 AR、VR、AI 和消费者硬件等。

Meta AI 是 Meta 旗下的人工智能实验室。2013 年，深度学习教授和图灵奖得主 Yann LeCun 创建 FAIR，其最初目标是研究数据科学、机器学习和人工智能。2014 年，统计学先驱 Vladimir Vapnik 加入 FAIR。2018 年，IBM 大数据集团前首席技术官 Jérôme Pesenti 担任 FAIR 总裁一职。2021 年，FAIR 更名为 Meta AI。2022 年，Meta AI 在两周内预测了 6 亿个潜在蛋白质的 3D 形状。2023 年，Zuckerberg 宣布将成立专注 AIGC 的顶级产品团队，由负责 AI 和机器学习的副总裁 Ahmad Al-Dahle 领导，队内共有数十名成员，汇聚了过去分散在公司各地团队的 AI 人才。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表29：Meta 核心成员和 Meta 人工智能实验室



资料来源：Meta 官网，华泰研究

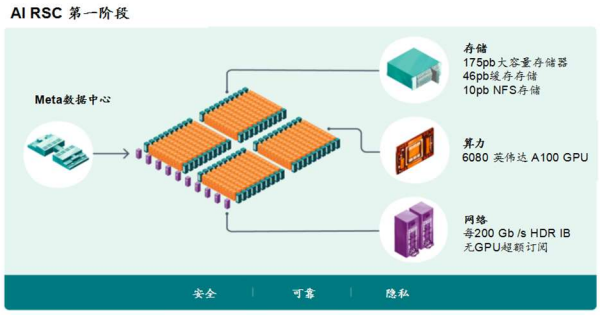
算力：全球最快 AI 超级计算机 AI RSC，推出自研 AI 芯片 MTIA

AI RSC 是“全球最快 AI 超级计算机”之一。AI RSC（AI Research SuperCluster）是 Meta 用于训练人工智能模型的 AI 超级计算机。2017 年，公司设计自研第一代算力设施，在单个集群中拥有 2.2 万个英伟达 V100 Tensor Core GPU，每天可执行 35,000 个训练任务。2020 年，公司决定加速算力增长的最佳方式是从头开始设计全新计算基础架构，以利用新的 GPU 和网络结构技术。公司希望新 AI 超算能够在 1 EB 字节大的数据集上训练具有超过一万亿个参数的模型，仅从规模上看，这相当于 36,000 年时长的高清晰度视频。2023 年 5 月，Meta 宣布已经完成了 RSC 的第二阶段扩建，使其成为世界上最快的 AI 超级计算机之一，其混合精度计算性能接近 5 exaflops。

图表30：AI RSC 比 Meta 当前基于 V100 的集群更快

资料来源：Meta 官网，华泰研究

图表31：AI RSC 第一阶段架构



资料来源：Meta 官网，华泰研究

Meta 推出自研 AI 芯片 MTIA，入场大模型军备竞赛。2020 年，Meta 设计了第一代 MTIA 芯片，采用 7 纳米工艺，可以从其内部 128 MB 内存扩展到高达 128 GB，并且在 Meta 设计的基准测试中，它处理了低等和中等复杂度的 AI 模型，效率比 GPU 高。2023 年 5 月，Meta 首次推出 AI 定制芯片——MTIA（Meta Training and Inference Accelerator），是加速 AI 训练和推理工作负载的芯片“家族”的一部分。MTIA 是一种将不同电路组合在一块板上的 ASIC 芯片，允许对其进行编程以并行执行一项或多项任务。

框架：PyTorch——深度学习领域最受欢迎的框架之一

PyTorch 是 Facebook 于 2017 年推出的一个基于 Python 的开源深度学习框架，具有灵活性、易用性和高性能，是深度学习领域最受欢迎的框架之一，在 GitHub 上获超 66,500 颗星。在 ML 领域的主要会议上提交工作成果的研究人员中，超过 80%使用了 PyTorch 框架。PyTorch 支持在 GPU 上进行高速计算，在训练时拥有更快的速度和更好的性能，从而能大规模提高研究效率。PyTorch 已成为亚马逊网络服务、微软 Azure 和 OpenAI 等众多公司和研究机构构建 AI 研究和产品的基础。2022 年，公司正式发布 PyTorch 2.0，它提供了相同的 eager mode 和用户体验，同时通过 torch.compile 增加了一个编译模式，在训练和推理过程中可以对模型进行加速，从而提供更佳的性能和对动态形状和分布式的支持。

更多免费行业报告 并购家 www.ipoipo.cn

模型：布局生成式 AI，开源语言、视觉等大模型

语言大模型#1：LLAMA 百亿参数模型在大多数基准上胜过 GPT-3。2023 年 2 月，Meta 推出了 LLaMA，目前提供 7B、13B、33B、65B 四种参数规模，都至少经过 1T token 的训练。LLaMA-13B 在多数基准测试下（如常识推理、闭卷问答、阅读理解、偏见性等）优于 GPT-3（175B），且可在单块 V100 GPU 上运行。LLaMA-65B 与目前 LLM 中最好的 Chinchilla-70B 和 PaLM-540B 旗鼓相当。LLaMA 不作商用目的，免费供给研究人员，目前 Meta 在 GitHub 上提供了精简版 LLaMA。

基于 LLaMa 衍生出多个开源模型，例如斯坦福的 Alpaca、伯克利的 Vicuna、Kaola 和 ColossalChat，以及哈工大基于中文医学知识的华驼。加州大学伯克利分校计算机科学教授 Ion Stoica 表示，免费的人工智能模型现在在性能上与谷歌和 OpenAI 的专有模型“相当接近”，大多数软件开发商最终都会选择使用免费模型。

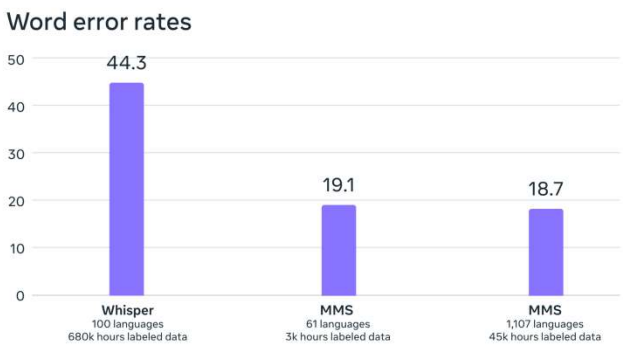
语言大模型#2：LIMA 无需 RLHF 就能对齐，媲美 GPT-4。2023 年 5 月，Meta 发布 650 亿参数语言模型 LIMA，仅在 1,000 个精心挑选的样本上微调 LLaMa-65B 且无需 RLHF，就实现了与 GPT-4 和 Bard 相媲美的性能。

语音大模型#1：MMS 模型可识别 1100+语言，旨在保护世界语种的多样性。MMS 模型在 2023 年 5 月推出。现有的语音识别模型覆盖约 100 种语言，仅为地球上已知的 7,000 多种语言的小部分，MMS 使用 wav2vec 2.0 的自监督学习，支持 1,107 种语言的语音转文字和文字转语音以及超过 4,000 种语言的语言识别。MMS 用圣经训练，与 OpenAI Whisper 相比，使用 MMS 数据训练的模型支持 11 倍以上的语言，且在 54 种语言上的平均错误率仅为一半。

图表32：MMS 支持的语言示意图

资料来源：Meta 官网，华泰研究

图表33：MMS 错误率仅为 Whisper 数据集的一半



资料来源：Meta 官网，华泰研究

CV 大模型#1：SAM 模型可“分割一切”，CV 领域迎来 GPT-3 时刻。2023 年 4 月，Meta 推出首个可“任意图像分割”的基础模型——SAM（Segment Anything Model），它具有强大的一键抠图功能，能从照片或视频中对任意对象实现一键分割，且能够零样本迁移到其他任务，为 CV 领域开启了新的篇章。Meta 还发布了有史以来最大的分割数据集 Segment Anything 1-Billion (SA-1B)，拥有超过 11 亿个分割掩码，掩码具有高质量和多样性。

CV 大模型#2：DINOv2 视觉大模型完全自监督，无需微调。2021 年 4 月，Meta 公开了 DINO 算法，通过自监督学习，DINO 可以从大量未标注的图像中提取视觉特征，这些特征对于各种下游计算机视觉任务非常有用，例如图像分类、物体检测和语义分割。2023 年 4 月，Meta 开源 DINOv2 版本，相比较原始的 DINO 模型，DINOv2 能够对视频进行处理，生成比原始 DINO 方法更高质量的分割结果。模型除了具备图像的识别、分类、分割处理等图像特征提取功能外，还具有语义分割，完善了以图搜图功能。

免责
更多免费行业报告 并购家 www.ipoipo.cn

多模态模型#1：ImageBind 用图像对齐六模态，旨在实现感官大一统。2023 年 5 月，Meta 开源 ImageBind 新模型，是一个像人类一样结合不同感官的新 AI 模型，能够同时从文本、图像/视频、音频、深度（3D）、热能（红外辐射）和惯性测量单元（IMU）等六种不同的模态中学习。ImageBind 可以使用文本、音频和图像的组合来搜索照片、视频、音频文件或文本消息。ImageBind 用于丰富的多媒体搜索、虚拟现实甚至机器人技术，可以和 Meta 内部的虚拟现实、混合现实和元宇宙等技术相结合。

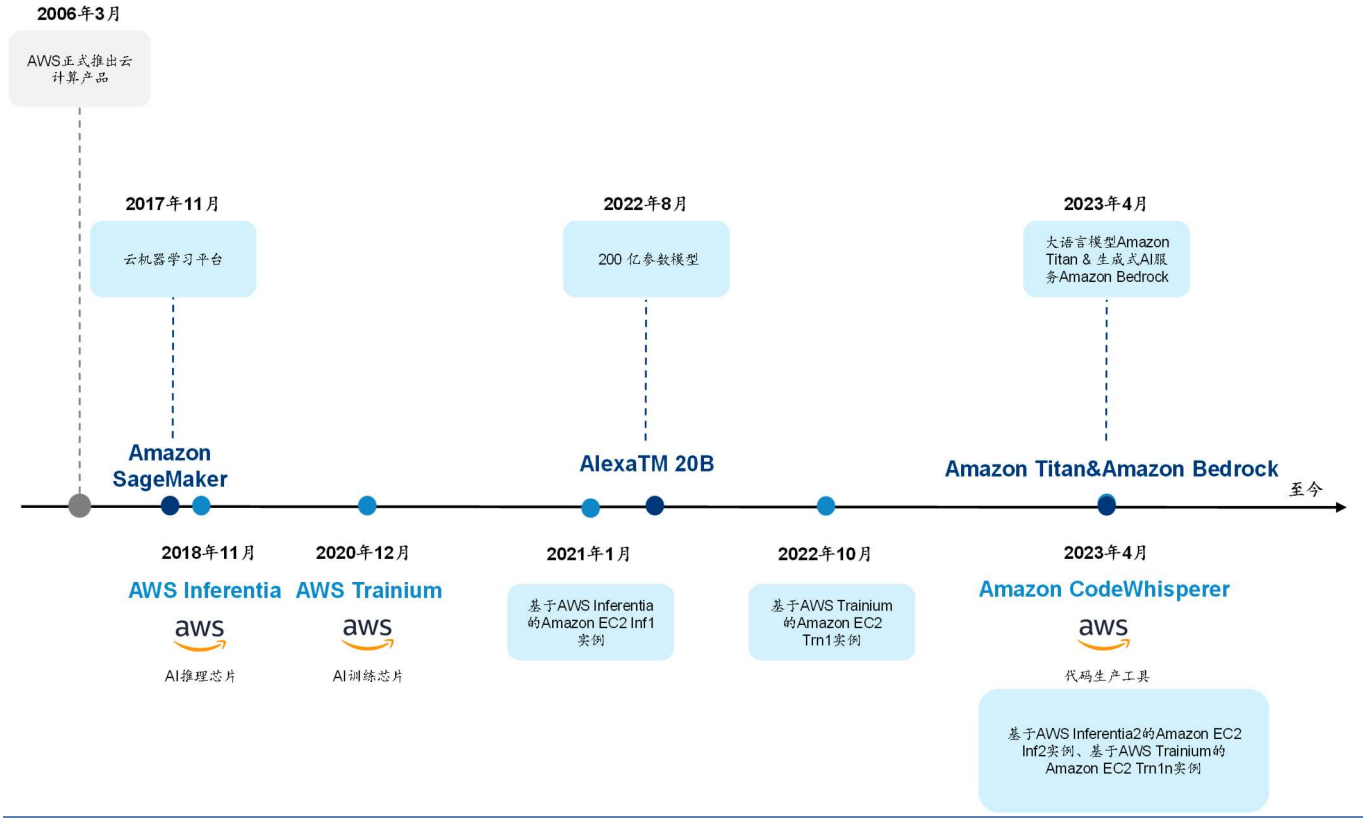
AWS：全球卓越的云服务平台发力 AIGC 市场

发展历程：全球领先的云服务平台强势入局 AIGC

亚马逊成立于 1994 年，是目前全球最大的互联网线上零售商之一。AWS（Amazon Web Services）于 2016 年正式推出，是亚马逊公司旗下的子公司，向个人、企业和政府提供按需即用的云计算平台以及应用程序接口，并按照使用量计费。2002 年 7 月，亚马逊的“Amazon.com Web Services”上线了首款 Web 服务。2006 年 3 月，AWS 推出了 Amazon S3 云存储，随后于 2006 年 8 月推出了 EC2。

AWS 四项技术创新助力 AIGC 发展。2023 年 4 月，AWS 正式入局 AIGC，推出自有基础模型 Titan 和 AIGC 服务 Bedrock，以及 AI 编程助手 Amazon CodeWhisperer，并宣布基于自研推理和训练 AI 芯片的最新实例 Amazon EC2 Trn1n 和 Amazon EC2 Inf2 正式可用。

图表34：AWS 发展历程



资料来源：AWS 官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

算力：提供具有成本效益的生成式 AI 云基础设施

Inferentia 是 ML/DL 推理（Inference）加速器。2018 年 11 月，AWS 发布首款云端 AI 芯片 Inferentia，旨在以极低成本交付高吞吐量、低延迟推理性能。2021 年 1 月，AWS 推出基于 AWS Inferentia 加速器的 Amazon EC2 Inf1 实例，与当前一代基于 GPU 的 Amazon EC2 实例相比，吞吐量最高可提高 2.3 倍，每次推理的成本降低多达 70%。2023 年 4 月，AWS 推出配备 Amazon Inferentia2 芯片的 Amazon EC2 Inf2 实例。与第一代 AWS Inferentia 相比，Inferentia2 的吞吐量提高了 4 倍，延迟低至前者的 1/10。

Trainium 是 ML/DL 训练（Training）加速器。2020 年 12 月，AWS 发布第二款定制的机器学习芯片 AWS Trainium，支持 TensorFlow、PyTorch 和 MXNet 等框架。2022 年 10 月，AWS 推出 Amazon EC2 Trn1。基于 Trainium 的 EC2 Trn1 实例与基于 GPU 的同类实例相比，可节省高达 50% 的训练成本。2023 年 4 月，AWS 推出配备 AWS Trainium 的 Amazon EC2 Trn1n 实例，相比于 Trn1 网络带宽提升了 1 倍，从之前的 800GB 左右提升到 1.6TB，旨在为大型的网络密集型的模型训练提供使用基础。

模型：推出 Titan 大模型及中立托管平台 Bedrock

2023 年 4 月，AWS 推出的自研语言大模型 Titan。Titan 系列模型分为两种，一种是用于内容生成的文本模型 Titan text，可以执行诸如撰写博客文章和电子邮件、总结文档和从数据库中提取信息等任务。另一种是可创建矢量嵌入的嵌入模型 Titan Embeddings，能够将文本输入（字词、短语甚至是大篇幅文章）翻译成包含语义的数字表达的大语言模型。

中立托管平台 Amazon Bedrock 让 AIGC 变得普惠。2023 年 4 月，AWS 推出 Amazon Bedrock，让企业级客户能快速、安全和高性价比地调用多种大模型构建自己的程序。区别于谷歌和微软已发布面向大众的产品，AWS 瞄准的是企业客户，并且期望作为一个“中立”的生成式 AI 大模型托管平台，不依赖于任何一家 AI 初创公司。借助 Bedrock，企业级客户能通过 API 调用来自不同提供商的 AI 模型，例如 AI21 Labs 的 Jurassic-2、Anthropic 的 Claude、Stability AI 的 Stable Diffusion 以及 AWS 自研模型 Titan。

产品：免费向个人开放 AI 编程助手 Amazon CodeWisperer

AI 编程助手 Amazon CodeWhisperer 提高开发者效率。2022 年，AWS 推出 Amazon CodeWhisperer 预览版，这是一款 AI 编程助手，通过内嵌的基础模型，可以根据开发者用自然语言描述的注释和集成开发环境（IDE）中的既有代码实时生成代码建议，从而提升开发者的生产效率。2023 年 4 月，AWS 将 Amazon CodeWhisperer 面向个人开发者免费开放，在适用 Python、Java、JavaScript、TypeScript 和 C#之外，新增支持 Go、Kotlin、Rust、PHP 和 SQL 等十种开发语言。开发者可以通过在 VS Code、IntelliJ IDEA、Amazon Cloud9 等集成开发环境中的 Amazon Toolkit 插件访问 CodeWhisperer。

国内大模型：百度、阿里、腾讯、华为、商汤、讯飞

国内大模型训练企业中，百度具备先发优势，芯片层、框架层、模型层和应用层布局完善。阿里紧随其后，在芯片层、模型层、应用层均有布局，并宣布将旗下所有产品接入通义大模型。腾讯步调稳健，持续研发，在今年以来的大模型热潮中发声克制，并未宣布更多模型、应用进展。华为的大模型布局更偏向基础设施，在芯片层、框架层和模型层均有布局，主要赋能 B 端客户。商汤的 AI 布局涵盖推理芯片、模型层和应用层，其以 CV 起家，应用层面不仅涵盖文字生成能力，同时包括多模态能力。讯飞的 AI 布局主要涵盖模型层、应用层。

免责
更多免费行业报告 并购家 www.ipoipo.cn

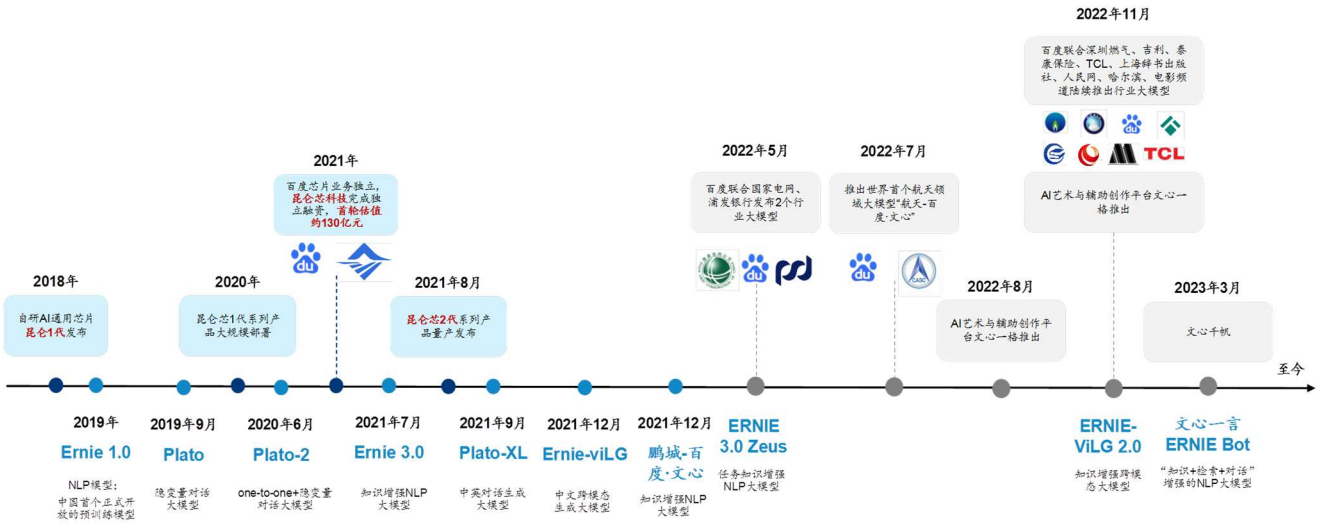
图表35：国内科技公司 AI 布局

资料来源：各公司官网，华泰研究

百度：昆仑芯+飞桨平台+文心大模型，构建广泛应用端生态

百度是当前唯一一家在芯片层、框架层、模型层和应用层都有领先产品布局的企业，四层架构相互协同优化，可以显著地降本增效。在芯片层，百度昆仑芯科技已实现两代通用 AI 芯片“昆仑”的量产及应用，为大模型落地提供强大算力支持。在框架层，“飞桨”是国内首个自主研发的产业级深度学习平台，集基础模型库、端到端开发套件和工具组件于一体，有效支持文心大模型高效、稳定训练。在模型层，“文心大模型”包括基础大模型、任务大模型、行业大模型三级体系，全面满足产业应用需求。在应用层，文心已大规模应用于百度自有业务的各类产品，并通过企业级平台“文心千帆”进一步推动生态构建。

图表36：百度大模型的发展历程



资料来源：昆仑芯科技官网、百度官网、华泰研究

团队：百度 CTO 王海峰领衔，技术大牛带队

百度 CTO 王海峰作为领头人推动百度的 AI 技术战略发展和生态构建。王海峰曾任职于微软，先后主持 Bing 语义搜索、微软小冰等项目；其于 2010 年加入百度，并在 2018 年升任百度 CTO，曾推出百度大脑、百度小度等一系列产品。其他带队的高管还包括百度集团副总裁兼深度学习技术及应用国家工程研究中心副主任吴甜、百度技术委员会主席吴华等。吴甜于 2006 年加入百度，目前负责百度 AI 技术平台和智能云 AI 产品，领衔研发为文心大模型提供支撑的飞桨深度学习平台。吴华曾在百度带领团队推出世界首个互联网 NMT（神经网络机器翻译）系统，目前主要带领 NLP 大模型的技术推进。

免责
更多免费行业报告 并购家 www.ipoipo.cn

算力：两代自研通用 AI 芯片“昆仑”支持大模型落地

百度旗下的芯片公司昆仑芯科技已实现两代通用 AI 芯片产品的量产及落地应用。昆仑芯 1 代 AI 芯片于 2020 年量产，是国内唯一一款经历过互联网大规模核心算法考验的云端 AI 芯片，当前已被广泛部署在互联网、工业制造、智慧城市、智慧交通、科研等领域。昆仑芯 2 代 AI 芯片于 2021 年 8 月量产发布，是国内首款采用 GDDR6 显存的通用 AI 芯片，相比昆仑芯 1 代 AI 芯片性能提升 2-3 倍，且在通用性、易用性方面也有显著增强。目前，昆仑芯已在百度搜索等业务场景落地，也为大模型训练提供底层算力支撑。百度之外，昆仑芯还可为客户提供大模型定制服务，凭借强大算力为大模型落地提供全流程支持。

框架：产业级深度学习平台“飞桨”为大模型构建提供有力支撑

我国首个自主研发、功能丰富、开源开放的产业级深度学习平台飞桨（PaddlePaddle）是百度大模型背后的有力支撑。飞桨以百度多年的深度学习技术研究和业务应用为基础，集深度学习核心训练和推理框架、基础模型库、端到端开发套件和丰富的工具组件于一体，有效支持文心大模型高效、稳定训练。截至目前，飞桨已和 22 家国内外硬件厂商完成了超过 30 种芯片的适配和联合优化，国产芯片适配第一。

文心大模型：“基础+任务+行业”大模型三层体系全面满足产业应用需求

百度文心大模型体系构建了文心大模型、工具与平台两层体系。在模型层，文心大模型包括基础大模型、任务大模型、行业大模型三级体系，打造大模型总量近 40 个，全面满足产业应用需求，涵盖电力、燃气、金融、航天等领域。基础大模型针对特定任务预训练构建任务大模型，任务大模型结合垂直领域数据和知识进一步形成行业大模型，行业大模型则在应用场景的数据反哺基础大模型优化。在工具与平台层，大模型开发套件、文心 API 以及提供全流程开箱即用大模型能力的 EasyDL 和 BML 开发平台，帮助全方位降低应用门槛。

图表37： 文心全景图

资料来源：文心大模型官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

应用：大模型生态逐步构建，推动 AI 落地产业

文心大模型是目前国内预训练大模型应用端生态最好的大模型之一。在百度自有业务中，文心已大规模应用于百度内部的各类产品，包含百度搜索、度小满金融、小度智能屏、百度地图等。例如，度小满的智能征信中台将文心 ERNIE 大模型应用在征信报告的解读上，能够将报告解读出 40 万维的风险变量，以更好地识别小微企业主的信贷风险。在百度业务之外，文心大模型联合国网、浦发、中国航天、人民网等企业推出了 11 个行业大模型，让大模型加速推动行业的智能化转型升级。

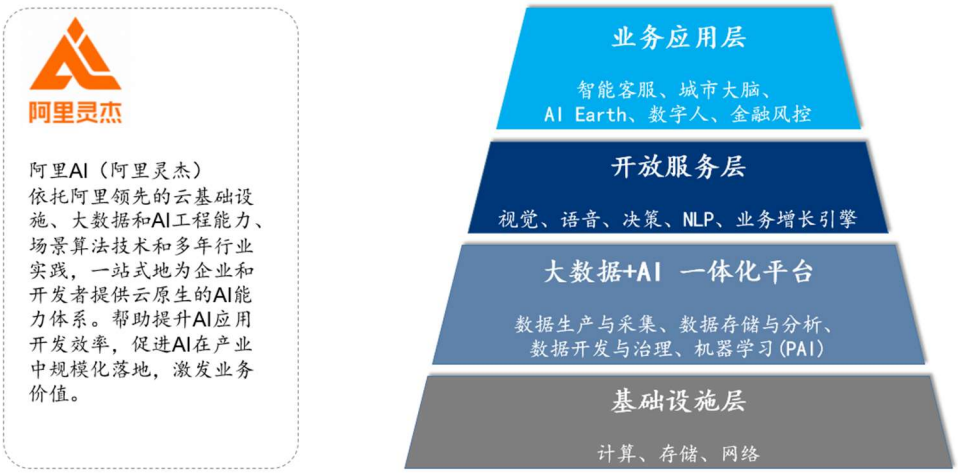
文心一言生态逐步构建，加速生成式 AI 应用需求落地。2023 年 3 月，百度率先发布国内第一个类 ChatGPT 生成式大语言模型“文心一言”，展示了其在文学创作、商业文案创作、数理逻辑推理、中文理解、多模态生成方面的能力，并宣布已有超过 650 家企业宣布接入文心一言生态。同月，百度智能云发布的一站式企业级大模型平台“文心千帆”，平台提供基于文心一言的数据管理、模型定制微调、推理云等服务，助力各行业的生成式 AI 应用需求落地，进一步推动文心一言生态构建。

阿里巴巴：云上赋智，拥抱消费互联网 AI 新时代

阿里巴巴长期深耕于云计算、AI、大模型、边缘计算等复杂系统工程的开发，我们认为其已形成从基础设施层到应用层完整的服务体系与服务能力。

阿里灵杰为企业和开发者提供一站式云原生的 AI 能力体系。依托阿里云的云基础设施、大数据和 AI 工程能力、场景算法技术和行业经验，阿里巴巴通过阿里灵杰一站式帮助企业和开发者提升 AI 应用开发效率，促成 AI 在产业中的规模化落地，激发业务价值。根据阿里灵杰官网介绍，阿里灵杰目前已在金融、教育、交通、新零售、政务和司法 6 大领域提供完整的 AI 应用方案，同时以商品形式出售人脸人体、图像/视频生产、图像分割等 12 大类服务，满足各项应用场景需求。此外，阿里灵杰已推出大数据+AI 一体化平台，包含云原生大数据计算服务 MaxCompute、数据开发与治理平台 DataWorks、机器学习平台 PAI 等，有效帮助客户实现“开箱即用”，提升 AI 工程开发效率。

图表38：阿里 AI 结构



资料来源：阿里灵杰官网，华泰研究

通义大模型——国内首个统一 AI 底座模型。2022 年 9 月，阿里巴巴发布通义大模型，该模型由统一底座层、通用模型层、行业模型层组成，是国内首个统一 AI 底座模型，而统一 AI 模型可以充分利用多种感知模态的数据，从而获得更加丰富和准确的信息，有效解决了传统大模型通用性和易用性欠缺的问题。

更多免费行业报告 并购家 www.ipoipo.cn

1）统一底座层：通义统一底座中的 M6-OFA 模型，在不引入新增结构的情况下，可同时处理图像描述、视觉定位、文生图、视觉蕴含、文档摘要等 10 余项单模态和跨模态任务。

2）通用模型层：覆盖自然语言处理、多模态、计算机视觉。包括 NLP 模型通义 AliceMind，多模态模型通义-M6，CV 模型通义-CV 三大模型。2023 年 4 月 7 日，阿里巴巴推出大规模语言模型通义千问，进一步丰富阿里巴巴在通用模型层的产品形态。

3）行业模型层：深入电商、医疗、法律、金融、娱乐等行业，通用与专业领域大小模型协同，让通义大模型系列可兼顾性能最优化与低成本落地。

图表39：阿里通义大模型结构

资料来源：2023 阿里云峰会，华泰研究

M6 大模型为超大规模跨模态预训练模型（参数 10 万亿+），提供语言理解、图像处理、知识表征等智能服务。根据 M6 大模型官网介绍，M6 以预训练模型的形式输出泛化能力，下游仅需提供场景化数据进行优化调整，即可快速产出符合自身行业特点的精准模型，具备“高精度”、“低门槛”、“多模态”三大特点，有效降低了 AI 使用门槛，推动各行业 AI 技术的落地和普及，已在电商、智能制造、金融等行业有丰富落地应用。

图表40：M6 大模型应用场景

资料来源：阿里云官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

多元应用场景已初步落地。根据阿里云官网介绍，M6 大模型具有强大的图像生成能力和创造力，例如在服装设计模型中，模型可以根据用户输入的款式要求，如“翻领 polo 简约开衫上衣”，快速完成服装的款式设计并生成白底图。在推荐理由生成功能模型中，M6 模型可以捕捉到图片中的主要特征并转换成文本表述，如在提取商品主图中的商品特点信息后，可自动展开扩写为一段流畅自然的商品推荐理由。除此之外，模型还包括视觉问答、图文检索等功能，其技术优势包括：一致性强（生成图像和文本保持一致）、清晰度高（细节逼真、纹理丰富）、多样性高（能够针对一个文本描述产出不同的图片）等。以电商场景为例，M6 大模型能够协助图文营销等部分基础工作，帮助商户降本增效。

M6 大模型已积累丰富用户案例，如每平每屋、斑马智行、支付宝等。根据 M6 大模型官网，M6 大模型与家居家装平台每平每屋合作，基于自身的多模态特征提取能力和模型中存储的丰富电商领域知识，为每平每屋业务提供了优质的家装类商品展示。在引入 M6 产出内容封面 embedding 后，线上精排场景取得了 pctcvr2.6%的提升。此外，专注于专业研发智能汽车操作系统底层基础软件技术的斑马智行算法团队使用 M6 的表征学习能力帮助推进语视对话模型在自动驾驶场景实现应用，语视对话模型在评测上实现 5%以上的绝对值提升，有效优化了自动驾驶体验。

图表41： M6 大模型用户案例

资料来源：阿里云官网，华泰研究

2023 年 4 月 11 日，张勇（阿里巴巴集团董事会主席兼 CEO，阿里云智能集团 CEO）在 2023 年阿里云峰会上指出，未来阿里巴巴所有产品都将接入大模型，进行全面升级，使其智能化水平达到全新高度。我们认为阿里巴巴丰富的消费互联网业务布局使其在 AI 大模型领域的投入具备长期的赋能空间，我们期待数据的持续积累与模型训练效果形成良性循环，并使得阿里巴巴在迈出早期技术和基础设施投资期后，在利润端收获果实。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表42：阿里巴巴所有产品未来将接入通义千问大模型

资料来源：2023 阿里云峰会，华泰研究

腾讯：大模型赋能长尾商家内容自制，提升整体创意和投放效率

腾讯以太极平台为底层，借力混元 AI 和广告两大模型，推出 4 大广告系统，覆盖广告内容理解、智能制作审核、广告推荐模型等环节，其他方案包括视频自动生成，模型定制化和动态投放等。根据腾讯广告公众号，在 DCO 系统加持下，广告曝光改善 5%，流水和 GMV 分别提升 10%和 9%。我们认为，未来腾讯或推出更多 AIGC 应用，视频号、朋友圈、小程序等都将受益于效率改善，或迎来更多提价空间。

技术分析：腾讯广告大模型的产品构架

一、以“太极机器学习平台”为底层技术，打造“混元AI大模型”+“广告大模型”

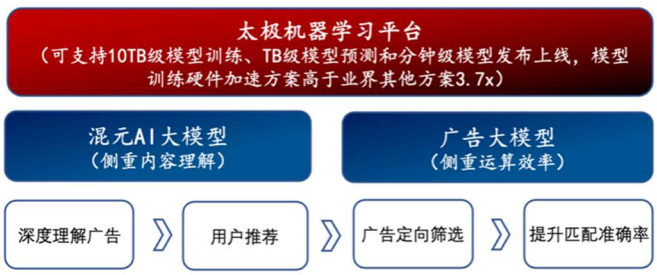
腾讯广告大模型的基本构架：腾讯以“太极机器学习平台”为底层技术依托，打造“混元AI大模型”和“广告大模型”两种具备不同功能的千亿参数模型，前者侧重广告内容理解，后者侧重平台运算效率。

如何理解“混元AI大模型”和“广告大模型”的区别和分工？

从系统处理的时间先后顺序看，混元AI大模型先助力系统深度理解广告内容，然后广告大模型进行推荐，进行广告定向筛选，提升广告和用户的匹配准确率。从大模型基本构架看，这两者均依靠“太极机器学习平台”作为底层技术依托。

从计算能力看，“混元AI大模型”的多模态理解能力，有效加深推荐系统对广告内容的理解，提高用户体验以及广告转化效果；“广告大模型”是广告系统运算能力的关键，在更短时间内实现更加高效精准的广告推荐，可有效提升数据运算的精度、维度与速度。从参数看，单模型推理参数达千亿级别，浮点数计算量最高每秒超过10亿次。

图表43：腾讯以“太极”为底层技术，依托“混元 AI +广告大模型”赋能广告匹配



资料来源：腾讯广告公众号，华泰研究

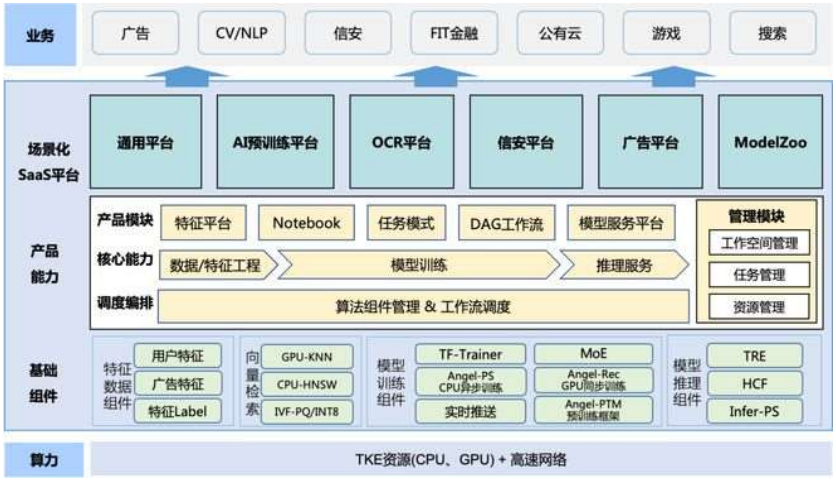
更多免费行业报告 并购家 www.ipoipo.cn

二、混元大模型赋能广告制作环节

太极机器学习平台：针对“大模型+高访问压力”自研搭建，支撑增设更多服务器和高算力需求模型。在大模型场景中，由于广告素材呈多倍数增长，内容对广告平台的承载和计算能力提出更高要求。太极机器学习平台技术层面已达业内顶尖水准，可支持 10TB 级模型训练，以及 TB 级模型推理和分钟级模型发布上线，使其在模型训练、推理等方面的性能达到行业领先水准。

混元大模型基于太极平台的加速效果达到业界平均水平的 3.7x。太极计算机学习平台采用先进的分布式参数服务器架构，该架构的优势在于协调模型参数存储和计算两项任务分布式运行，以便增设更多服务器，支持算力需求更高的模型训练。同时，太极机器学习平台还能够调用超大模型提供在线推理服务。太极机器学习平台在模型训练加速能力亦较为突出，以混元大模型的训练为例，基于太极的加速效果是业界平均水平的 3.7 倍。

图表44：太极机器学习平台逻辑框架



资料来源：腾讯官网，华泰研究

腾讯混元AI大模型：四大技术平台。以混元大模型为底座，腾讯细分研发出四大技术平台，分别为“巨阙”系统主攻广告内容理解、“乾坤”聚焦广告智能创作、“神针”侧重广告智能审核、“天印”主打广告指纹系统。

混元AI大模型能够自动将广告文案生成视频。目前混元AI大模型已能够实现广告智能制作，即通过AIGC自动将广告文案生成视频，大大地提高了系统的理解能力、行业特征的洞察力以及文案创意的生成效率，从而有效降低了广告视频制作的成本。

图表45：腾讯混元大模型具有较强的内容“一键生成”能力

功能	表现形式
图生视频	可以将静态的图片自动生成不同样式的视频广告
文案助手	可以为广告自动生成更恰当的标题，提升广告的效果
文生视频	未来只需要提供一句广告文案，就可以自动生成与之匹配的视频

资料来源：腾讯云开发者公众号，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表46：腾讯混元 AI 大模型：四大技术平台支持广告各环节效率提升

资料来源：腾讯云开发者公众号，华泰研究

三、精排大模型助力提升广告主15%GMV

腾讯大模型在广告中的实用效果：在广告大模型的帮助下，腾讯广告的整体投放水平累计提升20%以上。例如精排的单模型推理模型参数量高达千亿，经过序列化处理可达数百GB，其浮点计算能力也能够在10亿次以内快速完成，各项算法评价指标相较于百亿规模小模型均提升明显。目前，精排大模型能短时间内在太极平台上完成计算、提供可靠的广告匹配结果。

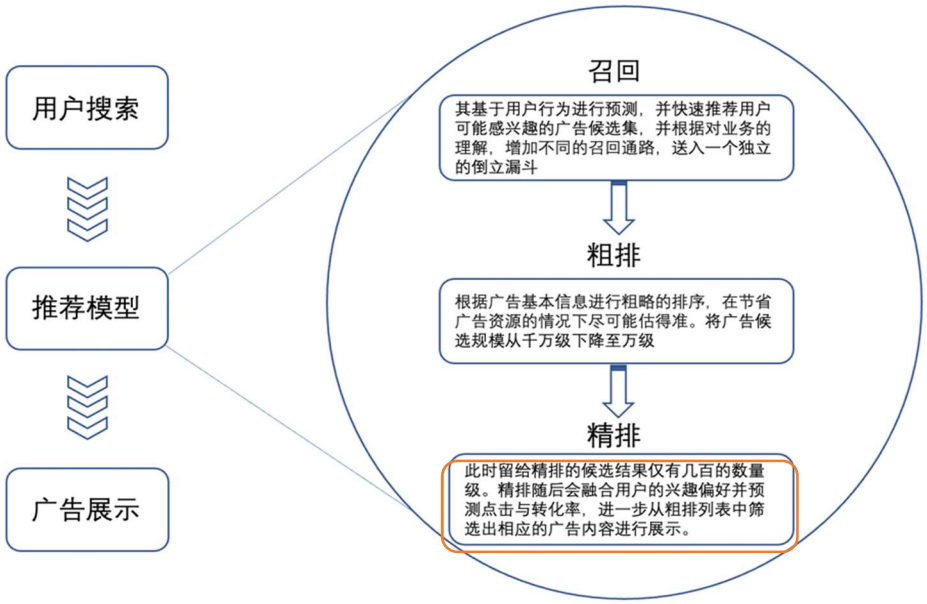
腾讯广告精排大模型：“高算力+强推理”助力广告高效匹配

何为精排？精排即为精确排序，在广告计算模型中，用户看到的广告先后顺序，涉及到精确排序环节。精排环节应用大模型分析后，将会获得显著的分析和匹配效果改善，尤其是涉及长尾广告的分析和匹配。换而言之，用户可以看到更加符合自身兴趣的广告，广告主获得更高的转化率。

腾讯的广告精排大模型，支撑了千亿的特征规模，模型体量达到TB级，各项算法指标相对于原来的模型，都有显著提升。上线后的评估表明，用户看到的广告与自己的相关性更高，广告主获得的用户生命周期总价值（LTV）也明显提高。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表47：精排大模型“千人千面”：精排环节应用大模型分析后，分析和匹配效果显著改善（尤其是长尾广告）。用户可以看到更加符合自身兴趣的广告，广告主获得更高的转化率。

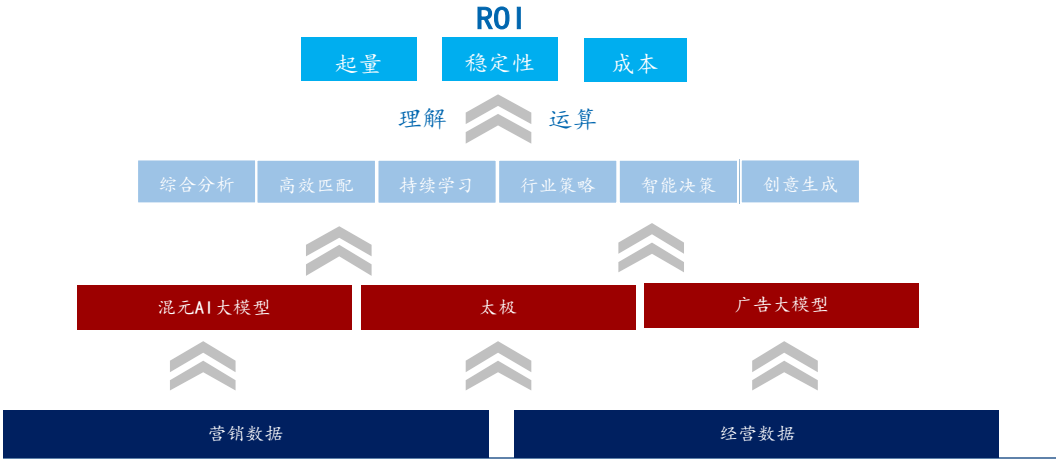


资料来源：腾讯云开发者公众号，华泰研究

精排大模型赋能下，广告主或可期待ROI的大幅抬升。相比百亿规模小模型算法，腾讯广告的"精排大模型"平均为广告主带来15%的 GMV 提升（截至22年6月），也初步验证了大模型的商业化潜力。未来，随着大模型平台系统和算法的持续升级，也会逐步向召回、粗排等其他环节辐射，不断优化投放效果实现高质匹配。

大模型如何更精确地分析腾讯视频素材、提升推荐精准度？腾讯视频号通过对海量视频打标签，让系统不仅能理解视频的尺寸、分辨率、大小、镜头数等物理特征，还能对播放场景、用户行为、内容等特征进行学习和训练，从而更好的理解视频素材，实现人群与视频素材的更精准匹配，加强视频类广告在粗排、精排等阶段的竞争力，最终实现效果的不断优化。

图表48：腾讯广告系统技术矩阵



资料来源：腾讯广告公众号，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

华为：昇腾芯片+MindSpore+盘古大模型，B 端应用场景落地可期

领导者：华为云 AI 首席科学家田奇

华为云 AI 首席科学家田奇是盘古大模型团队的负责人。田奇曾于 2008 至 2009 年，从大学调至微软亚洲研究院多媒体计算组进行研究工作。2018 年，田奇加入华为担任首席科学家，负责华为云相关业务，华为云是目前国内云服务市占率最高的厂商。当前，田奇正带领团队推动盘古大模型从科研创新走向产业应用，逐步构建大模型的应用生态。

算力：以自研 AI 芯片昇腾作为根基，打造 AI 产业平台

华为昇腾芯片包括用于训练的算力最强 AI 芯片昇腾 910，以及用于推理的全栈全场景 AI 芯片昇腾 310。2019 年 8 月推出的昇腾 910 算力超英伟达 Tesla V100 一倍，在同等功耗下拥有的算力资源达到了当时业内最佳水平的两倍。目前，“昇腾 AI”基础软硬件平台已成功孵化和适配了 30 多个主流大模型，为我国一半以上的原生大模型提供算力支撑。基于该平台，昇腾 AI 产业已与 20 多家硬件伙伴和 1100 多家软件伙伴建立了合作关系，并共同推出了 2000 多个行业 AI 解决方案，参与其中的开发者数量突破 150 万。

框架：深度学习框架 MindSpore 提升模型开发效率

深度学习框架 MindSpore 从训练推理部署全流程支撑模型高效开发，是国内社区中最活跃、关注度最高、被应用最多的框架之一。匹配昇腾 AI 处理器，MindSpore 有效克服 AI 计算的复杂性和算力的多样性挑战，打造面向端、边、云的全场景 AI 基础设施方案，让芯片的强大算力能够以最高效的方式被开发者们利用。面向大模型开发场景，MindSpore 提供系列工具及套件，发挥软硬件综合优势，从训练推理部署全流程支撑模型高效开发。同时，MindSpore 可兼容第三方 AI 框架生态，以实现模型的快速迁移。

盘古大模型：赋能千行百业的三阶段体系

华为云盘古大模型于 2021 年 4 月正式发布，到 2022 年聚焦行业应用落地，已经基于一站式 AI 开发平台 ModelArts，发展出包括基础大模型（L0）、行业大模型（L1）、行业细分场景模型（L2）三大阶段的成熟体系，让 AI 开发由作坊式转变为工业化开发的新模式，赋能千行百业。具体来说，盘古大模型的三层体系架构包含：1）L0 基础大模型：30 亿参数的 CV 大模型在业界首次实现模型按需抽取，千亿参数和 40TB 训练数据的 NLP 大模型在在 CLUE 榜单实现业界领先，科学计算大模型致力于解决各种科学问题、促进基础科学的发展；2）L1 行业大模型：涵盖矿山、气象、药物、分子、电力、海浪、金融等行业；3）L2 场景模型：包含金融 OCR、电力巡检、建筑能耗优化等场景。

应用：行业大模型逐步落地，B 端场景应用可期

盘古 CV 大模型可以赋能分类、识别、检测等视觉场景。华为已基于 CV 大模型推出矿山大模型、电力大模型等行业大模型，推动相关工业领域安全高效作业。在华为与能源集团合作推出的盘古矿山大模型中，模型能够解决 AI 在煤矿行业落地难、门槛高等问题。例如，在煤矿主运场景中，AI 主运智能监测系统的异物识别准确率达 98%，实现全时段巡检，避免因漏检造成的安全事故；在作业场景中，掘进作业序列智能监测的动作规范识别准确率超过 95%，保障井下作业安全。此外，盘古 CV 大模型还可应用在铁路轨道机车的缺陷检测中，识别机车中吊链、脱落、裂痕等潜在不安全因素，在郑州铁路段的 32000 多样本评测中，对缺陷和故障检测的准确度达 99%。

盘古 NLP 大模型可助力文档检索、智能 ERP、小语种等内容理解和文本生成场景。基于 NLP 大模型，华为与合作伙伴开发了支持千亿参数的阿拉伯语 NLP 大模型，语义理解准确率达到 95%。此外，盘古 NLP 平台凭借突破性的零样本 AI 建模技术，可帮助金融机构降低 AI 建模成本，提升 10 到 1000 倍 AI 建模效率，目前已经在多家银行、保险、证券等金融机构的数字化客户经营场景落地，全面助力客户沟通、销售管理、客户洞察等场景的降本增效。

免责
更多免费行业报告 并购家 www.ipoipo.cn

盘古药物分子大模型可辅助医药研发，突破新药研发的“双十定律”。西安交通大学第一附属医院刘冰教授团队在抗生素研发工作中，采用大模型辅助药物设计，突破性地研发出超级抗菌药 Drug X，突破新药从研发到上市平均需 10 年时间和 10 亿美元投入的“双十定律”，将先导药的研发周期从数年缩短至一个月，且研发成本降低 70%。此外，药物分子大模型还可应用于 AI 药物分子筛选，使得成药性预测准确率比传统方式高 20%。新技术大幅减少了人力筛选的时间与成本，充分释放了科研人员的创新活力。

图表49：通过华为云 AI 辅助药物设计服务平台进行药物分子筛选

资料来源：华为云官网，华泰研究

盘古气象大模型是全球首个天气预测精度超过传统方式的 AI 模型。传统是通过数字分析的方法，精度位列世界第一的是欧洲气象中心。区别于传统的数字分析方法，气象大模型基于一种 3D 高分辨率的 AI 气象预报方法，可以在秒级的时间内完成全球未来 1 小时到 7 天的天气预报，精度首次超过了欧洲气象中心的数字分析的方法，并且预测速度提升了 1 万倍以上。在自然灾害里面，例如台风轨迹预测，盘古的精度相对于世界第一的欧洲气象中心的方法提升了 20%以上。

商汤：SenseCore 大装置+日日新大模型

算力：基于 AI 大装置 SenseCore，以 AI 模型赋能四大业务

商汤基于 AI 大装置 SenseCore 辅助自身业务开展。AI 大装置 SenseCore 打通了算力、算法和平台之间的连接与协同，构建成一整套端到端的架构体系；基于 AI 大装置 SenseCore，商汤通过 22,000+商用 AI 模型，赋能智慧商业、智慧城市、智慧生活和智慧汽车四大业务。

大装置 SenseCore 拥有约 2.7 万块 GPU，约 5 exaFLOPS 算力，提供包括算力服务（IaaS）、开发工具和深度学习平台（PaaS）以及模型部署及推理（MaaS）在内的全栈式 AIaaS 服务。公司目前最大能够支持需要 4,000 张卡并行计算的千亿参数大模型训练，为 8 家外部大型客户提供大模型训练服务，客户涵盖互联网、游戏、商业银行，科研机构等多个领域。

日日新大模型：构建面向 AGI 的核心能力，驱动垂直行业降本增效

商汤是国内最早布局 AI 大模型的企业之一，已实现 CV、NLP、多模态等大模型的全面布局。2019 年已经发布了拥有 10 亿参数的图像大模型，2022 发布的视觉模型参数量达到 320 亿，是全球最大的通用视觉模型之一，能够实现高性能的目标检测、图像分割和多物体识别算法等功能。今年 3 月推出多模态大模型“书生 2.5”，具备图像描述、视觉问答、视觉推理、文字识别、文生图、文本检索视觉内容等功能，在国内处于领先地位。2023 年 4 月，商汤在技术交流会中正式发布“日日新 SenseNova”大模型体系，实现 CV、NLP、多模态等大模型的全面布局，并展示了其问答、代码生成、2D/3D 数字人生成、3D 场景/物体生成等 AI 模型应用能力。

更多免费行业报告 并购家 www.ipoipo.cn

1800 亿参数“商量”大模型赋能专业知识、代码生成、医疗等垂直场景。 商汤于技术交流会同时发布 1800 亿参数“商量”（SenseChat）语言大模型，主要能力包括：1）长文本理解：相比 ChatGPT，SenseChat 支持财务、法务等专业领域超长文本知识理解，能够基于用户上传的长文本 pdf 进行理解和对话。2）代码生成：Visual Studio Code 接入 SenseChat 插件，根据用户指令直接生成代码，在 humaneval 测试集上，一次通过率为 40.2%，据商汤表示该数据高于 Copilot；3）赋能行业：已落地新华医院“便捷就医服务”，辅助初步问诊、就医挂号建议等。此外，公司计划推出面向行业客户的类 ChatGPT 对话机器人服务。

应用：MaaS 模式可能成为重要新趋势

商汤日日新大模型开放 API 体系包含自然语言生成 API、图片生成 API、视觉通用感知任务 API 和标注 API。此外，商汤还提供了数据标注、模型训练及微调等一系列 MaaS 服务。近期，我们注意到，英伟达，百度，商汤等企业都提出了类似 MaaS 的新商业模式，其核心是利用自己已经拥有的通用大模型，帮助企业以专有数据创建专有模型。其中，英伟达提供基于其文字、图像和生物医药模型的大模型训练服务 AI Foundations，百度推出文心千帆大模型平台，表示未来云计算商业模式会变成 MaaS。拥有大模型的企业，从“卖算力”走向“卖模型”可能成为 AI 企业发展的一条新商业模式。

科大讯飞：重磅发布星火大模型+四大 AI 应用成果

科大讯飞在 AI 认知智能领域已有十几年的积累，具备认知大模型“智能涌现”的源头核心技术储备。14 年讯飞就推出讯飞超脑计划，明确提出要让机器像人一样具备“能理解、会思考”的能力；17 年讯飞智医助理在全球首次通过了国家执业医师资格考试，超过了 96.3%参加考试的医生；19 年在斯坦福大学发起的国际著名 SQuAD 机器阅读理解比赛中，讯飞模型在英文阅读理解中首次超过了人类平均水平；22 年在艾伦研究院组织的 OpenBookQA 科学常识推理比赛中，讯飞单模型首次超过了人类平均水平；22 年 1 月，讯飞正式宣布讯飞超脑 2030 计划，要让“懂知识、会学习、能进化”的通用人工智能技术，以机器人的实体形态或虚拟形态能够进入到每一个家庭；22 年 12 月，讯飞启动认知智能大模型的专项攻关；**23 年 5 月 6 日，讯飞星火认知大模型正式发布，同时发布了 4 大 AI 行业应用成果。**

图表50：讯飞为认知大模型所做的源头核心技术储备

资料来源：科大讯飞官网、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

讯飞星火大模型在 AGI 核心能力上表现出色

1）长文本生成：星火中文已经超过 ChatGPT 水平，英文能力接近并快速进步中；2）语言理解：星火在国内可测的现有系统中遥遥领先，跟 ChatGPT 相比还有细微的差别，且持续提升中；3）逻辑推理：讯飞星火大模型在常识、科学等推理方面表现较好；4）泛领域开放式知识问答：星火可以进行比较好的回答，在医学知识领域，甚至超越了 GPT-4 的能力；5）数学能力：在数学相关的 16 个类别详细任务中，讯飞在国内可测的同类系统中遥遥领先，现已经超过了 ChatGPT；6）代码能力：起步阶段，将在 8 月 15 日的升级版本中实现突破；7）多模态输入和表达能力：预计下一代大模型发布。

图表51：讯飞星火大模型六大核心能力

资料来源：科大讯飞官网、华泰研究

针对星火大模型有待完善的功能，讯飞给出了明确的升级时间表

讯飞星火大模型当前仍有待提升的能力包括开放式问答（实时性问题）、多轮对话能力、代码能力、数学能力、多模态等能力。针对此，讯飞给出了非常明确的升级时间点。6 月 9 日，突破开放式问答（接入搜索引擎）、多轮对话能力再升级、数学能力再升级；8 月 15 日，突破代码能力、多模态交互再升级；10 月 24 日，通用模型对标 ChatGPT(中文超越，英文相当)。

发布四大行业应用成果，率先实现 AI 产品化落地

1）星火+教育：主要针对学习机 T20Pro/T20 发布了 AI 新功能（老功能升级），包括 AI 中英文作文批改、英语 AI 口语对话、初中数学 AI 同步精准学）；

2）星火+办公：智能办公本 X2、智能录音笔（SR702）、讯飞听见、智能麦克风 M2 将具备会议纪要能力、语篇规整能力、一键成稿能力三大新功能；

3）星火+汽车：大幅提升人机交互能力，并接入插件功能；

4）星火+数字员工：虚拟人智能交互机能力全面提升，推出新一代生成式 PRA。

相关未上市公司：智谱、MiniMax、第四范式、聆心智能、澜舟科技、云知声、扩博智能

智谱华章（未上市）

北京智谱华章科技有限公司是认知智能领跑者，持续打造高性能千亿级普惠大模型。2019 年，公司成立，由清华大学计算机系知识工程实验室的技术成果转化而来，致力于打造新一代认知智能通用模型。公司合作研发了双语千亿级超大规模预训练模型 GLM-130B，并构建了高精度通用知识图谱，形成数据与知识双轮驱动的认知引擎，基于此模型打造了 ChatGLM。公司还推出了认知大模型平台 Bigmodel.ai，包括 CodeGeeX 和 CogView 等产品，提供智能 API 服务，链接物理世界的亿级用户、赋能元宇宙数字人、成为具身机器人的基座，赋予机器像人一样“思考”的能力。

更多免费行业报告 并购家 www.ipoipo.cn

MiniMax（未上市）

MiniMax 是国内首家多模态 AI 大模型创业公司。2021 年，公司成立，致力于成为通用人工智能时代基础设施建设者和内容应用创造者。公司自研了整套端到端 AGI 引擎系统：底层基础设施以极高效的并行计算能力支撑超大规模跨模型训练任务；文本、语音、视觉三个模态大模型算法能力均达到世界级领先水平；计算平台为通用模型提供推理优化能力，极大降低了 AIGC 产品研发门槛。公司首款 C 端产品 Glow 定位 AI 虚拟聊天社交软件，主打聊天、陪伴等情感功能。

第四范式（未上市）

北京第四范式智能技术股份有限公司是中国领先的企业级 AI 解决方案提供商。2014 年，公司成立，依托自动化机器学习、迁移学习、环境学习和自动化强化学习等核心技术，研发端到端企业级人工智能解决方案。通过统一的基于平台的交付模型，降低了人工智能跨行业应用的门槛，使行业用户能够在其平台上实现行业数据价值的最大化。产品广泛应用于金融、零售、制造业、能源电力、电信、医疗等领域。根据灼识咨询数据，在 2022 年以平台为中心的人工智能决策方面，公司国内市场份额排名第一。截至 2022 年底，公司拥有 104 名标杆用户（财富世界 500 强企业及公众上市公司）。

聆心智能（未上市）

北京聆心智能科技有限公司是大模型和对话系统的开发与应用领域的行业领先者。2021 年公司成立，致力于打造可控、可配置、安全的超拟人大模型，通过简单设置即可构造一个有知识、有个性、有风格的类人智能体。公司孵化自清华大学计算机系，团队核心成员均来自于清华大学、卡内基梅隆大学和谷歌等国内外顶尖高校及公司。2022 年，公司联合清华大学 CoAI 实验室共同发布开放域预训练对话模型 OPD，兼顾出色的闲聊能力与知识性，推动了中文对话领域的发展。同年，公司推出首个超拟人 AI 产品——“AI 乌托邦”，允许用户快速定制 AI 角色。

澜舟科技（未上市）

北京澜舟科技有限公司是业界领先的认知智能公司。2021 年公司成立，致力于以自然语言处理技术为基础，为全球企业提供新一代认知智能平台，助力企业数字化转型升级。同年，公司推出基于团队自研技术研发的大规模预训练语言模型“孟子预训练模型”，可处理多语言、多模态数据，同时支持多种文本理解和文本生成任务，能快速满足不同领域、不同应用场景的需求。它基于 Transformer 架构，仅包含十亿参数量，基于数百 G 级别涵盖互联网网页、社区、新闻、电子商务、金融等领域的高质量语料训练。

云知声（未上市）

云知声智能科技股份有限公司是头部物联网 AI 服务提供商。2012 年，公司成立，以智能语音识别、自然语言理解和知识图谱等全栈 AI 技术为核心，打造云服务和 AI 芯片，输出 PaaS 和 SaaS 平台服务，提供面向智慧物联、智慧医疗等场景的物联网智能化产品服务。公司精英云集，其研发团队占比 78%以上，主要核心成员来自盛大、IBM、华为、阿里和微软等国内外知名企业。2023 年，公司发布山海大模型，目标是年内通用能力比肩 ChatGPT，并在医疗、物联和教育等多个垂直领域能力超越 GPT4。

扩博智能（未上市）

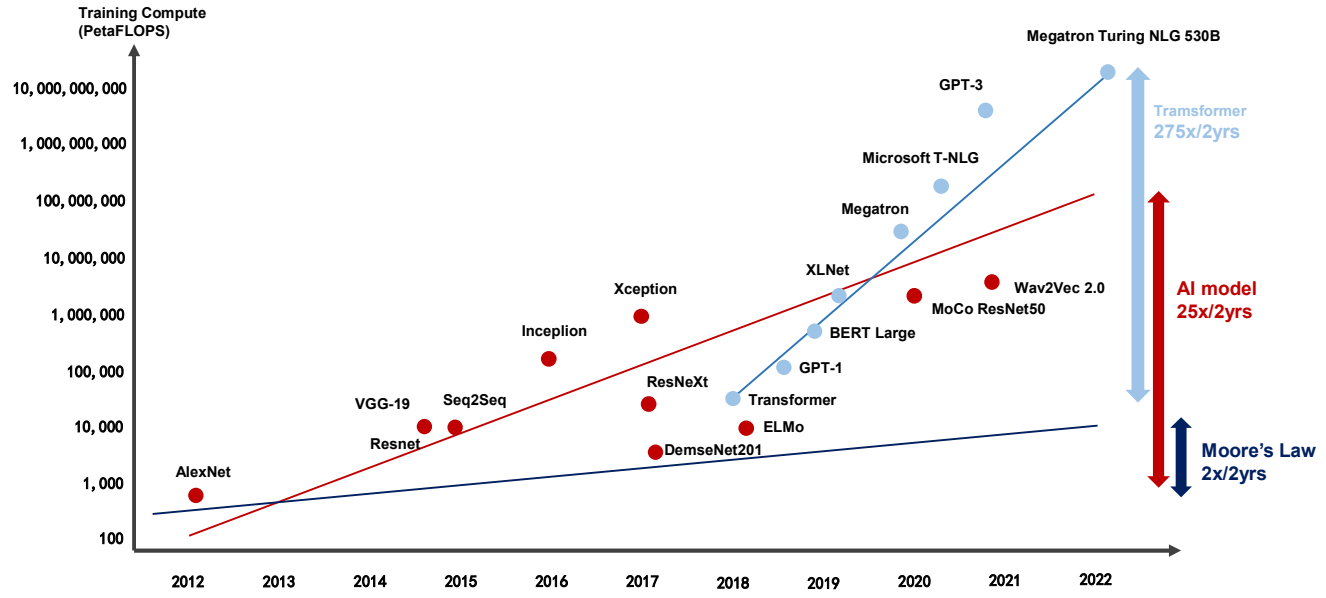
上海扩博智能技术有限公司是专注于风电和零售场景的人工智能行业专家。2016 年，公司成立，基于机器学习、IoT、边缘计算和计算机视觉等技术，为客户提供端到端的智能服务，有效提升传统行业运营效率，加快数字化变革。公司瞄准风电和零售行业，提供智慧风电全自动风机叶片巡检产品和智慧零售线下执行产品。公司自成立以来已提交 130 多项知识产权申请，凭借自身对技术的专注和探索吸引了众多优秀人才，其研发工程团队硕博人才占比 70%以上，主要核心成员来自微软、谷歌、脸书、百度和华为等知名企业。

更多免费行业报告 并购家 www.ipoipo.cn

算力：AI 应用如何拉动需求变化

随着 AI 应用的普及，以及以 ChatGPT 为代表的大模型的出现，AI 相关算力需求正成为高性能计算的主要增长点。当前，大模型最重要的特点是通过不断增加模型参数和训练的数据集的规模，来实现更高的预测精度和通用性。根据 OpenAI 统计，目前大模型训练所需算力的增速保持 3-4 个月/倍速度增长，远超摩尔定律 18-24 个月/倍。

图表52：大模型对半导体需求增长快于摩尔定律

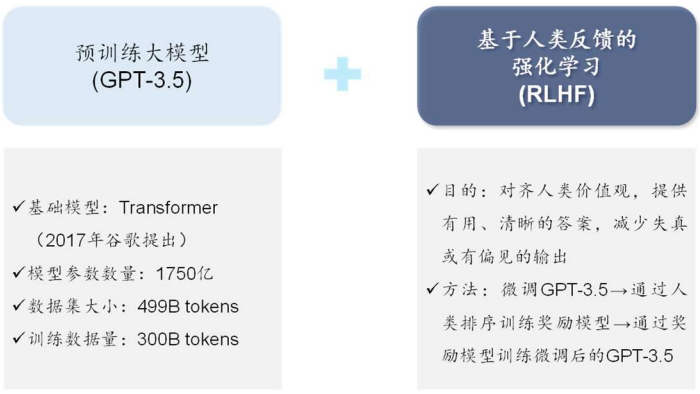


资料来源：NVIDIA Ian Buck statement (2018)，华泰研究

根据微软 2020 年披露，其服务 OpenAI 的计算集群采用超过 1 万张 GPU。据微软官网，与进入 TOP500 榜单的超算相比，这一计算集群排名前五。我们假设该集群使用英伟达 A100，经简略测算，在用于超算的 FP64 双精度下，其浮点计算能力约相当于我国最先进的超算中心太湖之光，是世界最大的超算中心 Frontier 的 9%。

以目前火热的 ChatGPT 为例，ChatGPT 主要是由（1）预训练大模型如 GPT-3.5 大模型，（2）对该大模型进行基于人类反馈的强化学习（RLHF）训练而来。GPT-3.5 模型未公布细节，但是其前代 GPT-3 拥有 1750 亿参数，数据集 499B tokens，训练数据量为 300B tokens。

图表53：ChatGPT 是 GPT-3.5+基于人类反馈的强化学习



注：GPT-3.5 的参数量、数据量参考 GPT-3 数据

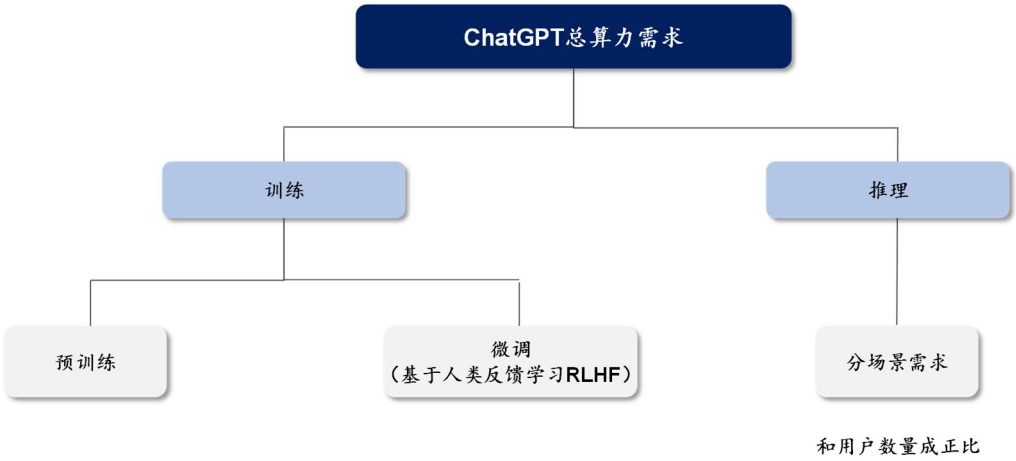
资料来源：《GPT-4 Technical Report》（OpenAI，2023），《Language Models are Few-Shot Learners》（Brown et al., 2023），华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

大模型对训练/推理算力需求的测算

我们认为 ChatGPT 对算力的需求主要包括三个方面：1）**初次训练：**通过大量数据训练出一个复杂的神经网络模型，使其能够适应一系列下游任务，具备一定的泛化性、通用性。训练一个模型本质上是需要持续训练，直到达到一个精确度水平。在一个模型被完全训练之后，可通过常规微调向其添加额外的层或额外的信息；2）**常规微调：**训练不止一次，包括前期调制参数与后期常规微调；**3）推理：**将训练出的模型投入具体场景使用，即使用已有模型进行运算，输入新数据一次性获得正确结论。推理高度独立于训练。对于训练，硬件看起来是一样的，取决于模型的大小，而推理依赖于任务是什么。

图表54：ChatGPT 训练+推理算力需求测算框架



资料来源：，华泰研究

训练阶段：GPT3 算力训练 3640PFlops/s-day，根据测算训练 30 天大约需要 1944 张 A100。目前的大语言模型大多基于 Transformer 结构，该架构由谷歌在 2017 年提出，摒弃了 CNN 和 RNN 的模型结构，完全基于注意力机制，并行程度高，模型训练速度快。ChatGPT 基于 GPT-3.5 模型，GPT-3.5 未公布细节，但其前代 GPT-3 拥有 1750 亿参数，数据集 499B tokens，训练数据量为 300B tokens。据 OpenAI 团队发表于 2020 年的论文《Language Models are Few-Shot Learners》，训练一次 1746 亿参数的 GPT-3 模型需要的算力约为 3640 PFlop/s-day（即假设每秒计算一千万亿次，需要计算 3640 天）。大模型训练一次 GPU 的需求量可以用训练所需浮点运算次数除以单位 GPU 单位时间有效浮点运算次数得到，目前训练端 GPU 多采用英伟达 A100，考虑：1）A100 Tensor Core 采用张量算力 TF32 稠密峰值 156 TFLOPS；2）实际 GPU 吞吐量与模型策略、Batch size、训练天数等都有很强的相关性，参考文献及实际案例给予 40%有效算力比率，最终确定单位 GPU 实际算力为 62.4 TFLOPS。根据《Language Models are Few-Shot Learners》论文中的算力数据，在模型训练时长为 30 天的假设下，对应需要 1944 张 A100。

图表55：模型训练 GPU 用量测算

$$\text{GPU 需求量} = \frac{\text{训练所需浮点运算次数}}{\text{单位 GPU 单位时间有效浮点预算次数}}$$

$$= \frac{\text{模型参数}\{1750\text{ 亿}\} \times \text{训练集大小}(300\text{B}) \times \text{每单词训练所需浮点数}(6)}{\text{单位 GPU 单位时间有效浮点预算次数(需考虑并行训练/网络数据交换带来的峰值算力折损)}}$$

资料来源：华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表56：各个模型参数以及 A100 用量（换算量）

	模型	总计算量(PF-days)	总计算量(flops)	参数量(百万个)	令牌数量(十亿)	处理每参数和词汇所需执行浮点运算数	一次更新参数所需矩阵乘法运算次数	前向传播时处理每个激活参数和每个输入词汇所需执行浮点计算次数	处理每个词汇时被激活的模型参数个数	A100 用量(张 30 天)
T5 模型	T5-Small	2.08E+00	1.80E+20	60	1000	3	3	1	0.5	1
	T5-Base	7.64E+00	6.60E+20	220	1000	3	3	1	0.5	4
	T5-Large	2.67E+01	2.31E+21	770	1000	3	3	1	0.5	14
	T5-3B	1.04E+02	9.00E+21	3000	1000	3	3	1	0.5	56
	T5-11B	3.82E+02	3.30E+22	11000	1000	3	3	1	0.5	204
BERT 模型	BERT-Base	1.89E+00	1.64E+20	109	250	6	3	2	1	1
	BERT-Large	6.16E+00	5.33E+20	355	250	6	3	2	1	3
	RoBERTa-Base	1.74E+01	1.50E+21	125	2000	6	3	2	1	9
	RoBERTa-Large	4.93E+01	4.26E+21	355	2000	6	3	2	1	26
GPT 模型	GPT-3 Small	2.60E+00	2.25E+20	125	300	6	3	2	1	1
	GPT-3 Medium	7.42E+00	6.41E+20	356	300	6	3	2	1	4
	GPT-3 Large	1.58E+01	1.37E+21	760	300	6	3	2	1	8
	GPT-3 XL	2.75E+01	2.38E+21	1320	300	6	3	2	1	14
	GPT-3 2.7B	5.52E+01	4.77E+21	2650	300	6	3	2	1	29
	GPT-3 6.7B	1.39E+02	1.20E+22	6660	300	6	3	2	1	74
	GPT-3 13B	2.68E+02	2.31E+22	12850	300	6	3	2	1	143
	GPT-3 175B	3.64E+03	3.14E+23	174600	300	6	3	2	1	1,944

注：A100 用量计算假设条件为训练 30 天

资料来源：OpenAI、Tom B. Brown et al. *Language Models are Few-Shot Learners* (2020)，华泰研究

微调阶段：微调（Fine-tuning）是在训练大模型时常见的技术，通常用于在一个已经经过预训练的基础模型上通过人类反馈机制进行进一步的训练，以适应特定的任务或领域，主要优势在于：1）预训练使得模型具备了一定的通用语言理解和生成能力，但预训练模型通常不具备针对特定任务或领域的专门知识。通过微调，可以进一步调整模型的参数，以使其适应特定任务的要求。2）微调使用任务特定的数据集来进一步调整模型，使其能够更好地捕捉任务相关的信息。3）通过对少量参数的调整，可以减少过渡拟合的风险。由于微调通常只需要数小时或数天，且在相对较小的特定数据集训练，微调所需算力通常小于一次预训练算力 3640 PFlop/s-day。

推理阶段：大模型推理的算力需求通常是在训练完成后的一个阶段，将已经训练好的模型应用于实际的任务和数据。由于推理算力需求直接和场景任务相关，我们参考 Fortune 杂志数据对 OpenAI 官网对话进行场景模拟，假设每次用户与 ChatGPT 互动产生的算力云服务成本约 0.01 美元，按照 SimilarWeb 统计 4 月全球访问量 17.6 亿次计算，则单月运营成本大概为 1760 万美元。根据 Lambda 数据训练 1750 亿参数的 GPT 3 大约需要的机时成本为 460 万美元，如果我们简单按照成本比例计算，则该场景至少需要算力近 13,905 PFlop/s-day。

AI 军备竞赛需要多大算力

训练算力：CSP 及 AI 创业公司带动训练算力需求快速增长

参考 2023 年 3 月 23 日发布的《从英伟达 GTC 看 AI 工厂的投资机会》，我们认为，在比较乐观的假设下，每家科技巨头（Big Tech）训练 GPT-4 及类似的多模态大模型需要 1.6 万张 H100 GPU 的算力。这对应约 5 亿美金/家的资本开支，其中 4 亿美金是 GPU 采购费用。此外，我们看到世界各地出现 AI 大模型创业热潮，我们认为 1600 张 H100/家是 AI 创业公司入门水平。

我们认为全球大约有 30 家科技巨头和 300 家 AI 初创企业进行相关投资。这对应近 300 亿美金的 GPU 需求，相当于英伟达 22 年数据中心收入的 2 倍。根据 IC Insight 预估 2022 年全球 MPU 市场 1104 亿美元的空间计算，相当于全球芯片市场的 26.4%。从数据中心机柜投资来看，我们假设单机柜投资金额为 4 万美金，整体相关 AI 数据中心投资规模约 114 亿美金。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表57： 训练资本开支测算

训练需求		公司数量	GPU 类型	GPU 数量	ASP (KUSD)	资本开支(mn USD)	单机柜投资金额 (KUSD)	数据中心资本开支 (mnUSD)	总资本开支 (mn USD)	总年运营成本（kUSD）	全年能耗（MWh）
CSP	非中国	15	H100	16000	25	6,000	40	1,200	7,200	288	1,471,680
	中国	15	A800	60000	10	9,000	40	4,500	13,500	1,080	3,153,600
AI 创业公司	非中国	150	H100	1600	25	6,000	40	1,200	7,200	288	1,471,680
	中国	150	A800	6000	10	9,000	40	4,500	13,500	1,080	3,153,600
合计						**30,000**		**11,400**	**41,400**	**2,736**	**9,250,560**

资料来源：英伟达官网，OpenAI，华泰研究预测

AI 算力缺口推动芯片快速迭代，高性能芯片功耗持续攀升。2020 年 5 月，英伟达发布了 A100 GPU，性能高达 19.5TFLOPS，相比普通 CPU 芯片算力提升多达 7 倍。后续 2022 年 4 月发布的 H100 性能在此基础上提升了 2.4 倍，达到 67TFLOPS。算力的提升伴随着功耗的增加，相比普通 CPU 芯片 85W 的 TDP，A100 功耗提升近 3 倍，达到 300W，两年后的 H100 在此基础上又翻倍达到 700W。

推理算力：需求和应用需求成长成正比，供应商呈现百花齐放态势

推理需求和用户数增长呈正比。通常来说我们采用“云端训练，端侧推理”模式，即将数据传到云侧进行训练模型，训练好的模型随后下放到端侧，端侧利用训练好的模型直接进行推理。在推理芯片的选用方面，相较于训练更多关注模型大小而言，推理更依赖于任务本质，以此决定所需芯片种类。当需要大量内容/图像 AI 生成式时，需要 GPU 进行推理计算（如英伟达主流的 T4 芯片）；而对于较简单的推理过程（语音识别等），CPU 有时也会成为较好的推理引擎（相关公司如英特尔），甚至在特定的场景 FPGA、ASIC 也能成为很好的推理芯片（相关公司如 Xilinx、Altera、Lattice 等），总之相比训练芯片推理芯片厂商所采用的路径更多元化。

从此前 IDC 等机构发布的统计结果来看，计算芯片中训练：推理需求的配比一般在 3：7 到 4：6 之间。由此，我们可假设在稳态下推理需求是训练需求的 2 倍，有望带来约 600 亿美金的投资。总体来看，我们预计 2025 年左右当生成式 AI 应用大规模落地后，数据中心用推理及训练芯片市场约 900 亿美元。

此外，能耗方面，据我们测算，AI 大模型全年训练阶段和推理阶段能耗分别将占全球数据中心耗电量 1.61%/0.77%。我们假设全球有 30 家科技巨头和 300 家 AI 初创企业参与 GPT-4 及类似的多模态大模型的训练和推理，并对其耗电量进行了测算。以同类型 GPU 芯片最高能耗配置为准，假设 GPU 芯片全年每天 24 小时无休运行，则估计全球 GPT-4 级别 AI 大模型训练阶段和推理阶段耗电量分别为 9251/4415GWH，总计约占全球数据中心耗电量的 2.38%。若全球科技巨头以此速度连续 5 年进行资本开支投入，则 5 年后 AI 大模型训练及推理阶段所需耗电量将占到全球数中心总耗电量的 11.9%。

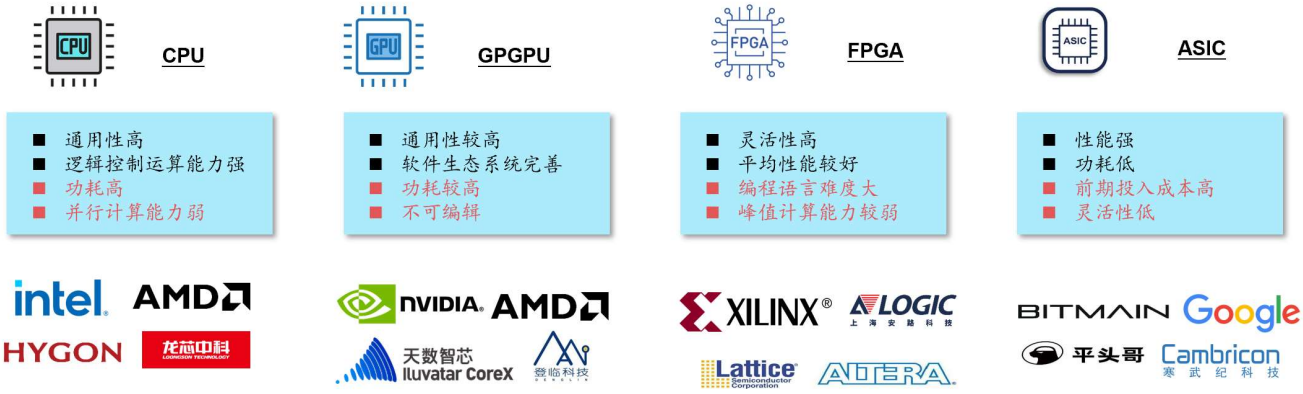
计算芯片：算力储备竞赛开启，推理初步具备进口替代能力

2022 年全球数据中心 CPU/GPU 市场规模约 249/166 亿美元

数据中心计算芯片主要包括 CPU 处理器和 GPU、FPGA、ASIC 协处理器。人工智能、5G、云计算等技术的发展推动全球数据量以指数形式快速增长，根据 IDC 预测，2025 年全球数据量将达到 175ZB，接近 2020 年数据量的 3 倍。面对数据中心海量数据处理工作，传统 CPU 计算结构越来越难满足需求，辅助 CPU 进行数据运算的协处理器应运而生。运算协处理器基于不同设计架构存在多条技术路线，包括通用 GPU（GPGPU）、FPGA、ASIC 等。GPGPU 凭借强通用性和完善软件生态系统成为该领域的主流解决方案。据 IDC 数据，2021 年我国 GPGPU 服务器销售额占据 AI 加速服务器市场的 88.4%，远超 ASIC、FPGA 等服务器。

免责
更多免费行业报告 并购家 www.ipoipo.cn

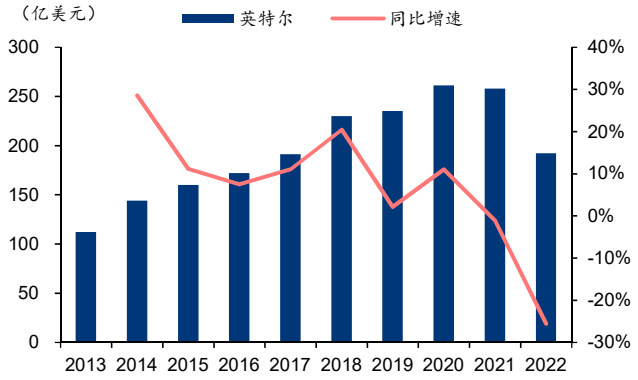
图表58：CPU 处理器与 GPGPU、ASIC、FPGA 协处理器



资料来源：各公司官网，华泰研究

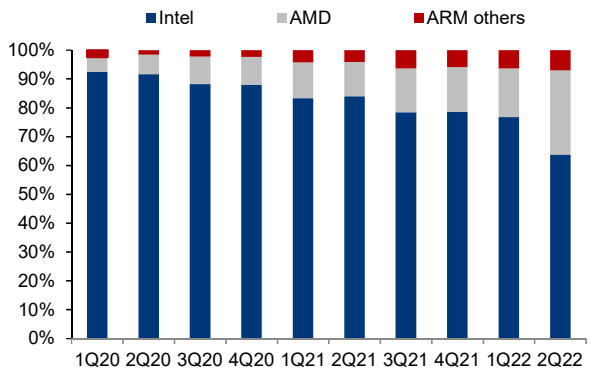
2022 年全球数据中心 CPU 市场规模约 249 亿美元，Intel、AMD 两家占据垄断地位。2022 年，英特尔数据中心事业部（DCG）收入 192 亿美元，AMD 数据中心业务收入为 60 亿美元，合计 252 亿美元。根据 DIGITIMES Research 数据，2022 年英特尔在服务器 CPU 市场出货量占比为 77.00%，据此我们测算全球数据中心 CPU 市场规模约 249 亿美元。从市场竞争格局来看，Intel 和 AMD 两家占据市场主要份额，其中英特尔凭借至强系列市场份额维持在 60%以上，AMD 市场份额持续提升。根据 IDC 数据，2Q22 英特尔和 AMD 服务器 CPU 全球出货量份额分别约为 63.7%和 29.2%，合计占据全球市场 89.9%的份额。

图表59：2013-2022 年英特尔数据中心收入



资料来源：，华泰研究

图表60：服务器 CPU 市场份额（按出货量）

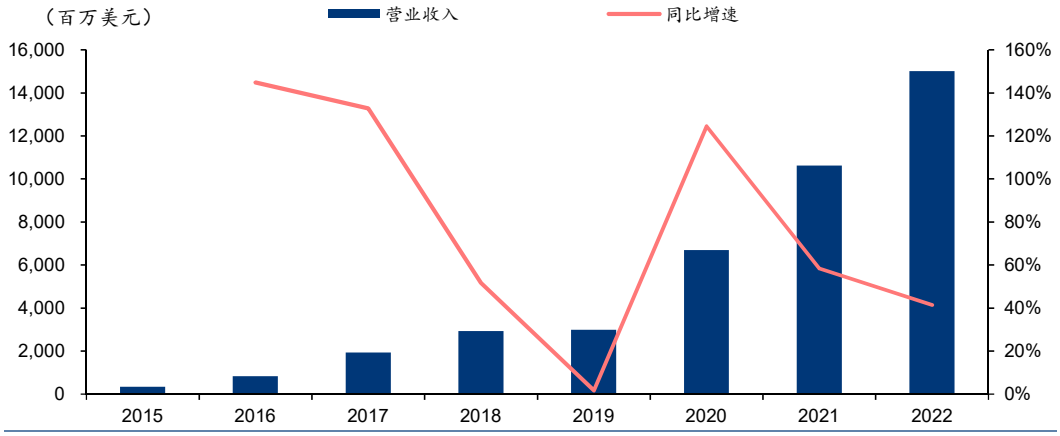


资料来源：IDC，华泰研究

2022 年全球数据中心 GPU 市场规模超 166 亿美元。NVIDIA 和 AMD 为 GPU 领域代表性厂商，二者占据市场绝大部分份额。2015 年 NVIDIA 数据中心业务收入仅为 3.4 亿美元，2022 年突破 150 亿美元，年复合增速达 71.86%，反映云端计算对 GPU 的旺盛需求。2022 年 AMD 数据中心业务收入为 60 亿美元，剔除服务器 CPU 收入约 39 亿美元（由 DIGITIMES Research 数据，2022 年 AMD 服务器 CPU 市占率 15.6%及全球服务器 CPU 市场规模 249 亿美元测算）和服务器 FPGA 收入约 5 亿美元（由 AMD 所披露的 Xilinx 营收 46 亿元及在数据中心占比约 11%测算），推算其数据中心 GPU 营收约为 16 亿美元。据此，我们测算 2022 年全球数据中心 GPU 市场规模超 166 亿美元。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表61：2015-2022 年 NVIDIA 数据中心业务收入

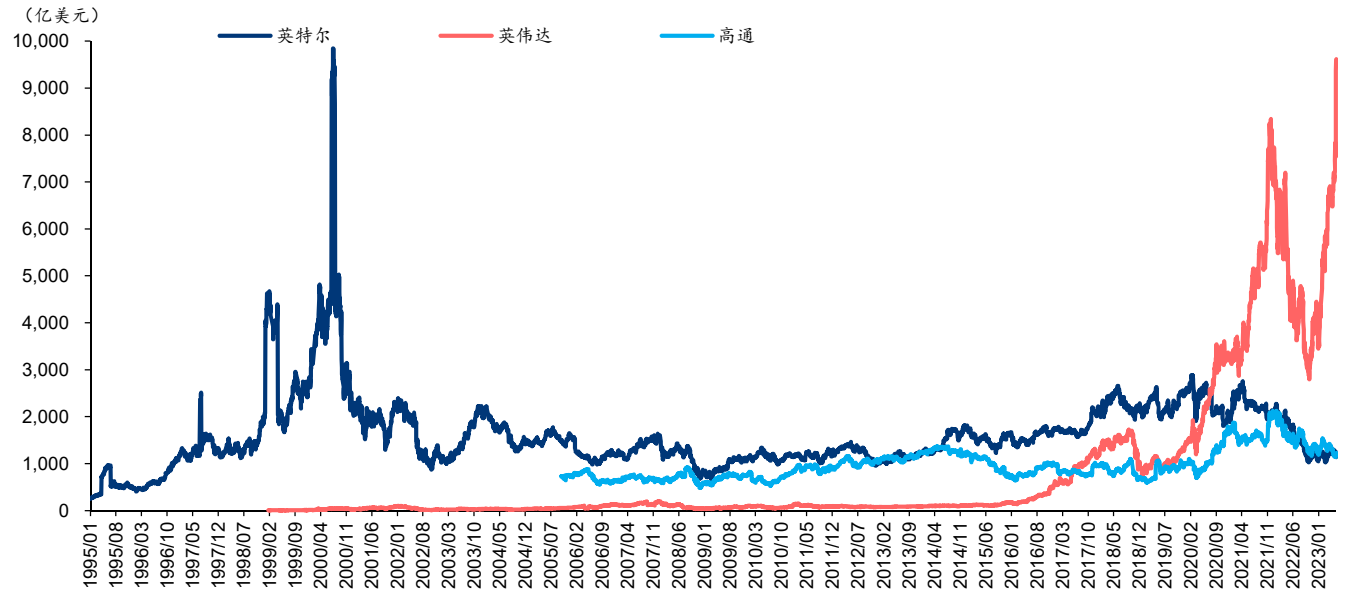


资料来源：Bloomberg，华泰研究

科技大厂陷入 FOMO(Fear of Missing Out)心理，算力储备竞赛开启

2023 年 5 月 24 日英伟达发布 1Q24 财报，受 AI 需求拉动业绩大超预期，带动市值提升至接近 1 万亿美金（市值约为 5 倍 AMD 和 8 倍英特尔），这是 2000 年 Intel 市值超过 1 万亿美金以来，芯片公司第一次提升至 1 万亿美金。业绩会上，英伟达 CEO 黄仁勋也表示全球价值 1 万亿美元的数据中心将从传统计算转向智能网卡、智能交换机和 GPU 等加速计算，且工作负载主要以生成式 AI 为主。我们认为，在大模型创新时代，当前主要科技企业陷入 FOMO(Fear of Missing Out)的心理状态，在 AIGC 领域积极部署，云厂商开启算力储备竞赛。

图表62：海外头部科技公司市值走势



资料来源：，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

推理初步具备进口替代能力，训练发展尚需时日

2022 年 8 月 31 日，英伟达发布公告称美国政府于 8 月 26 日对公司向中国（含中国香港）和俄罗斯出口高端 GPU 芯片加强限制。未来公司向中国客户出口 A100 和 H100 芯片需获得许可，该指令将适用于未来英伟达推出的峰值性能及接口性能大致相当/等于/优于 A100 的芯片及相关系统。尽管后续英伟达推出 A800 和 H800 作为替代 A100 和 H100 出口中国的产品，但在内存带宽上做了部分裁剪，在大量并行计算的情况下实际性能我们预估降低 30%-50%。美国对华 GPU 出口限制之后，又相继把海光信息、寒武纪等国内头部芯片公司列入实体清单，限制其芯片流片。我们认为美国对华出口限制，将致使中国先进工艺发展受限、国产替代方案下增加大模型训练成本/时间。但国内互联网、云计算大厂等下游客户出于供应链安全考虑加速芯片国产替代，或有助于国内计算芯片企业产品的商业化落地和软件生态打磨。

图表63：美国禁令限制中国半导体行业发展

资料来源：BIS，华泰研究

近年来，国产计算芯片取得长足发展，具体从 CPU、GPU、FPGA、ASIC 等领域来看：

1）CPU：国产 CPU 产业初具规模，现已形成以龙芯、飞腾、申威等为代表的具备竞争力的国产 CPU 力量。根据内核指令集路线，国产主流 CPU 厂商可分为三类：1）“指令集授权+自研”路线，代表厂商为龙芯和申威，分别基于 LoongArch 指令集和 Alpha 指令集，其他架构均自研。2）“指令集授权”路线，代表厂商为海思和飞腾，两者已获得 ARM V8 架构指令集永久授权，拥有较高自主发展权。3）“IP 内核授权”路线，代表厂商为获得 X86 内核层级授权的兆芯和海光，其自主可控程度最低，但生态更完善。

2）GPU：国产 GPU 厂商沿图像处理 GPU 和通用 GPU 两条路径切入，图像处理 GPU 侧重图形图像的渲染，GPGPU 则侧重处理以人工智能训练、推理为代表的通用计算任务。在人才、资本共同驱动下国产 GPU 产品落地加速，其中图像处理 GPU 代表厂商有景嘉微、芯动科技；通用 GPU 代表厂商包括壁仞科技、登临科技等；

3）FPGA、ASIC 等：依托互联网计算企业投资建设持续加码，国产 FPGA、ASIC 等 AI 加速芯片异军兴起。其中，昆仑芯、华夏芯等选择 FPGA 路线切入，由于 FPGA 产品兼顾灵活性与效率，其 CPU+FPGA 结构在应用和算法变化较多的场景优势明显；而寒武纪、燧原科技、平头哥等初创公司则选择 ASIC 技术路线，ASIC 类芯片专用性强，在特定场景中其针对部分算法的加速效果要显著优于 GPU，且能效比更高，其份额有望伴随生态环境进一步完善实现较大提升。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表64：国内计算芯片一览

公司	创立时间	团队背景	代表产品	类型	主要应用场景
海光信息	2014年	中科院计算所	海光3号，DCU8000系列	CPU/GPGPU	服务器，工作站
飞腾	2014年	中国电子	S2500，D2000	CPU	服务器、桌面等
龙芯中科	2001年	中科院计算所	龙芯3A5000	CPU	桌面、工控等
兆芯科技	2013年	上海市国资委	开胜 KH-30000	CPU/GPU	桌面
海思	2004年	华为	鲲鹏 920-7260	CPU/GPU	服务器等
申威	2016年	捷世智通科技	申威1621	CPU	超算
景嘉微	2006年	国防科技大学	JM5400、JM9231	GPU	军用市场、桌面办公等
芯动科技	2007年	Imagination、AMD	风华1号、风华2号	GPU	桌面PC等
芯瞳半导体	2021年	西邮	GenBu01	GPU	嵌入式设备、办公电脑等
瀚博半导体	2018年	AMD	SV100、SG100	GPU	云端推理
摩尔线程	2020年	英伟达	MTT S60	GPU	云端推理、边缘端
燧原科技	2018年	AMD	燧思1.0、燧思2.0	GPGPU	云端训练、云端推理
天数智芯	2015年	AMD	7nm GPGPU	GPGPU	云端训练、云端推理
登临科技	2017年	图芯,百度,英伟达,AMD,思科,博通	Goldwasser	GPGPU	云端训练、云端推理
墨芯人工智能	2018年	卡内基梅隆大学、Intel、Marvell和Oracle	Antoum 芯片	GPGPU	云端训练
壁仞科技	2019年	商汤科技、AMD	BP100	GPGPU	云端训练、云端推理
沐曦集成电路	2020年	AMD	MXN、MXC、MXG	GPGPU	数据中心、AI
寒武纪	2016年	中科院	思元590、290、370	ASIC	数据中心、安防
京微齐力	2017年	京微雅格、Skyworks	HME-H/P/R/M	FPGA	数据中心/电动汽车/5G
高云半导体	2014年	Lattice、Candance	小蜜蜂（GW1N）、晨熙（GW2A）、ASSP GoBridge	FPGA	通信、图像处理、工业控制、汽车电子、消费电子、人工智能
智多晶	2012年	Trident Microsystem、Broadcom、Lattice	Seagull 1000系列、Sealion2000系列、Seal 5000系列	FPGA	LED显示、工业控制、图像处理、仪器设备、高端医疗、基站通讯
云豹智能	2020年	博通、AMD、英特尔、海思	YB6480	DPU	云计算、数据中心
中科驭数	2018年	中科院	K1、K2	DPU	数据中心

资料来源：各公司官网，华泰研究

国产 AI 加速芯片在训练端和推理端进口产替代进程上有所差异：

1）中国 AI 训练芯片进展：仍与英伟达在性能和生态有一定差距。训练芯片需通过海量数据训练出复杂的神经网络模型，使其适应特定的功能，对性能和精度有较高的要求，并需具备一定的通用性。近年来，国内厂商训练芯片产品硬件性能不断取得突破，但与市场主流英伟达 A100 产品仍存在一定差距。寒武纪去年年底推出思元 590 系列在部分模型上由于其 ASIC 专用性表现出更优异的性能，理论性能已超过 A100，但仍需要后期适配和技术支持。此外，由于英伟达 CUDA 软件平台已经成为行业事实上的标准，国产 AI 训练芯片为兼容 CUDA 生态可能带来一定性能损耗。

图表65：国内主要云端训练芯片与英伟达主流训练芯片对比

主要云端训练芯片对比											
处理器名称	NVIDIA A100 80GB Pcie	NVIDIA H100 SXM	Google TPU 3.0	Habana（Intel）Gaudi	寒武纪思元290	寒武纪思元590	华为海思昇腾910	海光信息深算一号	摩尔线程MTT S300	燧原云遨T20	壁仞科技BR100
逻辑核心数	6912（CUDA cores）	14592（CUDA cores）	多核	多核	多核	多核	弹性多核	4096（64 CUs）	4096 MUSA	24 GCU-CARE	NA
训练计算性能	19.5 TFLOPS（FP32）	51 TFLOPS（FP32）	420 TFLOPS（FP32）	~60 TFLOPS（FP32）	256 TOPS（INT16）	500+TFLOPS（FP16）	320 TFLOPS（FP16）	NA	15.2 TFLOPS（FP32）	32 TFLOPS（FP32）	240 TFLOPS（FP32）
内存大小	40GB	80GB	128GB	32GB	32GB		NA	32GB	32GB	32GB	64GB
内存带宽	2039GB/S	3TB/S	NA	1TB/S	1228GB/S		NA	2014GB/S	448GB/S	1.6TBS	1.64TB/S
功耗	400W	700W	200W	140W	350W	400W	310W	350W	250W	300W	450-550W
能效比	0.05 TFLOPS/W	0.07TFLOPS/W	2.1 TFLOPS/W	~0.42TFLOPS/W	0.73 TFLOPS/W（INT16）		1.03 TFLOPS/W（FP16）	NA	0.06TFLOPS	0.11 TFLOPS/W	0.48 TFLOPS/W
工艺节点	7nm	5nm	NA	16nm	7nm	7nm	7nm	7nm	12nm	12nm	7nm

资料来源：各公司官网，华泰研究

2）中国 AI 推理芯片进展：推理芯片则是利用神经网络模型进行推理预测，对峰值计算性能要求较低，则更加注重单位能耗算力、时延、成本等综合指标，我国厂商多选择先从推理端切入。寒武纪、海思以及燧原等国内厂商产品已具备与市场主流的 Tesla T4 正面竞争能力。以寒武纪思元 370 为例，其 INT8 运算性能达到 256 TOPS，高于 T4 的 130 TOPS，且能效比和性价比均更优，已具备替代 T4 能力。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表66：国内主要推理芯片与英伟达主流推理芯片对比

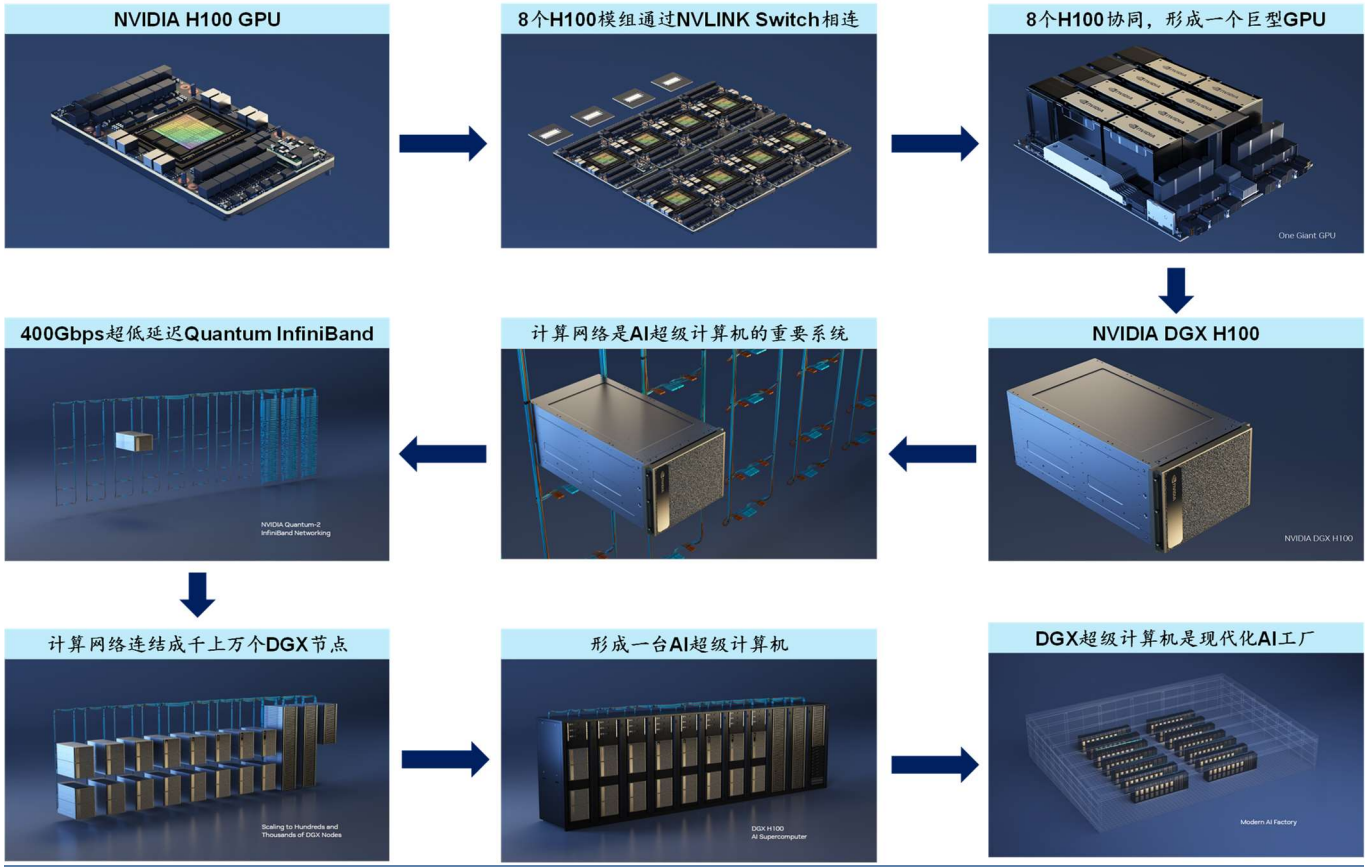
主要云端推理芯片对比							
处理器名称	NVIDIA Tesla T4	Google TPU 3.0	寒武纪 思元370	燧原 云燧i20	登临 Goldwasser L256	天数 智铠MR-V100	沐曦 MXN100
训练计算性能（TFLOPS）	8.1（FP32）	支持训练	支持训练	32（FP32）	64（FP16）		
推断计算性能（TOPS）	**130 INT8**	**支持推断**	**256 INT8**	**256 INT8**	**256 INT8**	**384 INT8**	**160 INT8**
功耗	70W	200W	75W	150W	45W	板级功耗 150W	
带宽	320 GB/S			819GB/S	-		
推断能效比	1.86 TOPS/W		3.41 TOPS/W	1.71 TOPS/W	5.69 TOPS/W	2.56 TOPS/W	
工艺节点	12nm		7nm	12nm	-		

资料来源：各公司官网，华泰研究

服务器：AI 服务器市场稳步增长，2021-2026E CAGR 预计达 17%

除了一系列 GPU 芯片以外，英伟达还进一步推出 NVIDIA DGX 超级计算机，成为大语言模型实现突破背后的引擎。下图展示了 DGX 超级计算机的典型结构。DGX 配有 8 个 H100 GPU 模组，H100 配有 Transformer 引擎，旨在支持类似 ChatGPT 的大模型。这 8 个 H100 模组通过 NVLINK Switch 彼此相连，以实现全面无阻塞通信。8 个 H100 协同工作，形成一个巨型 GPU。通过 400 Gbps 超低延迟的 NVIDIA Quantum InfiniBand 进行网络内计算，将成千上万个 DGX 节点连接成一台 AI 超级计算机，并不断扩展应用范围，成为全球客户构建 AI 基础设施的蓝图。

图表67：NVIDIA DGX 超级计算机结构



资料来源：GTC2023，华泰研究

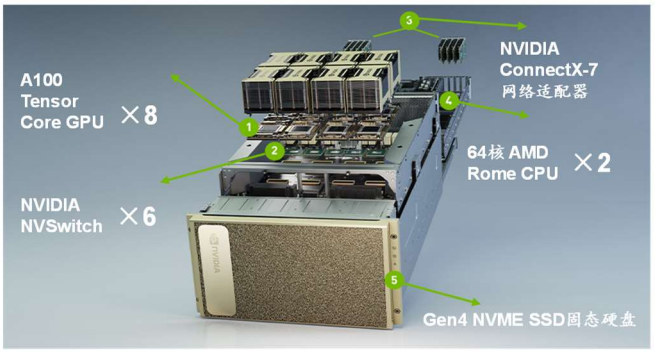
更多免费行业报告 并购家 www.ipoipo.cn

AI 大模型训练和推理催生 GPU 需求上涨。目前 AI 训练主要采用：1）英伟达在 AI 训练 GPGPU：包括 V100/A100/H100 以及美国限制出口后英伟达推出的裁剪版 A800。2）AMD 推出的 AI 训练芯片包括 MI 250/250X/300。AI 服务器采用异构式架构，NVIDIA DGX A100 服务器 8 个 GPU+2 个 CPU 的配置远高于普通服务器 1~2 个 CPU 的配置，能够满足 AI 大模型需求。IDC 预计，全球 AI 服务器市场规模稳健增长，2021-2026E CAGR 预计达 17%。我们认为，AI 大模型将带动算力需求增长，催生对 AI 服务器的需求。工业富联、联想、中兴通讯、紫光股份等公司有望把握住 AI 大模型带来的机会，受益于 AI 服务器需求的持续提升。

在推理芯片的选用方面，相较于训练更多关注模型大小而言，推理更依赖于任务本质，以此决定所需芯片种类。当需要大量内容/图像 AI 生成式时，需要 GPU 进行推理计算（如英伟达主流的 T4 芯片）；而对于较简单的推理过程（语音识别等），CPU 有时也会成为较好的推理引擎。

图表68：AI 服务器和普通服务器的区别

AI服务器：NVIDIA DGX A100



芯片	价格	数量
A100 Tensor Core GPU	USD14,999	8
64 core AMD Rome CPU	~USD7,000	2

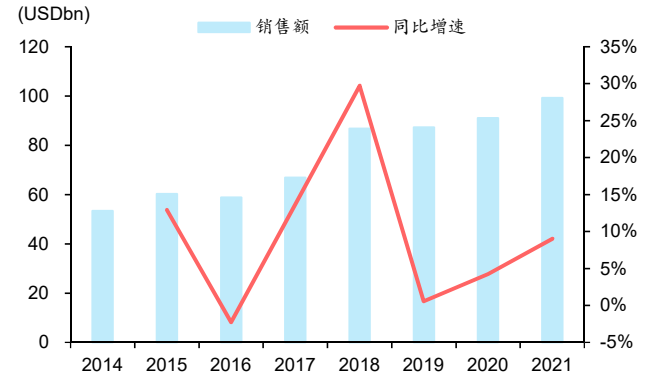
普通服务器：浪潮英信服务器NF5280M6

芯片	价格	数量
英特尔第三代Xeon处理器	RMB64,000	1~2

资料来源：英伟达官网，浪潮官网，华泰研究

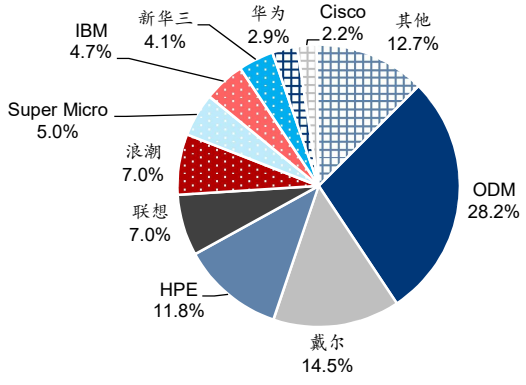
根据 IDC 数据，2014-2021 年，全球服务器厂商销售总额总体呈上升趋势，2021 年全球服务器市场规模达到 USD99bn。其中，戴尔、HPE、联想、浪潮市场份额排名前列，合计占比超过 40%。IDC 认为 2021-2026E 全球 AI 服务器厂商市场规模将稳健增长，2026 年市场规模有望达 USD35mn，预计 2021-2026E CAGR 达到 17%。1H21 全球 AI 服务器厂商竞争格局相对集中，其中浪潮信息占比最高（20.2%），其后是戴尔、HPE、联想。

图表69：2014-2021 年全球服务器厂商市场规模



资料来源：IDC，华泰研究

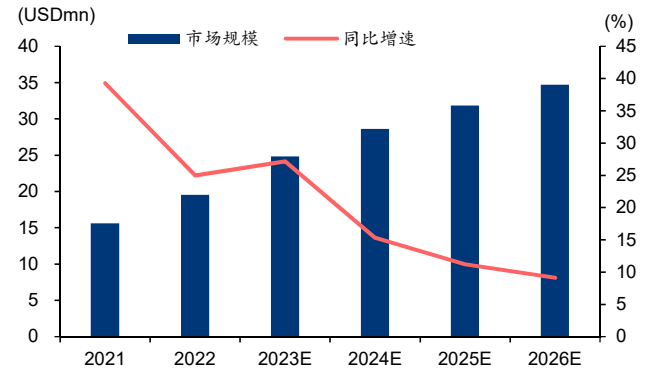
图表70：4Q22 全球服务器厂商竞争格局



资料来源：IDC，华泰研究

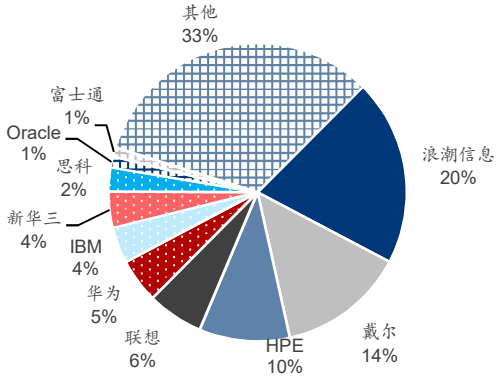
更多免费行业报告 并购家 www.ipoipo.cn

图表71：2021-2026E 全球 AI 服务器厂商市场规模



资料来源：IDC，华泰研究

图表72：1H21 全球 AI 服务器厂商竞争格局



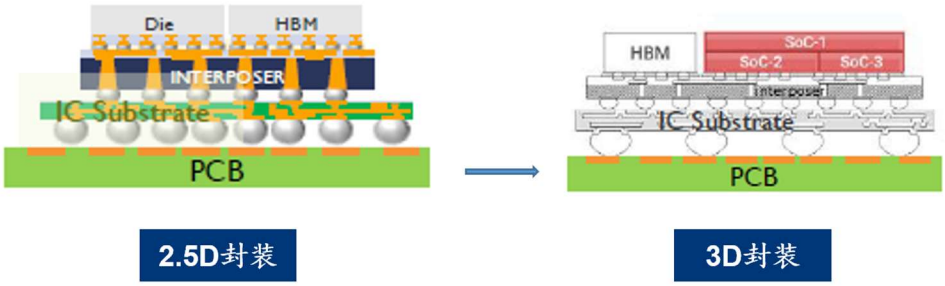
资料来源：IDC，华泰研究

先进封装&材料：高算力需求带动先进封装需求

高算力需求带动先进封装需求，ABF 载板有望充分受益。以 AI 大模型为基础的 AIGC 新应用快速兴起，对算力的要求提高，服务器单机将搭载更多的 CPU/GPU 等芯片，解决多芯片间高速互连的先进封装成为关键。

产业将从 2.5D 逐步走向 3D 封装，先进封装助力高性能计算芯片实现超越摩尔定律。高性能计算领域，先进封装是超越摩尔定律的一项重要路径，其能提供更高的系统集成度与芯片连接密度，芯片系统将持续实现性能提升，同时兼顾功耗和面积。目前行业正从芯片/Chiplet 在平面上通过中介层、硅桥、高密度 RDL 等方式连接的 2.5D 封装，逐步走向把存储、计算芯片在垂直维度进行堆叠的 3D 封装。我们看到 2.5D/3D 封装领域主要方案提供者为：台积电（3D Fabric 平台），英特尔（EMIB/Foveros）、三星（3D TSV/X-Cube）、日月光（VIPack）等。

图表73：先进封装：从 2.5D 走向 3D 封装

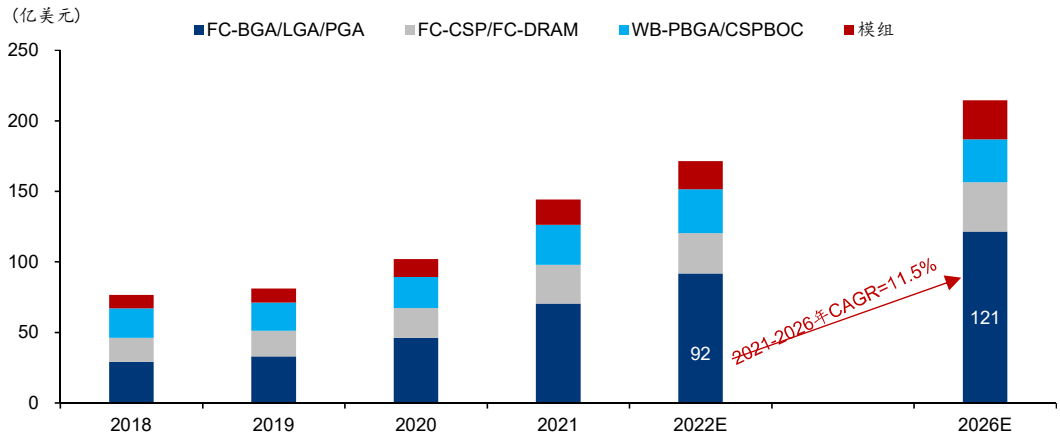


资料来源：ISSCC，华泰研究

而 IC 载板作为先进封装的关键材料，有望进一步打开价值空间。尤其 ABF 载板，相较于 BT 载板具备层数多、面积大、线路密度高、线宽线距小等特点，更能承载 AI 高性能运算，在 CPU、GPU、FPGA、ASIC 等高运算性能 IC 放量预期下将更充分受益。据 Prismark 预测，ABF 载板为 IC 载板行业规模最大、增速最快的细分领域，预计全球 ABF 载板市场规模 2026 年将达到 121 亿美元，2021-2026 年 CAGR 为 11.5%。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表74：全球 IC 载板行业细分领域市场规模



资料来源：Prismark（含预测），华泰研究

长电科技等国内公司已在量产 2.5D 封装领域实现初步突破，未来将逐步进入 3D 封装领域。2.5/3D 封装核心在于以微小线宽距和微小中心距的微凸点为特点的高密度中介层互联，包括 TSV（硅通孔）、RDL（重布线堆叠）、FO interposer（扇出型中介）、嵌入式芯片基板。我们看到国内公司正在相关技术方面取得初步突破。长电科技 2022 年已突破带 2.5D 硅通孔 MCM 的大尺寸 FCBGA 技术，并进入小量产。未来其有望在 2.5D 封装领域逐步成熟，并走向 3D 封装领域。通富微电多层堆叠 NAND Flash 及 LPDDR 封装实现稳定量产，并于 2022 年完成基于 TSV 技术的 3DS DRAM 封装开发。

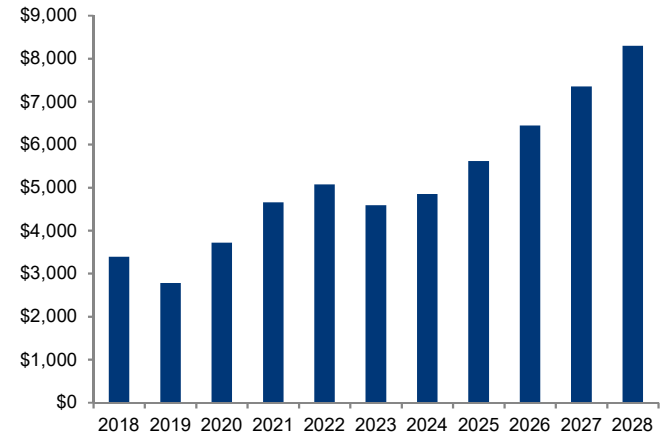
设备方面，封装设备国产替代空间广阔，测试设备长期受益于 Chiplet 等趋势。当前后道封装设备主要由 Besi、Disco、K&S 等公司占据，国产化率较低。我们看到光力科技子公司 ADT、LP 及 LPB 可提供涵盖传统封装及先进封装领域的划片机产品，大族激光可提供半导体激光切割设备。此外，我们看到 Chiplet 技术可将不同工艺的小芯片集成，提升芯片系统的良率及降低成本，正在成为行业趋势。我们认为未来成熟的 Chiplet 替代传统 SoC 方案，或将提升模拟、数模混合等测试机需求，华峰测控在积极布局。

光模块：数据中心市场空间广阔，2023 年有望迎 800G 放量元年

全球光模块市场稳健增长，数据中心光模块市场空间广阔。根据 Lightcounting 于 2023 年 3 月发布的数据，2019~2021 年，全球光模块市场规模从 89 亿美元增长到 114 亿美元，2022 年全球以太网光模块出货量达 6349 万只，同比增长 11%，出货金额达 50.72 亿美元，同比增长 9%；Lightcounting 预计 2028 年全球以太网光模块出货量合计将达到 8576 万只，出货金额 82.97 亿美元，对应 2023-2028 年复合增长率分别达 5%、9%。全球光模块市场结构方面，CWDM/DWDM 光模块和用于数据中心的以太网光模块占比较高，二者出货金额合计占比从 2019 年的 74%提升至 2021 年的 79%。

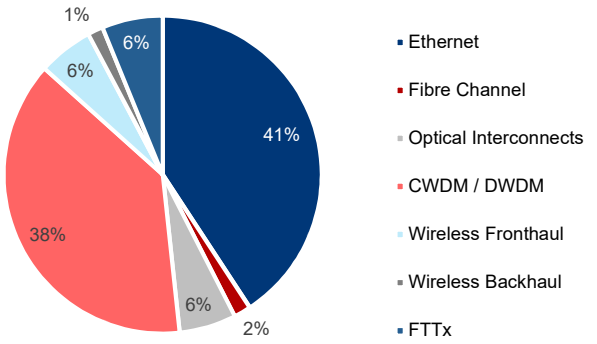
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表75：全球光模块市场规模及预测（亿美元）



资料来源：LightCounting，华泰研究

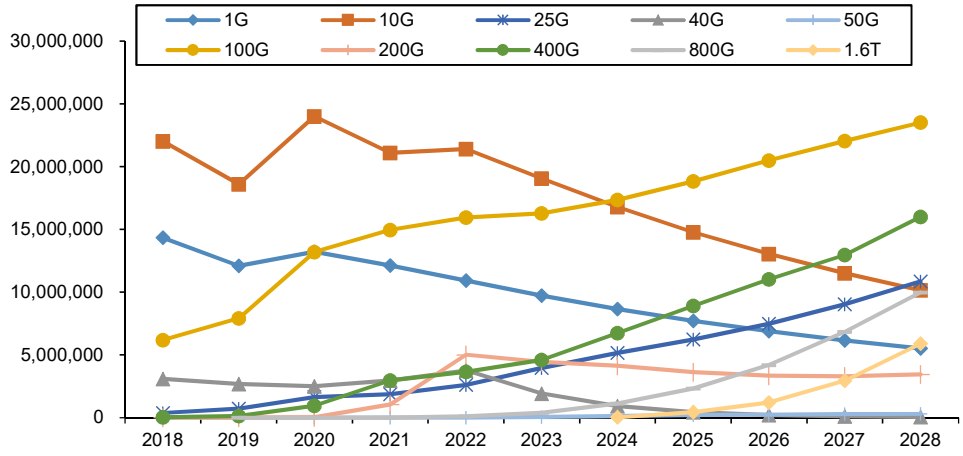
图表76：2021 年全球各品类光模块出货金额占比



资料来源：LightCounting，华泰研究

800G 有望迎接放量元年。随着云计算、AI、元宇宙等技术产业的快速发展和传统产业数字化的转型，全球范围内数据快速的增长对网络带宽提出新的要求，有望催生光模块速率升级需求。LightCounting 指出，AI 应用等带来的数据流量的增长，超预期的数据中心带宽需求以及光模块厂商技术的迭代都是推动 800G 应用的动力。我们认为随着 800G 光模块有望迎来放量元年，国内头部光模块以及上游光器件、光芯片厂商有望迎发展机遇。根据 Lightcounting 预测，2028 年 800G 出货量有望达 998 万只，对应 2023-2028 年复合增长率为 72%。

图表77：各速率全球以太网光模块出货量及预测



资料来源：LightCounting，华泰研究

国内光模块厂商彰显全球竞争力。根据 Lightcounting 于 2022 年 5 月发布的统计数据，2021 年全球前十大光模块厂商，中国厂商占据六席，分别为旭创（与 II-VI 并列第一）、华为海思（第三）、海信宽带（第五）、光迅科技（第六）、华工正源（第八）及新易盛（第九）；相比于 2010 年全球前十大厂商主要为海外厂商，国内仅 WTD（武汉电信器件有限公司，2012 年与光迅科技合并）一家公司入围，体现出十年以来国产光模块厂商竞争实力及市场地位的快速提升。

免责
更多免费行业报告 并购家 www.ipoipo.cn

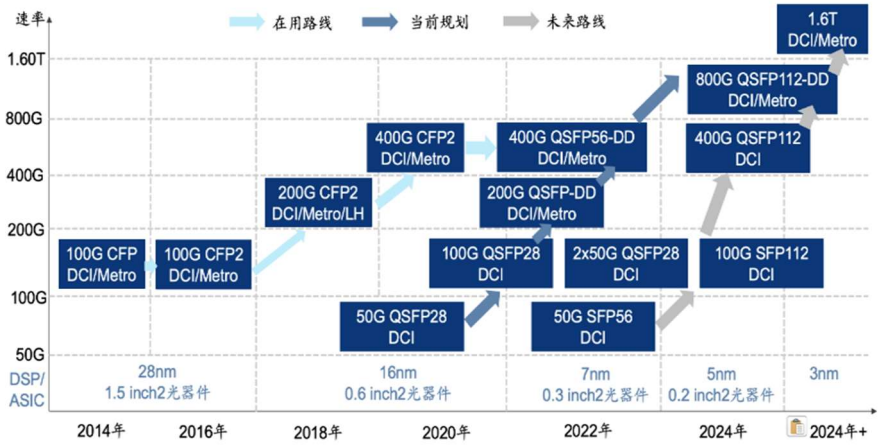
图表78：2021 年全球 Top10 通信光模块国产厂商占六席

厂商	2010	2016	2018	2021
1	Finisar	Finisar	Finisar	II-VI &旭创科技
2	Opnext	海信宽带	旭创科技	
3	Sumitomo	光迅科技	海信宽带	华为海思
4	Avago	Acacia	光迅科技	Cisco(Acacia)
5	Source Photonics	FOIT（Avago）	FOIT（Avago）	海信宽带
6	Fujitsu	Cclaro	Lumentum/Oclaro	光迅科技
7	JDSU	旭创科技	Acacia	Broadcom(Avago)
8	Emcore	Sumitomo	Intel	华工正源
9	武汉电信器件	Lumentum	AOI	新易盛
10	Neophotonics	Source Photonics	Sumitomo	Molex

资料来源：LightCounting，华泰研究

数据中心网络架构的演进带动高速光模块用量增长。随着数据流量与算力的提升，数据中心网络架构与容量都面临更为严苛的挑战，多平面 CLOS 网络架构正逐步发展成为数据中心网络架构的主流。根据张璋《数据中心演进发展对光通信器件模块的影响》的数据，从 10G 网络架构向多平面网络架构演进时，盒式交换机用量提升，带动光模块应用数量迅速增长，以收敛比 1:3 计算容量，多平面 CLOS 架构光模块应用相较于传统 10G 网络架构，光模块应用数量快速增长。

图表79：光模块的发展趋势



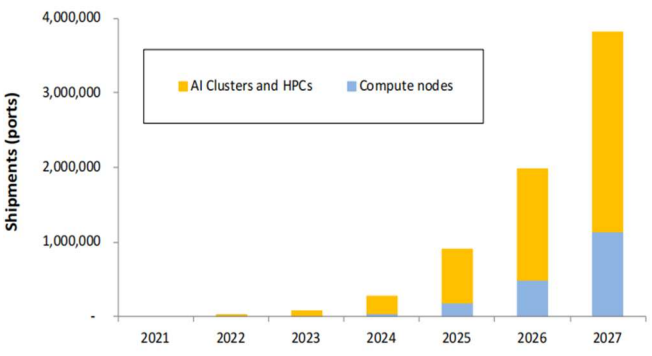
资料来源：ITPUB，华泰研究

传统可插拔光模块技术演进难以支撑高算力背景下数据中心的可持续发展。根据 Cisco 数据，在高算力背景下交换机交换芯片、光模块功耗增长迅速，传统可插拔光模块技术演进难以支撑数据中心的可持续发展，主要由于：1）SI（电传输）的实现遭遇材料瓶颈，PCB 传输高速电信号，在应用传统可插拔光模块时，信号传输距离长、传输损失大。更低损耗的可量产 PCB 材料面临多重技术困难；2）功耗问题，满载 1.6T 模块设备功耗巨大，散热设计难度大，3）成本问题，整机成本相应提升，会加大网络建设的初期投入；4）端口设计密度无法持续提升，传统可插拔光模块系统在支持 128 端口时需要非常复杂的系统设计，模块散热难度较大。

CPO 技术在 AI 集群和 HPC 中渗透率将逐步提升，预计在 2027 年部署的 800G 和 1.6T 端口中，CPO 端口将占到近 30%。根据 LightCounting 预测，未来 5 年时间内，AI 集群与 HPC 将会是 CPO 技术主要的两大切入点，其背后的驱动力也是高算力、高带宽的迫切需求。LightCounting 预计可插拔设备将在未来 5 年甚至更长时间内继续主导市场，然而在 2027 年部署的 800G 和 1.6T 端口中，CPO 端口将占到近 30%。Communications Industry Researchers (CIR)指出，2025 年全球共封装光学元件(CPO)的销售额将超过 13 亿美元，并预测至 2028 年全球共封装光学器件的销售额将会翻一番，达到 27 亿美元。

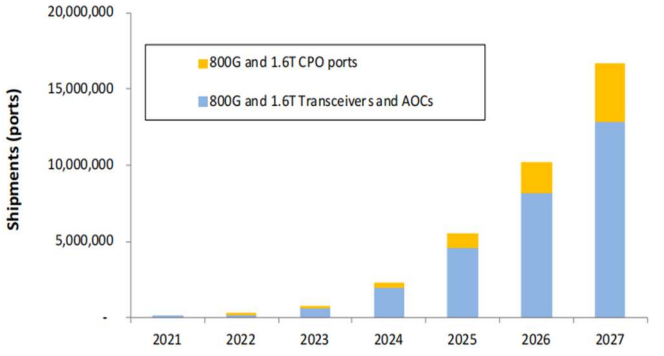
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表80：CPO 技术应用领域端口数量对比



资料来源：LightCounting，华泰研究

图表81：2021-2027 年 800G 与 1.6T 端口技术类型



资料来源：LightCounting，华泰研究

服务器产业链相关公司：工业富联、环旭电子

工业富联：服务器代工龙头厂商，深度受益 AI 服务器市场扩容带来的机会

工业富联作为提供服务器代工服务的龙头厂商，将深度受益于 AI 服务器市场扩容带来的机会，有望凭借高性能的 AI 服务器迎来营收增长。客户方面，相较于 2021 年，2022 年工业富联服务器营收中云服务商比重由 35%上升至 42%，来自云服务商客户的营收占比持续提升；产品方面，伴随着 AI 硬件市场迅速成长，公司 AI 服务器及 HPC 出货增长迅速，在 2022 年云服务商产品中占比增至约 20%。算力时代的开启为高效 AI 服务器提供了更广阔的发展空间，据公司 2022 年报，相关新产品将在 2023 年陆续研发推出。2022 年，工业富联云服务设备占总营收比例超过 40%。

环旭电子：服务器和高速交换机主机板供应商，协助客户推出边缘 AI 计算服务器

环旭为客户提供 ODM/JDM/EMS 伺服器、存储、NAS 和 SSD 产品以及制造服务，并提供 L10 系统设计服务，包括主机板、固件 BIOS 和 BMC、子卡(背板、附加卡等)、外壳和散热设计以及系统集成。AI 服务器方面，环旭提供设计制造(JDM)服务，协助品牌客户推出边缘 AI 计算服务器。根据环旭电子 2022 年报，云端及存储类业务营收达 69.89 亿元，同比增长 45.8%，占公司营收比例由 2021 年的 8.7%增长至 2022 年的 10.2%。我们认为，受益于 AI 带来的云端及边缘计算新需求，公司云端及存储类产品营收有望实现长期稳健增长。

计算芯片产业链相关公司：燧原科技、壁仞科技、寒武纪、京微齐力、芯动科技、沐曦集成电路、海光信息、登临科技、中科驭数、摩尔线程、天数智芯、高云半导体、墨芯人工智能、智多晶、云天畅想

燧原科技（未上市）：专注于人工智能云端领域，训练、推理产品全覆盖

燧原科技成立于 2018 年 3 月，公司专注于人工智能云端领域，致力于成为中国领先的数据中心基础设施提供商。公司创始人兼 CEO 赵立东曾历任 AMD 产品工程部高级总监以及紫光集团副总裁，有近二十多年的 CPU/GPU/APU 研发经验；创始人兼 COO 张亚林在 AMD 亦有长达十年的任职经历，期间曾领导开发全球最大融合芯片 APU，产业经验丰富。2021 年公司接连发布第二代训练产品云燧 T20/T21 和第二代推理产品云燧 i20，其中云燧 i20 单精度 FP32 峰值算力达到 32TFLOPS，整型 INT8 峰值算力达到 256TOPS，分别较第一代推理产品提升 1.8/3.6 倍，存储带宽高达 819GB/s，领先于行业同类产品水平。公司成立至今已完成多轮融资，累计融资超 30 亿元，其中 2022 年 8 月完成 C+轮融资，大基金二期入股。腾讯作为公司战略投资方，为其提供广泛应用场景，双方针对 AI 的多种应用场景进行深度合作，有利推动公司产品研发突破进程。

免责
更多免费行业报告 并购家 www.ipoipo.cn

壁仞科技（未上市）：致力于打造通用计算体系和软硬件平台，多次刷新融资纪录

壁仞科技创立于 2019 年，公司致力于开发原创性的通用计算体系，建立高效的软硬件平台，同时在智能计算领域提供一体化的解决方案。发展路径上，公司首先聚焦云端通用智能计算，逐步在人工智能训练和推理、图形渲染等多个领域赶超现有解决方案，实现国产高端通用智能计算芯片的突破。2022 年 8 月 9 日，壁仞科技在上海发布首款通用 GPU 芯片 BR100，16 位浮点算力达到 1000T 以上、8 位定点算力达到 2000T 以上，单芯片峰值算力达到 PFLOPS 级别，标志着中国的通用 GPU 芯片迈上“每秒千万亿次计算”新台阶。成立以来，公司在短时间内完成了由顶级投资机构启明创投、IDG 资本、华登国际中国基金、高瓴创投等领投的多轮融资。公司于 2020 年 6 月完成 A 轮融资，总额为 11 亿元，创近年芯片设计领域新纪录。2021 年 3 月，公司完成 B 轮融资，累计融资额超 47 亿元，继续刷新芯片设计领域融资速度及融资规模纪录。

寒武纪：国内 AI 芯片先行者

寒武纪成立于 2016 年，专注于人工智能芯片产品的研发与技术创新，可为云端、边缘端、终端提供全品类系列化智能芯片和处理器产品。公司研发团队实力雄厚，创始人陈天石曾在中科院计算所担任研究员、博士生导师，在人工智能和处理器芯片等相关领域从事基础科研工作十余年，具备坚实的理论功底和丰富的研发经验。截至 2022 年底，公司研发人员数量达 1205 人，占公司总人数 79.49%。公司已推出思元 100、思元 220、思元 270、思元 290、思元 370 等多款处理器芯片，面向互联网、金融、交通、能源、电力和制造等领域的复杂 AI 应用场景提供算力支持。2018-2022 年，公司营收从 1.17 亿元增长至 7.29 亿元，年均复合增长率达到 57.99%。

京微齐力（未上市）：国产 FPGA 领军厂商，立足消费市场开拓高端

京微齐力成立于 2017 年，是国内最早进入自主研发、规模生产、批量销售通用 FPGA 芯片及新一代异构可编程计算芯片的企业之一。公司拥有超 200 件专利，具备独立完整的自主知识产权，涵盖 FPGA 内核设计、SOC 架构设计、芯片开发、EDA 软件开发、IP 开发与集成等全栈技术领域。公司制定了四大产品方向：面向数据中心/电动汽车/5G 的高端 FPGA 芯片；面向人工智能的 AiPGA (AI in FPGA) 芯片；面向新基建/工业物联网的异构计算 FPGA 和 HPA (Heterogeneous Programmable Accelerator) 芯片；面向嵌入式应用的 eFPGA (embedded FPGA) 核。目前，公司产品在 65/55/40nm 工艺节点上全面实现量产，22nm 产品从 2022 年开始规模量产，产品应用场景涵盖通信、工业、安防、电力、医疗、消费等领域。

芯动科技（未上市）：国内一站式 IP 和芯片定制领军企业

芯动科技是拥有 17 年历史的中国一站式 IP、芯片定制及 GPU 领军企业，聚焦计算、存储、连接等三大赛道，提供全球各大工艺厂从 55 纳米到 5 纳米 FinFET 全套高速 IP 和高性能定制芯片解决方案。公司 17 年以来已赋能全球数百家知名客户，授权逾 100 亿颗高端 SoC 芯片进入规模量产，拥有 100%成功率以及超过十亿颗 FinFET 定制芯片成功量产的骄人业绩。并先后推出了国内第一款 4K 级多路服务器 GPU、第一款 4K 级 4 屏桌面和嵌入式 GPU 并快速进入商用，性能强劲，跑分领先，功耗低，自带智能计算能力。2023 年 5 月 20 日，芯动科技在全国科技活动周暨北京科技周活动透露正在研发新一代的风华 3 号 GPU 将支持光线追踪等先进技术。公司历史客户群涵盖中兴通讯、瑞芯微、全志、君正等国内前十设计公司，以及高通、微软、AMD、亚马逊、安盛美等全球知名企业。

免责
更多免费行业报告 并购家 www.ipoipo.cn

沐曦集成电路（未上市）：国产高性能 GPU 芯片解决方案供应商

沐曦集成电路成立于 2020 年，致力于为异构计算提供全栈 GPU 芯片及解决方案。公司创始团队产业经历丰富，创始人陈维良曾任 AMD 全球 GPGPU 设计总负责人，两位 CTO 均为前 AMD 首席科学家，目前分别负责公司软硬件架构，核心成员平均拥有近 20 年高性能 GPU 研发经验。目前公司已推出 MXN 系列 GPU（曦思）用于 AI 推理，MXC 系列 GPU（曦云）用于 AI 训练及通用计算，以及 MXG 系列 GPU（曦彩）用于图形渲染，可满足数据中心对高能效和高通用性的算力需求。产业合作方面，公司已与服务器 OEM、大数据中心、互联网、运营商等行业客户建立紧密的合作关系，并与清华大学、浙江大学等众多知名高校和研究机构密切开展产学研合作，持续推进产业上下游生态的建设。

海光信息：国产 CPU 及 DCU 领军企业

海光信息成立于 2014 年，专注于高端处理器的研发与设计，是国内重要的高性能 CPU 及 DCU 产品供应商。公司拥有实力雄厚的研发团队，截止 2022 年底，研发技术人员达到 1283 人，其中硕士及以上学历人员占比达 75.84%，多数核心研发人员具有二十年以上高端处理器研发经验，为公司不断取得产品突破打下坚实基础。公司 CPU 产品包括 7000 系列、5000 系列和 3000 系列，分别面向高端、中低端、工作站和边缘计算服务器市场。公司 CPU 产品基于市场主流 X86 指令集，具有成熟丰富的应用生态，当前在浪潮、联想、新华三、同方等多家国内知名服务器厂商产品中得到广泛应用。此外，海光信息 DCU 系列产品深算一号已经实现商业化应用，深算二号正处于研发阶段。2022 年公司实现营业收入 51.25 亿元，同比增长 121.83%；归母净利润 8.04 亿元，同比增长 145.65%。

登临科技（未上市）：致力于开发兼容现有生态的创新通用人工智能芯片

登临科技成立于 2017 年，是一家专注于自主创新，构建基于 GPGPU 核心技术的云端 AI 计算平台公司。公司核心团队成员来自图芯、S3、英伟达、AMD、阿里等全球知名的芯片设计公司。公司产品采取兼容 CUDA 的技术路线，保证客户业务可继承现有生态上的投入，且相比传统 GPU 在 AI 计算上性能和能效均有明显提升，大大降低外部带宽的需求，显著降低客户成本。2020 年 6 月公司 Goldwasser 系列产品在台积电 12nm 工艺上 FullMask 流片成功。2021 年 7 月，百度飞桨适配登临 GPU+系列芯片，加速 AI 应用落地；同年 11 月公司 Goldwasser 系列产品已实现量产。作为国内首个实现规模化商业落地的 GPU 企业，登临首款基于 GPU+的创新 AI 加速器-Goldwasser（高凛）已规模化运用在边缘至云计算的各个应用场景，成功填补国内高性能 GPGPU 领域技术、产品及商业方面的空白。今年，登临的新一代 Goldwasser（高凛）产品针对基于 Transformer 的网络和生成式 AI 类大模型的应用在性能有大幅提升，对标国际大厂的产品有明显的能效比和性价比的优势。近日，登临完成新一轮融资，将进一步夯实其产品与研发能力。此前，登临科技更是获得了多家产业、知名头部企业及基金的投资。

中科驭数（未上市）：国内领先 DPU 芯片提供商

中科驭数成立于 2018 年，聚焦于专用数据处理器的研发与设计，主要产品包括 DPU 芯片、超低时延 DPU 网卡、数据查询加速 DPU 卡等产品和解决方案。公司为国内较早进行 DPU 芯片研发的团队，创始人兼 CEO 鄢贵海博士、联合创始人兼 CTO 卢文岩博士及首席科学家李晓维博士均来自中科院计算所，在芯片架构领域有 20 年的技术积累。目前公司已累计申请 157 项发明专利，其中 83 项已获得授权。中科驭数 DPU 芯片基于自主研发的 KPU 芯片架构，目前公司已经完成了四类 KPU 芯片架构设计，包括 1）KPU-Swift 针对网络协议处理设计；2）KPU-Conflux 针对时间序列/大数据分析设计；3）KPU-Trusy 针对安全领域处理设计，4）KPU-FlexFlow 针对智能计算设计，并在网络、数据库等应用领域积累了 80 余类功能核。

免责

更多免费行业报告 并购家 www.ipoipo.cn

摩尔线程（未上市）：核心团队产业经验丰富，商业化应用持续加速

摩尔线程创立于 2020 年 10 月，专注于研发设计全功能 GPU 芯片及相关产品，支持 3D 图形渲染、AI 训练与推理加速、超高清视频编解码、物理仿真与科学计算等多种组合工作负载。创始人张建中为原英伟达全球副总裁、中国区总经理，公司团队其它成员来自微软、英特尔、AMD、ARM 等海外芯片巨头，具有深厚产业经验。2021 年 11 月，摩尔线程宣布首颗国产全功能 GPU 芯片研制成功，开创了国产 GPU 研发速度的先河。公司已发布两颗基于其 MUSA 统一系统架构打造的多功能 GPU 芯片“苏堤”和“春晓”，以及系列 GPU 软件栈与应用工具，并已迅速将多款 MTT S 系列显卡推向市场，覆盖桌面、边缘和数据中心等多个场景。2022 年公司联合数家 OEM 合作伙伴斩获大型央企和大型国有银行台式机采购大单，上万台搭载摩尔线程 MTT S 系列显卡的国产 PC 将广泛应用于我国电信、金融行业的数字办公业务场景，将显著提升国产 PC 的应用体验。2023 年 5 月，公司宣布了一系列新产品与技术更新，涵盖数字办公、娱乐与创作、AI 与云计算以及元宇宙等 GPU 重要应用场景，标志着摩尔线程为用户提供的高品质、易部署、创新性应用型解决方案取得重大进展。

天数智芯（未上市）：专注高端通用 GPU 设计，赋能千行百业实现算力升级

天数智芯于 2018 年成立并启动 7 纳米通用并行计算芯片（通用 GPU）设计，是中国领先的通用 GPU 高端芯片及超级算力系统提供商。公司研发实力雄厚，研发团队人数占比 80%以上，既包含一批行业经验超过 20 年的世界级技术专家，同时也包含一大批具备 5-15 年业界经验的高执行力技术菁英。2020 年 12 月，公司成功点亮国内第一款 7nm 云端训练通用 GPU 产品天垓 100，并于 2021 年 3 月正式对外发布，2021 年三季度规模量产，是国内第一家实现通用 GPU 量产的硬科技企业。2022 年 12 月，公司通用 GPU 推理产品智铠 100 发布，标志着天数智芯成为云边协同、训推组合的完整通用算力系统全方案提供商。2022 年 7 月，公司宣布完成超 10 亿元人民币的 C+及 C++轮融资，本轮融资将助力公司量产 AI 推理芯片智铠 100，开发第二三代 AI 训练芯片天垓 200 及 300。

高云半导体（未上市）：三大产品系列聚焦民用 FPGA 市场

高云半导体成立于 2014 年，是一家专业从事现场可编程逻辑器件（FPGA）研发与设计的国产集成电路设计公司，提供从芯片、EDA 开发软件、IP、开发板到整体系统解决方案的一站式服务。公司核心研发成员来自于 Lattice，Cadence 等全球知名公司，拥有国内极少数全面掌握 FPGA 设计与产业化核心技术的研发团队，高云半导体致力于 FPGA 芯片的正向开发，坚持自主创新及国产替代并行策略，目前已形成小蜜蜂、晨熙、Arora V 三大产品序列，迄今已发布百余款芯片，产品均拥有自主知识产权。

高云半导体于 2015 年一季度推出第一款 55nm 工艺 400 万门的中密度 FPGA 芯片和第一版 EDA 工具云源软件；2016 年第一季度顺利推出国内首颗 55nm 嵌入式 Flash SRAM 的非易失性 FPGA 芯片；2017 年产品进入市场应用，开始批量出货；2019 年，发布国内第一颗 FPGA 车规芯片，FPGA 芯片年出货量突破千万片级别；2022 年高云发布其最新工艺节点，即采用 22nm SRAM 工艺的晨熙家族第 5 代（Arora V）高性能 FPGA 产品，除了工艺的优势，还集成了自研的核心模块，包括高速 SerDes 模块，MIPI 硬核，为通讯及 AI 运算加速的全新架构的 DSP 模块等。

免责

更多免费行业报告 并购家 www.ipoipo.cn

墨芯人工智能（未上市）：稀疏计算推动大模型发展

墨芯成立于 2018 年，总部位于中国深圳，在上海、北京、硅谷设有办公室。拥有来自学界与业界的全球顶级创始团队，团队包括卡内基梅隆大学 AI 科学家、世界顶尖半导体公司(Intel、Marvell 和 Oracle 等) 核心高量产芯片研发团队的核心成员，深耕行业二十余年，具有数十款产品、超 50 亿主流芯片量产经验。墨芯人工智能是稀疏化计算引领者，致力于提供云端和终端 AI 计算平台和服务。在最近的 MLPerf 评测中，墨芯凭借软硬协同的稀疏计算技术在 ResNet50 模型上蝉联冠军。其 S40 计算卡以 127,375 FPS 获得单卡算力全球第一；S30 计算卡以 383,520 FPS 算力获整机 4 卡算力全球第一。墨芯 AI 计算卡系列是基于其稀疏计算 AI 芯片 12nm Antoum，性能超越了 4nm 工艺的 GPU，展现出稀疏计算的强大优势。稀疏计算能够为 AI 大模型带来数十倍的加速性能，有力推动 AI 在各行业、各领域的广泛应用。

智多晶微电子（未上市）：国产 FPGA 领军企业，从 LED 走向工控、通讯广阔市场

智多晶微电子成立于 2012 年，专注可编程逻辑电路器件技术的研发，并为系统制造商提供高集成度、高性价比的可编程逻辑器件、可编程逻辑器件 IP 核、相关软件设计工具以及系统解决方案。智多晶的创始人及核心团队由国际知名 FPGA 公司产品开发骨干以及学术界资深 FPGA 科研人员组成。依托多年来建立的业内领先 EDA 软件、FPGA 硬件设计能力，公司已经实现了 55nm-28nm 工艺 FPGA 的规模量产，并已开始 14nm 产品的研发工作，逻辑资源覆盖 5K、12K、25K、30K、100K，相关产品广泛应用于 LED 视频处理、工业控制、通讯、消费电子等领域，并针对性推出了内嵌 Flash、SDRAM 等集成化方案产品。主要产品包括 Seagull 1000 系列（0.162um 工艺）、 Sealion 2000 系列（55nm 工艺）、Seal 5000 系列（28nm 工艺）。此前，智多晶已完成多轮融资，过往投资方包括盛宇投资、超越摩尔资本、临芯投资、深创投、宏燚投资、华御投资、华天科技、龙鼎投资等。

云天畅想（未上市）： 先进算力服务提供商，从云游戏向 AI 计算拓展

云天畅想创立于 2016 年，专注于为数字世界提供端到端的先进算力基础设施及服务，应用于云游戏、AIGC、元宇宙、自动驾驶等实时交互、海量算力需求应用场景，下游客户包括运营商、云服务商、互联网公司、游戏开发商及政企单位等超百家企业机构。云天畅想是中国三大运营商 5G 边缘云解决方案的主要供应商，同时在图形图像计算领域拥有领先的市场地位，为多个爆款游戏提供算力解决方案和云服务；云天畅想也在不断拓展应用场景和海外市场，在虚拟现实、AI 推理和自动驾驶等领域提供算力解决方案，并于今年 6 月与沙特卓越控股公司签约了战略合作协议，推广海外业务落地。作为创新技术驱动的硬核科技公司，云天畅想搭建了全球化的创始团队，拥有来自 Intel、SIE、腾讯的创新基因，成员均为 GPU 芯片、服务器架构、3D 渲染、视频云计算、通讯网络等领域专家。公司员工近 400 人，其中 80%为技术研发人员，并在中国深圳、北京和日本东京设有研发中心。2023 年 5 月，云天畅想宣布获得数千万美元的 C+轮融资，巩固了行业领跑者的地位。

光通信产业链相关公司：长光华芯、源杰科技、云岭光电、苏州熹联、天孚通信、中际旭创

长光华芯：光耀中国芯，激光芯片国产化领军者

深耕高功率半导体激光芯片，纵向+横向打造激光产业“中国芯”。公司系国内少数具备高功率激光芯片量产能力的企业之一，实现了高功率激光芯片的国产化与进口替代。根据公司招股书测算，2020 年公司占国内高功率半导体激光芯片市场份额的 13.4%，位居国产厂商第一位，2021 年公司已实现 30W 单管芯片量产，技术实力全球领先。在产品布局上，一方面，公司通过纵向延伸打通芯片→器件→模组→半导体激光器产业链条，实现了在高功率单管系列、高功率巴条领域的全产业链布局；另一方面，公司基于高功率半导体激光芯片的技术优势以及设计和量产能力，切入 VCSEL 及光通信芯片赛道，已建立了 VCSEL 产品包含外延生长、条形刻蚀、端面镀膜、划片裂片、特性测试、封装筛选和芯片老化的完整工艺线。

更多免费行业报告 并购家 www.ipoipo.cn

源杰科技：国内领先的高速率激光芯片厂商，IDM 模式铸就核心竞争力

优质高速率半导体光芯片供应商，进入国内外知名运营商网络供应体系。公司聚焦于光芯片行业，主要产品包括 2.5G、10G、25G 及更高速率激光器芯片系列产品等。经过多年研发和产业化积累，公司已拥有多条覆盖 MOCVD 外延生长、光栅工艺、光波导制作等全流程自主可控的生产线，实现向客户 A1、海信宽带、中际旭创、博创科技、铭普光磁等国际前十大以及国内主流光模块厂商批量供货，产品最终应用于中国联通、中国电信、AT&T 等国内外知名运营商网络中，已成为国内领先的光芯片供应商。由于下游客户在选择光芯片产品时需要经过较长的验证过程，公司率先进入供应商体系，建立了较高的客户资源壁垒。

低速率激光芯片市场实现差异化，并率先攻克高速率激光芯片市场。国内光芯片市场中，2.5G、10G 激光器芯片市场国产化程度较高，但不同波段应用场景不同，工艺难度差异大，公司凭借长期技术积累为光模块厂商提供全波段、多品类的产品，同时提供更低成本的集成方案，实现差异化竞争，据 CIC，公司在 2.5G 及以下光芯片市场中发货量占比为 7%；25G 及更高速率激光器芯片国产化率低，公司凭借核心技术和 IDM 模式，率先攻克技术难关，打破国外垄断，并实现 25G 激光器芯片系列产品大批量供货。

云岭光电：致力于成为国际一流的光芯片供应商

专注中高端光芯片市场，拥有自主知识产权。武汉云岭光电有限公司由国际领先的芯片专家团队与华工科技于 2018 年 1 月共同发起设立，承载近 20 年的技术探索与市场实践，专注于中高端光通信半导体光芯片产品，是拥有完全自主知识产权，具备全流程生产能力的 IDM 光芯片企业。

云岭光电主营业务为 2.5G/10G/25G 全系列激光器（LD）和探测器（PD）光芯片及封装类产品。公司建有光芯片研制平台，拥有 MOCVD 材料生长，器件工艺制备，和后端测试封装的完整生产线，从海外引入具有国外光芯片大厂多年工作经验的核心技术团队，掌握 10G/25G 高速率光收发芯片设计、高质量材料生长、关键工艺制备、和封装测试等核心技术。公司具备年产芯片 7500 万颗、TO 7200 万只的生产能力，致力于成为世界一流的光芯片企业，为全球光通信企业提供优质全系列光芯片。

苏州熹联：以硅光科技构建数字化未来

高效能光通信解决方案提供商。苏州熹联光芯微电子科技有限公司成立于 2020 年 7 月 20 日，由半导体、硅光及金融等领域多位资深专家领头，致力于打造硅光领先技术平台，努力推动全球 5G、数据中心及数字化进程。为了快速实施在硅光领域的战略布局，熹联光芯于 2021 年 10 月份完成了对德国 Sicoya GmbH 的 100%股权并购。公司以苏州为总部，以上海和柏林研发中心为双引擎，结合国内制造基地强大的量产实力，将快速实现推动硅光科技的发展。

全球硅光科技引领者，占据高端。公司在硅光领域有 10 多年的技术积累和储备，拥有硅光领域完整的自有设计器件 IP 组合，同时拥有光电一体全集成化的硅光芯片技术。公司拥有完善的专利体系，核心技术能够涵盖多种产品和应用领域，在无线通信、数据通信（光引擎&光模块）、服务器网络、智能驾驶、激光雷达、生物传感等多领域拥有技术储备和商业合作。目前在全球范围拥有 50 多项硅光授权发明专利，同时还有 50 多项正在申请中。公司目前拥有多家全球行业一流的客户，均在稳定的批量交货和商业技术合作中。

免责

更多免费行业报告 并购家 www.ipoipo.cn

天孚通信：光器件整体解决方案引领者

致力成为光器件领域引领者，打造全球光网络优质链接。苏州天孚光通信股份有限公司（简称 TFC）2005 年成立，是业界领先的光器件整体解决方案提供商。公司于 2015 年登陆中国创业板，目前已成为光器件细分行业龙头企业。产品广泛应用于光纤通信及光学传感，智能汽车，生命科学等领域。天孚通信致力于成为引领光器件领域发展的国际一流企业，为全球光网络畅通提供优质连接。公司上市以来总营收实现稳定增长，光有源器件板块营收总体高增长。公司上市后加码研发投入力度，不断推出新的产品，拉动营收增长。2020 年公司在光引擎研发方面取得重大进展，光有源板块同比增长迅速，推动公司整体营收与净利润快速提升。

中际旭创：全球光模块行业龙头，800G 产品有望快速上量

全球光模块龙头，AI 带动公司 800G 产品快速上量。公司集高端光通信收发模块的研发、设计、封装、测试和销售于一体，在云数据中心领域，为客户提供 100G、200G、400G 和 800G 的高速光模块，在电信领域，为客户提供 5G 前传、中传和回传光模块以及应用于骨干网和核心网传输光模块等高端整体解决方案。据 Lightcounting 统计的全球光模块市场份额排名，公司与 II-VI 市场份额并列全球第一位，体现出公司卓越的技术研发能力、成本及交付能力优势。2023 年以来，以 ChatGPT 为代表的 AIGC 应用驱动数据流量持续增长，加速光模块向 800G 及以上产品迭代，据 Lightcounting，800G 光模块产品有望拉动新一轮行业增长，公司作为 800G 光模块的主力供应商有望深度受益。

云计算产业链相关公司：道客云、鹏云网络、StreamNative

道客云：国内领先容器云平台和企业级解决方案提供商

国内领先的企业级 PaaS 解决方案提供商。DaoCloud 道客科技（未上市）成立于 2014 年底，是一家容器云平台和企业级解决方案提供商，提供包括持续集成、持续交付、发布管理、镜像托管等在内的基于容器技术的全套开发运维生命周期服务，同时与国内外 IaaS 云计算平台对接，实现云端应用从开发、测试、部署到运维的完整容器化工具链。

高兼容技术适配不同业务类型，深耕垂直行业打造标杆项目。公司产品技术实现跨云和大规模部署，同时兼容适配各种开发方向，能够适应各行业企业的架构特点。成立迄今，公司已在金融科技、先进制造、智能汽车、零售网点、城市大脑等多个领域深耕，标杆客户包括交通银行、浦发银行、上汽集团、东风汽车、海尔集团、金拱门（麦当劳）。

鹏云网络：深耕云原生存储，技术能力领先

十年深耕软件定义存储，国内率先大规模商用。鹏云网络（未上市）创立于 2012 年，创始团队集结了来自亚马逊、EMC 等全球知名企业专家骨干，并率先在国内开展大规模商用软件定义存储产品。历经近十年发展，鹏云网络凭借完全自主知识产权，成为行业领先的软件定义分布式智能存储软件与服务提供商。鹏云网络云原生存储平台 ZettaStor HASP，通过自研的高性能用户态文件系统、跨异构存储多副本冗余保护、存储资源统一灵活编排、与容器平台紧密集成的四大核心技术特性，满足了容器化应用对存储系统性能及可靠性的多样化需求。

目前鹏云网络服务对象已经广泛覆盖电信、金融、能源、政府、军队、公安、制造业、医疗、教育、科研等 20 多个行业，主要客户包括华泰证券、辽宁移动、福建联通、中国电子 32 所、航天科技、CEIEC、龙江环保等，客户服务网络遍布全国 31 个省市及自治区。

免责
更多免费行业报告 并购家 www.ipoipo.cn

StreamNative：云原生批流融合数据平台

围绕 Apache Pulsar，StreamNative 打造下一代批流融合数据平台。StreamNative（未上市），成立于 2019 年，由 Apache Pulsar 与 Apache BookKeeper 创始团队成员组建，围绕 Apache Pulsar 打造开源云原生批流融合数据平台，助力企业挖掘实时数据价值。公司为 Apache Pulsar 的商业化公司，也是全球唯一的 Apache Pulsar 技术原厂，不断推动 Apache Pulsar 和 BookKeeper 社区与生态建设。

作为大数据平台基础设施组件，流数据平台的重要性日益提升。Apache Pulsar 是下一代云原生分布式消息流平台，集消息、存储、轻量化函数式计算为一体，采用计算与存储分离架构设计，支持混合公有云部署，可覆盖 AWS、Microsoft Azure、Google 云、阿里云等全球范围的主流云计算平台，是云原生时代解决实时消息流数据传输、存储和计算的最佳解决方案。目前项目和社区正在快速发展，已有众多国内外大型互联网和传统行业公司采用 Pulsar，使用案例分布在人工智能、金融、电信运营商、直播与短视频、物联网、零售与电子商务、在线教育等多个行业。随着未来数据基础设施的普及，流数据的接收、存储和计算会成为企业的共性需求，流数据平台作为基础软件的重要地位日益提升。

免责

更多免费行业报告 并购家 www.ipoipo.cn

AI 赋能百业：产业变革的奇点

自 OpenAI 发布 ChatGPT 掀起 AI 浪潮以来，国内外科技巨头陆续在 AI 大模型领域加快布局，人工智能也因此催生出许多创新性领域以及传统领域多个环节的变革，有望实现下游应用百花齐放。其中，**企业服务领域**，AI 赋能办公行业生产力、沟通和协作工具，持续带动生产效率及沟通协作效率提升；**汽车领域**，AI 模型的引入增强了智能驾驶在感知、预测和决策方面的能力，加速无人驾驶的应用与落地；**工业领域**，AI 加码助力工业研发、生产、管理及服务等全环节，极大解放生产力，加速工业智能化水平；**金融领域**，AI 在银行、保险、资管、投研、投顾等多个细分领域正在落地或拥有潜在落地场景，帮助金融机构降本增效。我们认为，伴随技术和各大模型的升级迭代，AI 将赋能百业，催生出更多环节的创新型变革。

AI+企业服务：赋能生产力/沟通/协作工具

数字化时代，办公行业生产力、沟通、协作工具不断演进，持续带动生产效率及沟通协作效率提升。以 Microsoft 365 Copilot 为例，Copilot 通过把 GPT-4 提供的内容生成功能，与存储在 Graph 数据库中的企业数据，以及 Word、Powerpoint、Excel、Outlook、Teams 等办公工具相结合，提供包括内容（包括文字、图片）创作、数据分析、辅助决策等一系列新功能，功能丰富度上，远超 notion.ai 等现有的基于 AI 的办公软件。

图表82：数字化办公行业发展历程

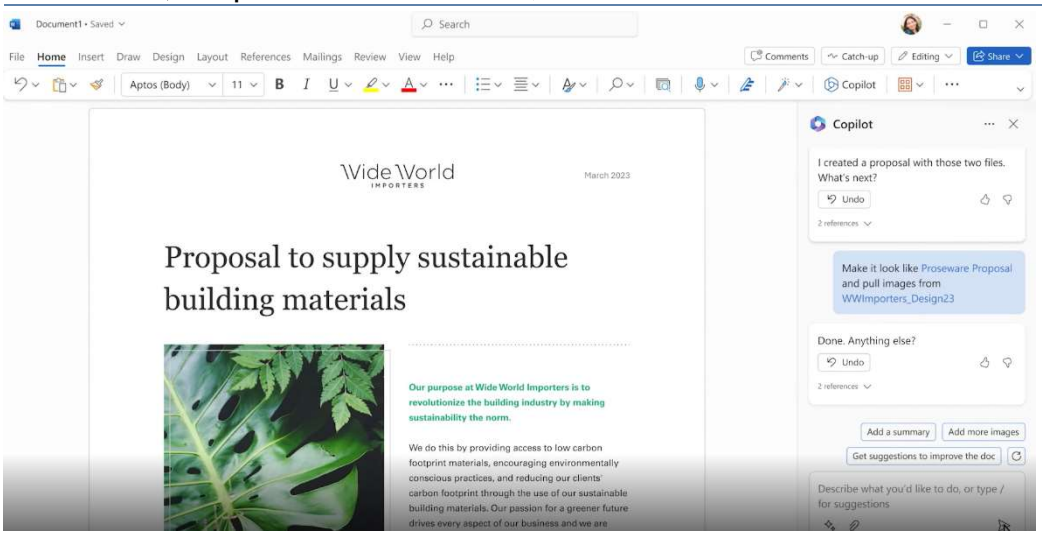
资料来源：各公司官网、华泰研究

AI+生产力工具：强化分析、总结、创作能力

融合 GPT 能力的新版 Word 可与用户一起写作、编辑、总结、创作。据微软 Copilot 365 发布会，融合 GPT 能力的 Word 已具备自动生成、自动排版、摘要生成、常见问答生成等功能。**1）自动生成：**根据指定的内容、指定的话题、指定的文档/表格，自动生成文本内容；**2）自动排版：**根据用户的展示需求，对文本进行指定参考风格的自动美化；**3）摘要生成：**根据已有文本生成大纲/摘要/问答等总结性内容，并支持二次编辑；**4）风格改写：**根据已有文本进行语言风格、文本类型改写，适应不同的业务场景。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表83：Word 中的 Copilot 能够根据指令自动生成文件初稿

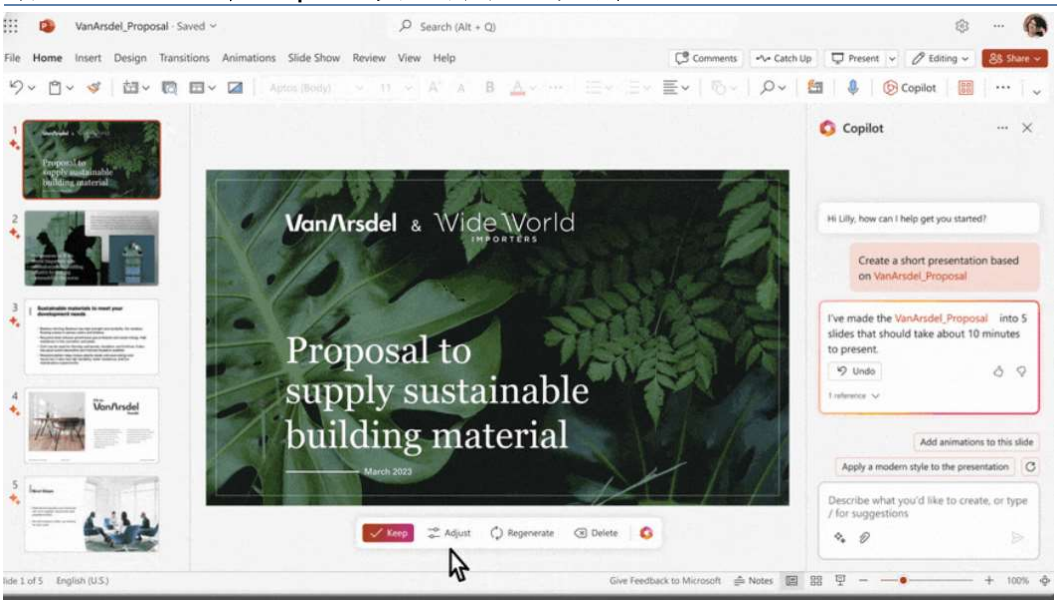


资料来源：微软官网、华泰研究

融合 GPT 能力的新版 Excel 可帮助用户分析数据趋势、实现数据可视化。根据发布会展示，融合 GPT 能力的 Excel 已具备自动数据分析、基于自然语言的深度加工、预测、图表可视化等功能。**1）自动数据分析：**根据已有数据进行自动化数据分析并给出关键信息总结；**2）深度分析：**根据用户的深入分析需求，新建总结的 sheet 和图表并展示加工后的数据，并支持自然语言优化；**3）深度预测：**根据用户需求进行数据的预测并自动生成新的工作区域；**4）图表可视化：**根据数据与自然语言描述，自动生成符合用户期待的可视化图像。

融合 GPT 能力的新版 PowerPoint 可根据用户的想法自动生成演示文稿。根据发布会展示，融合 GPT 能力的 PowerPoint 已具备演示文稿自动生成、设计美化、note 生成等功能：**1）演示文稿自动生成：**根据已有文稿自动生成 PPT 初稿（可指定页数、图片）；**2）自然语言设计美化：**根据用户需求，使用自然语言命令来增减页数、调整布局、重新格式化文本和制作动画；**3）Note 生成：**根据用户需求/页面变化自动生成 PPT 的备注内容，方便用户进行 PPT 演示讲解。

图表84：PowerPoint 中的 Copilot 能够根据指令自动生成幻灯片



资料来源：微软官网、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

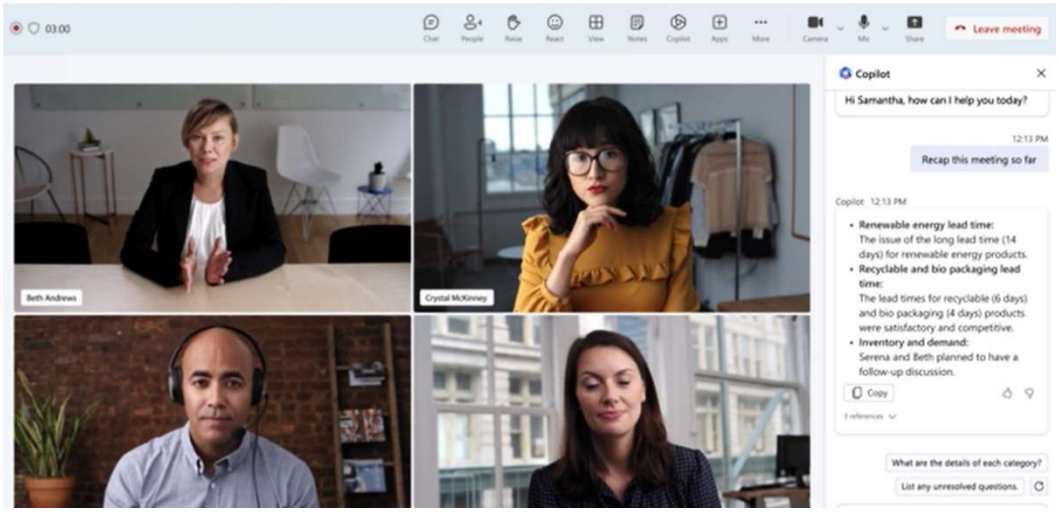
AI+沟通工具：实现自动沟通管理，赋能不同场景的内容生成

融合 GPT 能力的新版 Outlook 可帮助用户对邮件进行管理、汇总、分类、内容起草。根据发布会展示，融合 GPT 能力的 Outlook 已具备邮件要点总结、邮件撰写、风格转换等功能：**1）要点总结：**根据用户需求自动总结（未读/长/多线程）邮件的核心要点，帮助用户处理纷繁复杂的邮件信息，并支持移动端；**2）邮件撰写：**根据用户简短的自然语言提示/其他邮件提取，自动生成格式化邮件初稿，协助邮件撰写；**3）风格转换：**根据用户需求，快速切换邮件的语气、风格，满足不同的工作场景需求。

AI+协作工具：自动总结+跟进+回顾，扩大团队协作智能辅助的外延

融合 GPT 能力的新版 Teams 可帮助用户实时总结会议的相关信息、推进任务的执行进度。根据发布会展示，融合 GPT 能力的 Teams 已具备会议要点总结、流程跟进、纪要生成、智能回顾等功能：**1）要点总结：**根据用户需求自动总结会议要点，各参会人员观点、共识、分歧等内容；**2）流程跟进：**根据用户需求提供“未解决问题提示”、“会议议程创建”等工作内容；**3）纪要生成：**自动生成会议纪要、要点和任务模板；**4）智能回顾：**根据用户 follow 的会议自动生成回顾邮件，并根据用户提问解答会议相关内容。

图表85：Teams 中的 Copilot 能够实时跟进会议进度并提供相关信息



资料来源：微软官网、华泰研究

融合 GPT 能力的 Business Chat，是企业内部的智能搜索与协同办公平台。根据发布会展示，最新推出的 Business Chat 实现了企业内部的搜索与知识管理，可以看作是工作场景下的 ChatGPT：**1）跨应用总结：**根据用户需求自动构建企业内部资料的知识图谱（文档、会议、邮件等）并提供要点总结；**2）业务流程安排与分析：**根据用户需求自动化创建业务流程，为用户提供高可执行性的工作安排；**3）业务问答：**根据用户提问与需求，自动生成符合企业智识的解决方案与措施建议，实现特定问题的智能化解答。

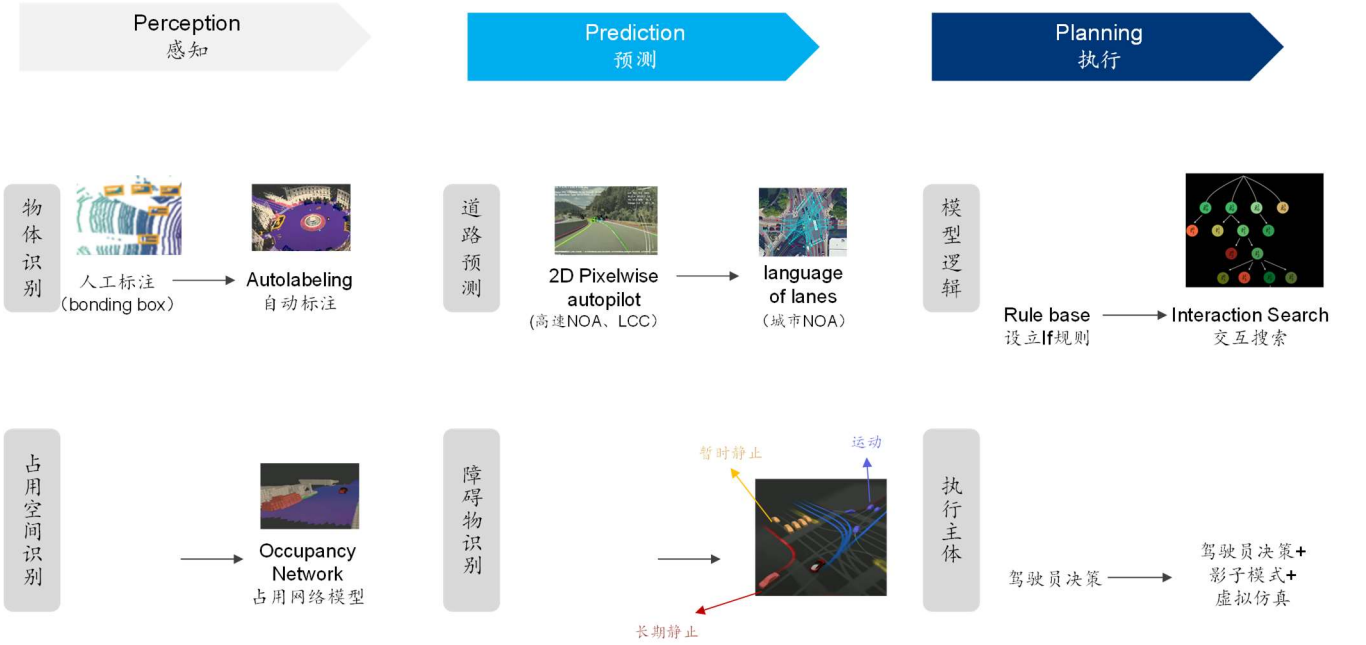
AI+汽车：AI 大模型加速无人驾驶发展

“无人驾驶”看上去很美，但一直很难落地。“无人驾驶”是最早被提出的人工智能应用场景之一，谷歌、苹果、特斯拉、百度等海内外科技巨头从 2016 开始就积极布局，但直到现在仍很难实现大规模商用落地。我们认为，1）多维度数据的获取和标注成本高，2）对小概率事件的决策准确度和人类还存在较大差距，3）事故时法律权责归属不明确，是制约其发展的部分原因。**我们认为，以 ChatGPT 和 SAM 为代表的 AI 大模型的引入（自动标注、算法迭代），从行业整体层面可以加速智能驾驶的量产落地。**

本节，我们参考特斯拉 FSD，将自动驾驶模型算法按流程分为感知（Perception，识别道路和物体）、预测（Prediction，预测周围车辆和行人的行为）、执行（Planning，控制车辆速度方向等）三个阶段进行分析。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表86：AI+汽车：感知决策执行三阶段



资料来源：特斯拉官网，华泰研究

感知层面：物体识别、占用空间识别

感知是模型的起点。我们暂且将感知分为两大类：**1）物体识别：**对物体的种类进行识别（区分车辆、行人、道路等）；**2）占用空间识别：**无需判断物体种类，仅对是否占用前方空间进行判断，典型代表是特斯拉的占用网络模型（Occupancy Network）。**简言之，我们认为在感知层面，大模型的引入可以：1）通过自动标注大幅降低模型识图成本；2）通过占用网络模型等提高模型“举一反三”的能力。**

#1 物体识别：自动标注大幅降低系统成本

数据体量与复杂度提升，标注需求增加。随着智能驾驶的成熟，激光雷达 3D 点云信息、摄像头采集的 2D 图像信息的增加，道路场景的丰富，自动驾驶的数据标注类型与数量在不断增加。

人工标注成本高、效率低，自动标注是 AI 大模型赋能智能驾驶最直接的应用，能大幅降低数据标注的成本。当前智能驾驶领域以人工识别为主。根据恺望数据产品项目副总裁张鹏在 2023 年 2 月的介绍，目前 95%的数据标注是以人工标注为主，机器标注为辅。而据 2023 年 4 月毫末智行 DriveGPT 发布会显示，目前要得到诸如车道线、交通参与者、红绿灯等信息，行业人工标注的成本约每张图为 5 元，毫末 DriveGPT 的成本为 0.5 元。我们认为科技公司大模型训练成熟后，单张图自动标注的边际成本趋近于 0，平均成本有望进一步下降。

以特斯拉为例，特斯拉 FSD 通过自动标注优化系统效率，通过“多重轨迹重建”技术自动标注车辆行驶轨迹。目前在集群中运行 12 小时即可完成 10,000 次行驶轨迹标注，取代 500 万小时人工标注。通过机器的自我训练，减少了人力标注成本高、效率低的问题。自动标注技术成熟，人工标注团队规模缩小。2021 年人工标注团队为 1000 多人，2022 年该团队裁员 200 余人。

占用空间识别：“不认识”物体，亦不影响行车决策

国内新势力与智驾解决方案供应商多使用基于二维点格（pixel）的鸟瞰图（BEV，Bird Eye View）模型，使用矩形框（bonding box）来框定物体，进行物体识别。**特斯拉创新性的提出了占用网络（Occupancy Network）模型，无需训练模型“认识”物体，只需判断物体是否占用前方空间，即可进行行车决策。**

免责
更多免费行业报告 并购家 www.ipoipo.cn

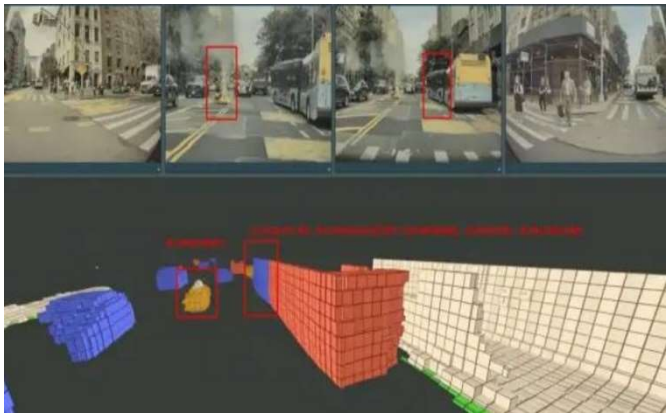
占用网络（Occupancy Network）模型，将三维空间点格化，每个 3D 点格即为一个 voxel，在 8 颗摄像头采集到的平面信息基础上添加时间、空间信息，即可输出：

1）该 3D 点格被占用的概率（例如：区分静止的车辆/运动的车辆）；
2）语义信息（Semantics output；例如：区分静止的车辆/路牙）；
3）表面信息（Surface output；例如：坡度、泥坑、积水）。

占用网络相较于之前在障碍物识别和行驶路径预判方面有了明显提升，具体来看：

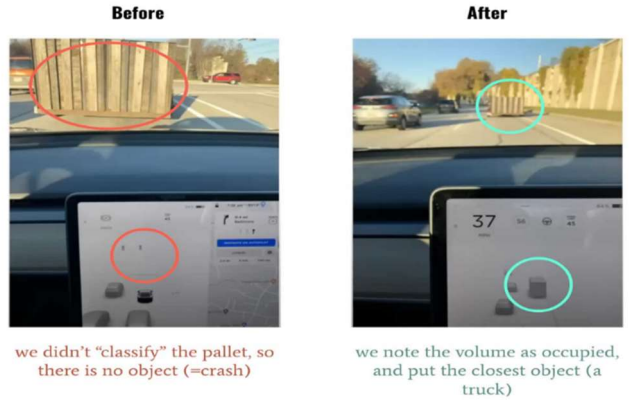
1）通过占用网络，只需分析物体的空间内栅格占用情况，不需对物体本身进行检测识别，规避传统视觉算法中对物体识别失败带来的车祸风险。尤其在面对静态障碍物、与周围环境类似的障碍物、训练模型中未涵盖到的障碍物时，可以更大程度的规避风险。

图表87：占用网络对双节公车动态（蓝色）静态（红色）精确识别



资料来源：特斯拉官网，华泰研究

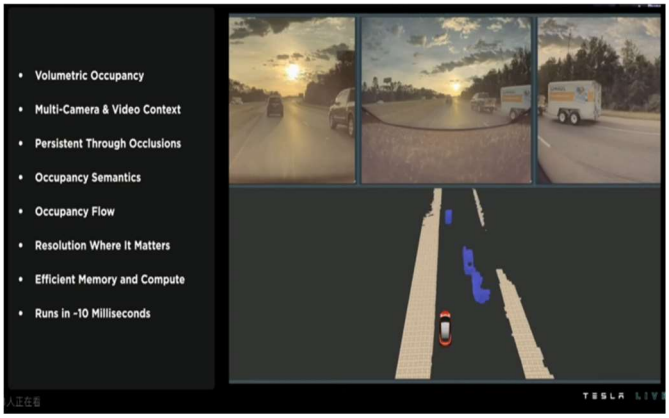
图表88：占用网络对“不认识”的静态障碍物的识别



资料来源：特斯拉官网，华泰研究

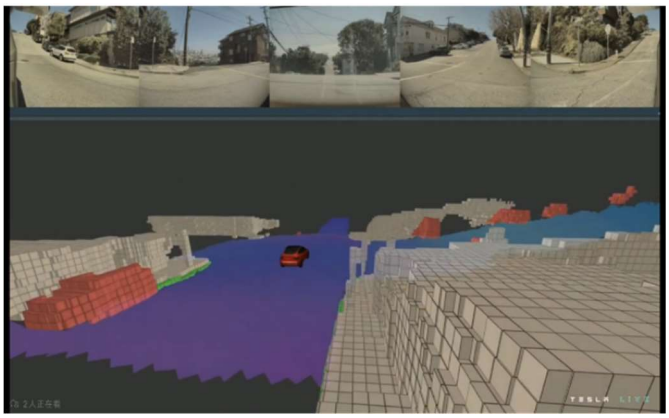
2）占用网络解决传统视觉算法难题及还原道路坡度和曲率，辅助行车更优决策。通过计算几何空间的体积占用率，可以在占用网络中精确地还原物体本身形状。占用网络塑造的 3D 世界还可以还原道路的坡度和曲率，让车辆根据实际道路情况提前预测加速和减速判断，进一步提高行车安全性和舒适度。

图表89：通过占用网络可以检测多厢卡车甩尾问题



资料来源：特斯拉官网，华泰研究

图表90：占用网络可以还原道路坡度和曲率



资料来源：特斯拉官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

预测层面：道路拓扑关系预测、障碍物预测

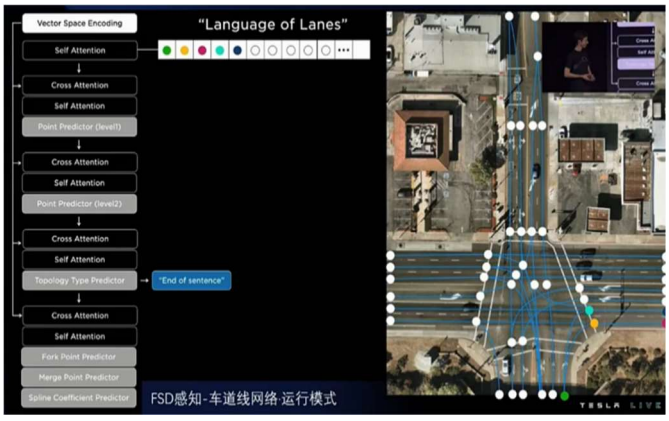
预测分两种，一种是道路信息的预测（lane prediction），另一种是障碍物的预测。

道路信息：基于大模型勾勒拓扑关系，摆脱对高精度地图的依赖

道路信息的预测包含：1）语义信息，2）连接信息。最初 autopilot 使用的传统 link prediction，只能预测比较简单的道路，比如高速公路，基于此已经可以实现 LCC 等 L2 的功能。要实现更加复杂的城市道路的拓扑关系预测，需要基于：1）高精度地图；或者 2）导航地图+神经网络预测。特斯拉基于基础的硬件配置（摄像头+导航地图）+自创的 language of lanes 模型，来通用化地勾勒整个世界的道路信息。

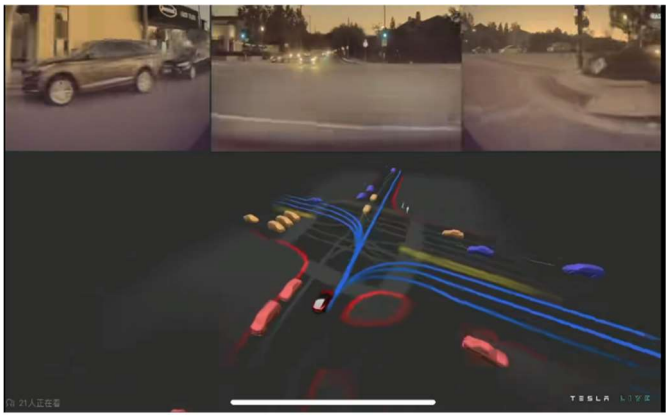
车道线网络模型辅助进行车辆行驶路径的预判。车道线网络模型通过车道语言（Luange of Lanes）可以在车载摄像头及地图数据所形成的图像上，将道路数据标注成一系列节点并赋予不同语义（起始点、延续点、交叉点、终点等），并通过组合不同语义的“单词”形成“句子”，自动勾绘出一条条车道线。这套“车道语言”，可以在小于 10 毫秒的延迟内，思考超过 7500 万个可能影响车辆决策的因素，运行这套语言的功耗只需要 8W，较大的提升了特斯拉 FSD 对车辆行驶路径的预判能力。

图表91： 特斯拉车道线网络模型示意



资料来源：特斯拉官网，华泰研究

图表92： 特斯拉对可能的行进轨迹进行预判并绘制成车道线



资料来源：特斯拉官网，华泰研究

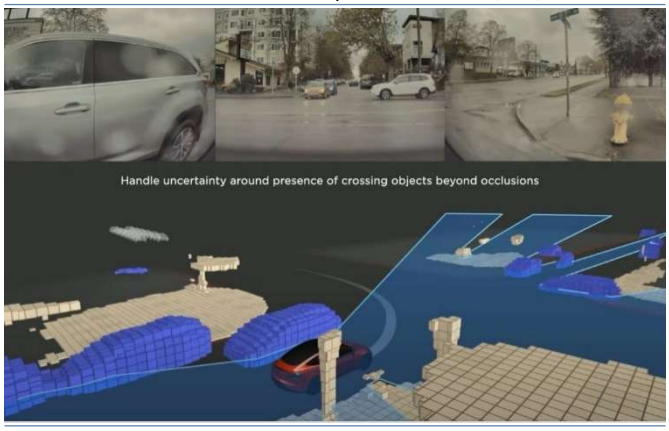
物体信息：基于大模型预测动静信息，为行驶决策提供支持

物体的预测包含动、静概率信息，再结合道路拓扑信息，为最终的形式决策提供支持。特斯拉的 Occupancy Network 中红色代表长期静止的车辆，黄色代表临时停车，蓝色代表运动，可对物体的动静状态及其概率进行预测。在一些特殊情景下，例如左转摄像头被左侧大货车遮挡，无法判断左向是否有来车，模型会自动生成虚拟车辆，假设左侧有被遮挡的来车，基于此进行决策，更贴近人类驾驶员的思维模式。

基于占用网络可以预测道路上其它物体的行进轨迹。通过对栅格进行光流估计来检测物体运动并预测其短期行进运动轨迹，并标注上丰富的语义（红色—静止，蓝色—加速，黄色—减速等），从而在特斯拉车辆行驶过程中规划最优行驶路径进行避让，保证驾驶安全性。

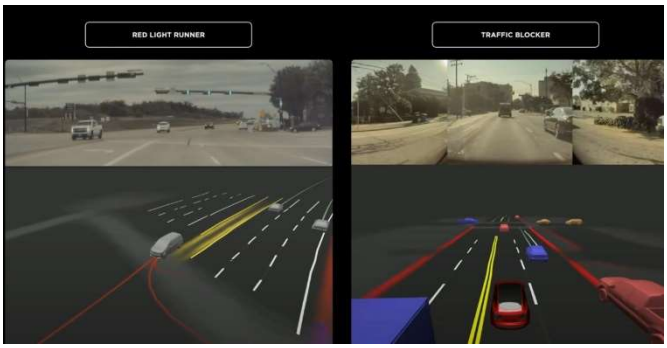
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表93：特斯拉左转摄像头被遮挡，模型假设左侧有来车



资料来源：特斯拉官网，华泰研究

图表94：特斯拉对闯红灯车辆、道路阻碍者的预判



资料来源：特斯拉官网，华泰研究

决策层面：交互搜索模型，影子模式，虚拟仿真

决策的难点在于多方的交互与对路权的博弈，计算的效率是至关重要的。目前业内普遍在50-100毫秒之间完成一轮计算。受车端算力与计算效率的限制，目前决策层面的模型可分为两类：

1）Rule base 的模型（类似 if 程序）：提前设定了某些情境下的反应机制，例如左转选择超车还是避让前车，更大程度的保障了安全性，但一定程度上也较为僵化，不符合“老司机”的驾驶习惯。

2）特斯拉的交互搜索的模型：基于神经网络模型，自主搜索目标和备选方案（maneuver trajectory），query base 的条件下可缩短单次计算时间至 100 微秒。

此外，影子模式、虚拟仿真的引入也有望提高驾驶的智能化程度和流畅度。

1）影子模式是指大模型在车辆后台运行，模拟决策而不实际控车，不会对驾驶者及车辆产生任何干扰。但是在有异常场景或模型与人类驾驶员的决策不同时，触发数据采集及回传，对机器做出的决策进行“纠正”。

2）虚拟仿真是指通过 AI 自动生成道路场景、车辆、行人等信息，对模型进行训练。特斯拉的模拟仿真可以通过建立一个虚拟世界，以提供现实中难以获得/数据量不足的情景，目前特斯拉可以在 5min 内自动生成一个复杂路口 3D 虚拟场景。进一步，还可以通过道路中的场景，创建更多变种场景，帮助决策模型进行不断优化迭代。

更多免费行业报告 并购家 www.ipoipo.cn

AI+工业：ChatGPT 引领 AI 赋能工业，工业智能化水平或将加速提升

ChatGPT 引领 AI 大模型突破，随着 AI 技术和工业领域深度融合，工业 AI 应用或将迎来高速发展窗口。随着 2023 年 3 月 OpenAI 正式发布大型自然语言处理模型 ChatGPT-4，同月微软将其旗下 Azure 云服务 Azure、Microsoft365 办公软件、Bing 搜索引擎、Power Platform 低代码开发平台等产品与 ChatGPT 深度融合，生成式 AI 在文本、代码、视频等一系列应用领域的前景逐渐明晰。

AI 大模型+工业所形成的核心产品、方案与服务，是 AI 赋能工业的主要载体。其中主要包括四个方面，按层级关系可分为基础软硬件、智能工业装备、自动化与边缘系统、平台/工业软件与方案。其中，基础软硬件是指各类芯片/计算模块、AI 框架、工业相机等相对通用的软硬件产品；智能工业装备是指融合智能算法的机器人、AGV（自动导向机器人）、机床等工业生产制造设备；自动化与边缘系统是指融合了智能算法的工业控制系统；平台/工业软件与方案则是指各类具有 AI 能力的工业互联网平台及其衍生解决方案和应用服务。**以上应用通过识别类应用、数据建模优化类应用及知识推理决策类应用三种模式被广泛用于工业研发、生产、管理及服务等全环节。**

图表95：“AI+工业”应用场景

资料来源：各公司官网，华泰研究

AI 大模型+工业研发设计：增进芯片设计、CAD 设计及仿真优化效率

AI 大模型+工业软研发设计软件将加快推动工业数字化进程，极大解放生产力。工业数字化进程大致始于 20 世纪 50-70 年代，以图形化、自动化为代表性技术，将计算机设计、调度能力首次带融入工业生产领域，替代人工操作，解放生产力的同时提升生产效率，这一时期的典型代表有 50-60 年代出现的 CAD 软件，60-70 年代出现的 CAE、EDA、PLC 软件，极大地提升了生产效率。在当前时间点，AI 与工业研发设计软件积极相融合，借助云计算，进行超大数据量的推理训练，进一步优化软件工作效率，简化研发流程和复杂度，帮助企业提升效率的同时还能优化成本结构。

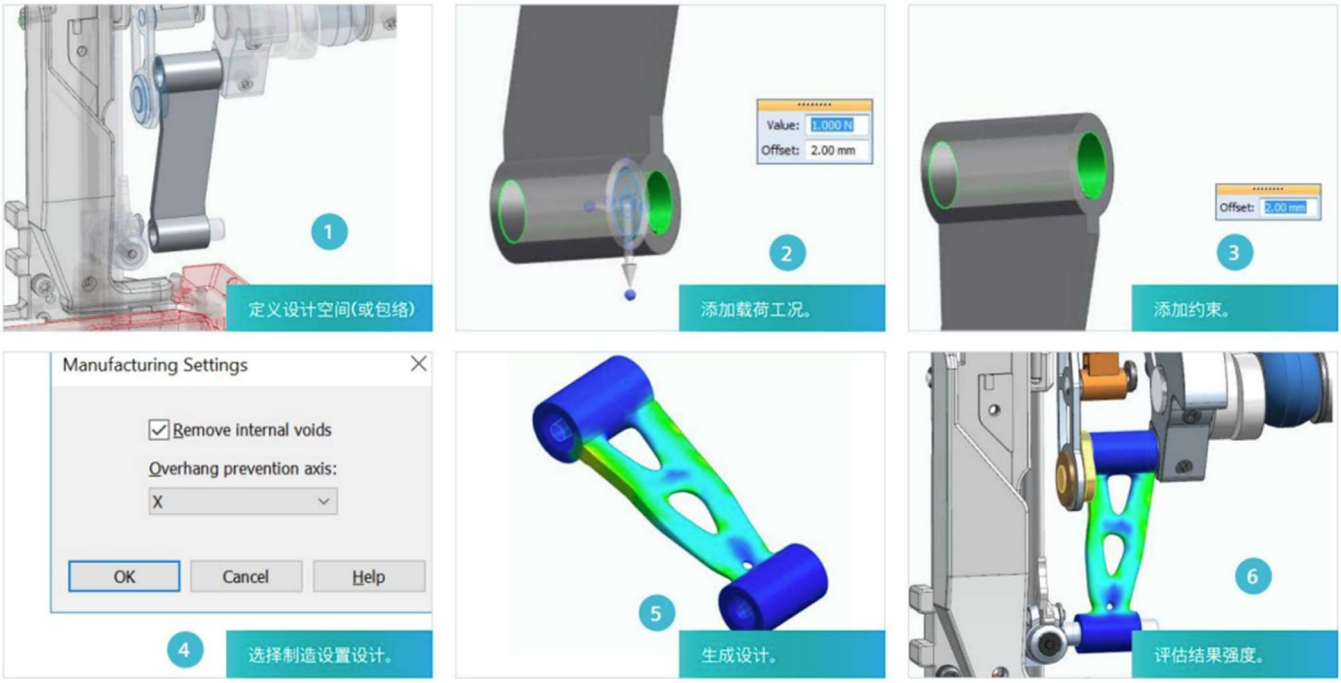
AI 大模型+EDA 将成为未来芯片设计软件的主流趋势，大幅加快芯片研发设计的进度。2023 年 4 月 7 日，美国 Cadence 公司推出 Allegro X AI technology 新一代系统设计技术。本技术依托于 Allegro X Design Platform 平台，利用云端的扩展性实现物理设计自动化，在提供 PCB 生成式设计的同时，确保设计在电气方面准确无误，并可用于制造。本项新技术可自动执行器件摆放、金属镀覆和关键网络布线，并集成了快速信号完整性和电源完整性分析功能。使用本技术的生成式 AI 功能，客户可以简化系统设计流程，将 PCB 设计周转时间缩短至原来的十分之一。国内方面，概伦电子在 3 月的业绩会上表示正在关注 AI+EDA 软件的发展机遇，并将未来在该领域加大投入。

更多免费行业报告 并购家 www.ipoipo.cn

AI 大模型+CAE 有效提升仿真效率，优化代码生成能力。CAE（computer aided engineering），指用计算机辅助求解分析复杂工程和产品的结构力学性能，以及优化结构性能等，把工程（生产）的各个环节有机地组织起来，其关键是将有关的信息集成，使其产生并存在于工程（产品）的整个生命周期。CAE 软件可作静态结构分析，动态分析；研究线性、非线性问题；分析结构（固体）、流体、电磁等。因为 CAE 已经在工业领域得到了多年的大量应用，积累了大量的数据，可以作为生成式 AI 的训练数据来源。未来基于生成式 AI 的新型 CAE 建模将基于历史设计数据，通过大数据挖掘不同影响因素之间的关系，而不再过分追求从物理理论生成模型，从而能大幅提升建模的效率。

AI 大模型赋能创成式设计，生成式 AI 未来或可为 CAD 软件提供大量可供选择的模型。创成式设计是一种利用 AI 技术根据一系列系统设计来自主创建优化设计的 3D CAD 功能。其特点在于能在设计师给定的约束条件和目标下，借助 AI 的能力来快速生成满足要求的目标模型，供设计师从中选择合适的模型进行进一步的设计优化，从而提升设计效率，降低设计成本。从目前来看，生成式 AI 在严谨理性的数学和逻辑领域的能力相对文字领域仍然比较有限，但未来随着以 GPT 为代表的通用大模型逐渐成熟，工业设计数据库不断丰富，CAD 有望一方面借助生成式 AI 对设计进行参数优化，另一方面借助大量的设计模块数据库生成推荐的设计草图。

图表96：创成式设计的工作流程



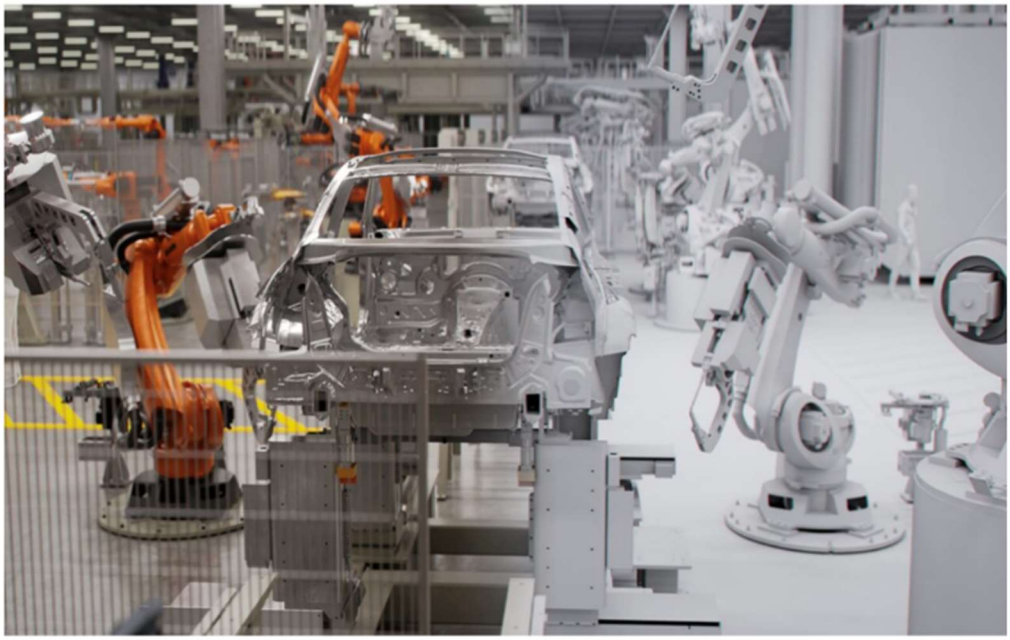
资料来源：西门子官网，华泰研究

AI 大模型+数字孪生概念，仿真优化生产流程以改善全球工厂协作能力。数字孪生是指充分利用物理模型、传感器更新、运行历史等数据，集成多学科、多物理量、多尺度、多概率的仿真过程，在虚拟空间中完成映射，从而反映相对应的实体装备的全生命周期过程。目前已经有企业将数字孪生技术运用在模拟工厂领域，例如 BMW 依靠英伟达的 Omniverse 平台，对整座工厂模型中的所有元素进行模拟，并将来自不同制造商的几种设计和规划工具的数据汇总在一起，最终在协作环境中生成物理级逼真的实时仿真；虚拟工厂通过规划工具集成了规划数据和应用程序，从而实现了无兼容性限制的实时协作，宝马全球数千名工程师依托 Revit, Catia 等软件以及云平台，在同一个 3D 虚拟环境中进行协作，从而可以在早期规划阶段最大化评估变化和调整，并利用 NVIDIA Omniverse 构建工厂端到端数字孪生，将规划流程的效率提高了 30%。目前宝马已计划将 Omniverse 推广到其整个全球生产网络，使得其汽车生产更容易跨地点、跨时区协作，跨流程合作，强化全球各地工厂的合作能力。

免责
更多免费行业报告 并购家 www.ipoipo.cn

数字孪生反哺 AI 大模型，生成数据供以 AI 模型训练。我们认为不同于金融等其他行业，工业场景相对来说数据样本量较小，AI 训练相对困难，在此背景下，工业数字孪生可以通过仿真的形式生成大量数据，帮助 AI 模型深度优化。目前，微软通过与 Ansys Twin Builder 合作，其 Project Bonsai 可同时运行数百个机器或者应用的虚拟模型，并将这些数字孪生生成的数据，直接输入大脑对其进行优化，从而不断克服各种局限性。使用大量虚拟模型（在不减少物理模型的数量的条件下）可以缩短训练大脑的时间，降低成本，还能让工程师在虚拟环境中引入可能对物理机器造成潜在危险或损坏的极端情况，这样大脑在投入运行之前，就可以学习了解所有可能遇到的情况。具体到应用来看，CPCS（机舱压力控制系统）可以将数字孪生技术与 Bonsai 进行集成，工程师在 Bonsai 中可以通过图形选择和连接控制代码的功能块构建 AI 大脑，并在 Twin Builder 中对系统进行建模，之后利用简单的 Python API，将模型移植到数字孪生工作流中，并用于训练 Bonsai 大脑及创建控制器。

图表97：宝马使用 Omniverse 建设的数字化虚拟工厂



资料来源：英伟达官网，华泰研究

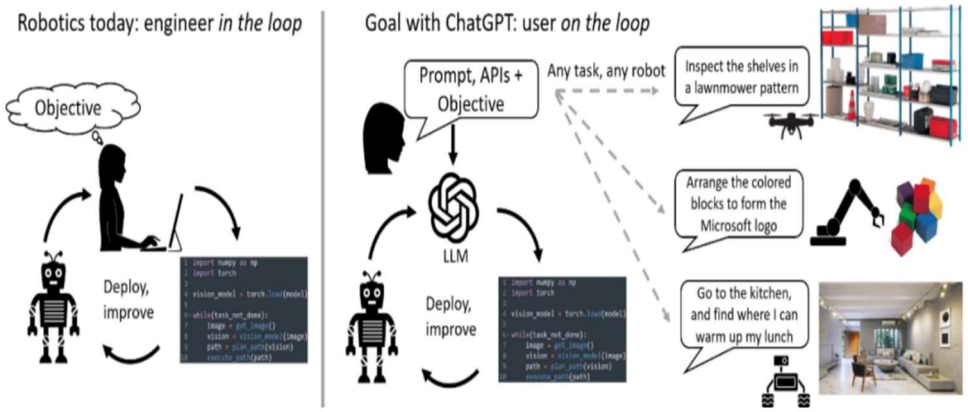
AI 大模型+工业生产制造：强化工业机器人的信息处理、感知执行等能力

自然语言处理工具 ChatGPT 的出现，可以帮助机器逐渐实现真正像人类一样交流、执行大量任务。生成式 AI 的发展将真正促进制造环节向智能化和数字化转型，而工业机器人和自动化工厂作为智能制造的核心载体，将作为生成式 AI 与智能制造的中间桥梁。根据微软发布的《ChatGPT for Robotics: Design Principles and Model Abilities》，目前生成式 AI 主要通过两个层面对工业机器人进行辅助，第一，ChatGPT 作为预训练语言模型，可以被应用于人类与机器的自然语言交互。机器通过 ChatGPT 理解人类的自然语言指令，并根据指令进行相应的动作。第二，GPT 可以帮助机器在执行路径规划、物体识别等任务时做出相应的决策。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表98：“无代码编程”工作流程图



资料来源：微软官网，华泰研究

生成式 AI 可进一步加强工业机器人的信息处理和分析能力。ChatGPT 可以与数据分析师进行交互，以帮助分析和可视化工业生产数据，或者可以运用于生产线中帮助普通操作员完成多种任务。此外 ChatGPT 还可以用于监控生产质量，并给出改进质量的建议，以提高产品质量。生产出现故障时，ChatGPT 还可以提供快速诊断和解决方案，从而提高生产效率。**从具体应用来看，国外方面**西门子将 ChatGPT 技术与其现有的自然语言技术相结合，通过将生成式 AI 嵌入智能交互系统的架构中，有效实现了操作者与系统自然语言的交互，目前其自动化生产 SIMATIC IT 软件已经引入 ChatGPT；**国内方面，**2022 年底百度联合 TCL 搭建文心电子制造行业模型，TCL 产线检测 mAP 指标平均提升 10%以上；此外新产线的冷态起动效果可以提高至原来的 3 倍，产线投产的研发时间缩短了 30%。

生成式 AI 可与机器视觉融合赋予工业机器人更为强大的感知执行能力。

1） **感知方面，**根据佐思汽车研究，常用的图像数据拓展多采用平移、变换、拷贝粘贴等策略，多数只是对图像或者目标进行局部处理，很难在保持语义信息和多样性之间找到平衡，而基于 Stable Diffusion 模型的 SIP 模型可以拓展使用生成式 AI 大模型，通过图像的标签和标题指导生成 Image-To-Image 高清图片，对比常用算法性能有所升。从应用层面来看，2023 年 4 月 10 日海康机器人发布相关新产品：AI 读码器、AI 智能相机、2D 工业相机、3D 相机和软件(Vison Master 3D 算法平台和机器人视觉引导平台)。

2） **执行方面，**微软研究显示，ChatGPT 具备构建“感知-执行”循环模型的能力，该模型先通过 API 库输出拥有循环结构代码，之后使用图像采集和目标检测等感知函数提取机器人导航和控制的相关信息，并通过文本对话将信息输入到 ChatGPT，最终 ChatGPT 将解析输入的信息流并输出相关的动作指令。

生成式 AI 依托工业大模型打造工程制造引擎，不断提高生产力。例如，国内方面已有企业开始针对此领域落地应用。例如国内“AI+制造”解决方案供应头部玩家创新奇智，已于 2023 年 4 月 13 日发布“奇智孔明”AIGC 引擎（AInnoGC），该产品是针对制造业垂直场景的 AIGC 工程算法引擎，依托于创新奇智自研 MMOC 人工智能技术平台，以工业预训练大模型为核心。奇智孔明拥有五大能力，分别是内容生成、智能问答、多轮对话、推理能力、代码生成，能够满足制造行业企业的个性化需求。本引擎将着重应用于制造业、工业软件领域，有效解决行业中大客户的专属需求，如交互式动态业务报表生成，智能产线设计等，打破垂类信息孤岛，提升生产力，实现更全面的数字化转型。

免责
更多免费行业报告 并购家 www.ipoipo.cn

AI 大模型+工业管理与服务：或成为 ChatGPT 最易实现的工业应用场景

相较于研发设计和生产制造的环节，管理服务环节或因其具备更强的通用性而成为生成式 AI 最易突破的工业应用场景。2023 年 3 月，微软陆续推出 GPT 互动式 AI 能力其商业产品 Dynamics 365 Copilot 及 Microsoft 365 Copilot，大幅提升用户在经营管理类软件上的工作效率，后续此模式或将被推广至供应链管理、客服服务和市场营销等其他应用场景。

AI 大模型+CRM 帮助销售人员快速相应客户需求，提升市场营销的效率。例如，国外 Dynamics 365 Customer Service 中，Copilot 能帮助服务专员提供更好的客户体验，其能够针对聊天对话和电子邮件中的问题，撰写出符合上下文语境的答案，其互动式的对话基础来自知识库和过往案例，从而帮助销售人员快速响应客户需求。此外销售人员可以使用该平台用自然语言与客户数据平台进行对话以获得高度定制化和目标明确的客户分类，从而准确定位特定客户群体。国内方面，AI 领先企业第四范式也在大模型研发上也取得了阶段性进展，其已经上线的企业级 4Paradigm SageGPT，将大型生成式语言模型与垂直领域专业知识进行融合，并具备企业级场景下的多模态及 Copolit 能力。**以上案例充分证明生成式 AI 可以在工业管理服务环节发挥降本增效的作用，具有极强的应用落地的可行性。**

AI 大模型+ERP 软件，多场景发挥产品能力。ERP 管理软件是以管理会计为核心的信息管理系统，其将企业内部所有资源整合在一起，对采购、生产、成本、库存、分销、运输、财务、人力资源进行规划，从而达到最佳资源组合，取得最佳效益。基于 AI 辅助的 ERP 系统，可以通过庞大的数据训练，帮助企业做出最有效的决策。具体应用有，国内汉得信息已经基于 AI 技术打造了多场景应用解决方案，其所开发的数字员工通过积累多家企业的数字化和智能化财务系统的经验，在 AI 的加持下，帮助客户培养用于审核、计税、支付、稽核等多场景的数字化员工，满足全天候工作需求。

AI 大模型+供应链与物流管理推动智能调度、智能跟踪和智能预警。因为供应链和物流管理涉及从原始供应商到客户终端的全链条管理，具有监控难度高，人为调节难度大等特点，生成式 AI 的出现可有效降低人员对此的管理难度。**需求端，**生成式 AI 可以通过对更多数据的训练，包括经济周期、地缘政治、时事新闻和天气等外部因素，从而对需求实现更准确、更及时的预测，帮助企业增效降本。**仓库和物流端，**生成式 AI 可以分析仓库的布局和工人的移动模式，以优化产品布局，此外还可以开发提高仓库效率的策略，例如存货运输路线规划，最大限度地减少移动时间，节省大量时间成本。另一方面，AI 可以根据业务场景选择最高效的仓储地址，合理布局仓网。目前，国内旷视科技已开始在视觉大模型领域布局，希望从“感知-决策-执行-反馈”全链条探索生成式 AI 在仓储物流领域的可行性。

AI+金融：坐拥数据富矿，有望成为 AI 大模型率先落地的垂直领域之一

2023 年 3 月以来，多家金融机构、金融服务机构发布其 AI 大模型：彭博发布支持金融领域的自然语言处理（NLP）任务的 BloombergGPT，中国农业银行推出类 ChatGPT 的大模型应用 ChatABC，中国工商银行发布了基于昇腾 AI 的金融行业通用模型。我们认为垂直行业的高价值量数据对于 AI 大模型的训练和垂直领域应用至关重要，金融行业数字化程度领先，拥有数据富矿，有望成为 AI 大模型率先落地的垂直领域之一。应用端，我们看到生成式和理解式大模型在银行、保险、资管、投研、投顾等多个细分领域正在落地或拥有潜在落地场景，帮助金融机构降本增效。

免责
更多免费行业报告 并购家 www.ipoipo.cn

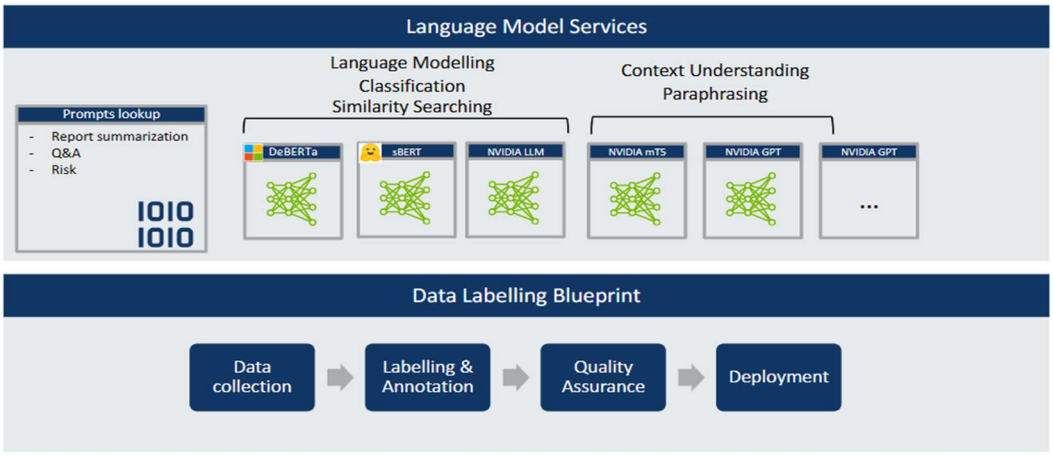
图表99：大模型逐渐渗透银行、保险、资管、投研、投顾等多个金融细分领域应用

资料来源：各公司官网，华泰研究

银行：理解式大模型助力数据洞察能力提升，生成式大模型赋能高质量顾问式金融服务

在银行领域，理解式大模型可以用在信贷风险管理、智能获客和产品识别，通过提升银行的数据洞察理解能力，来更好地识别客户需求以及评估客户信用风险。例如，度小满的智能征信中台将大型语言模型 LLM、图算法应用在征信报告的解读上，能够将报告解读出 40 万维的风险变量，更好地识别小微企业主的信贷风险。此外，英伟达和德意志银行也正在合作测试一系列名为 Financial Transformers (Finformers)的大模型，这些模型能够从非结构化数据中提取关键信息，更快地检索并识别数据质量问题，以提供风险的早期预警信号。

图表100：Finformers 底层大模型有多个，分别擅长不同的任务

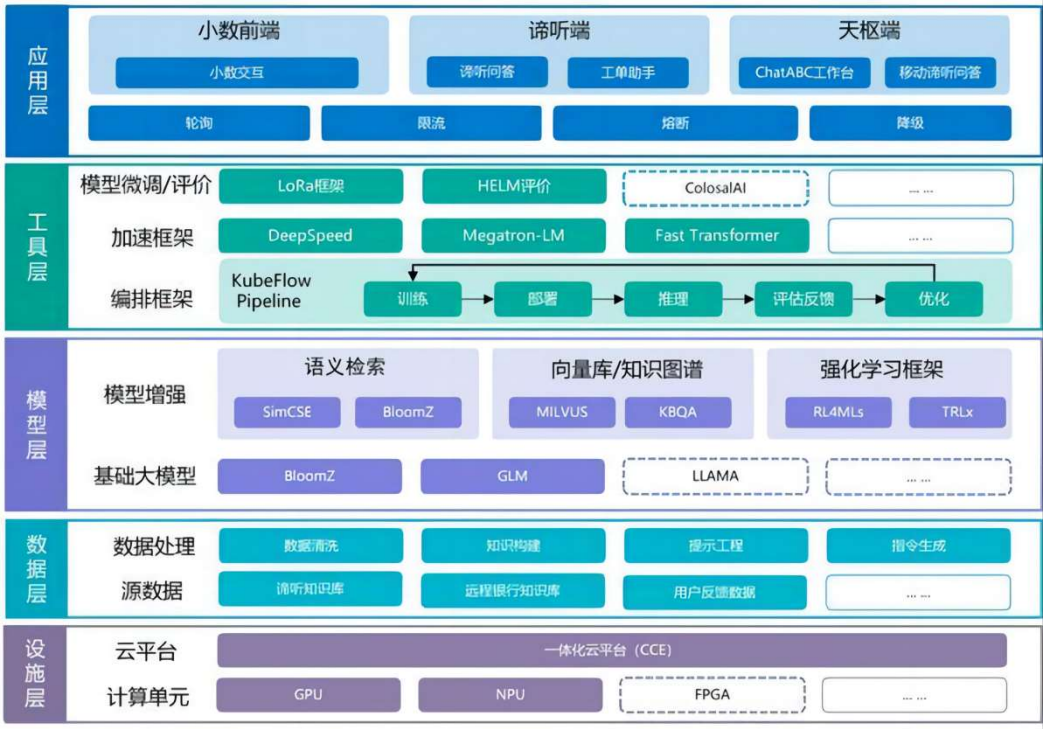


资料来源：英伟达 GTC 2023 AI 开发者盛会官网，华泰研究

生成式大模型应用于银行领域则可提高客户服务质量，智能客服既能与用户进行多轮对话，还能提出具体可行的解决方案。比如，中国农业银行近期推出的 AI 大模型 ChatABC 1.0，拥有在金融领域的知识理解能力、内容生成能力以及安全问答能力，当前已通过平台问答助手、工单自动化回复助手等形式面向内部员工试用。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表101： 中国农业银行 ChatABC 1.0 五位一体 AI 大模型创新应用架构



资料来源：中国农业银行，华泰研究

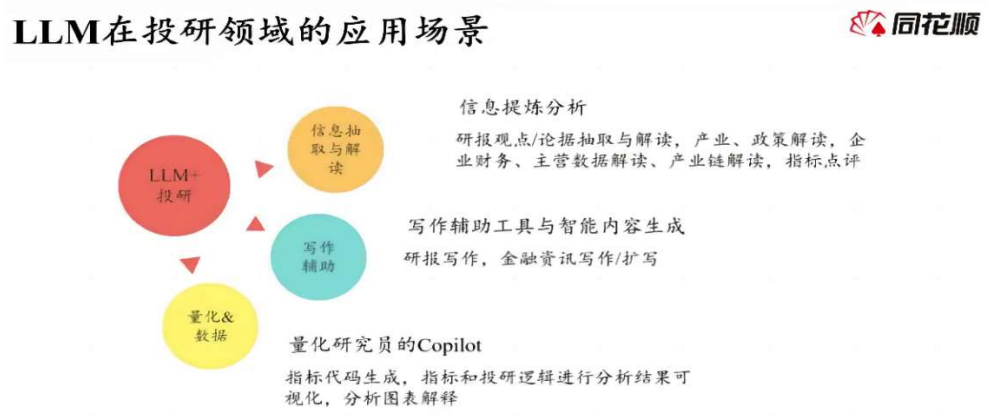
中国工商银行最近发布了基于华为昇腾 AI 的金融行业通用模型，该模型兼具理解式和生成式大模型的功能，赋能风险防控、客户服务以及运营管理。在风险防控上，对融资项目建设的进度进行监控，使得监测精准度提升约 10%、研发周期缩短约 60%。在客户服务上，该模型支撑的智能客服在识别客户来电诉求和情绪识别的准确率上有显著提升，能够更精准有效地响应客户需求，并且使得维护成本大幅缩减。在运营管理上，模型的应用帮助智能提取期限、利率等信贷审批书核心要素，提升了信贷审批效率。

投研：“全能助理”辅助信息了解、提炼和挖掘，实现研报自动生成

在投资研究领域，大模型可以成为投研从业者的“全能助理”。理解式大模型不仅可以帮助了解国内外的宏观政策、行业信息、公司和产品信息，并将关键信息进行抽取与提炼，还能通过对海量非传统数据进行挖掘来发现另类投研因子。生成式大模型则能实现研报的自动生成和翻译。例如，中信证券近期正在内测基于生成式预训练大模型的智能助手 CiticsAI，CiticsAI 的训练结合证券行业和中信证券内部特有的语料数据，应用涵盖文本生成、代码生成、图像理解与生成等场景。同花顺的算法总监马天翼在 2023 年 3 月 18 日举行的大语言模型与资产管理论坛中也指出，大模型在投研领域可以用于研报观点抽取和解读、研报生成和翻译、量化指标代码生成及解释等方面。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表102：LLM 在投研领域的应用场景



资料来源：同花顺，华泰研究

投顾：全方位分析客户需求，自动化定制化投资建议

在投资顾问领域，大模型不仅能帮助全方位分析客户需求和市场趋势，还能提供自动化的投资建议。比如，同花顺应用 LLM 来构造合规、准确的投顾助手，有 4 个方面的尝试：1）基于 Text2SQL 的问句-查询条件匹配：将自然语言描述转化成对应的 SQL 查询语句，从而实现用自然语言查询数据库的目的；2）投资对话意图的 Double Check：用于投资对话中的歧义消解，更好地锁定投资者意图；3）通过语言模型构建金融事理图谱，图谱反过来增强语言模型准确性；4）多轮对话：通过对话技术了解用户画像，设定投资目的及风险承受度，并提供自动化的投资建议。

摩根士丹利已经接入 OpenAI，基于自身庞大的研究资源和数据库为财富管理顾问提供帮助。摩根士丹利拥有数十万页涵盖投资策略、市场研究以及分析师见解的内容库，主要以 PDF 形式呈现，财富管理顾问需要浏览大量信息才能找到特定问题的答案。因此，摩根士丹利 300 名顾问正在测试基于 OpenAI 的 AI 工具，以帮助约 1.6 万名财富管理顾问利用庞大的研究和数据库，并计划在未来几个月内广泛推广。

财经新闻：理解财经新闻背后的市场“情绪”，辅助金融资讯写作

在财经新闻领域，理解式大模型帮助理解和判断财经新闻文章中的情感走向，生成式大模型助力更准确的金融问答和资讯写作。例如，彭博社近期发布的 BloombergGPT 得益于大规模金融垂直领域的文件、行业新闻、社交媒体等文本数据集的训练，能够理解财经新闻背后的市场“情绪”，辅助金融资讯写作，这解决了通用 NPL 模型在金融领域应用的痛点

图表103：BloombergGPT 在金融垂直领域的能力突出，在通用领域的能力接近 GPT-3

Finance-Specific	BloombergGPT	GPT-NeoX	OPT-66B	BLOOM-176B
Financial Tasks	**62.51**	51.90	53.01	54.35
Bloomberg Tasks (Sentiment Analysis)	**62.47**	29.23	35.76	33.39

General-Purpose	BloombergGPT	GPT-NeoX	OPT-66B	BLOOM-176B	GPT-3
MMLU	**39.18**	35.95	35.99	39.13	**43.9**
Reading Comprehension	**61.22**	42.81	50.21	49.37	**67.0**
Linguistic Scenarios	**60.63**	57.18	58.59	58.26	**63.4**

资料来源：BloombergGPT: A Large Language Model for Finance，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

AI 赋能传媒：AIGC 应用鹏程万里

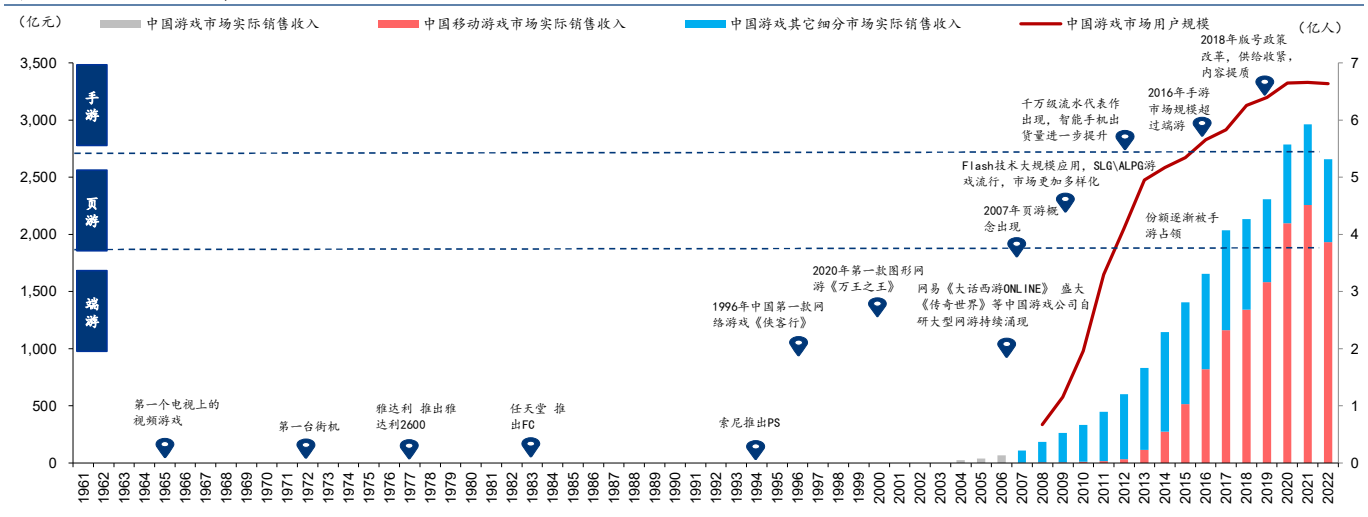
22H2 以来海外大模型快速更迭，ChatGPT/GPT-4 向开发者开放 API 且价格下探，带动应用层面持续落地，国内如文心一言、通义及天工等众多大模型也陆续发布，从文字向图文、视频等多模态拓展。同时 AI 在多个领域的应用已初具成效：游戏方面，AI 在素材原画生成、代码辅助编写等领域落地，AI+玩法也在不断尝试创新；电商方面，AI 可以用于智能选品、智能投放、智能客服等，有效提升工作效率；营销方面，AI 能够创新广告形式，提高营销内容生产效率；影视方面，AI 能够赋能前期筹备、中期拍摄、后期制作及宣发全流程；办公方面，AI 办公应用交互性增强，效率较大提升。我们认为，伴随技术的发展与产品的更迭，AI 将在各领域取得更为深入的发展，AIGC 应用仍有较大空间。

AI+游戏：降本增效持续验证，AI 新玩法尝试创新

游戏产业市场空间广阔，游戏形式不断更新。按照产品类型划分，可将游戏分为端游、手游及云游戏。由于计算机硬件的发展，1997 年起主机游戏进入衰退期，在电脑上进行的端游开始逐渐兴起；到 2005 年，浏览器相关技术已相当成熟，网页游戏由此诞生；随着 3G-4G 移动通信技术升级和智能手机等硬件普及，手机游戏开发技术不断创新，渐渐主导了游戏市场。云计算及相关技术的出现让游戏娱乐场景的更广覆盖成为可能，云游戏是基于云端技术与 5G 技术发展而形成的一种不同的游戏运行模式，其脱离了终端限制，可以跨终端，实现了在云端服务器运行游戏。随着元宇宙赋能游戏行业，承载游戏的终端设备也将随之迭代，游戏行业或将进入以 VR/AR 等技术为核心的可穿戴设备时代。

AI+游戏提升生产力，有望打破“不可能三角”。据 a16z，游戏行业存在着不可能同时攻破成本、质量或速度三项的“不可能三角”。然而，在 AIGC 的加持下，人工制作的时间有望缩短，能够大大提升游戏策划、音频、美术、程序等环节生产力，有望打破“不可能三角”。

图表104：游戏业态发展历程



资料来源：中国音数协游戏工委、华泰研究

在素材原画/代码编写/开发测试等领域显著降本增效

游戏作为涵盖文本、图像、声音、视频、3D 内容及交互性的复杂娱乐形式，本身具有相对较高的创作准入门槛及制作成本。AIGC 在游戏领域的应用有望全面带动游戏行业生产革命。从 UGC 到 AIGC 的转型，不仅较大程度地丰富了游戏内容，为玩家参与游戏创造提供了无限的想象空间，提高了玩家游戏参与度，建立用户粘性，还通过 AI 生成图像、音频、视频等高质量游戏内容进一步提升了游戏表现力，最终使游戏玩家体验升级。

免责
更多免费行业报告 并购家 www.ipoipo.cn

1）AI 辅助角色原画生成

AI 与人工配合下，角色创意和场景概念图等高耗时环节的生产效率有望被大幅提升，已有工作室通过应用 AI 创作角色，节省成本超 80%。据 Gamelook，俄罗斯游戏工作室 Lost Lore 通过在角色设计阶段应用 AI 相关技术后，将角色创作的时间及开发成本大幅缩短。**制作 100 名角色由 5 万美元的开发成本及 6 个月的工时缩短至 1 万美元及 1 个月。**该工作室选用 Midjourney，采用 AI 文字输入调整配合人工调整的方法，利用 AI 快速生成角色造型。在加载了已经画好的部分角色作为参考的基础上，使用详细的文字指令（Prompt）来规定新创作角色的特征，包括主体颜色、道具、角色姿势、背景元素等等。比如生成左下图的指令为“一只行军之中的邪恶灰熊，穿着铁制的盔甲和头骨装饰，动态的攻击姿势，古老的头盔和防毒面具，后启示录风格”。随后，美术画师基于 AI 生成结果对画面细节进行进一步监修，大幅节省前期创意时间。

图表105：基于已画好角色及详细文字指令生成的新角色形象

资料来源：Bearverse 官网，华泰研究

2）AI 助力游戏代码编写

游戏代码编写方面，AI 已取得较大突破，OpenAI 等大厂持续布局。GitHub Copilot 功能源于 OpenAI Codex Models。OpenAI Codex Models 模型是强大的代码生成工具，能根据用户输入的自然语言指令生成相应的程序代码。由数十亿代码进行训练的模型可以精确处理原本耗时长的重复性工作，从而极大解放程序员的工作压力，使其将更多精力运用到创造性工作上。GitHub 与 OpenAI 合作，将该模型的能力代入 GitHub 的环境中，最终提供 GitHub Copilot 服务：代码生成、完善、优化服务。根据用户提供的指令（可以是文字或语音），GitHub Copilot 能自动生成相应环境下的代码语言，对已有代码在用户的指令下进行完善，并提供相应的代码优化服务。GitHub Copilot 开启了开发环境的新时代，较大地提升了开发效率，释放创意。

更多免费行业报告 并购家 www.ipoipo.cn

3）AI 提升游戏开发测试效率

AI 能够低成本、大规模生成 NPC，辅助英雄的开发测试效率。据腾讯官网，王者荣耀的研发存在几大难题：地图庞大、操作序列空间大、需要多智能体配合协作、需要大量复杂的即时与长期博弈。2017 年，腾讯 AI Lab 启动“绝悟”研发。经过多次迭代演进，绝悟 AI 飞速成长。2021 年 7 月，腾讯 AI Lab 在 2021 游戏开发者大会亮相，首次全面展示了以“绝悟”为代表的 AI 技术在游戏产业全链路中的研究与应用能力。

图表106：绝悟 AI 能力演进路线

资料来源：腾讯 AI 实验室、华泰研究

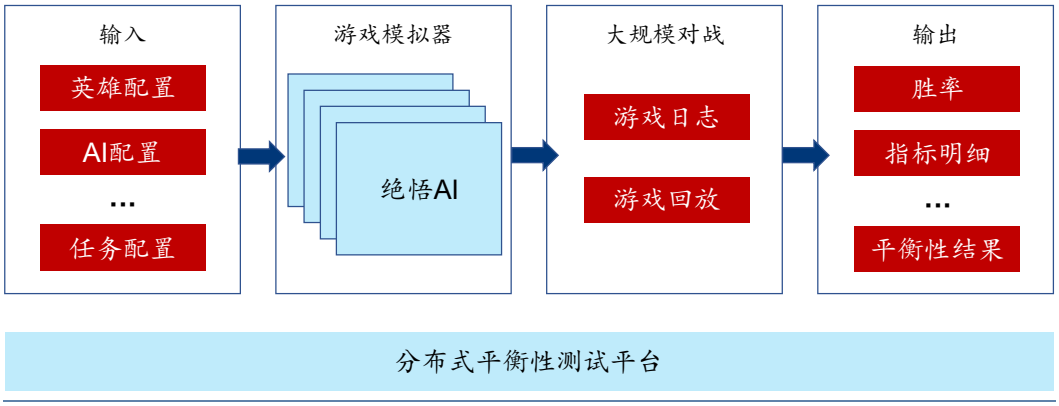
低成本、大规模生成 NPC。传统方法生成 NPC 的成本很高，制作一分钟人物动画就需要耗费数天，使用 AI 可以高效低成本地实现生产。在人物外观上，可以通过 AI 对抗生成网络实现「千人千面」，快速生成高保真 AI NPC，搭配不同的发型和妆容，为创建海量角色奠定基础；在人物行为上，通过整合文字转语音的前沿研究，生成自然语音，同步驱动嘴型、表情等面部变化，达到高度逼真。利用相位神经网络技术控制人物运动，将原本数周工作量减少到几小时。AI 在大量的运动数据集上学习人体动作，将捕获的运动数据与场景地形数据相匹配，完成准确、流畅的动作反应，如行走、跑步、跳跃等。

辅助英雄平衡性测试。据腾讯 AI 实验室介绍，MOBA 游戏核心玩法在于多样化的英雄选择和战斗体验，过强或过弱的英雄都会破坏玩家体验，因此保持每个英雄胜率 50%左右是王者荣耀平衡性系统的目标。1）在测试准确性方面，按照传统测试方法，由于正式服和测试服玩家的真实能力和游戏心态不一致，会导致测试结果不准确。“绝悟”AI 通过强化学习的方法来模仿真实玩家，包括发育、运营、协作等指标类别，以及每分钟手速、技能释放频率、命中率、击杀数等具体参数，让 AI 更接近正式服玩家真实表现，将测试的总体准确性提升到 95%；2）在测试效率方面，传统的平衡性测试是调整数值后在测试服上验证，周期较长，每个英雄测试大约需要 1-2 个月。

“绝悟”AI 通过分布式平台和推理优化节省了大量的时间。分布式平台支持大规模并发测试，开发者只需要输入英雄的修改数值，平台就会在短时间内自动输出测试报告。推理优化则通过模型剪裁、量化、自研推理框架来减少模型的计算量和加快推理速度，实现 3 小时完成 20 万场对局。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表107：绝悟 AI 分布式平衡性测试平台示意图



资料来源：腾讯 AI 实验室、华泰研究

4）AI 助力营销图文视频素材生产及创意

游戏营销行业开发、转化效率已遇瓶颈，未来“AI+人”创作状态将持续深化。游戏营销行业经过多年的发展，已能较好实现投放精准和人群精准。目前，从业者逐渐将精力投入到内容创意的生产研发，试图从投放精准和人群精准层面深化发展为内容精准，从而引起受众的注意，该过程需要 AI 提供较大辅助。未来游戏营销工作中“AI+人”状态将持续深化，最终呈现由工作人员输入创意思路，AI 依据相应方向输出精美图片、视频等营销素材的模式。AI 将助力素材生成效率的不断提升，降低相应成本，实现广告的针对性批量铺设。

增加用户付费和留存，AI+创新玩法探索

AI 提升游戏 NPC 对话效果，辅助用户精准预测与留存，助力运营效率提升。一方面，AI 的应用有望实现游戏 NPC 的智能交互升级，提升游戏的交互性及自由度，实现千人千面的游戏体验；另一方面，数据分析贯穿游戏生命周期。游戏用户数量监测、用户维持、流失预警等都包含在数据分析中。AI 技术增强了数据分析的能力，游戏开发者利用 AI 技术在数据分析方面的能力推进用户维持。

1）创新游戏玩法，实现 NPC 实时开放性互动

逆水寒实装国内首个游戏 GPT，NPC 智能交互提供全新游戏体验。与市面上大多类“ChatGPT”产品定位于对话服务、搜索协助等不同的是，“逆水寒 GPT”定位为首个能在具体情境中应用、与游戏机制深度融入结合的、用来丰富游戏虚拟世界的完全由 AI 驱动的高智能 NPC 系统。据逆水寒手游公众号，作为国内 2023 年第一款生成式 AI 游戏，其自由度、开放度高，重点体现在：

1）无设定自由交互。削弱游戏“主线剧情”概念，智能 NPC 成为虚拟世界的一份子，除了能和 ChatGPT 一样进行智能对话，还能进行游戏式互动，并赋予 NPC 内在性格特质，被赋予“灵魂”，与玩家之间的差异进一步缩小；

2）游戏内容无限生成。智能 NPC 可以基于对话语义进行无限延展，并记住该对话，从而使被玩家改变的 NPC 在再次相遇时维持被改变的特性。此外，NPC 之间会互相交流、影响，如与其畅聊“高启强卖鱼”的故事，NPC 也觉精彩，便会出现人传人、传遍城，隔天汴京城人人都在寻找传说中“高启强的鱼档”。NPC 之间形成了社会关系网，玩家的每一个行为都可能触动这个网，并产生蝴蝶效应；

3）玩家 0 门槛创作内容。游戏推出打字捏脸，玩家可通过输入文字得到自生成脸部模型，此外，逆水寒端游还有一款功能强大的“大宋映画”游戏视频编辑器，玩家能通过文字指令拍摄视频。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表108：《逆水寒》手游

资料来源：逆水寒手游官网，华泰研究

图表109：《逆水寒》手游

资料来源：逆水寒手游官网，华泰研究

2）AI 通过分析数据，提升用户留存率

数据分析贯穿游戏生命周期。游戏用户数量监测、用户维持、流失预警等都包含在数据分析中。AI 技术增强了数据分析的能力，游戏开发者利用 AI 技术在数据分析方面的能力推进用户维持。据微软官网，如 Azure PlayFab 能够通过获取 21 天游戏运营数据和 7 天玩家数据，AI 根据数据利用已有模型对用户进行打分与分类，确定相应玩家的流失风险类别：高风险、中等风险和低风险。在获取高流失风险用户后需对其进行减缓策略，PlayFab 细分模块中缓解用户流失的行为包括发送邮件、发放金币与礼物、定制个性化模块与功能等。

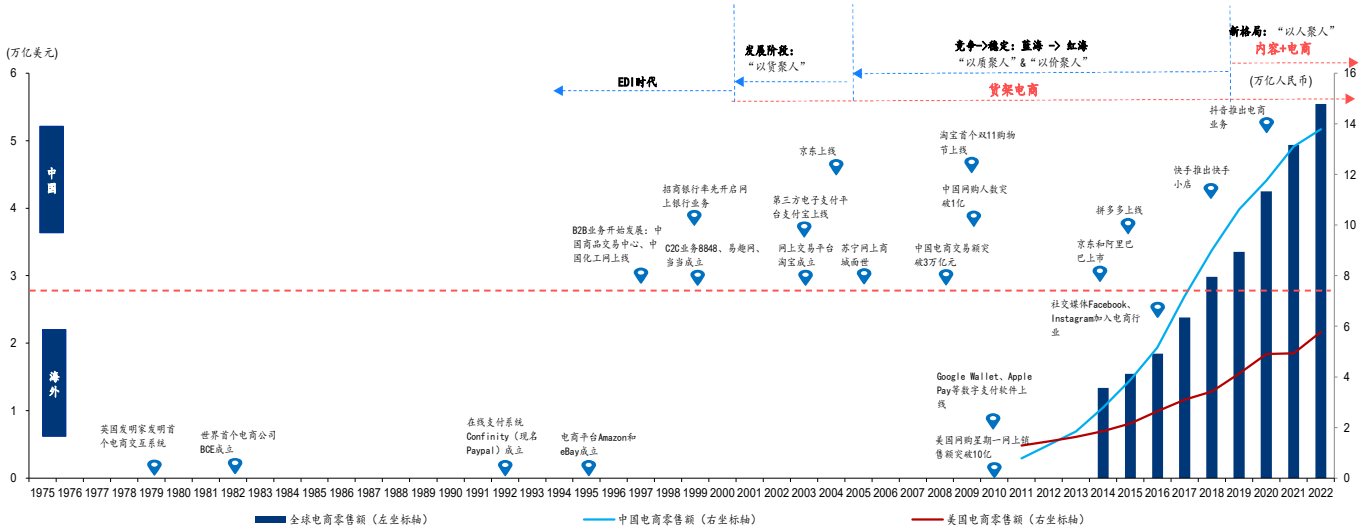
AI+电商：多场景应用，降本增效提升体验

内容电商形式出现打破传统货架电商格局。电商最早起源于英国发明家 Michael Aldrice 发明的电商交互系统，早期以电子数据交换（EDI）为主要商业模式，进入 90 年代，基于互联网的电子商务逐渐被开发，海外电商龙头企业亚马逊和 eBay 均于 1995 年成立，中国最早的电子商务公司中国商品交易中心和中国化工网于 1997 年上线。随后进入中国电商元年 1999 年，大量电商平台成立，货架电商进入“以货聚人”发展阶段。2003 年，电商巨头淘宝和京东相继成立，C2C 成为主流模式，中国电商进入初期竞争阶段。2010 年，中国电商高速增长，电商市场逐渐由蓝海变为红海，进入稳定发展阶段，为寻求突破，传统货架电商开始通过“质”与“价”的优势吸引消费者。2018 年，随着社交电商拼多多的迅速崛起，短视频软件快手开始布局电商行业推出快手小店，中国电商开始出现新格局，内容电商时代由此打开，突破了中国长达 20 年的货架电商模式，电商行业开启“以人聚人”新模式。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表110：全球电商行业发展



资料来源：知网、eMarketer、商务部、华泰研究

电商行业分为“人-货-场”，互联网助力达成“千人千面”。我们认为，电商行业分为“人-货-场”，不同于传统零售，电商的“场”主要表现为 APP、小程序、官网等线上渠道。随着“内容电商”时代到来，商品能够通过短视频、直播等形式被分发给消费者，实现“货找人”的发现式购物方式。同时，互联网助力实现了“千人千面”的精细化运营手段。在电商购物过程中，用户点击、收藏等数据将被记录在数据库中，平台依据此对用户进行针对性分析，针对性推送客户感兴趣的内容，提高商品曝光机会，大大增加用户粘性。

电商行业天然具有海量的交易数据资源、多样的营销策略，可以快速受益于 AI 技术所带来的效率提升。在优化选品决策、智能客服工作、图文素材制作、广告智能投放、精准关联推荐、处理图像素材等领域，AI 持续赋能降本增效，覆盖多应用场景，基本上可以贯穿电商运营的每个环节，带来进一步的用户体验提升。

优化选品决策，准确识别标签，助力图像生成

预判顾客选品需求是电商行业的关键一环，也是 AIGC 应用的关键场景。海外电商龙头亚马逊早在 2021 年 4 月已首次推出官方 AI 选品工具，通过大数据及人工智能算法，使用 600 多个数据维度，个性化预测卖家在不同国家所售产品的全球需求。生成式 AI 推出后，点峰数字科技的牛卖浏览器为亚马逊卖家免费打造全新 AI 工具——ChatGPT 亚马逊运营助手，具备 AI 生成、AI 解析、AI 优化、AI 检测等功能。在选品决策方面，该工具可深度挖掘产品优缺点、深度分析客户购买动机。

AI 能够准确识别图片标签，提升图片利用率。阿里云的视觉智能开放平台提供广告素材分析接口，可以对素材图片中的人物（明星、素人、CG 人物）、场景等打上标签信息，并可支持数千个内容标签；素材分析处理功能可以对包含明星、素人、CG 人物等复杂场景进行准确打标，将物体、人物相关标签同步返回并精准标记标签位置。**百度也拥有“定制化图像识别技术”**，该技术可以在很短的时间内训练出若干分类标签，然后对数百万张图片进行识别。根据百度 AI 开发平台给出的来自家图网案例的数据，这项技术使图片的利用率获得从 30%到 80%的提升。百度通过对相似图片推荐的改进，使图片点击率上升了 30%；百度将训练结果应用到商品识别和购买推荐上，品类识别准确率从 50%，上升到 95%。

此外，AI 依靠多模态创作技术，可实现文字、图片、音频、视频等素材的自动生成，具有一定的创造力。通过使用 AIGC，可以在短时间内批量生成大量营销内容，提升营销内容产出效率的同时，也可以提供营销创意方向，激发创作者灵感。在 OpenAI 开放 API 接口后，ChatGPT 的能力也可以被使用在更多的场景当中。

免责
更多免费行业报告 并购家 www.ipoipo.cn

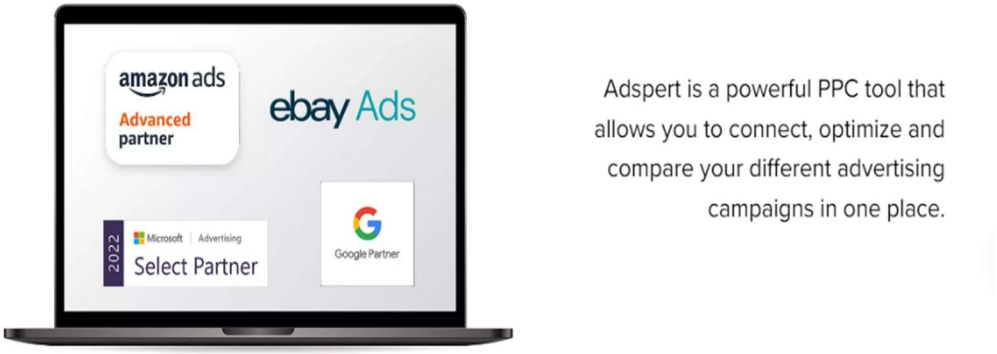
万兴播爆助力出海营销，一键生成“真人”短视频。万兴科技旗下“万兴播爆”是其布局 AIGC 文生视频赛道的最新应用，在 2023 年 4 月的中国跨境电商大会中亮相。据公司官网，万兴播爆通过对抗生成、唇形迁移、口型矫正等 AI 技术，打造媲美真人的 60+多国籍数字人形象，涵盖多人群、多肤色。此外，万兴播爆联动海外和本地化设计专家制作上百套专业级的爆款视频模板和 10 万+资源素材，满足各类行业需求，具有高人效、短周期、低成本等特点。

广告智能投放，精准关联用户，提高购买意愿

泛品类跨境电商由于 SKU 规模庞大，存在单位推广成本高、转化率低的痛点，大部分卖家依赖自然流量销售商品，在付费推广方面投入较小。AI 技术和相关算法可实现广告的智能投放，降本增效。

国外方面，Adspert 是一款广告全自动化 AI 工具，是谷歌、微软、亚马逊等公司广告的官方合作伙伴，主要服务于跨境电商的卖家客户群体。通过人工智能，Adspert 利用广告利润最大化原理优化亚马逊广告竞价投放，增加销售额，提升投产比。同时，通过大数据云计算技术，该工具能够每天根据搜索词报告向账户推荐或者自动添加否定和肯定关键字以及 ASIN。

图表111：Adspert 合作伙伴



资料来源：Adspert 官网、华泰研究

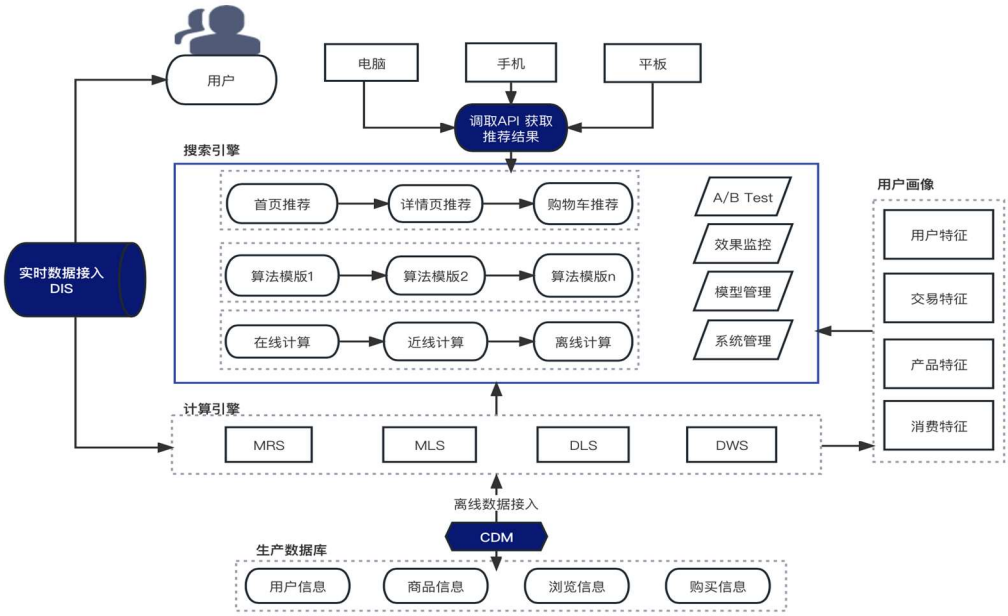
国内方面，华凯易佰自主研发的“智能广告投放运行监控分析统计系统”针对不同业务场景和时间阶段的广告需求，设置差异化投放策略，并对广告数据进行跟踪，对销售业绩的快速增长有直接促进作用。根据公司 2023 年 03 月 27 日的投资者互动回复，以亚马逊为例，2022 年公司在亚马逊的广告订单金额在总销售收入的占比约 31.4%，即通过付费推广在自然流量基础上带来的增量业绩，有效提升了广告投放效率。

精准推荐中的常用模型有：交叉销售模型、关联推荐匹配模型、互联网信用模型、电商动态定价模型、信息聚合分类模型等等。以近年流行的信息聚合分类模型为例，先对来源分散的信息资源进行筛选、组织、整合，再分别分类。AI 模型可以通过社交网站采集数据并进行特征归类，自动将每一条信息划归最合适的分类区；与此同时，模型通过定期对数据进行自学习与更新，产生新的分类建议。此外，AIGC 还可以根据个人喜好生成个性化内容，在此基础上进行虚拟商品的精准推荐，将提高购买率。

免责
更多免费行业报告 并购家 www.ipoipo.cn

华为云电商智能推荐解决方案，实现用户精准洞察。基于对用户海量访问、交易等数据的深度挖掘分析，华为云推出了电商智能推荐解决方案，以帮助电商企业建立精准的用户模型。借助华为云大数据服务，推荐系统可以实现秒级更新用户画像、物品画像、召回候选集等。可实时从海量数据中挖掘用户兴趣，及时捕捉用户的兴趣变化，确保可在合适的时机对客户进行精准推荐。该智能推荐方案实现了数据模型的自动更新，可以自动发现特征，并根据客户实时变化的信息动态调整数据模型，让推荐更为精确。同时，该方案还支持获取单个用户的推荐历史记录及用户画像，运营人员可以通过分析具体用户的产品体验找出改进的机会，持续提升商品的提袋率，降低库存积压。

图表112：华为云电商智能推荐解决方案架构



资料来源：华为云官网、华泰研究

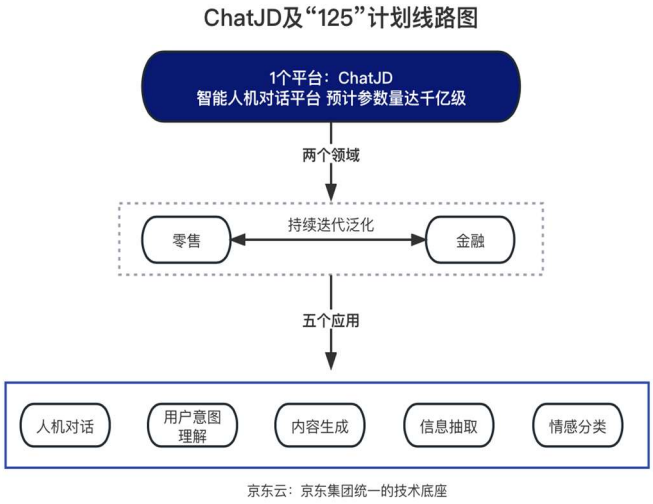
智能客服工作，提升产业化服务水平

国内方面，京东云旗下言犀人工智能应用平台推出 ChatJD，定位于应用高频的产业版 ChatGPT。ChatJD 以“125”计划作为产品的应用路线图，具体包括一个平台、两个领域、五个应用。其中，一个平台即 ChatJD 智能人机对话平台，该平台将广泛应用于京东云在零售和金融领域多年积淀的知识体系和子系统，表现为人机对话、用户意图理解、内容生成、信息获取和情感分类五种高频的应用场景。

此外，京东云在多种实体场景中已经逐步得到应用。京东云言犀团队目前已打造并应用于零售和金融多个场景的智能应用，零售方面包括京东智能客户服务系统、京小智平台商家服务系统、言犀数字人等，而金融领域则将智能金融服务大脑应用于行业之中。据京东云 2023 品牌增长峰会，目前公司已与超过 5.8 亿终端客户及 17.8 万京东第三方商家开展多场景领域的深度合作。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表113：ChatJD 及“125”计划路线图



资料来源：京东云官网、华泰研究

焦点科技推出 AI 外贸助手“麦可”打造外贸企业的“数智大脑”。2023 年 4 月 4 日，在公司投资者互动平台回复中，明确目前公司研发端接入的是 ChatGPT 3.5 Turbo，ChatGPT-4.0 正在接入审核中。4 月 25 日，焦点科技旗下中国制造网（以下简称“MIC 国际站”）推出 AI 外贸助手，以品牌 IP“麦可”的 AI 形象呈现给网站用户，为外贸企业提供 AI 智能化服务，赋能企业开拓海外市场。在线上店铺运营过程中，产品推广效果差、开发客户效率低、不懂小语种错失大单等难题，常常使得外贸企业的获客成本不断攀升。针对这些难题，麦可为外贸企业提供了一系列解决方案。**在营销推广场景**，麦可是每一位供应商的“AI 营销专家”，可以实现从市场洞察到营销文案的综合产出，多场景撰写外贸专业开发信、智能改写产品描述，并实时进行 SEO 检测和优化提醒；**在品牌出海场景，**麦可又能成为供应商的“AI 创意大脑”，能以企业品牌形象打造、社媒贴文创意策划、Blog 自动生成等功能驱动品牌内容生产自动化；**在采洽沟通场景，**麦可承担了“AI 全能助理”的重要角色，当收到国外买家的询盘时，麦可能智能提炼买家的诉求和关键点，并实现对话内容的多语种翻译，贴合不同对话场景进行自动回复，辅助外贸人员工作。据焦点科技官网，麦可后续有望持续迭代升级，未来将具备数据的整合、挖掘及分析能力，帮助外贸企业洞察国际市场的行业趋势，清晰了解买家的背景实力，提高国际贸易中的风险防范能力，为企业提供智能化的决策辅助和策略参考，最终成为外贸企业的“数智大脑”。

5 月 15 日，中国制造网召开 AI 麦可发布会。麦可展示了全方位的能力，包括：多语种实时翻译，助力外贸企业开发多语市场；秒速处理产品图片，解决外贸企业缺少专业美工的痛点；一键生成开发信、社媒贴文等多场景内容，帮助外贸企业提高营销推广效率；智能改写产品描述，并能实时进行 SEO 检测和优化提醒，促进外贸企业多方位提质增效。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表114：焦点科技 2022 年度 AI 研发投入项目

主要研发项目名称	项目目的	项目进展	拟达到的目标	预计对公司未来发展的影响
智能销售系统	提升产品性能及市场竞争力	已完成	提升中国制造网等业务平台 的推荐能力，实现多形式、多场景的精准推荐	提升产品性能及市场竞争力，提高用户体验和满意度
基于深度学习的智能推荐系统	提升产品性能及市场竞争力	已完成	提升中国制造网等业务平台的推荐能力	提升产品性能及市场竞争力，提高用户体验和满意度
智能定向推广系统	提升产品性能及市场竞争力	进行中	丰富定向推广的投放形式和渠道，增加新颖的视频展示方式，提高定向推广流量和曝光，实现智能定向推广	提升产品性能及市场竞争力，提高用户体验和满意度
产品智能发布系统	提升公司内部办公效率	已完成	提高平台运营人员产品编辑效率	提升公司内部办公效率，更好的为客户服务，提升用户体验和满意度
智能图片设计系统	实现自动抠图、图片合成功能，满足中国制造网等平台图片智能化设计的需求。	进行中	自主研发智能图片设计系统，实现自动抠图、图片合成功能，提供个性化 Banner 图片生成服务。	提高业务平台图片生产效率，提升平台用户体验，未来还可以向供应商直接开放，创造一定的收益。

资料来源：公司年报、华泰研究

国外方面，HubSpot 已接入 OpenAI 的 ChatGPT 模型。HubSpot 作为知名的海外营销、销售和服务平台（CRM），将自身的 CRM 技术与 ChatGPT、DALL-E 2 和谷歌文档应用程序相结合，于 2023 年 3 月推出生成 AI 智能营销助手 ChatSpot。ChatSpot 的推出可以简化操作流程、智能回答问题、智能生成文案，并通过 ChatGPT 接口调用外部数据。具体而言，之前客户在 HubSpot 的 CRM 菜单中需要逐一点击后输入客户信息，引入 ChatGPT 之后仅需将信息发送给 ChatSpot 即可，简化操作流程并提升共组效率。此外，ChatSpot 可以辅助进行市场研究，同时可以根据上下文生成邮件回复客户的短信息，能够优化客户服务，提升交互能力。此外，ChatSpot 除可以调用 HubSpot 内部数据外，还可以通过训练 ChatGPT 大模型并通过接口调用外部数据，对于宏观数据进行调用分析市场总体情况。

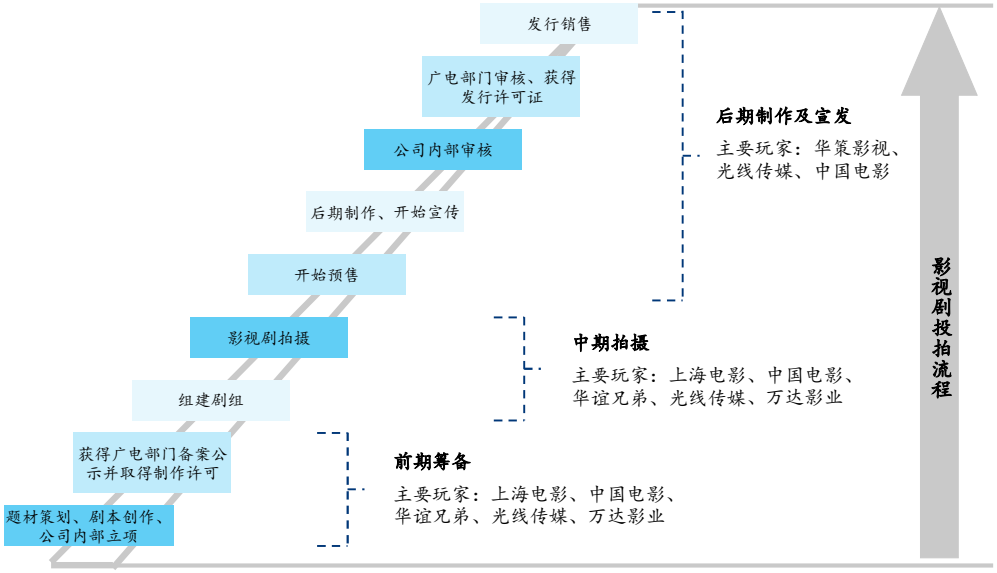
AI+影视：驱动影视创作升级

影视剧的制作流程主要包括前期筹备、中期拍摄、后期制作及宣发阶段。前期筹备阶段包括题材策划、剧本创作、公司内部立项及获得广电部门备案公示并取得制作许可；中期拍摄阶段包括组建剧组及影视剧拍摄；后期制作及宣发阶段包括预售、制作、宣传、公司及广电部门审核、获得发行许可证后发行销售。目前 AI+影视的应用仍处早期摸索尝试阶段，重点在于完善技术以及探索需求，但随着 AI 生成视频领域不断地取得技术突破，AI 技术将逐步赋能影视创作的各个生产环节，其作用包括：

1）提高制作效率。AI 能够代替人工自动化处理影视制作过程中的任务，包括视频剪辑、声音录制、图像渲染、特效制作等方面，降低影视制作的成本，缩短影视制作周期，提升制作性能。**2）增强创意发挥，**扩张影视行业产能。AI 能够快速地生成创意想法，提供创意灵感。例如，相比于人工编剧，AI 具备更广阔的知识储备，能够增强创意发挥，帮助编剧打破创作的产能瓶颈，扩张行业产能。**3）为影视制作带来新模式和新思路。**全球众多用户的口味和个人并非千篇一律的，每一个影视创作者要去综合、满足大部分人的口味需求，而 AI 可以辅助分析和预测市场诉求，抓住观众需求点以制定发行策略，为观众提供个性化推荐，打造智能化和个性化的影视作品。**4）提高观众的体验。**AI 能帮助影视制作人员更好地了解市场近期热点和需求偏好形态，从而提供更具吸引力和趣味性的内容，为观众带来更佳的观看体验。

更多免费行业报告 并购家 www.ipoipo.cn

图表115： 影视剧投拍流程



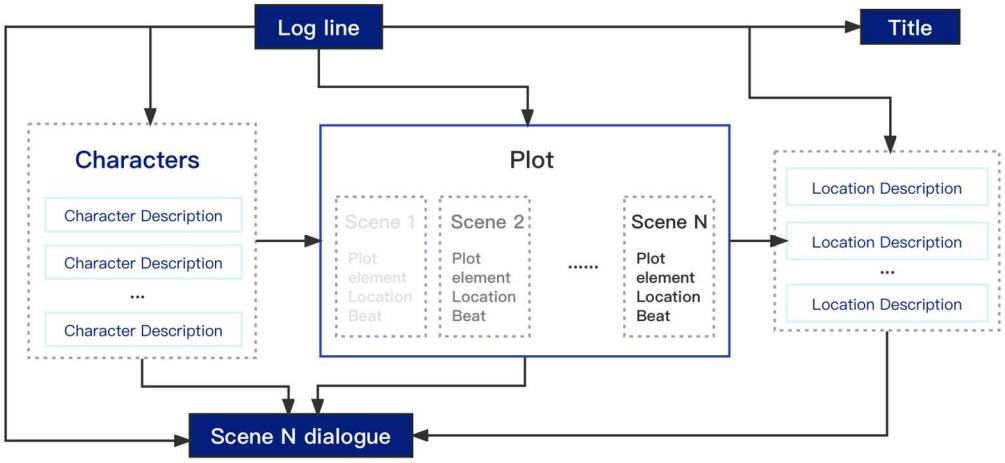
资料来源：华策影视官网、华泰研究

AI+前期筹备：助力剧本生成和数据分析

我们认为，AI 在影视前期筹备阶段将从两方面发挥作用：**1）剧本生成：**通过对海量剧本进行学习，AI 可以自动生成批量化的剧本，并按照任务性格生成对话，激发作者创作灵感，缩短创作周期。**2）数据分析：**通过对同类型电影、用户反馈、剧本内容等进行分析，AI 可预测票房回报等数据，为选择演员、拍摄地点、预算投入提供依据。

由 DeepMind 开发的应用语言模型 Dramatron 可根据用户提供的中心句生成标题、角色、场景描述和对话。据 DeepMind 研发专家 PIOTR MIROWSKI、KORY W. MATHEWSON 在 2022 年 9 月 29 日发布的学术研究《Co-Writing Screenplays and Theatre Scripts with Language Models An Evaluation by Industry Professionals》，Dramatron 使用分层故事生成（hierarchical story generation）方法生成脚本和剧本。从输入中心句（logline）开始，Dramatron 可以生成一个完整的剧本，包括标题、角色列表、情节、位置描述和对话，用户可以在分层生成的任何阶段进行干预，实现人类和 AI 在交互过程中共同编写脚本。剧作家们认为 Dramatron 可以为人类提供灵感、构建故事背景及生成内容，并可能用于文学分析。

图表116： Dramatron 的分层连贯故事生成



资料来源：DeepMind《Co-Writing Screenplays and Theatre Scripts with Language Models An Evaluation by Industry Professionals》（2022 年 9 月 27 日）、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

海马轻帆“剧本智能评估”功能可从剧情、场次、人物维度评估剧本。据海马轻帆，由海马自主研发的“剧本智能评估”3.0 功能可对院线电影、网络电影、电视剧、网剧等影视作品的剧本进行智能数据化分析。项目方上传剧本文档后，“剧本智能评估”可对影视剧本从剧情、场次、人物维度提供 AI 自动化数据分析。**1）剧情分析维度**，该服务可智能生成剧情曲线，展示对应剧本内剧情冲突的高低起伏，并借此分析剧情整体节奏、重要剧情设置的合理性。**2）场次分析维度**，该服务能够识别重要场次及整体分布比例，从而判断该剧本的文本结构、冲突节奏分布、重要角色场次排布的合理性。**3）人物分析维度**，该服务可以按照剧本角色互动关系生成人物关系网，计算主要角色之间的互动及戏份占比情况，并从人物命运转折维度分析人物成长性。

海马轻帆“剧本智能评估”功能对剧本可商业开发性提出建议。“剧本评估服务”还可应用于剧本商业化测评、指导改编、指导开发、艺人筛选剧本等多个场景，帮助影视公司解决剧本初期筛选工作的高耗时低产出、剧本可商业开发的标准判断与风险规避、剧本内容质检与后期制作开发等问题。

AI+中期拍摄：关注虚拟场景与虚拟人生成

在影视中期拍摄过程中，AI 能够帮助生成虚拟场景，同时在影视中引入虚拟人等，提升创作效率，降低时间及人力成本。一方面，AI 可帮助生成动画中的场景，帮助完成高重复性工作，或现实生活中难以实拍或成本过高的场景。另一方面，利用 AI 技术能够实现演员年龄的跨越、重现已故演员及同台呈现 CGI 角色和真人演员。

AIGC 技术辅助商业化动画片的首支发行级别作品《犬与少年》开播。据 Netflix，2023 年 1 月 31 日，Netflix 与微软小冰、WIT STUDIO 共同创作的首支 AIGC 动画短片《犬与少年》正式发布。《犬与少年》是一款典型的日式风格动画，主要讲述了一只机械狗与主人从幼年到暮年，相互陪伴、相互守望的故事。在这部电影中，AI 主要负责部分动画场景的绘制。

AI 在动画制作中承担繁琐工作，提升动画创作效率。在动画结尾，制作团队展示了 AI 辅助场景制作的过程，共分为布局（layout）、一次 AI 生成、二次 AI 生成、成稿（Final BG）四步。画师先手绘大致的场景，然后交给 AI 进行 2 次生成，最后画师在 AI 生成内容的基础上进行修改。其中，繁琐的第二步和第三步完全交由 AI 操作。《犬与少年》摄影导演田中宏侍认为 AI 将提升动画创作效率，使用 AI 能为同时参与多部作品制作的创作者提供更加充裕的时间。

图表117：《犬与少年》动画片画面

资料来源：Youtube、华泰研究

图表118：《犬与少年》场景制作第二、三步交由 AI 完成

资料来源：Youtube、华泰研究

CGI 技术在影视拍摄过程中广泛采用。目前 CGI（Computer-Generated Imaging，电脑生成动画）已是影视作品中不可或缺的一部分，贡献出多个经典的角色。比如，《阿凡达》里潘多拉星球的土著“纳美人”，漫威宇宙里作为强大战力输出的绿巨人，及《指环王》中全 GC 角色咕噜。据《阿凡达》幕后制作纪录片，拍摄时演员穿着带有标记点的动作捕捉服，头上佩戴配备摄像头的特殊装置进行表演，计算机能够实时创建演员的骨骼，用以驱动 CG 角色的运动。演员头上离面部只有几英寸距离的微缩高清摄像头能够实时计算出面部肌肉的走势和运动，以此得到最真实的面部细节和表情变化。

更多免费行业报告 并购家 www.ipoipo.cn

图表119：《阿凡达》实地拍摄场面与虚拟摄像机实时预览画面对比

资料来源：Youtube、华泰研究

AI+后期制作：在画质增量、人脸替换等领域落地

后期制作阶段需完成对影视作品进行画面剪辑、配音或修音、特效制作、配乐创作、海报及预告片宣发制作等环节，而我们观察到，AI 技术能够成熟地运用到配音、译制、剪辑、修复、转制以及后续宣发上，实现后期制作阶段的降本增效。

1）画质修复及增强

AI 能够修复还原受损的影视图像，进而保障影像资料的清晰度及画面质量。例如，由中影基地与中科大共同研发的“中影·神思”人工智能图像处理系统，能够对影视项目的图像视频素材进行去场纹、修复、超分增强等自动处理，相较于传统的处理方法，“中影·神思”系统处理影像的效率及质量均更优。据中国电影网，中影基地使用“中影·神思”系统后，仅在四个月内便处理了《厉害了，我的国》超过三十万帧的低清图像素材，电影成片中共计去场增强画面 74 分 25 秒，完成了传统数字修复手段不可能实现的任务，为《厉害了，我的国》成为国产纪录电影票房冠军提供了坚实的技术保障。

图表120：“中影·神意”AI 画面增强效果对比

资料来源：中国电影网、华泰研究

图表121：“中影·神意”AI 胶片修复去重影效果对比

资料来源：中国电影网、华泰研究

2）人脸替换及人声替换

AI 能够辅助演员面部的识别匹配，助益后期制作。如果前期演员拍摄过程中有所不足需要补拍，可以应用 AI 技术直接在后期弥补，避免重复工作。若视频中出现“劣迹艺人”等敏感人员也可通过 AI 技术对视频进行编辑，进行“人脸替换”。例如，在《了不起的儿科医生》拍摄完成后，扮演雷皓一角的演员由于陷入学历造假争议，全程被“AI 换脸”，继而减少了影视剧制作方及投资方的损失。

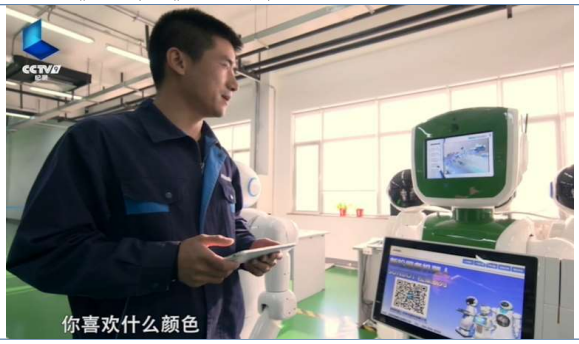
更多免费行业报告 并购家 www.ipoipo.cn

同时，在后期制作流程时，AI 能够通过语音识别、语音合成技术来处理人声，辅助进行音色微调或修音，也能够为影视作品合成配音。例如，科大讯飞采用智能语音及 AI 技术，通过采集已逝的著名配音艺术家李易老师的声音资料，让李易老师的声音重现荧幕，完成了纪录片《创新中国》的配音。我们认为，随着 AI 配音技术逐渐成熟，AI 有望替代部分配音演员，实现高效的自动化配音，为电影制作省去一定的配音成本。

图表122：《了不起的儿科医生》采用 AI 换脸劣迹演员

资料来源：优酷、华泰研究

图表123：《创新中国》纪录片中采用 AI 技术合成已故艺术家的配音



资料来源：《创新中国》纪录片、华泰研究

3）辅助创作影视配乐

AI 音乐公司“AIVA”通过深度学习技术，能够对莫扎特、巴赫、贝多芬等名家作品进行学习，提取出古典音乐特征，并进一步根据这些特征建立数学模型，来辅助创作丰富的乐曲，用户只需选择预定义的样式或上传原始声音文件，即可在 AIVA 中生成音乐。我们认为，AI 音乐生成器将辅助影视配乐制作，能够按照预定风格批量化生成配乐，提高配乐创作效率，缩短创作周期。

4）特效生成

AI 能够辅助后期特效制作，不仅能够缩短特效场景的制作时长，也能够为影视剧带来更好的视觉效果。例如，在科幻电影《瞬息全宇宙》中，视觉效果团队采取了 Runway 公司的技术来创建特效镜头，在短时间内便尝试了多种特效效果，在特效制作过程中借助 AI 技术实现了降本增效。

图表124：《瞬息全宇宙》特效画面

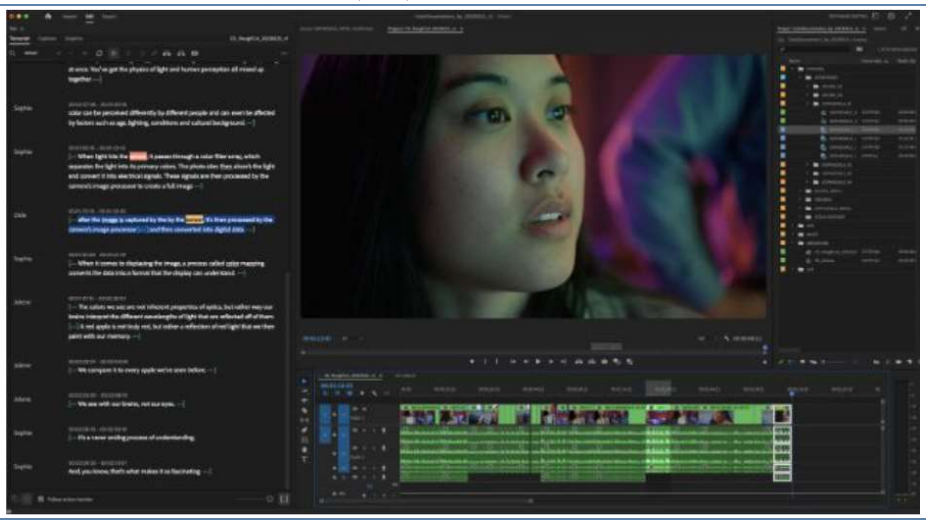
资料来源：《瞬息全宇宙》电影、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

5）辅助后期视频剪辑

AI 能够自动识别视频、填充或删除相关内容，辅助提升后期影视剪辑效率。例如，Adobe Premiere Pro 推出了 AI 文字编辑视频功能，允许用户通过修改自动生成的视频文本，匹配特定的单词和短语来快速编辑视频，大幅提升了视频编辑效率和速度。

图表125： Adobe Premiere Pro 推出 AI 文字编辑视频功能

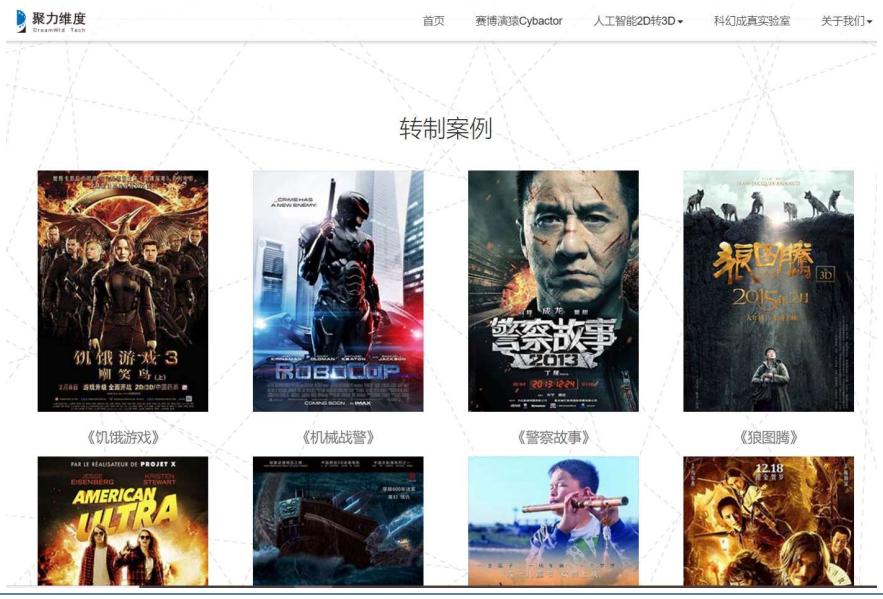


资料来源：Adobe Premiere Pro 官网、华泰研究

6）辅助影视内容由 2D 向 3D 转制

传统的 3D 影视制作模式一般由实拍+后期合成，或者纯后期 2D 转 3D 而成，但实拍模式成本高、制作流程复杂、人力消耗多、时间周期长，同时纯后期 2D 转 3D 工序对每一个制作环节要求细致，也需耗费大量的人力及资金成本。而 AI 技术推出后，能够实现将影视内容快速由 2D 向 3D 自动转制，大幅提升影视作品 3D 转制效率。例如聚力维度推出的人工智能 3D 内容自动制作平台“峥嵘”，目前已在云端实现 2D 转 3D 功能，能够随时随地将普通的 2D 图片、视频转制成影院级 3D 视效的图片、视频。据聚力维度官网，“峥嵘”相比于人类立体设计师具备精确、高效、稳定的优势，在 3D 转制时效上，是人类立体设计师的 1250 倍，仅用一周时间就能够独立制作一部 3D 电影，目前已成功转制了《狼图腾》、《机械战警》、《饥饿游戏》等一系列影视作品。

图表126： “峥嵘”转制的影视作品案例



资料来源：聚力影视官网、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

产业链相关公司：昆仑万维、汤姆猫、蓝色光标、焦点科技、华策影视等

昆仑万维：发布天工大模型，布局图像、音乐、文本等 AIGC 应用

2022 年 12 月昆仑万维正式发布全系列算法与模型“昆仑天工”并宣布模型开源，推出包括天工巧绘 SkyPaint、天工乐府 SkyMusic、天工妙笔 SkyText、天工智码 SkyCode，覆盖图像、音乐、文本、编程等多领域。2023 年 2 月 27 日，公司宣布旗下的 Opera 浏览器与 OpenAI 达成合作，通过 API 在浏览器中正式接入 ChatGPT 功能。在地址栏中，Opera 新增“Shorten”功能，能够利用 AI 来生成网页和文章的总结或摘要，此外 AI 内容生成服务也将添加到浏览器侧栏。2023 年 4 月 17 日，昆仑万维开启对标 ChatGPT 的双千亿大语言模型“天工 3.5”的邀请测试，公司后续将持续迭代模型，保持领先优势。

汤姆猫：以“会说话的汤姆猫家族”IP 为核心，接入海外 GPT 提升用户粘性

汤姆猫通过接入 GPT-3.5 API，增强用户的粘性，持续探索新变现模式。汤姆猫是一家围绕“会说话的汤姆猫家族”IP 为核心的、线上与线下协同发展的全栖 IP 生态运营商，围绕核心 IP，在全球范围研运了 20 余款游戏，全球月活超 4 亿，其中 80%来自于海外。基于过往的经验，汤姆猫已经尝试应用 GPT-3.5 模型进行 AI 语音互动产品功能原型测试，并已经将汤姆猫 IP 形象植入产品功能原型当中，加入语音识别、语音合成、性格设定、内容过滤等功能，并测试了语音交互、连续对话等产品性能，初步验证了相关技术实现的可行性。

蓝色光标：AIGC 板块布局广泛，商业模式全流程覆盖

作为我国营销行业的龙头，蓝色光标积极探索各类新技术在营销领域的应用。

携手微软，开启 GPT-4 元宇宙营销新纪元。蓝色光标 3 月 9 日公布与 Microsoft 广告达成战略合作，并成为中国区官方代理商。根据蓝色光标官网，基于蓝色宇宙已有业务版图与 GPT-4 的合作可能性包括：

1）虚拟人：GPT-4 将为蓝色宇宙打造一批具有个性和智慧，且对话内容详实安全的虚拟人物，作为品牌形象或者社交媒体影响者，能够更好地为各类客户提供安全可控、人性化、可定制的人际交流和服务。

2） 虚拟物：GPT-4 将为蓝色宇宙设计一系列具有创意和美感的虚拟物品，作为营销礼品或者收藏品，不仅能够根据图像和文本输入生成适合的描述和故事，还能够根据用户的行为和反馈进行动态变化。

3）虚拟空间：GPT-4 基于其强大的游戏逻辑生成能力，将为蓝色宇宙构建一些具有主题和氛围的虚拟空间，作为营销活动或者社交聚会的场地。这些虚拟空间不仅能够根据图像和文本输入生成适合的布局和风格，还能够根据用户的位置和动作进行实时渲染。

同时公司在 AIGC 板块布局还包括分身有术、蓝标智播和销博特等，其中分身有术聚焦虚拟人物，蓝标智播从事虚拟直播，销博特涵盖虚拟空间。具体而言，分身有术是一款高性价比的真人数字分身驱动平台，主要是为各行各业各领域的关键人通过建设数字分身虚拟人，通过低成本和高效率解决“分享”场景下的写真需求；蓝标智播通过建设虚拟直播间作用于长尾时间段，帮助商家实现 24 小时全天带货直播，解决了真人无法长时间值守的困难，服务于现阶段直播电商增效；销博特基于云端聚焦营销场景，功能覆盖全流程营销设计，包括策划案撰写、人群画像、创意灵感、slogan 生成和海报绘制等板块。

免责
更多免费行业报告 并购家 www.ipoipo.cn

焦点科技：推出 AI 外贸助手全面赋能外贸营销、采洽沟通，打造外贸企业“数智大脑”

公司始终致力于将传统行业与互联网深度融合，先后将互联网引入外贸、保险、企业采购、教育、医疗等多个领域，研发并运营了中国制造网（Made-in-China.com）、开锣（Crov.com）、百卓采购网（Abiz.com）、新一站保险网（xyz.cn）等电子商务平台，以及智慧教育、移动医疗等互联网项目。凭借多年来的不断积累，旗下中国制造网已成为全球采购商寻找中国供应商以及贸易伙伴的重要网络渠道。

推出 AI 外贸助手“麦可”打造外贸企业的“数智大脑”。4 月 25 日，焦点科技旗下中国制造网（以下简称“MIC 国际站”）推出 AI 外贸助手，以品牌 IP“麦可”的 AI 形象呈现给网站用户，为外贸企业提供 AI 智能化服务，赋能企业开拓海外市场。在营销推广场景，麦可是每一位供应商的“AI 营销专家”，可以实现从市场洞察到营销文案的综合产出，多场景撰写外贸专业开发信、智能改写产品描述，并实时进行 SEO 检测和优化提醒；在品牌出海场景，麦可又能成为供应商的“AI 创意大脑”，能从企业品牌形象打造、社媒贴文创意策划、Blog 自动生成等功能驱动品牌内容生产自动化；在采洽沟通场景，麦可承担了“AI 全能助理”的重要角色，当收到国外买家的询盘时，麦可能智能提炼买家的诉求和关键点，并实现对话内容的多语种翻译，贴合不同对话场景进行自动回复，辅助外贸人员工作。据焦点科技官网，麦可后续有望持续迭代升级，未来将具备数据的整合、挖掘及分析能力，帮助外贸企业洞察国际市场的行业趋势，清晰了解买家的背景实力，提高国际贸易中的风险防范能力，为企业提供智能化的决策辅助和策略参考，最终成为外贸企业的“数智大脑”。

华策影视：接入百度“文心一言”，AIGC 助力影视制作效率提升

华策影视重视科技导向，落实 AI 技术应用落地。据公司公告，2021 年，公司提出“影视+科技”，全面部署“精品内容开发+创新科技应用”的发展道路，并于 2022 年提升科技数字增长中心为一级部门，进一步重视科技导向，牵头研究、落实 AI 技术在公司日常运营和项目开发中的应用落地。并鼓励全体员工在工作领域积极使用各类 AI 工具，培养善于使用 AI 的技术人才。

华策影视接入百度“文心一言”，利用 AIGC 技术提升影视行业全管线效率。据公司公告，华策影视已全面接入百度文心一言，将和包括百度在内的战略合作伙伴们共同将 AIGC 技术应用于内容创意、内容制作、内容宣发等影视生产全链路。公司将以科技化数字化赋能产品创新，提升生产效率，在技术与内容融合中找到新的价值增长点，充分利用 AIGC 技术提升影视行业全管线效率，包括影视剧本创作、AI 换脸及换声、场景及分镜制作、数字人模拟虚拟角色、预告片剪辑、特效处理、AI 海报制作等。

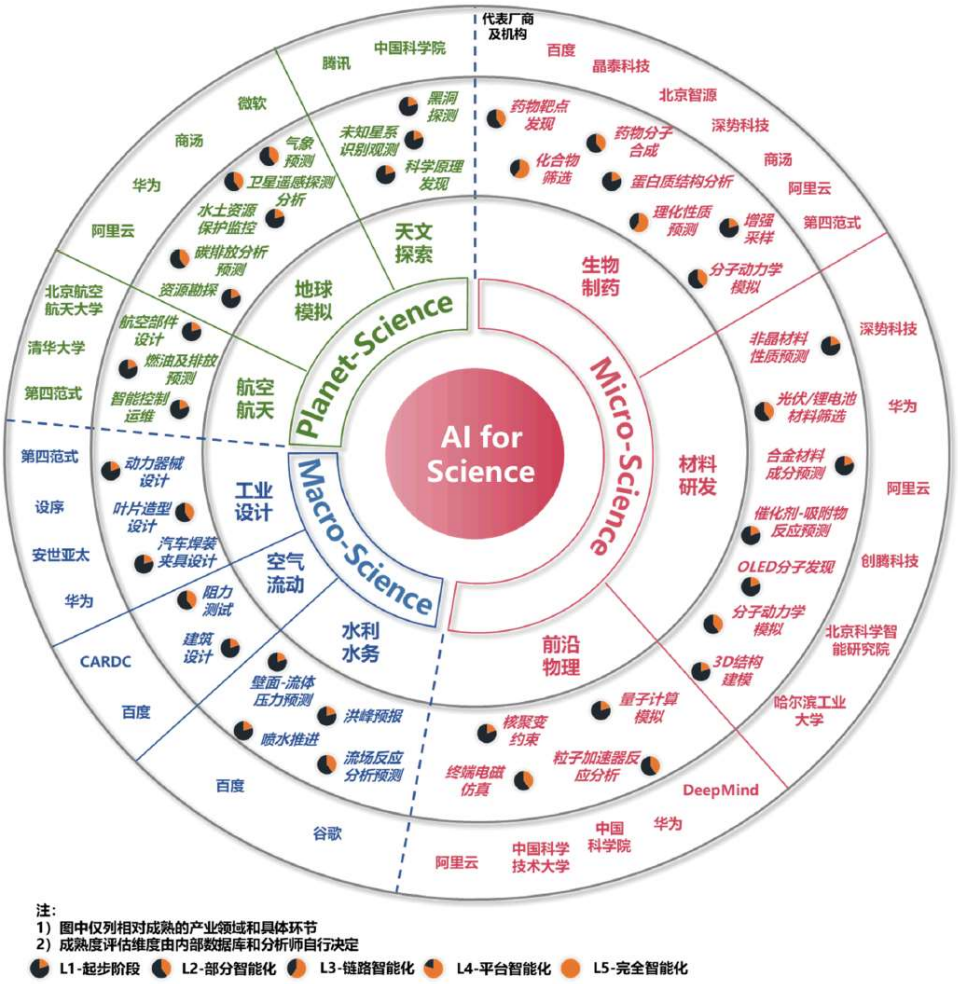
华策影视加强对上海华剧汇控制，布局 AIGC。2023 年 4 月 16 日，华策影视发布公告宣布收购景宁元晖、景宁元鑫企业份额，加强对上海华剧汇的股权控制。华策影视表示，此次收购将助力上海华剧汇积极扩充影视剧版权数量、加快制作优质视频素材、加速创新商业模式落地、探索基于华策元视界的 AIGC 技术和场景应用，构建中国最大的影视素材版权运营平台。

免责
更多免费行业报告 并购家 www.ipoipo.cn

AI for Science：人工智能助力科学探索的新范式

AI for Science 是指以机器学习、深度学习等人工智能技术分析处理多维度、多模态、多场景下的模拟和真实数据，解决复杂推演计算问题，加快基础科学和应用科学的发现、验证、应用，打造下一代科学范式。如下图所示，AI for Science 主要包括：（1）蛋白质结构分析等生物制药，（2）新材料研发，（3）核聚变等前沿物理，（4）气候预测等地球模拟和天文探索，（5）飞机引擎，汽车动力结构，建筑等工业设计领域。

图表127： AI for Science 全景图

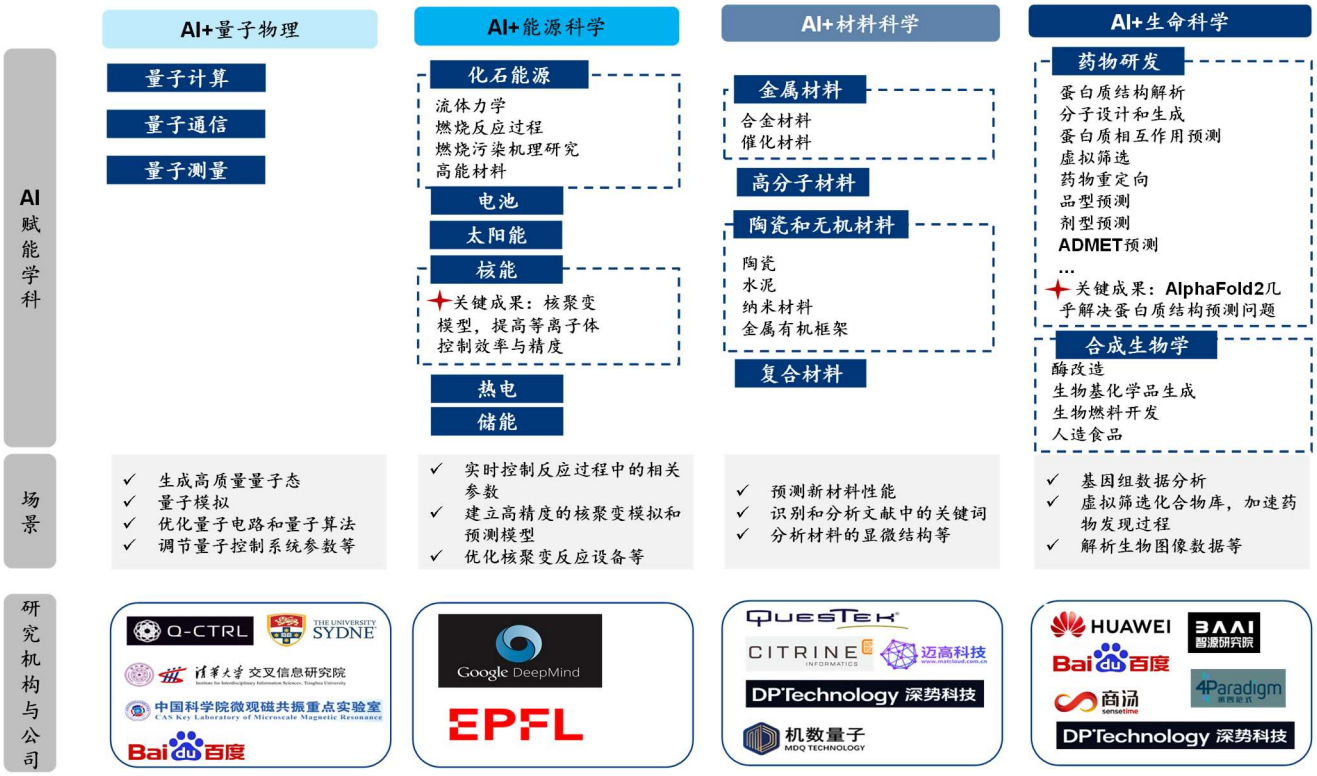


注：百度（BIDU US）、晶泰科技（未上市）、北京智源（未上市）、深势科技（未上市）、商汤（00020 HK）、阿里（BABA US）、第四范式（未上市）、华为（未上市）、创腾科技（未上市）、谷歌（GOOGL US）、CARDC、安世亚太（833714 CH）、设序（未上市）、微软（MSFT US）、腾讯（00700 HK）
资料来源：IDC，华泰研究

具体来说，**在生命科学领域：**AI 为研究人员提供了强大的工具和方法来处理和分析大规模的生物学数据，包括解读基因组数据和预测基因功能、预测蛋白质的三维结构和加速药物发现过程、医学影像分析和病理学诊断、建立个体化医疗模型和提供定制化医疗方案等。**在量子物理领域：**人工智能赋能量子物理主要包括量子态模拟和优化、量子态重构、量子控制和优化等应用，这些环节往往在极短时间内就产生大量数据，而人力识别需要耗费大量的时间成本，因此需要 AI 来建立仿真模型，实时处理实验数据。**在能源科学领域：**涵盖材料设计、能源系统优化、智能电网、能源储存与转换以及节能环保等方面，特别是在核聚变方面，AI 技术可以挖掘数据中的模式和趋势，提供对核聚变过程的精确建模和预测，还可以通过智能化的控制策略和自适应优化方法，提高核聚变设备的性能和效率。**在材料科学领域：**材料研发技术流程与药物研制类似，涉及材料理化性质预测、结构筛选、建模仿真与高性能算力支持等。IDC 认为在基于人工智能技术的材料智能搜索中，AI 帮助在庞大的搜索空间中寻找最佳配比，优化预测和设计新材料的体系。另外，AI 也可以采集学习实验失败数据，有效提高模型精度和数据利用率。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表128： AI+量子/核聚变/材料科学/生命科学具体应用



注：QUESTEK（未上市）、CITRINE（未上市）、迈高科技（未上市）、机数量子（未上市）、HUAWEI（未上市）、BAIDU（BIDU US）、商汤（00020 HK）、第四范式（未上市）

资料来源：各公司官网，华泰研究

各国相继出台 AI for Science 政策措施，要求加快人工智能促进科学研究创新发展。具体来说，中国发布了《“十四五”国家信息化规划》（2021）《量子安全技术白皮书（2022 年 1 月修订版）》（2022）《“十四五”数字经济发展规划》（2022）等政策，大力加强基础科学、前沿科学和应用基础研究。美国发布《国家量子倡议(NQI)法案》（2018）《美国量子网络战略构想》（2020）支持 AI 在量子物理领域的研究，发布《材料基因组计划战略规划》（2021）和《国家生物技术和生物制造计划》（2022）分别支持 AI 在材料科学和生命科学领域的应用。英国发布《国家量子战略》（2023）《迈向聚变能源：英国聚变战略》（2021）《生命科学产业战略》（2017），提及推动 AI 在量子物理、核聚变、生命科学领域的应用。此外，日本发布了《量子未来社会愿景》(2022)，并公布了首个核聚变能源开发战略方案。

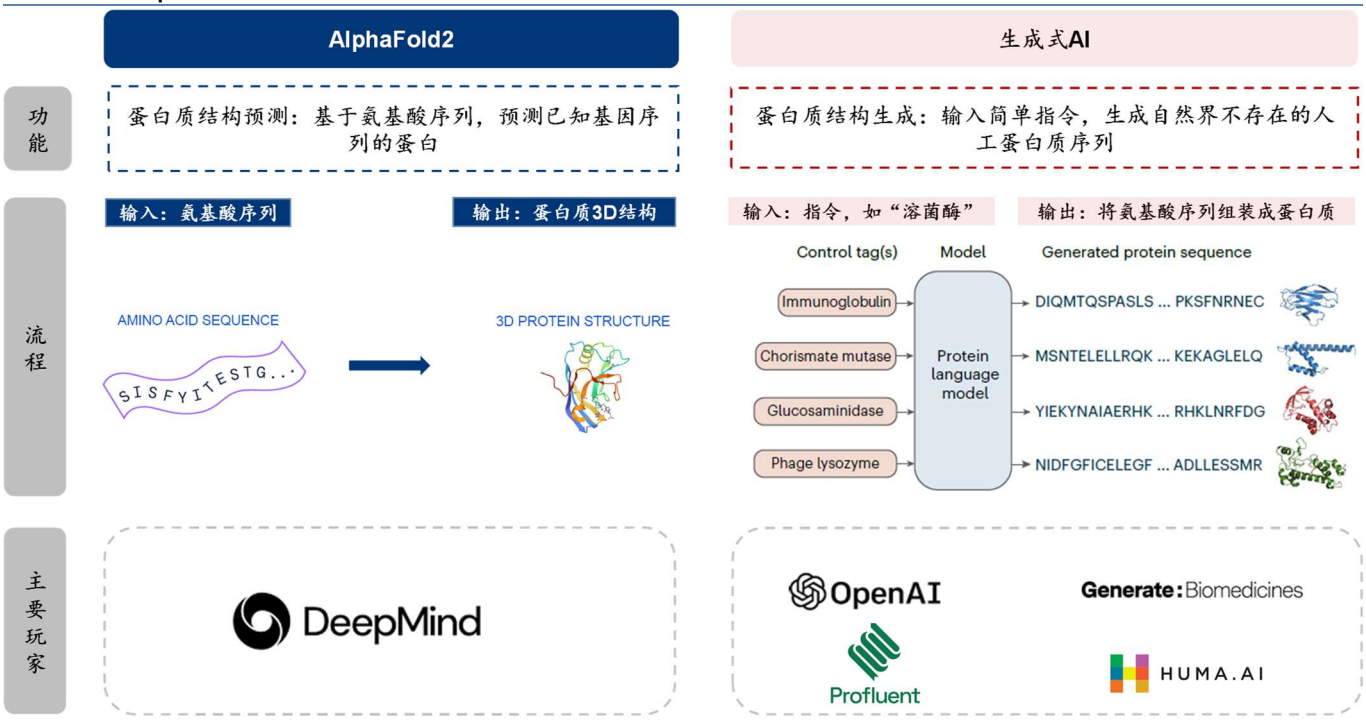
AI+新药开发：根据功能需求设计/优化蛋白质，加速新药探索速度

DeepMind 两代 AlphaFold 引领行业进入高速发展期。梳理 AI+制药的发展历程，我们发现 AI 与制药的结合逐步深入，成为生物制药企业的常态化工具。经历了包括 Exscientia、Atomwise、英矽智能、晶泰科技在 AI 新药研发领域的早期探索（2014-2017 年）、最早一批 AI+新药企业开始获得临床前候选药物(Pre-clinical candidate, PCC)一类的验证性成果（2018-2019 年）后，DeepMind 推出的 AlphaFold 和 AlphaFold2 实现了蛋白质空间结构预测的重大突破，引领行业进入高速发展期。

近期，我们看到类 ChatGPT 的生成式 AI 开始出现，印证 AI 与药物发现的结合逐渐深入。比如 Salesforce Research、Tierra Biosciences 和加州大学研究团队发布的蛋白质语言模型 ProGen，通过在数百万个原始蛋白质序列上的训练，可生成跨多个家族和功能的人造蛋白质。不同于 AlphaFold2 利用 AI 的手段辅助理解自然界已有的蛋白结构，以 ProGen 为例的类 ChatGPT 生成式 AI 的目标是生成自然界不存在的人工蛋白质序列。虽然生成式 AI 在制药领域的应用仍处于起步阶段，但天壤 XLab 负责人苗洪江博士认为，**从 AlphaGo、AlphaFold2 到生成式 AI 是一个从分析工具走向解决方案、并落地解决实际问题的过程。**长期来看，生成式 AI 有望赋能生物医药源头创新。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表129：从 AlphaFold2 到类 ChatGPT 生成式 AI

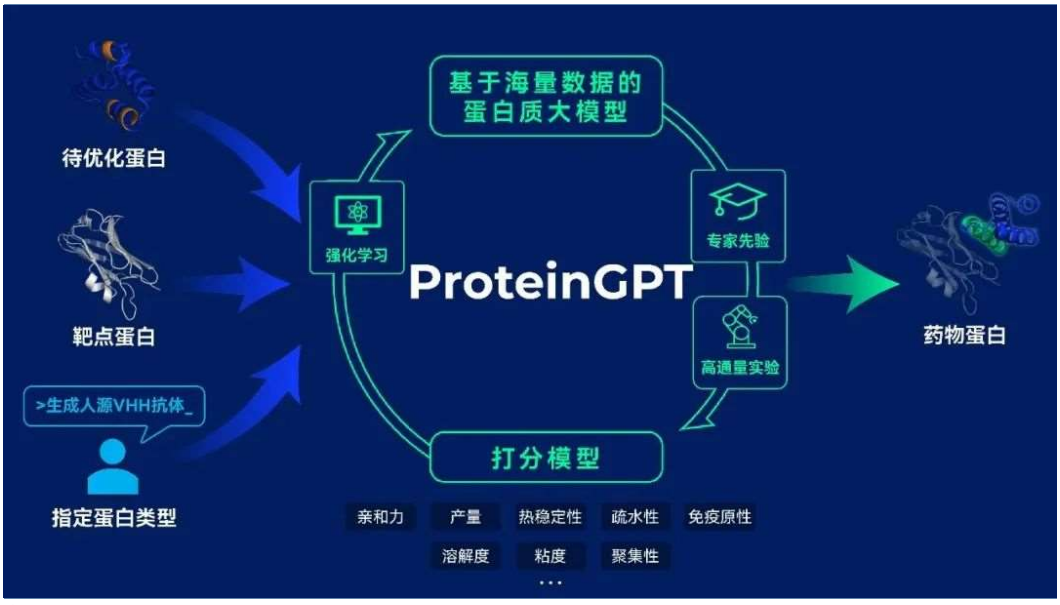


资料来源：GTC2023，*Large language models generate functional protein sequences across diverse families*，华泰研究

国内，包括晶泰科技、英矽智能、英飞智药、百图生科在内的公司均在探索生成式 AI 在制药领域的应用。

晶泰科技通过无标记的蛋白质序列数据（约 2.8 亿条）+抗体序列数据训练出的 ProteinGPT 可以一键生成符合要求的蛋白药物，应用案例包括：1）根据给定的靶点一键生成 Binder 蛋白；2）根据指定要求一键生成抗体文库；3）对性质欠佳的抗体进行改造，一键生成优化后的抗体。

图表130：ProteinGPT 模型架构

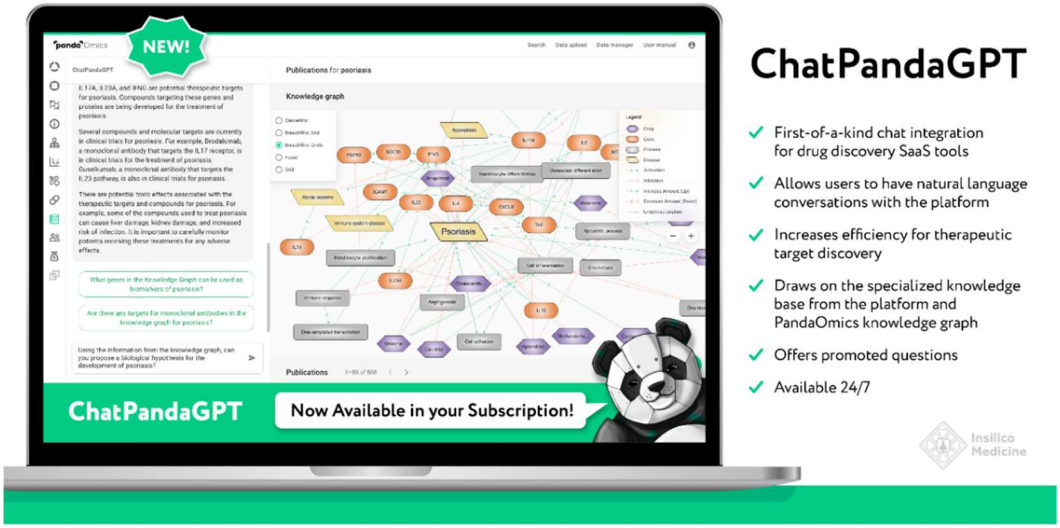


资料来源：晶泰科技公众号，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

英矽智能宣布在其靶点发现平台 PandaOmics 上整合了 AI 问答功能“ChatPandaGPT ”，支持研究人员在浏览和分析大型数据集的同时，高效开展基于自然语言的问答，促进更便捷发现潜在靶点和生物标志物。

图表131： 英矽智能 ChatPandaGPT



资料来源：公众号，华泰研究

AI+量子：优化软硬件设计，提升量子技术应用的效率与准确性

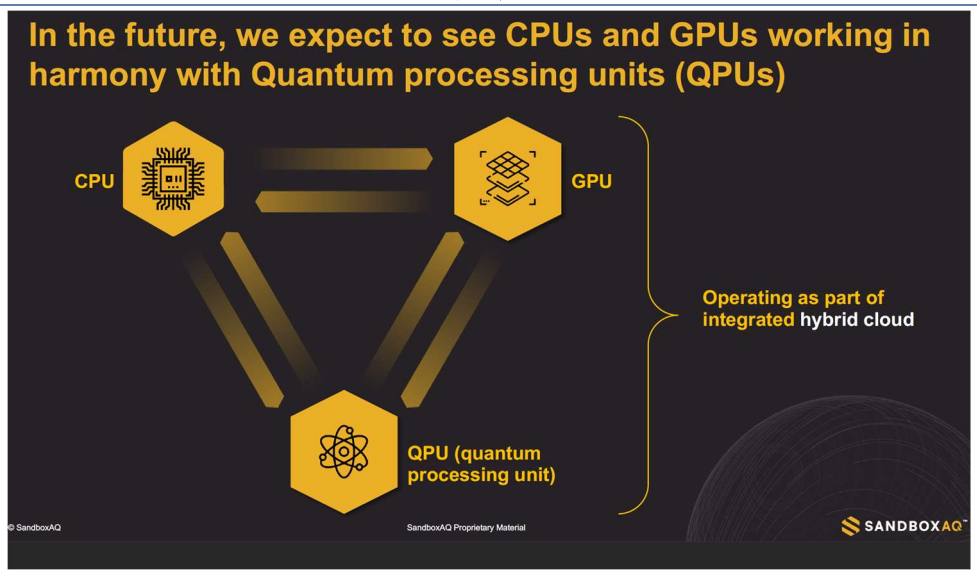
量子科技已成为全球各国科技政策布局与市场投资热点，据 WFE 数据，2022 年全球对量子技术研发的投资总额达 355 亿美元。市场普遍认可量子科技发展通过减轻算力束缚对 AI 大模型发展的促进作用，我们认为 AI 技术同样可用于量子系统的深入研究，如采用不同的监督和无监督学习算法解决量子多体问题、重建高维量子系统的密度矩阵、促进自校准量子传感器的发展等，反向推动量子科技实现突破。目前，业界和学术界均已实现 AI 技术在量子计算、量子通信以及量子精密测量三大量子信息技术领域的应用。

量子计算方面，AI 可用于优化量子计算的硬件和软件算法设计。1）硬件是量子计算应用的基础，面临的挑战之一便是如何提升量子比特质量。由于量子比特的特殊性质，量子计算机容易受到环境噪声和量子退相干的影响，导致计算错误，AI 技术可以通过深度学习算法来学习和识别错误，从而优化量子计算中的错误纠正。2）解决实际问题，离不开云平台、软件和算法的发展。AI 技术在量子计算的软件的应用包括设计和优化量子算法、量子编译器、量子仿真等。例如，通过机器学习技术训练神经网络可以设计更高效的量子算法，使用遗传算法可以优化量子算法的参数。

英伟达（NVDA US）将量子计算机与现代科学计算生态系统无缝集成，CPU、GPU、QPU 将会集成到一个混合云系统中，以智能和协调的方式工作。QPU 永远不会取代经典计算，而是被视作某些特殊应用的加速器，即没有 QPU 就无法有效运行的量子算法。大多数人和企业现在或将来都不需要拥有自己的量子计算机，未来的计算世界将是混合的，CPU、GPU、QPU 会集成到一个混合云系统中，量子计算机将在云端访问。混合算法将根据需要在 CPU、GPU 和 QPU 上无缝运行，但这个过程必须是智能和协调的，QPU 只应该在被需要的时候调用。

更多免费行业报告 并购家 www.ipoipo.cn

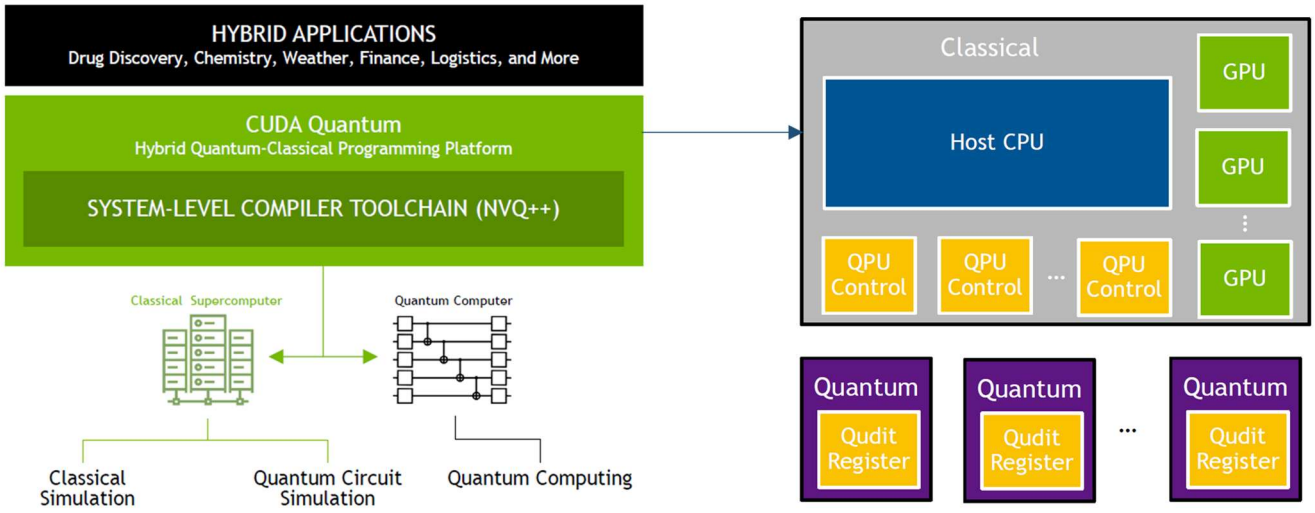
图表132：CPU、GPU、QPU 会集成到一个混合云系统中



资料来源：Sandbox AQ 官网，华泰研究

英伟达推出 CUDA Quantum 和 DGX Quantum，致力于无缝集成 CPU、GPU 和 QPU，并用 CPU 和 GPU 提高 QPU 的性能。CUDA Quantum 是一个混合量子-经典编程平台，DGX Quantum 是一个集成了量子与 GPU 超级计算的系统，具备如下特点：1）将量子与 GPU 超级计算紧密集成；2）量子比特不可知——支持不同的量子比特模式；3）GPU-QPU 延迟降低 1-2 个数量级，从而使 GPU 加速进行量子纠错，并实现更快的校准、实时控制和混合算法的使用；5）可扩展支持更多 GPU 和更大的 QPU。

图表133：DGX Quanum 和 CUDA Quantum 结构示意图



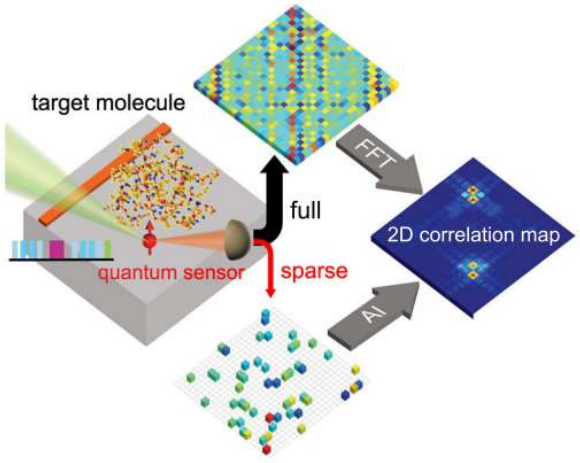
资料来源：NVIDIA 官网，华泰研究

量子通信方面，AI 可用于训练生成量子通信协议。现阶段量子通讯协议的设计和优化依然停留在人工处理的阶段，从研发到实验验证时间周期长、成本高，引入 AI 技术搭建系统化模拟平台可有效缓解该问题。2021 年 2 月，百度量子机器学习平台“量桨”新增分布式量子信息模块 LOCCNet，支持量子纠缠蒸馏、量子态分辨、量子隐形传态等核心量子信息协议的便捷开发。基于量桨搭载 AI 量子通信模拟平台不仅可以训练出传输单量子比特量子态的量子隐形传态协议，还可以训练出适用于多量子比特情形下的隐形传态协议，大大提升量子通信协议的开发效率。

更多免费行业报告 并购家 www.ipoipo.cn

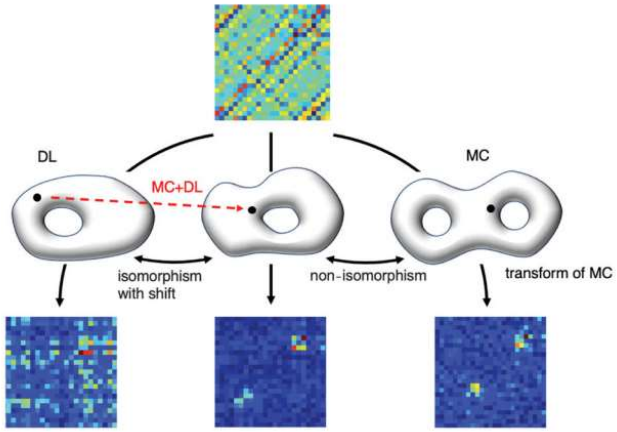
量子精密测量方面，AI 可用于提高二维纳米核磁共振谱的探测效率。核磁共振是目前少有的对人体没有任何伤害的安全、快速、准确的临床诊断方法，在结构生物学和临床医学上具有广泛应用，金刚石氮-空位（NV）色心作为一种出色的量子传感器，可以实现纳米级甚至单分子中的核磁共振。然而，必须使用高效的数据积累和处理方法，才能使高空间分辨率的氮空位传感器实现纳米尺度二维核磁共振谱。2020 年 9 月，中科大微观磁共振重点实验室杜江峰团队与爱荷华大学巫晓东教授合作，利用深度学习算法成功将基于金刚石量子精密测量技术的纳米核磁共振二维谱的探测效率提高近一个量级。深度学习算法可以从少量的信息中提取复杂的纳米核磁谱线信息，从而大大提高实验测量效率。

图表134：深度学习算法可从少量信息中提取复杂的纳米核磁谱线信息



资料来源：Xi Kong et al. *Artificial intelligence enhanced two-dimensional nanoscale nuclear magnetic resonance spectroscopy* (2020)，华泰研究

图表135：深度学习结合矩阵填充算法在保持重构能力同时可去除掉偏倚



资料来源：Xi Kong et al. *Artificial intelligence enhanced two-dimensional nanoscale nuclear magnetic resonance spectroscopy* (2020)，华泰研究

AI+核聚变：精准调控核聚变控制过程，降低开发周期

核聚变是指将两个轻核聚合成一个重核的过程。在这个过程中，轻核融合在一起并释放出能量。由于核聚变需要极高温度，一旦某一环节出现问题，燃料温度下降，聚变反应就会自动中止，因此 AI 在核聚变中的应用十分重要。目前我们看到：1）**模拟和预测：**AI 可以用于模拟和预测核聚变反应的行为。通过对已知反应的数据进行训练，可以开发出能够预测和优化核聚变反应的 AI 模型。2）**实时监测和控制：**通过使用机器学习算法，可以从传感器数据中提取有用的信息，例如温度、压力和辐射水平，并用这些信息来控制反应堆的运行参数。这有助于确保反应堆在安全的条件下运行，并优化反应堆的性能。3）**数据分析和处理：**通过使用机器学习算法，可以从数据中发现聚变反应隐藏的模式和趋势。4）**聚变反应堆设计：**通过使用机器学习算法，可以优化反应堆的几何形状和运行参数，以提高聚变反应的效率和稳定性。

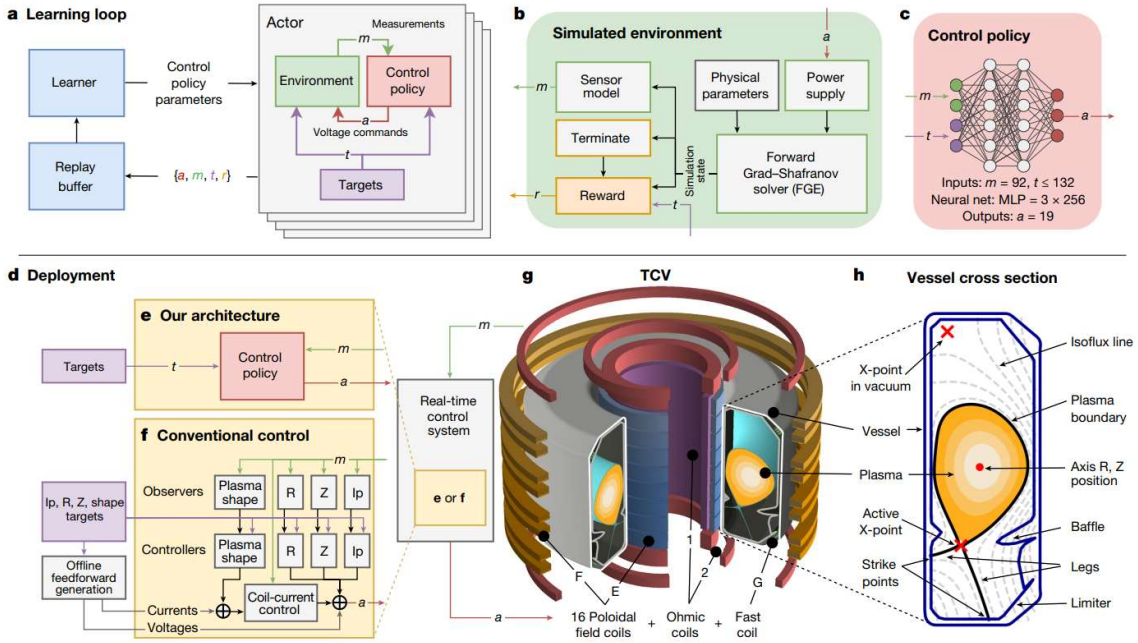
在核聚变过程中存在“托卡马克磁控制问题”。托卡马克装置的作用是可以长时间的约束高温的等离子体，让核聚变的反应可以持续并可控。为了优化等离子体的稳定性、封闭性和能量排放，需要在托卡马克内设计反馈控制器以精确控制磁耦合的线圈来操纵磁场，以达到理想的等离子体电流、位置和形状。这个问题也就是著名的“托卡马克磁控制问题”。

深度强化学习可以作为全新的处理复杂核聚变方式，用于解决“托卡马克磁控制问题”。随着核聚变的反应堆增大，托克马克装置越来越复杂，对其内部配置的运行设置要求越来越高。而且这个控制过程是非线性、实时变化、多变量的，非常复杂，需要大量的工程、设计和专业知识，同时还要进行复杂的平衡估计来实时计算调控的电流。传统的计算机模型和模拟方式不一定能够优化该过程，精确性也无法保障。而通过强化学习设计的磁性控制器，可以自主学习设计整个控制线圈的电流，既可以满足苛刻的物理操作约束条件，也能够实现稳定的输出控制，有效减少了磁约束控制器的开发周期，加速了科研人员对替代性等离子体配置的研究。

更多免费行业报告 并购家 www.ipoipo.cn

2022 年 2 月 16 日，来自人工智能公司 DeepMind、瑞士洛桑联邦理工学院的研究团队，通过与托卡马克模拟器的互动学习生成控制策略，并将以神经网络表示的控制策略直接在硬件上进行托马克磁控制，不仅实现了对等离子体平衡的基本品质的精确控制，而且具备复杂配置能力，实现在放电实验所有相关阶段准确的等离子体控制。

图表136：AI 控制核聚变的学习控制和训练架构



资料来源：Jonas Degrave et al. *Magnetic control of tokamak plasmas through deep reinforcement learning* (2022)，华泰研究

目前许多国家政府都在支持核聚变技术的研发和应用：1）中国：已经建成了 EAST（Experimental Advanced Superconducting Tokamak）聚变实验装置，还积极推进聚变能源产业的发展，并计划在 2030 年前建成第一座商业聚变电站。2）日本：已经建成了 JT-60SA（Japan Torus-60 Super Advanced）聚变实验装置，并在 3 月 14 日正式敲定了首个核聚变能源开发战略方案，提出要推进企业参与研发实验用的核聚变反应堆，并争取在 2050 年左右实现核聚变发电。3）欧盟：已经建成 ITER（International Thermonuclear Experimental Reactor）聚变实验装置，是目前正在建设的世界上最大的实验性托卡马克核聚变反应堆。4）美国：已建成 NSTX-U（National Spherical Torus Experiment-Upgrade）聚变实验装置，正计划在未来几年内建造一座商业核聚变电站。

AI+材料科学：逆向设计是 AI 助力材料研发创新的主要范式

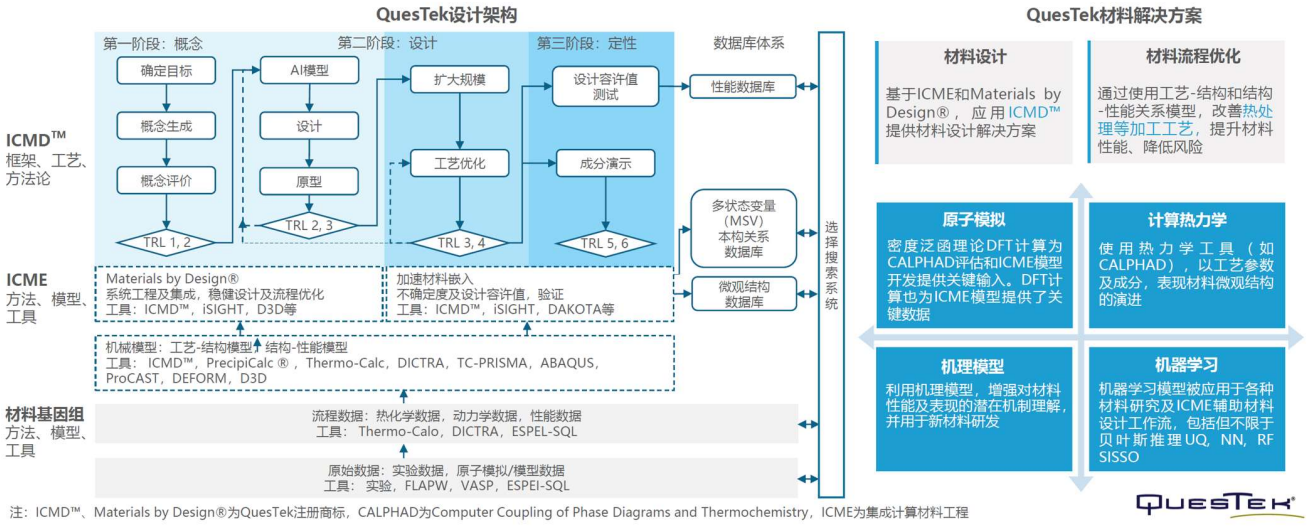
AI 技术助力解决新材料研发周期长、产业化速度慢、投入成本高等难题。新材料“发现/设计-开发-制造-应用”环节通常需要 10-20 年时间，导致新材料企业难以快速盈利，并且带有研发项目无法实现应用和投资回报的风险。人工智能擅长建立数据间的关系，研究人员可利用 AI 模型逆向推导目标材料的组成成分、结构和流程工艺，这种逆向设计已经成为 AI 赋能材料研发创新的主要范式。相较正向设计，逆向设计可以更好地分析已有数据，更快更好地设计出满足性能和生产要求的材料结构和制造工艺，从而加快新材料的研发速度和效率，降低研发成本。机器学习更是凭借其强大的预测性能，广泛应用于材料科学各领域，例如：1）高效材料特性预测的代用模型开发；2）使用变异自动编码器 VAE 和生成对抗网络 GAN 的生成性材料设计；3）通过将实验设计算法与自动机器人平台结合，实现自主材料合成；4）深度学习用于原子尺度材料成像数据的精确表征。

美国企业 QuesTek、Citrine，国内深势科技、鸿之微科技（未上市）、机数科技、迈高材云科技等均在探索 AI 在材料科学领域的应用。

免责
更多免费行业报告 并购家 www.ipoipo.cn

QuesTek 致力于集成计算材料工程（ICME）技术，通过对铸造、增材制造、热处理等关键工艺过程中材料组织-性能演变进行建模和计算，开发出用于航空航天、超临界发电机组、海工结构等领域的高性能新材料。区别于传统试验试错方法，QuesTek 的研发核心是“大设计、小试验”，通过多尺度计算和数据库辅助设计，将材料成熟度提升至 5-6 级，再付诸实验室或工程试验，大幅度提升研发效率。QuesTek 已成功应用 ICME 设计了 Ferrium S53 飞机着陆架用齿轮钢，并承担了美国空军、海军、能源部等材料研发项目。

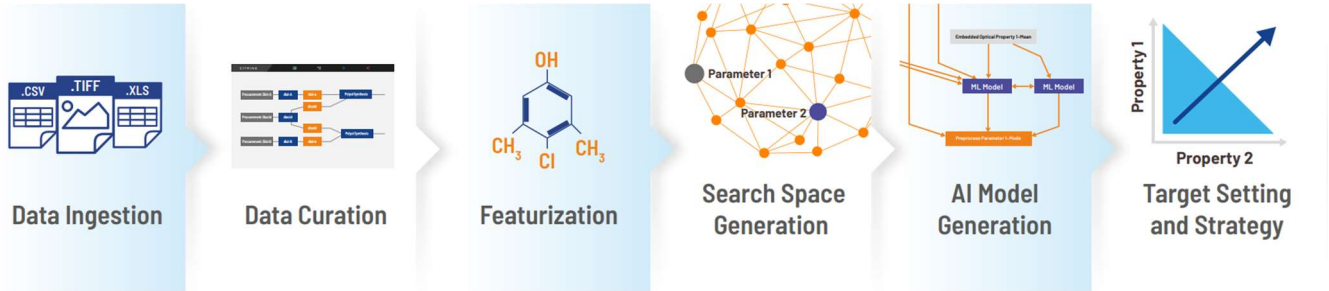
图表137： QuesTek 设计架构和材料解决方案



资料来源：亿欧智库，QuesTek，华泰研究

Citrine 智能材料平台基于尖端 AI 工具和智能数据管理基础架构搭建而成，可用于数据驱动的材料和化学品开发。该平台通过各种来源吸纳结构性和非结构性材料数据，并利用人工智能引擎来识别数据中的重要信息，结合用户企业的行业知识来预测材料在各种配方下的性能，加速材料研发工作。Citrine 服务于先进材料研发企业，如 3M、BASF、Showa Denko、LANXESS、AGC 等；科研机构，如加州大学伯克利分校；政府机构，如美国能源部等，涵盖汽车、航空、消费品、电池和电子等领域。

图表138： Citrine 平台使用 AI 快速筛选聚合物流程



资料来源：公司官网，华泰研究

深势科技提出“多尺度建模+机器学习+高性能计算”的科学研究新范式，推出了 Bohrium®微尺度科学计算云平台、Hermite®药物计算设计平台等微尺度工业设计基础设施，对合金材料、锂电池材料、半导体材料、储氢材料等新材料进行高效准确的性质计算与预测，并与晶泰科技、汉森制药、宁德时代、金羽新能等公司达成合作。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表139： Bohrium®微尺度科学计算云平台功能

资料来源：公司官网，华泰研究

AI for Science 产业链相关公司：国仪量子、本源量子、讯飞医疗、BrainCo、脑虎科技

国仪量子：全球少数具备量子计算整机交付能力的企业

国仪量子成立于 2016 年，源于中国科学技术大学，是国内第一家以量子精密测量为核心技术的国家高新技术企业，主要从事量子精密测量、量子计算和高端科学仪器等技术和相关产品的研制、生产与销售。公司自成立以来各项指标均保持逐年翻番的增长速度，2022 年销售收入超 2 亿元，获得科大讯飞、高瓴资本等多家知名投资机构的投资，累计获得社会资本投入超 10 亿元，估值近 70 亿元，获得“独角兽企业”称号。

公司秉承“为国造仪”的理念，积极为解决“卡脖子”问题贡献力量，陆续研制并发布了国际首台量子钻石原子力显微镜、金刚石量子计算教学机，国内首台商用脉冲式电子顺磁共振谱仪、场发射扫描电子显微镜；发布了量子测控系列产品锁相放大器、任意波形发生器，比表面吸附系列产品比表面及孔径分析仪、高温高压气体吸附仪等 20 余款新产品，在关键性能指标上实现了超越。公司产品已经交付到清华大学、上海第九人民医院、华为公司等数百家客户，并在德国、美国、以色列等国家完成海外交付。公司同时致力于用量子技术赋能各行各业，已经在石油勘探、生命科学、先进材料、电力电网等领域实现了示范应用。

免责
更多免费行业报告 并购家 www.ipoipo.cn

本源量子：可出货真实工程化量子计算机的行业龙头

本源量子成立于 2017 年，是国内首家将量子计算推向商用领域、国内唯一可实际交付使用量子计算机的企业，专注量子计算全栈式开发，主营业务涵盖量子计算机整机、量子芯片、量子测控系统、量子软件、量子云平台及量子教育产品等。目前已实现多个“国内第一”，拥有首个量子计算机控制系统、首个自主研发的量子计算软件系统、首个量子计算云平台、首款国产量子计算机操作系统、联合晶合科技共建我国首个量子计算芯片联合实验室等。本源悟源 24 比特超导量子计算机获得安徽省首台套重大技术装备认定。本源量子已申请量子计算领域专利、软著、商标共计 1000 余项，2022 年 10 月跻身“全球量子计算技术发明专利排行榜（TOP100）”全球第六，专利引证量全球第十，是国内唯一双指标进入前十的量子计算公司。先后获得国家级高新技术企业、安徽省量子计算工程研究中心、安徽省院士工作站、安徽省新型研发机构、安徽省博士后科研工作站、合肥市技术创新中心、2020 全球 “50 家聪明公司”、中国 500 家隐形独角兽、全国颠覆性创新大赛最高奖优胜奖、世界互联网大会最具创新奖、集成电路一等奖等众多荣誉。

讯飞医疗：深度融合人工智能技术与医疗行业

讯飞医疗科技股份有限公司是科大讯飞的子公司，成立于 2016 年。基于科大讯飞二十余年积累的国际领先的智能语音和人工智能技术，讯飞医疗在智慧医院和基层医疗 AI 应用领域，利用语音识别、图像识别、自然语言理解等核心技术，实现 AI 导医导诊、语音电子病历、影像识别、辅助诊断、外呼随访等功能模块，在提升医院信息化、智能化水平，提高基层医生诊疗能力方面取得了良好效果。截至 2023 年 4 月 30 日，讯飞医疗已覆盖疾病种数达 1600 种，覆盖医疗机构 3 万家，AI 辅诊次数 6.07 亿次，AI 提醒后修正诊断的病历达 126 万+，语音外呼累计服务人次 10.97 亿+，识别准确率高达 98%。

BrainCo：致力于成为全球领先的非侵入式脑机接口技术解决方案供应商

BrainCo 强脑科技创立于 2015 年，是首家入选哈佛大学创新实验室(Harvard Innovation Lab)的中国团队，致力于成为全球领先的非侵入式脑机接口技术解决方案供应商，在康复、人机交互等领域具有领先优势。BrainCo 强脑科技以技术为核心，以科研为导向。拥有北美领先的科学家团队，其中来自哈佛、MIT 等全球顶尖学府的优秀校友在核心研发团队中占比超过 70%，目前拥有脑机接口领域核心专利数百余项，其中 60%左右为发明专利。

2022 年，BrainCo 强脑科技实现全球首个高精度脑机接口产品单品 10 万台量产，突破了消费级脑机接口设备的工程和技术难题。2022 年 8 月，开星果脑机接口社交沟通训练系统入选工信部人工智能医疗器械创新任务揭榜挂帅项目。

脑虎科技：侵入式脑机接口设备龙头

脑虎科技是一家国内领先的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。公司成立于 2021 年 11 月，聚焦侵入式脑机接口设备研发。公司自主研发微创侵入式高通量柔性脑机接口技术，打破国外技术垄断，并在柔性电极、生物材料、芯片设计、核心算法、数字脑库、植入方式、临床应用及生态建设等多环节全面领先。在商业化落地方面，公司取得快速进展，公司自主产权的首批产品已实现落地，下游客户包括医疗机构、高校和科研院所，相关产品已经开始获得商业化收入，并保持快速增长趋势。

2022 年 12 月 28 日，脑虎科技 NeuroXess 正式宣布完成数亿元人民币 A 轮融资。本轮融资由中平资本领投，轻舟资本、国生资本（成都生物城）、上海赛领、演化博弈参与跟投，现有投资方盛大、联新资本、红杉中国持续追加投资。

免责
更多免费行业报告 并购家 www.ipoipo.cn

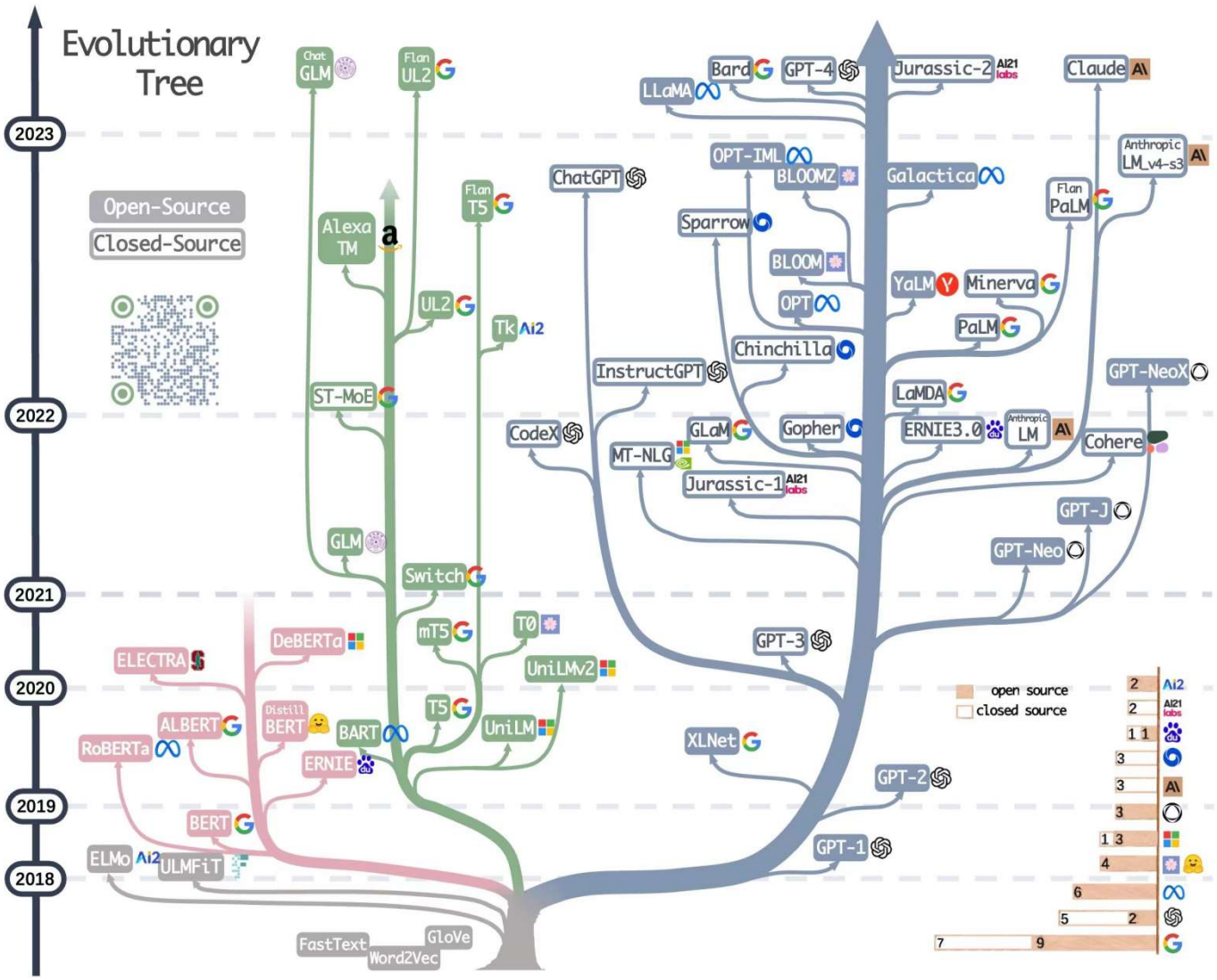
附录：AI 大模型算法的前世今生

Transformer 开启大语言模型发展新时期，为通用人工智能铺平了道路。大语言模型（LLM）是在大量数据集上预训练的，且没有针对特定任务调整数据，其在处理各种 NLP（自然语言处理）任务方面显示出了较大潜力，如自然语言理解（NLU）、自然语言生成任务等，甚至为通用人工智能（AGI）铺平了道路。而近年来的 LLM，基本都脱胎于 2017 年 Google 提出的 Transformer 架构，Transformer 开启了 LLM 发展的新时期。

基于 Transformer 架构的三条演进路线中，解码器路线发展最为迅速。从 LLM 近年的发展情况来看，其路线主要分为三种：1）编码器路线；2）编解码器路线；3）解码器路线。从发展特点来看：1）解码器路线占据主导，编码器路线趋于落寞，归因于 2020 年 GPT-3 模型表现出的优异性能；2）GPT 系列模型保持领先，或归因于 OpenAI 对其解码器技术道路的坚持；3）模型闭源逐渐成为头部玩家的发展趋势，这一趋势同样起源于 GPT-3 模型，而 Google 等公司也开始跟进；4）编解码器路线仍然在持续发展，但是在模型数量上少于解码器路线，或归因于其复杂的结构，其在工程实现上没有明显的优势。

本部分将从神经网络出发，梳理深度学习方法的发展情况，引出注意力机制与 Transformer 架构，并对全球 LLM 的典型代表——GPT 系列模型进行详细拆解。

图表140：大语言模型（LLM）演化树



资料来源：《Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond》，Yang Jingfeng（2023）、华泰研究

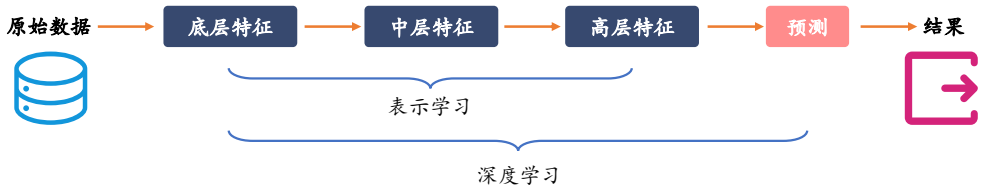
免责
更多免费行业报告 并购家 www.ipoipo.cn

溯源：从经典神经网络到 Transformer 架构

深度学习是基于神经网络的机器学习

深度学习可以概括为特征的抽象和结果的预测。深度学习是将原始的数据特征通过多步的特征转换得到一种更高层次、更抽象的特征表示，并进一步输入到预测函数得到最终结果。深度学习需要解决的关键问题是贡献度分配问题（Credit Assignment Problem，CAP），即一个系统中不同的组件（component）或其参数对最终系统输出结果的贡献或影响。深度学习的目标是让模型自动学习出好的特征表示，从而最终提升预测模型的准确率。

图表141：深度学习的数据处理流程

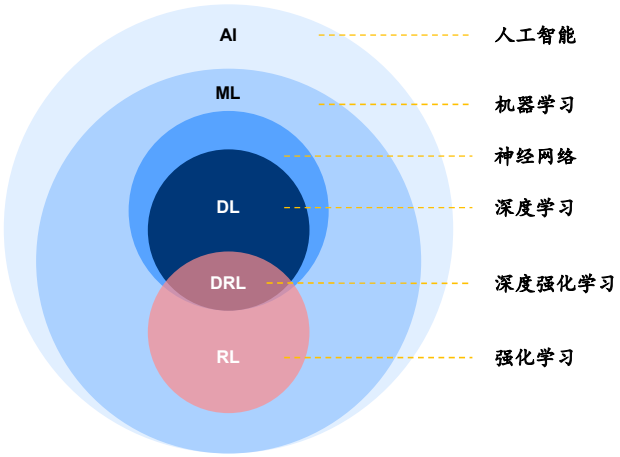


资料来源：《神经网络与深度学习》（邱锡鹏），ISBN: 9787111649687、华泰研究

#1 关系梳理：人工智能>机器学习>深度学习

人工智能包括机器学习，机器学习包含深度学习。人工智能（AI）的重要组成部分是机器学习（ML），而神经网络（NN）是 ML 的一个子领域。深度学习可以看做是神经网络的分支，使用了更复杂和更深层次的网络结构，如卷积神经网络（CNN），来处理更高维度和更抽象的数据，如图像、声音和文本。此外，逐渐兴起的强化学习也是机器学习的一个分支。深度学习与强化学习可以相互结合，称为深度强化学习（DRL）。

图表142：人工智能、机器学习、深度学习关系梳理



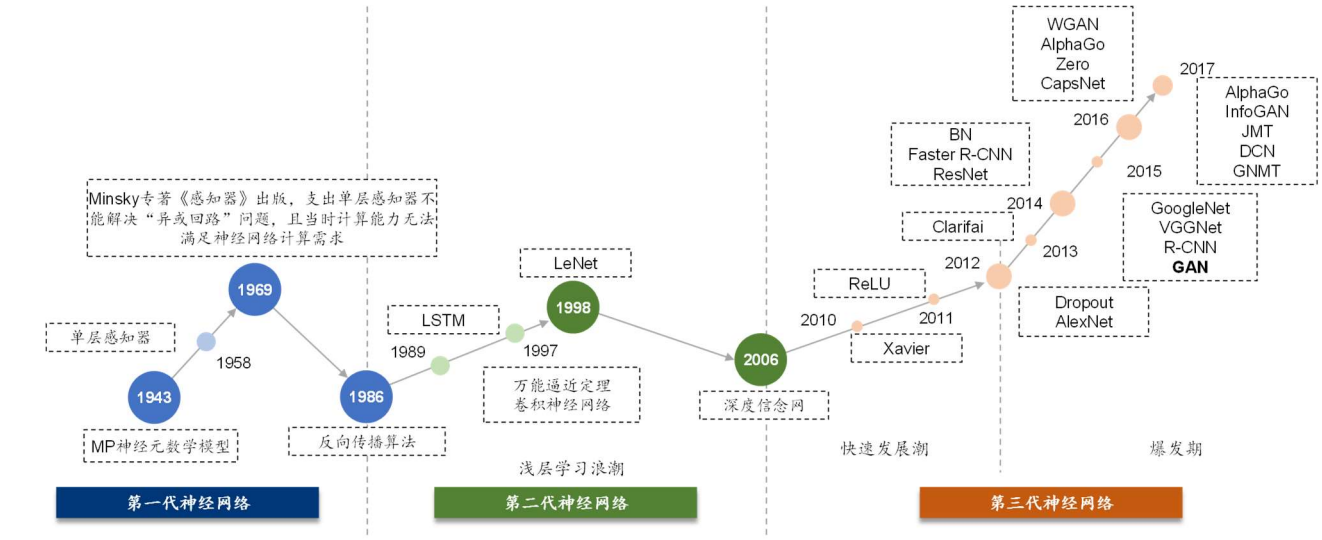
资料来源：《The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches》，Zahangir（2018）、《Deep Reinforcement Learning》，Yuxi Li（2017）、华泰研究

#2 深度学习与神经网络发展历史（1943-2017）

神经网络诞生于 1943 年，深度学习崛起于 2006 年。1943 年，心理学家 McCulloch 和数学家 Pitts 最早提出了一种基于简单逻辑运算的人工神经网络—MP 模型，开启了人工神经网络研究的序幕。1986 年，Hinton 提出第二代神经网络，并利用误差的反向传播算法来训练模型，该算法对神经网络的训练具有重要意义。1989 年，Yann LeCun 等提出用卷积神经网络（CNN）来识别手写体。1997 年，循环神经网络（RNN）的改进型长短时记忆网络（LSTM）被提出。2006 年，Hinton 首次提出了深度置信网络（Deep Belief Network，DBN）和深度学习的概念。2014 年，生成对抗网络（GAN）的提出是深度学习的又一突破性进展。2016 年，AlphaGo 击败围棋世界冠军李世石，将深度学习推向高潮。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表143：深度学习发展历史



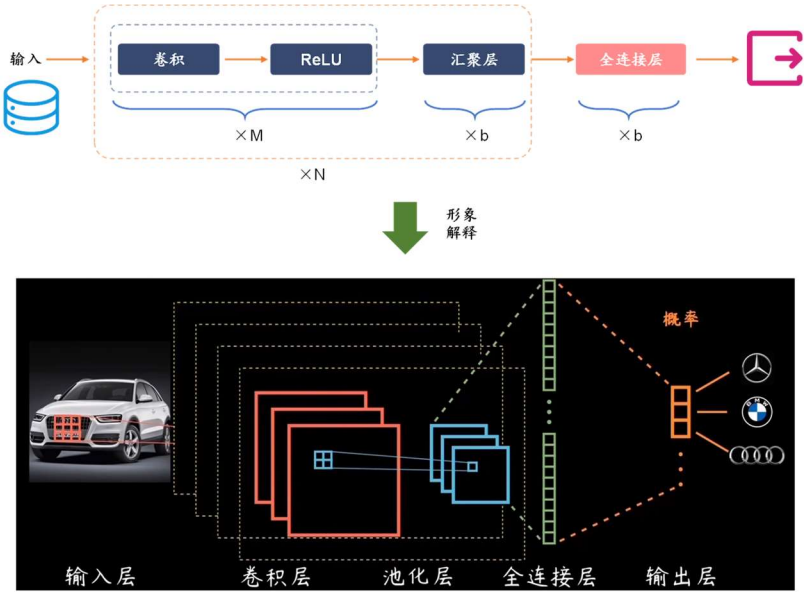
注：最大圆圈表示深度学习热度上升与下降的关键转折点，实心小圆圈的大小表示深度学习的突破大小，斜向上的直线表示深度学习热度正处于上升期，斜向下的直线表示下降期。

资料来源：《深度学习研究综述》，张荣（2018）、华泰研究

CNN：卷积+汇聚+全连接

典型的卷积神经网络是由卷积层、汇聚层、全连接层交叉堆叠而成。其中，一个卷积块包括连续 M 个卷积层和 b 个汇聚层（M 通常设置为 2~5，b 为 0 或 1）。一个卷积网络中可以堆叠 N 个连续的卷积块，然后在后面接着 K 个全连接层（N 的取值区间比较大，比如 1~100 或者更大；K 一般为 0~2）。CNN 能够有效的将大数据量的图片降维成小数据量，并有效的保留图片特征。

图表144：常用的卷积网络整体结构



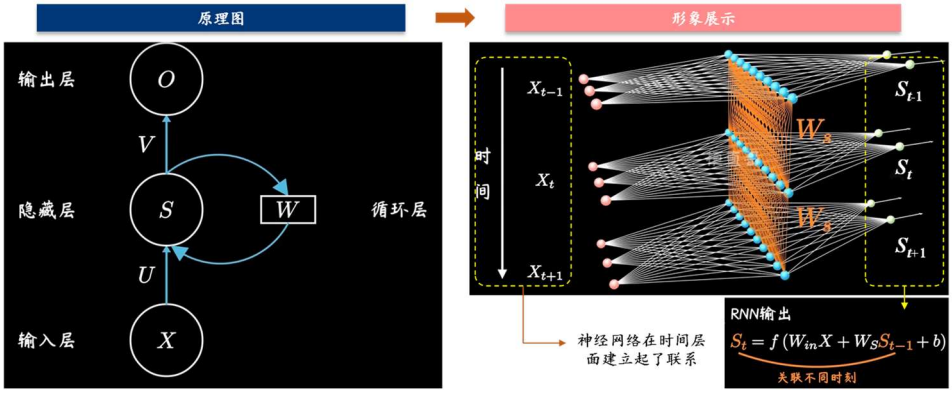
资料来源：《神经网络与深度学习》（邱锡鹏），ISBN: 9787111649687、华泰研究

RNN：具有短期记忆的神经网络

循环神经网络（RNN）是一类具有短期记忆能力的神经网络。在循环神经网络中，神经元不但可以接受其他神经元的信息，也可以接受自身的历史信息，形成具有环路的网络结构。和前馈神经网络相比，循环神经网络更加符合生物神经网络的结构。循环神经网络的短期记忆能力提供了一定的存储功能，因此其计算能力更为强大。理论上，循环神经网络可以近似任意的非线性动力系统。

更多免费行业报告 并购家 www.ipoipo.cn

图表145：循环神经网络引入了短期记忆功能



资料来源：《神经网络与深度学习》（邱锡鹏），ISBN: 9787111649687、华泰研究

LSTM：将短期记忆进一步延长

LSTM 引入门控机制，解决 RNN 的长程依赖问题。循环神经网络在学习过程中的主要问题是梯度消失或爆炸问题（也称为长程依赖问题），可以简单理解为参数无法正常更新（梯度变得很大或变为 0），很难建模长时间间隔状态之间的依赖关系。长短时记忆网络（Long Short-term Memory，LSTM）引入门控机制，很好了解决了 RNN 的上述问题。

LSTM 网络特点在于新内部状态和门控机制的引入。1）新的内部状态：与 RNN 相比，LSTM 网络引入一个新的内部状态（internal state）c_t，专门进行线性的循环信息传递。在每个时刻 t，LSTM 网络的内部状态 c_t 记录了到当前时刻为止的历史信息，相当于延长了网络的记忆特性。**2）门控机制：**引入遗忘门、输入门和输出门。其中，遗忘门控制上一个时刻的内部状态需要遗忘多少信息；输入门控制当前时刻的候选状态有多少信息需要保存；输出门控制当前时刻的内部状态有多少信息需要输出给外部状态。参数训练方面，虽然相比 RNN，LSTM 引入了更多参数矩阵，但依然可以使用梯度下降算法进行参数更新。

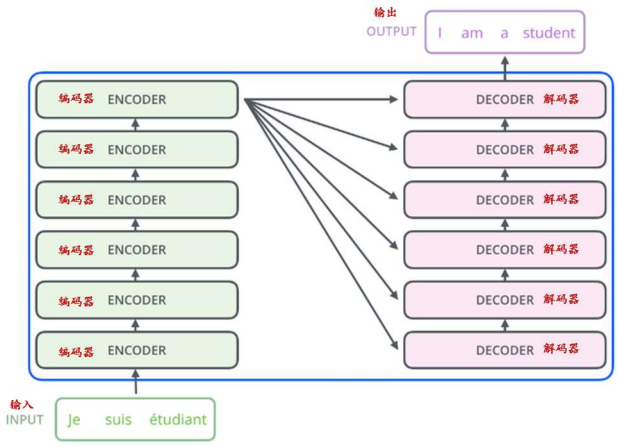
循环神经网络的隐状态 s 为短期记忆，新内部状态 c 具有更长的记忆。循环神经网络中的隐状态 s 存储了历史信息，可以看作一种记忆。在简单循环网络中，隐状态每个时刻都会被重写，因此是一种短期记忆。在神经网络中，长期记忆可以看作网络参数，隐含了从训练数据中学到的经验，其更新周期要远远慢于短期记忆。在 LSTM 网络中，记忆单元 c 可以在某个时刻捕捉到某个关键信息，并有能力将此关键信息保存一定的时间间隔，其保存信息的生命周期要长于短期记忆 s，但又远远短于长期记忆，因此称为长（的）短期记忆。

Transformer：基于自注意力机制的新兴架构

Transformer 是一种基于自注意力机制的编码器-解码器深度学习模型。2017 年 6 月，Google Brain 在神经信息处理系统大会（NeurIPS）发表论文“Attention is all you need”，首次提出了基于自我注意力机制（self-attention）来提高训练速度的 Transformer 模型，将其用于自然语言处理。Transformer 架构由编码器和解码器（Encoder-Decoder）两个部分组成，每个部分包含多个子层，如多头自注意力、前馈神经网络、和归一化等。Transformer 架构的优点是：1）可以并行处理输入序列的所有元素，大大提高训练速度；2）不需要依赖循环或卷积结构，能够捕捉长距离的依赖关系，3）可以通过预训练和微调的方式适应不同的任务。因此，Transformer 一经问世便大有取代传统 CNN、RNN 的趋势。

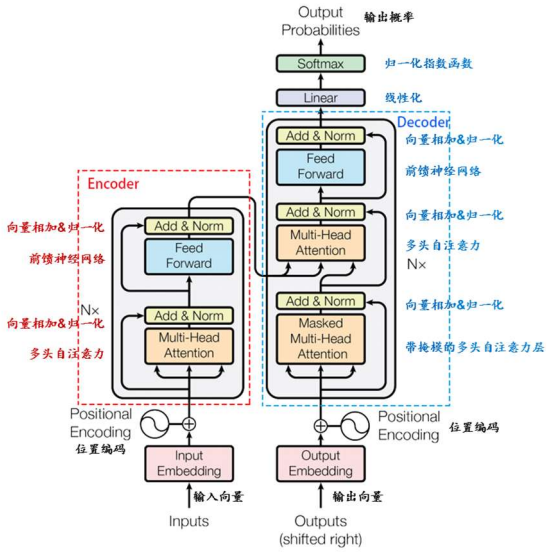
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表146： Transformer 模型中的 encoer-decoder 架构



资料来源：《Attention is all you need》，Ashish Vaswani（2017）、华泰研究

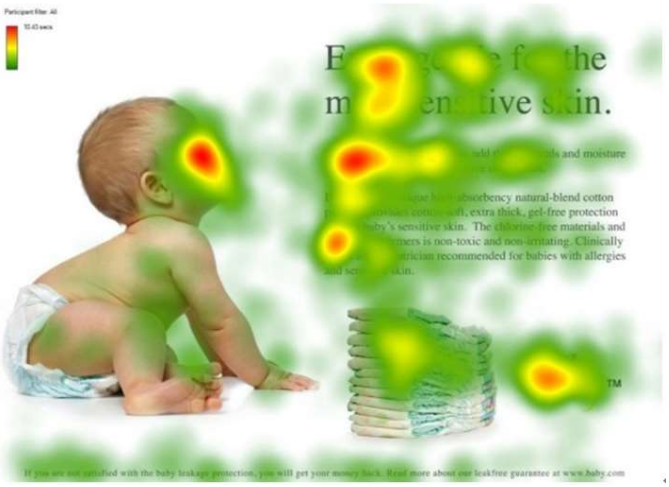
图表147： Transformer 模型架构



资料来源：《Attention is all you need》，Ashish Vaswani（2017）、华泰研究

自注意力机制是 Transformer 的核心理念，注意力机制帮助机器在学习过程中聚焦在重要信息上。深度学习中的注意力机制从本质上讲和人类的选择性视觉注意力机制类似，核心目标是从众多信息中选择出对当前任务目标更关键的信息，忽略大多不重要的信息。

图表148： 人类的注意力机制（深色代表注意力聚焦点）

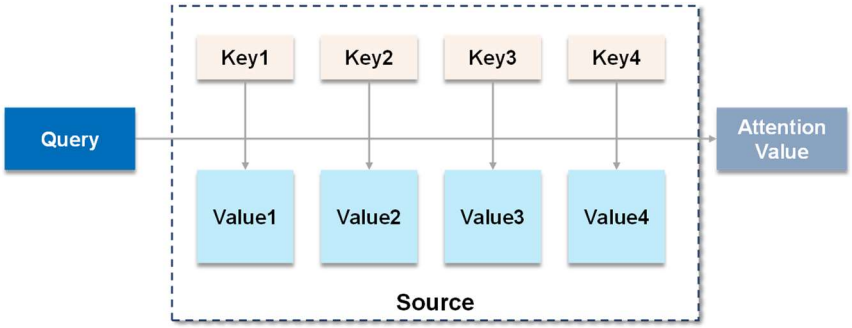


资料来源：CSDN、华泰研究

注意力模型可以用 Query、Key 和 Value 模型进行描述。本质上 Attention 机制是对 Source 中元素的 Value 值进行加权求和，而 Query 和 Key 用来计算对应 Value 的权重系数。以大脑读图为例，Value 可以理解为人眼视网膜对整张图片信息的原始捕捉，不受“注意力”所影响；Key 与 Value 相关联（<Key,Value>数据对），是图片原始信息所对应的关键性提示信息；读者大脑有意识地向图片获取信息，即发起了一次 Query，包含了读者的意图等信息。在一次读图过程中，Query 与 Key 之间计算出每个 Key 对应 Value 的权重系数，得到最具有吸引力的部分，然后对 Value 进行加权求和，即得到了最终的 Attention Value 数值。

更多免费行业报告 并购家 www.ipoipo.cn

图表149： Attention 机制的本质思想



资料来源：CSDN、华泰研究

Transformer 的自注意力机制是特殊情况下的注意力机制。在一般任务的 Encoder - Decoder 框架中，输入 Source 和输出 Target 内容是不一样的，例如对于英-中机器翻译来说，Source 是英文句子，Target 是对应的翻译出的中文句子。注意力机制发生在 Target 的元素 Query 和 Source 中的所有元素之间。而自注意指的不是 Target 和 Source 之间的注意力机制，而是 Source 内部元素之间或者 Target 内部元素之间发生的注意力机制，也可以理解为 Target = Source 这种特殊情况下的注意力计算机制。具体计算过程相同，只是计算对象发生变化。**自注意力机制更容易捕获句子中长距离的相互依赖的特征，且对于增加计算的并行性也有直接帮助作用。**

发展：从 GPT-1 到 GPT-4，开启大模型新纪元

GPT 系列模型贯彻 Decoder-only 路线，通过增大模型参数、训练集和 RLHF 等技术，不断提高模型表现。OpenAI 的 GPT 系列模型，均是基于 Google 提出的 Transformer 模型的 Decoder（解码器）架构，每代模型仅对架构进行微调。

从模型参数量和训练数据集维度看，在 GPT-3 模型之前，参数量和训练数据量均呈现快速增长态势，尤其是 GPT-3 模型参数量为 1750 亿，达到阶段性巅峰；而从 ChatGPT 开始，OpenAI 不再公布模型细节。

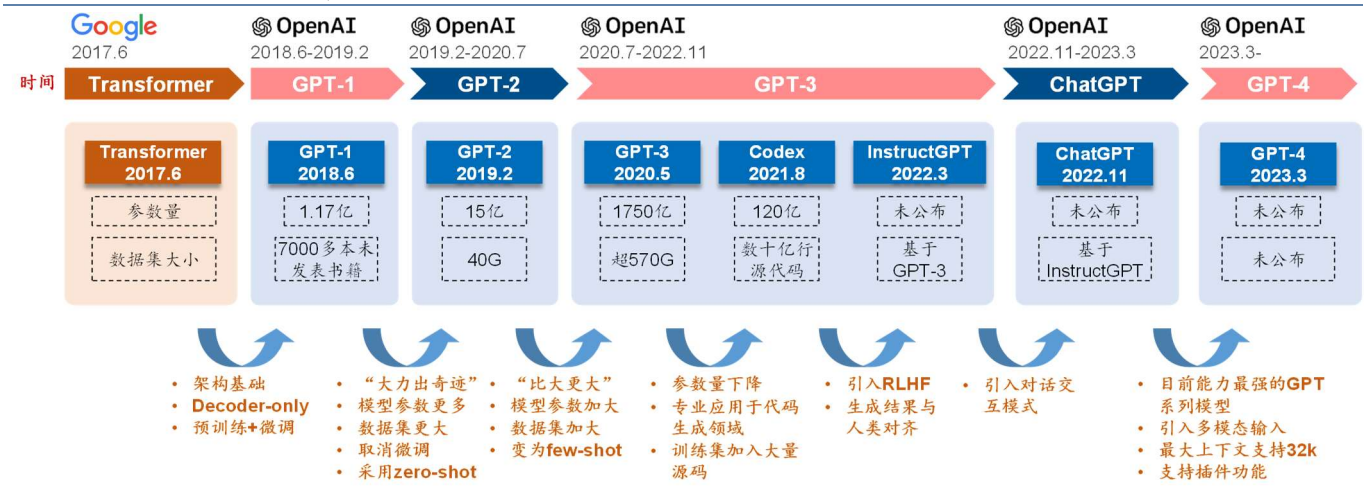
从技术维度看，OpenAI 十分注重模型的泛化能力，从 GPT-2 开始探究 zero-shot（零样本学习）和 few-shot（少样本学习）等技术，证实了模型在未训练过的数据上依然有较好的表现，泛化能力凸显。

从应用维度看，基于 GPT-3 的 Codex，是 GPT 系列在代码专业领域应用上的重要探索，通过在训练数据集中加入专业的代码，实现了较强的代码生成能力；ChatGPT 则是 GPT 系列应用的重要里程碑，通过用户友好的对话式方式，将 GPT 技术带入大众视野，成为近来 AI 飞速发展的重要导火索。

从进化维度看，目前的最新模型 GPT-4 是能力最强的 GPT 系列模型，在吸收并强化过往模型的语义理解、逻辑推理、代码生成等能力的基础上，增加了多模态输入的能力，进一步扩展了模型的边界，并通过支持更长的上下文窗口，增强模型对更长的提示词（prompt）的支持能力，能够让长问题的回答更加连贯和统一。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表150：从 GPT-1 到 GPT-4 发展历程

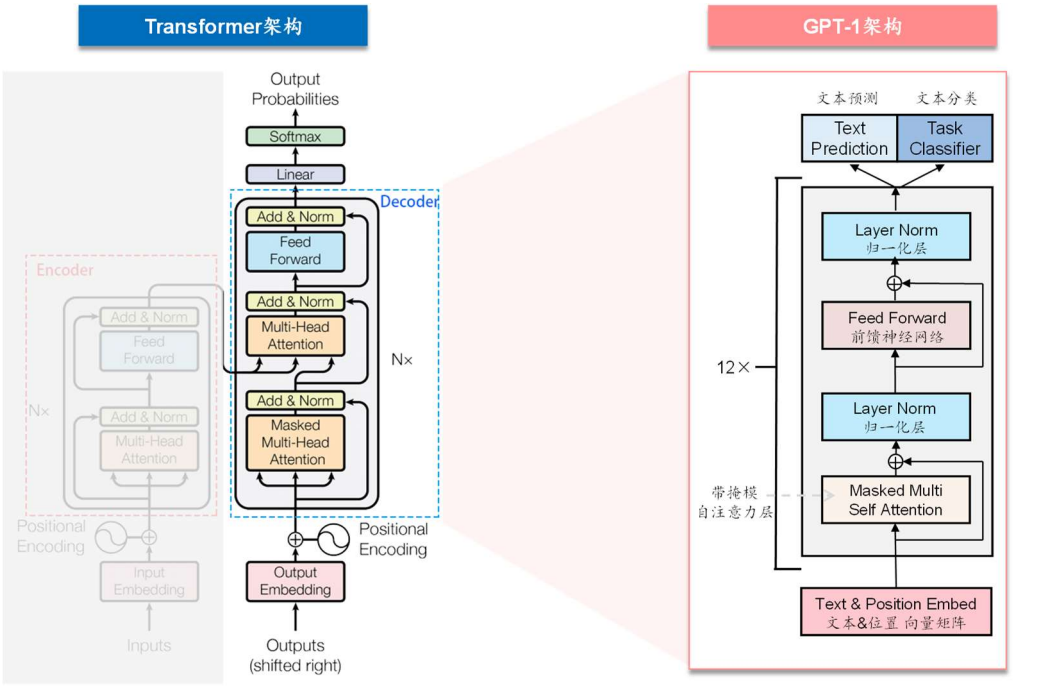


资料来源：公司官网，华泰研究

GPT-1：确定生成式预训练模型基本路线

GPT-1 是预训练+微调相结合的语言理解任务半监督模型。GPT 全称生成式预训练（Generative Pre-training，GPT），于 2018 年 6 月由 OpenAI 在论文《Improving Language Understanding by Generative Pre-Training》中提出。从架构上看，GPT-1 基于 Transformer 架构，但是仅保留了架构中的解码器（Decoder）部分。

图表151：GPT-1 的架构



资料来源：《Improving Language Understanding by Generative Pre-Training》，Alec Radford（2018）、华泰研究

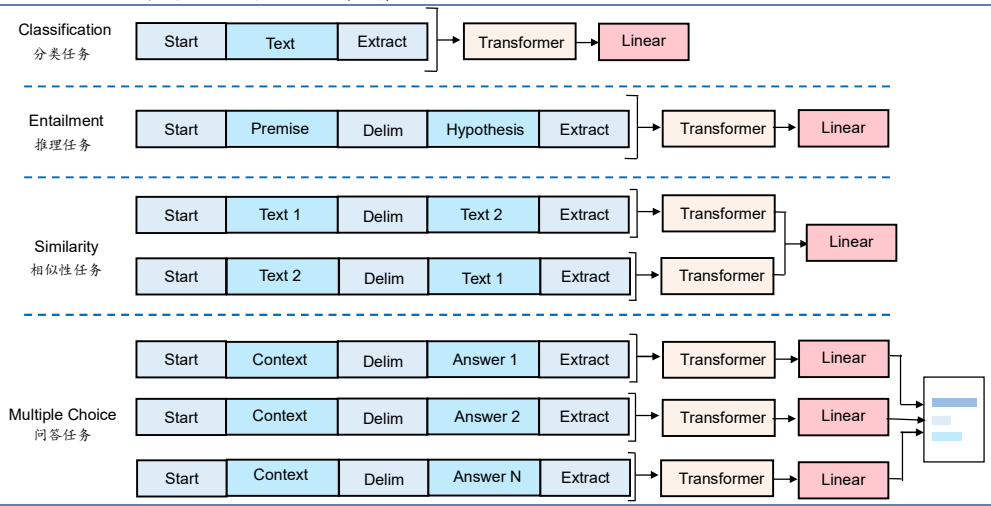
GPT-1 的训练过程分为无监督训练和有监督微调。GPT-1 的训练过程分为两个阶段：第一阶段为非监督预训练阶段，即在大型文本语料库上学习高容量语言模型。第二阶段为监督微调阶段，即在第一阶段训练的基础上，将参数调整为受监督的目标任务，进行进一步优化。

1）第一阶段：采用多层 Transformer 模型的解码器（Decoder）为语言模型，进行基于大文本语料库的高容量无监督学习。

免责
更多免费行业报告 并购家 www.ipoipo.cn

2）第二阶段：在微调时，若输入的问题具有结构化特征，例如有序的句子对或文档、问题和答案的三元组，则需要首先将问题转化成特定的序列模块，再进行训练。例如，对于文本蕴涵任务，将前提和假设序列相连接，并在中间插入分隔符；对于相似性任务，则将两个需要比较的句子按照不同顺序排列，分别进行处理后再做比较；对于更复杂的问题回答和常识推理任务，则需要将文本、问题和可能的答案分别连接成不同序列，并对所有序列做独立的模型处理，将结果归一化得到可能答案的输出分布。

图表152：GPT-1 架构以及用于微调不同任务的输入转换模块



资料来源：《Improving Language Understanding by Generative Pre-Training》，Alec Radford（2018）、华泰研究

无监督训练采用大数据集，有监督微调采用针对任务的数据集。在无监督训练阶段，采用了 12 层解码器（Decoder）堆叠的 Transformer 模型，共包含 12 个掩蔽自注意力头（masked self-attention head）。训练数据集方面，采用 BooksCorpus 数据集，该数据集包括 7000 多本来自风格不同的未出版书籍，内含长段的连续文本。也可使用备选数据集 1B Word Benchmark。在无监督训练阶段，提高解码器的层数能够有效提高训练效果。在有监督微调阶段，可根据不同任务类型选择不同的数据集。

图表153：GPT-1 在有监督微调阶段使用的部分数据集

任务	数据集
自然语言推理	SNLI，MultiNLI，Question NLI，RTE，SciTail
问题解答	RACE，Story Cloze
句子相似性	MSR Paraphrase Corpus，Quora Question Pairs，STS Benchmark
分类	Stanford Sentiment Treebank-2，CoLA

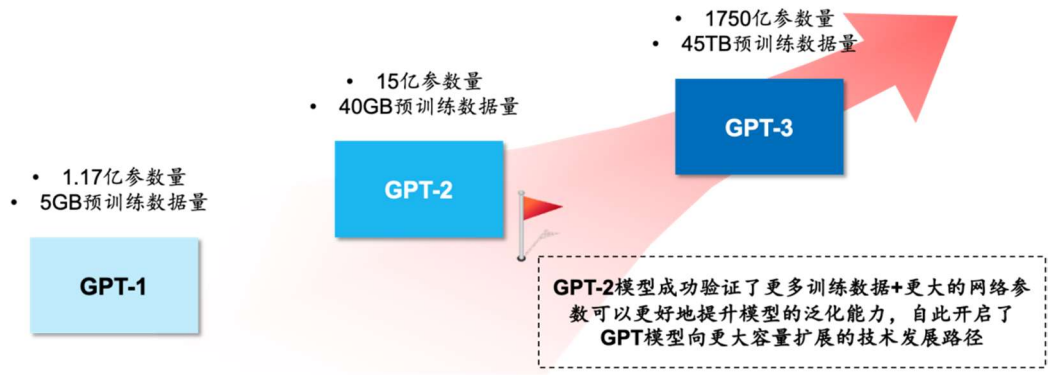
资料来源：《Improving Language Understanding by Generative Pre-Training》，Alec Radford（2018）、华泰研究

GPT-2：关注模型的零次学习 Zero-shot 能力

理解 GPT 系列论文是我们理解 GPT 模型技术路径演绎的重要方法。GPT-2 论文的核心重点在于验证是否可以通过更多的训练数据+更大的网络参数让模型具备零次学习（Zero-shot Learning）能力；即通过更广泛学习+更强的学习能力，加速 GPT 从“专才”走向“通才”。实验结果证明，当训练数据足够多、模型参数足够大时，就可以通过训练一个模型，处理多类任务；自此确立了 GPT 模型向更大容量演进的发展目标。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表154：GPT 模型演进路径



资料来源：OpenAI 官网、华泰研究

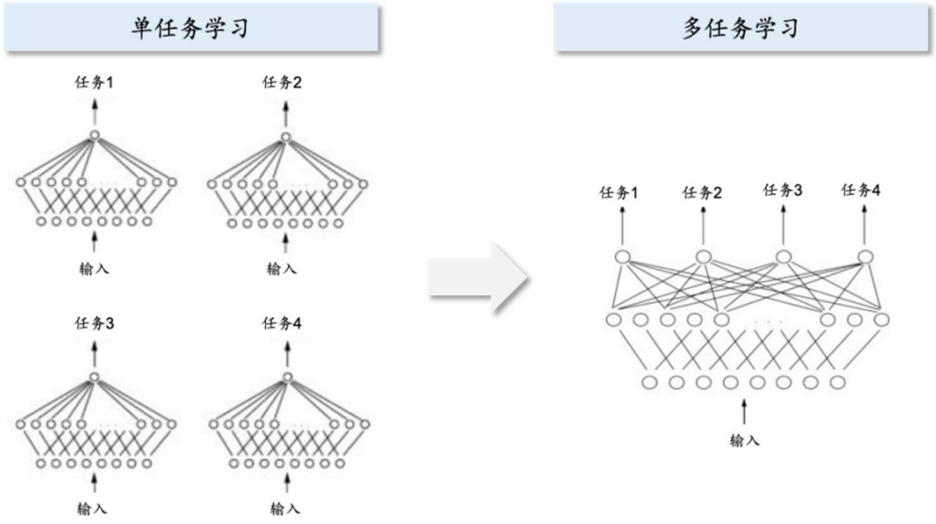
《Language Models are Unsupervised Multitask Learners》论文发表，GPT-2 开启“一个模型处理多类任务”的新时代。GPT(Generative Pre-trained Transformer)系列预训练语言模型，早在 2018 年在论文《Improving Language Understanding by Generative Pre-Training》中被提出，初代 GPT-1 模型验证了“通过无监督预训练，针对特定任务微调，处理有监督任务”模型的有效性。2019 年 2 月，论文《Language Models are Unsupervised Multitask Learners》正式发表，基于“无监督学习的全局最小也必定是监督学习的全局最小”的理论基础，GPT-2 模型去掉 fine-tune 层，不再针对不同任务分别进行微调建模，而是通过使用更多的网络参数和更大的数据集来强化模型的泛化能力。

GPT-2 模型成功验证了更多训练数据+更大的网络参数可以有效地提升模型的泛化能力，实现了“训练一个模型可以解决多类任务”的效果，不再需要针对不同任务进行模型微调。GPT-2 的成功自此开启了 GPT 模型向更大容量扩展的技术发展路径。

特点一：预训练阶段采用多任务学习方式

GPT-2 在预训练阶段采用多任务学习方式。不同于单任务学习，单次仅针对一个任务学习，以及模型参数独立的特点；多任务学习，单次可同时学习多个任务，同时各个任务之间的模型参数共享，可以实现已学习信息的共享、补充，有效提升了模型的学习效率。

图表155：GPT-2 预训练采用多任务学习方式

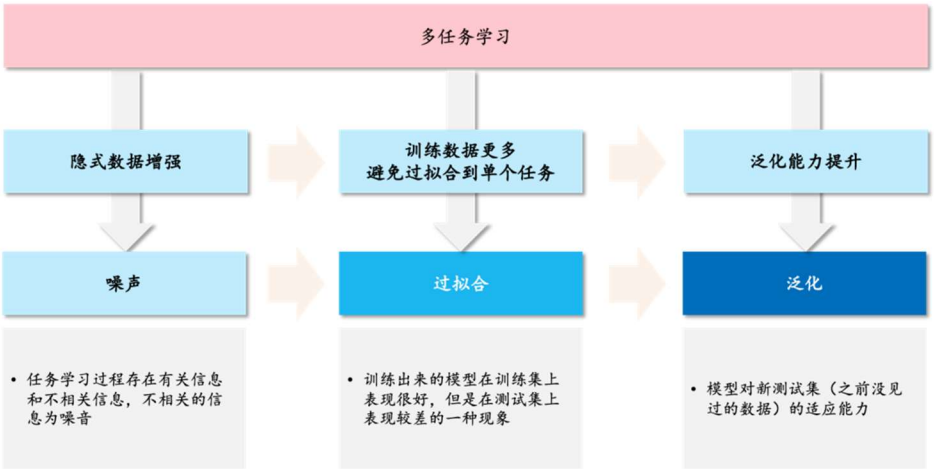


资料来源：《Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts》，Alex Kendall (2017)、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

多任务学习有效防止模型过拟合，提升模型的泛化能力。为理解多任务学习对模型泛化能力的提升，首先需要明晰噪音、过拟合、泛化这几个概念。**1）噪音：**任务学习过程不相关的信息为噪音；**2）过拟合：**训练出来的模型在训练集上表现很好，但是在测试集上表现较差的一种现象；**3）泛化：**模型对新测试集（之前没见过的数据）的适应能力。三者之间的联系是，当把噪声作为有效信息或训练数据集较小时，容易导致模型过拟合，模型过拟合导致模型的泛化能力变差。多任务学习提升模型泛化能力的原因主要有三点：1）训练数据集更大；2）多任务之间有一定相关性，多任务学习相当于隐式数据增强；3）多任务学习的共享模块需要兼顾所有任务，一定程度避免了模型过拟合到单个任务的训练集。

图表156：多任务学习可有效提升模型的泛化能力

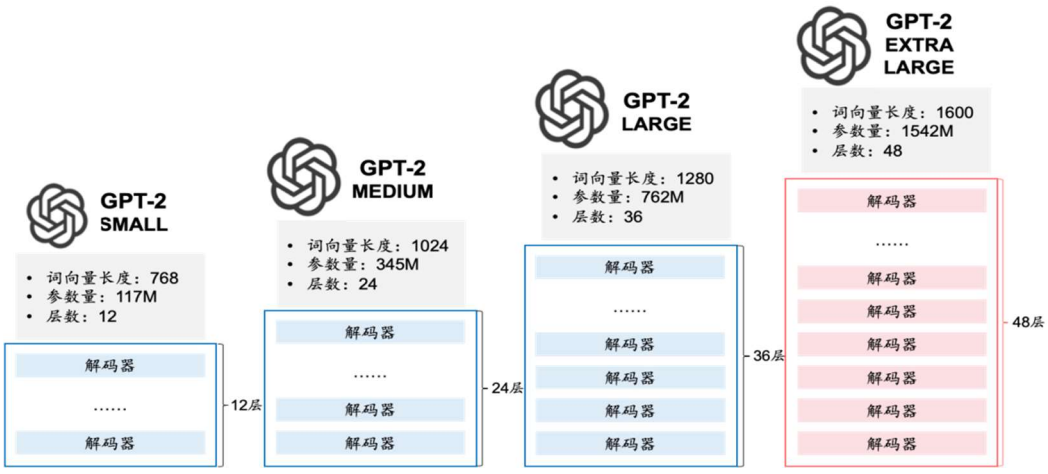


资料来源：《神经网络与深度学习》，邱锡鹏（2020）、华泰研究

特点二：GPT-2 拥有较大的模型容量

GPT-2 最大模型（GPT-2 EXTRA LARGE）的层数达到 48 层，词向量长度到 1,600。1）层数：理解层数对 GPT 模型的影响，需要理解 GPT 的模型架构。GPT 模型由层层堆叠的 Transformer 模块组成，因此可以理解为模型层数越多，模型可以完成的非线性映射就越复杂，模型的学习能力更强；**2）词向量长度：**词向量长度即隐藏层的神经元个数，每一层 Transformer 都是一个隐藏层，隐藏层的神经元个数越多，即可以产生越复杂的组合，拟合越复杂的情况，解决越复杂的问题，计算量也随之越来越大。GPT-2 论文（《Language Models are Unsupervised Multitask Learners》）提及四种规模的 GPT-2 模型，GPT-2 的模型层数最多为 48 层，词向量长度（隐藏层维度）最大为 1,600，模型的最大参数量多达 15 亿。

图表157：GPT-2 模型层数最大达 48、词向量长度最大达 1,600



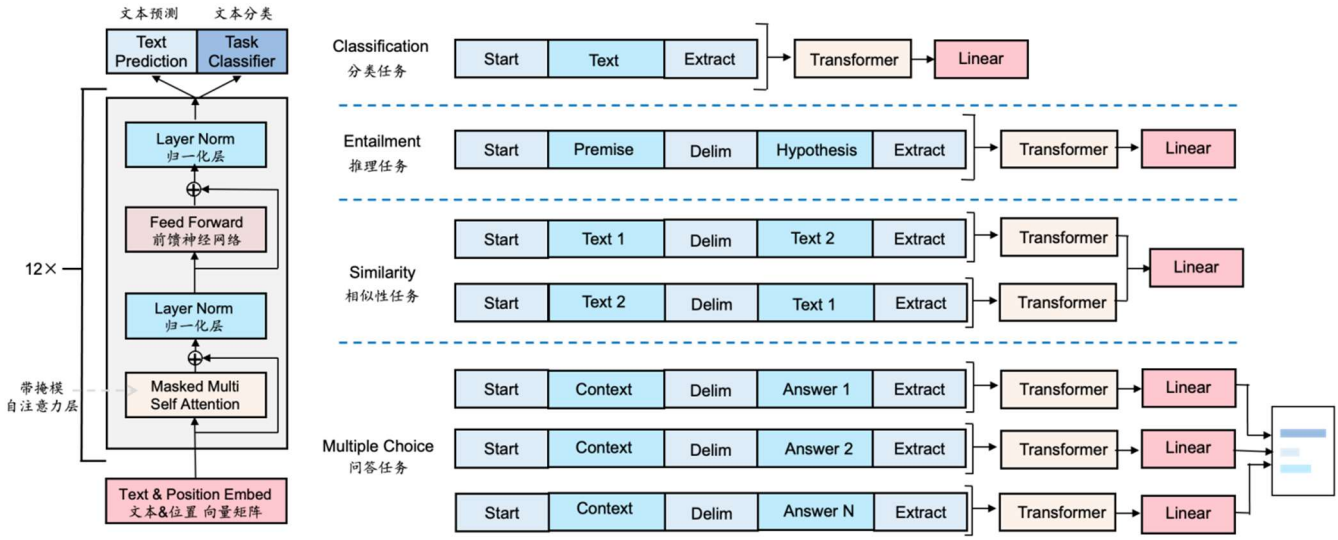
资料来源：《Language Models are Unsupervised Multitask Learners》，Alee Radford（2019）、华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

特点三：GPT-2 取消 fine-tune 微调建模层

GPT-2 旨在“训练一个模型，解决多类任务”，因此取消 fine-tune 微调建模层。GPT-2 聚焦模型的泛化能力，旨在通过多任务学习训练一个模型，训练好的这一个模型可以解决多类任务，无需再根据特定任务进行模型微调。因此，GPT-2 模型在训练任务中取消了 fine-tune 任务，只保留了语言模型部分的训练，同时将诸多下游任务根据各自的特点，规划成语言模型的训练，使得最终的语言模型学到的东西更丰富，拥有更多样的应用，奠定了模型的 Zero-shot Learning（零次学习）的基础。

图表158：GPT-1 四类任务微调建模示意图



资料来源：《Improving Language Understandingby Generative Pre-Training》，Alee Radford（2018）、华泰研究

GPT-3：模型参数达 1750 亿，few-shot 进一步提高性能

GPT-3 将训练参数量扩大到 1750 亿。2020 年 5 月，OpenAI 发表论文《Language Models are Few-Shot Learners》，提出 GPT-3 模型。GPT-3 模型的参数量达 1750 亿，是 GPT 史上最大的参数规模。在 GPT-2 的 zero-shot learning（零次学习）基础上，GPT-3 在 zero-shot、one-shot（一次学习）和 few-shot（少量学习，通常 10-100 次）方面均进行了评估，取得了较好的训练结果：例如，GPT-3 在 zero/one/few-shot 设置下 CoQA（测试阅读理解能力）分别达到 81.5、84.0、85.0F1（F1 分数是统计学中用来衡量二分类模型精确度的一种指标，值越大意味着模型越好）；在 zero/one/few-shot 设置下 TriviaQA（测试闭卷问答能力）分别达到 64.3%、68.0%、71.2%的准确度，部分结果甚至超过微调模型。OpenAI 指出，GPT-3 也可以在传统的微调设置中进行评估。

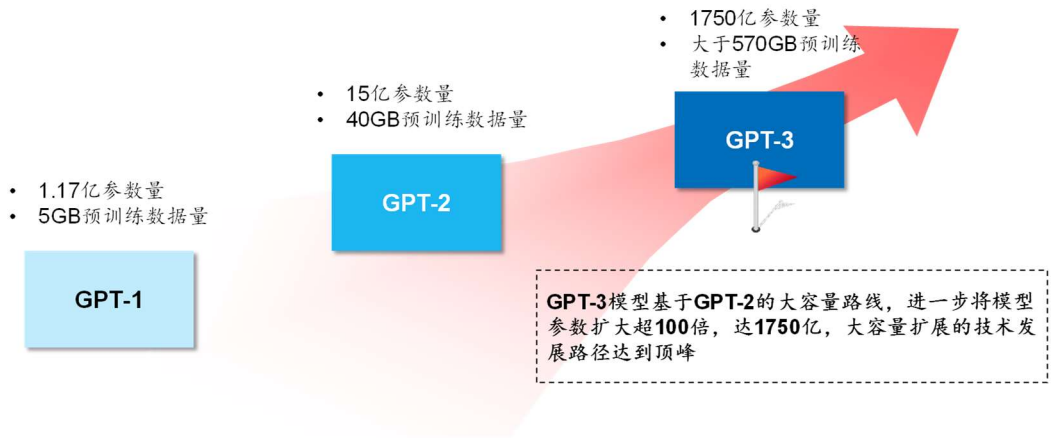
GPT-3 模型延续了 GPT-2 模型“更多训练数据+更大的网络参数可以有效地提升模型的泛化能力”的理念，将模型参数提升到新高度。同时与 GPT-2 一样舍弃了针对不同任务进行的微调，转而用 few-shot 来实现下游任务，兼顾了训练效果与成本。

特点一：架构基本不变，参数量大大增加

GPT-3 延续了 GPT-2 的大容量发展路线。GPT-2 已经证明，通过更多的训练数据+更大的网络参数能够让模型具备 zero-shot 学习能力，并取得了一定的效果，由此确立了 GPT 模型向更大容量演进的发展路线。GPT-3 延续了这一路线，将模型参数由 GPT-2 的 15 亿增加到 1750 亿，提升幅度超 100 倍。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表159：GPT 模型演进路径—GPT-3

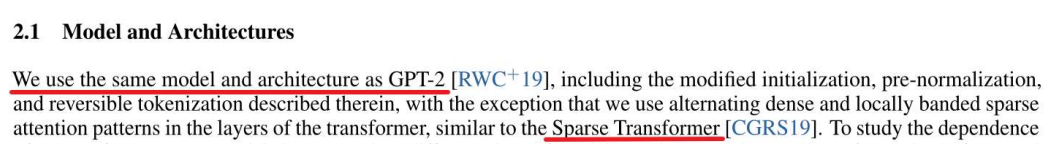


资料来源：OpenAI 官网、华泰研究

大训练参数能明显提高 GPT-3 模型准确性。从上下文学习能力训练结果准确率来看，对于 GPT-3 模型，当模型参数量较低时，模型准确率较低，例如参数量 1.3B 的模型准确率不超过 10%。当模型参数提升 10 倍至 13B 时，最高准确率接近 30%。当模型参数达到最大 175B 时，准确率最高接近 70%，进一步验证了 GPT-2 大容量路线的正确性。

GPT-3 和 GPT2 采用几乎同样的架构。GPT-3 和 GPT-2 模型和架构基本相同，包括修改后的初始化、预归一化和可逆标记化，唯一区别在于将 transformer 中注意力模式替换成了类似 Sparse Transformer 的稀疏注意力模式。

图表160：GPT-3 采用和 GPT-2 同样的架构



2.1 Model and Architectures

We use the same model and architecture as GPT-2 [RWC+19], including the modified initialization, pre-normalization, and reversible tokenization described therein, with the exception that we use alternating dense and locally banded sparse attention patterns in the layers of the transformer, similar to the Sparse Transformer [CGRS19]. To study the dependence

资料来源：《Language Models are Few-Shot Learners》，Tom Brown（2020）、华泰研究

特点二：采用 few-shot 学习完成下游任务

Few-shot 取代 zero-shot，训练效果得到进一步加强。GPT-2 模型中，在下游训练时采用 zero-shot 学习理念，在执行各类子任务时不给任何样例，属于较为极端的情况。虽然在 GPT-2 在某些测试中取得了不错的效果，但在一些任务上结果不达预期。因此，OpenAI 引入了 few-shot，即对于特定任务仅给予少量的样例（10-100 个），没有任何梯度更新或微调，任务和 few-shot 示例均通过与模型的文本交互指定。从多种训练基准综合训练结果来看，few-shot 能够取得比 zero-shot 更好的准确度。

特点三：预训练数据集达 570GB 以上

GPT-3 训练数据集为多种数据集的混合。数据集以 Common Crawl 为基础，其大小为 45TB。但 Common Crawl 数据集的质量低于更精确的数据集，为了保证数据集质量，一方面对 Common Crawl 进行过滤，大小压缩到 570GB，另一方面增加了质量更高的数据集，如 WebText2、Wikipedia 等。在训练期间，数据集的采样并不与其大小成比例，而是质量更高的数据集采样频率更高，因此 CommonCrawl 和 Books2 数据集在训练期间采样次数少于一次，其他更高质量的数据集采样次数为 2-3 次，以换取更高质量的训练数据。比较来看，GPT-1 和 GPT-2 的训练数据集大小分别为 5GB 和 40GB，GPT-3 数据集大小超 GPT-210 倍以上。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表161：GPT-3 训练数据集组合

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. "Weight in training mix" refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

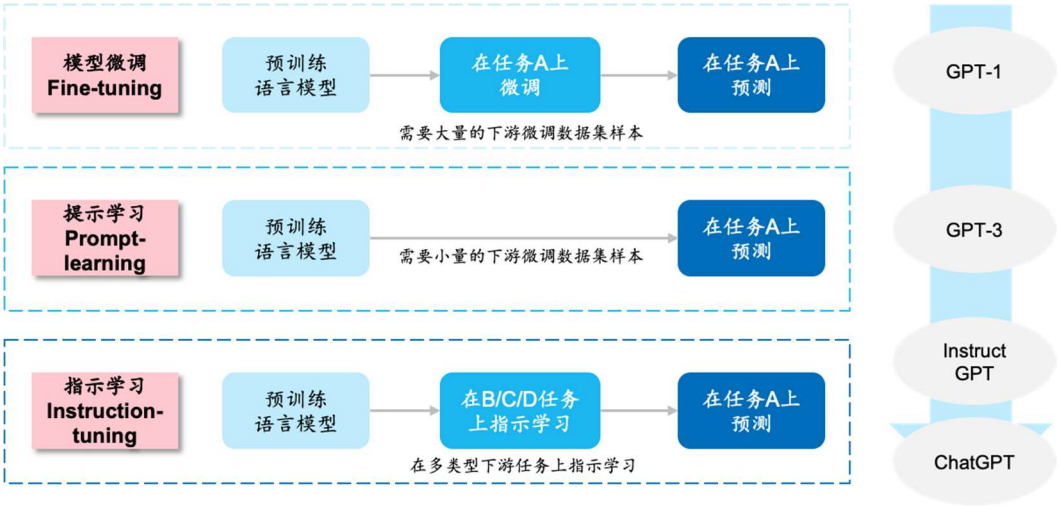
资料来源：《Language Models are Few-Shot Learners》，Tom Brown（2020）、华泰研究

ChatGPT / InstructGPT：增加人类反馈强化学习（RLHF）

#1 优化目标：从提示学习到指示学习，聚焦需求对齐（alignment）

从模型微调到提示学习再到指示学习，GPT 模型零次学习能力不断提升。2022 年论文《Finetuned Language Models Are Zero-Shot Learners》提出指示学习（Instruction-tuning）思想，以提升模型的零次学习能力。对比来看：**1）模型微调（Fine-tuning）：**以 GPT-1 为代表，需要大量的微调数据集样本，特定模型解决特定任务；**2）提示学习（Prompt-learning）：**以 GPT-3 为代表，需要少量的微调数据样本，模型小样本学习（few-shot Learning）能力更强；**3）指示学习（Instruction-learning）：**以 FLAN、InstructGPT、ChatGPT 为代表，模型通过下游多类型任务的指示微调，零次学习（zero-shot Learning）能力更强；同时，提示学习与指示学习均为一个模型解决多类任务。

图表162：模型微调（Fine-tuning）、提示学习（Prompt-learning）、指示学习（Instruction-tuning）对比

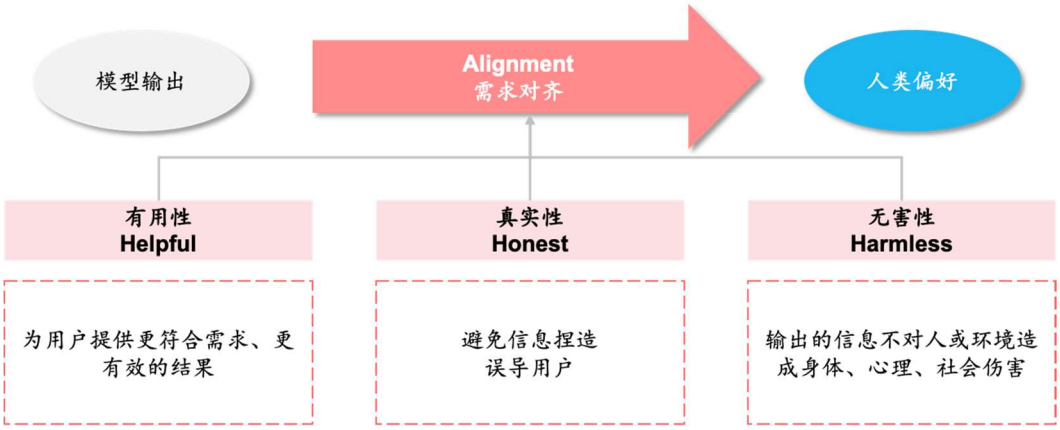


资料来源：《Finetuned Language Models are Zero-shot Learners》，Jason Wei（2022）、华泰研究

聚焦模型输出与人类需求对齐，引入人类反馈强化学习。GPT 系列模型发展至 GPT-3，已经在翻译、问答、摘要等任务上取得了良好的性能，研究者开始更加关注模型生成内容的有用性（Helpful）、真实性（Honest）、无害性（Harmless），希望实现模型输出内容与人类偏好内容的需求对齐（alignment）。为提升模型的"3H"特性，从 InstructGPT 开始，InstructGPT、ChatGPT 均引入人类反馈强化学习（Reinforcement Learning from Human Feedback，RLHF），实现基于人类反馈的模型微调，让模型输出更符合人类期待。

更多免费行业报告 并购家 www.ipoipo.cn

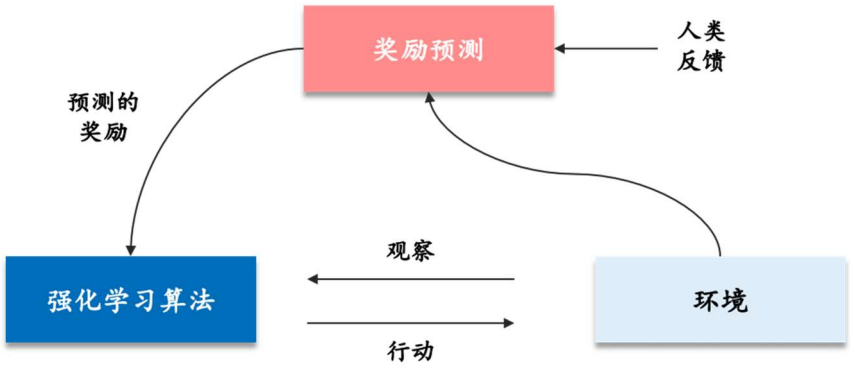
图表163：InstructGPT 的“3H”优化目标



资料来源：《Training language models to follow instructions with human feedback》，Long Ouyang（2022）、华泰研究

人类反馈可作为强化学习的奖励，通过引入语言模型微调，可使模型输出与人类需求对齐。从技术原理来看，强化学习通过奖励（Reward）机制来指导模型训练，奖励机制可视为传统训练机制的损失函数；同时，奖励的计算要比损失函数更灵活、多样（例如 AlphaGO 的奖励是对局的胜负），代价是奖励计算不可导，不能直接用来做反向传播；强化学习的思路是通过对奖励的大量采样来拟合损失函数，从而实现模型的训练。**类似的，人类反馈也不可导，也可以作为强化学习的奖励，从而产生基于人类反馈的强化学习。**

图表164：人类反馈强化学习（RLHF）基本原理



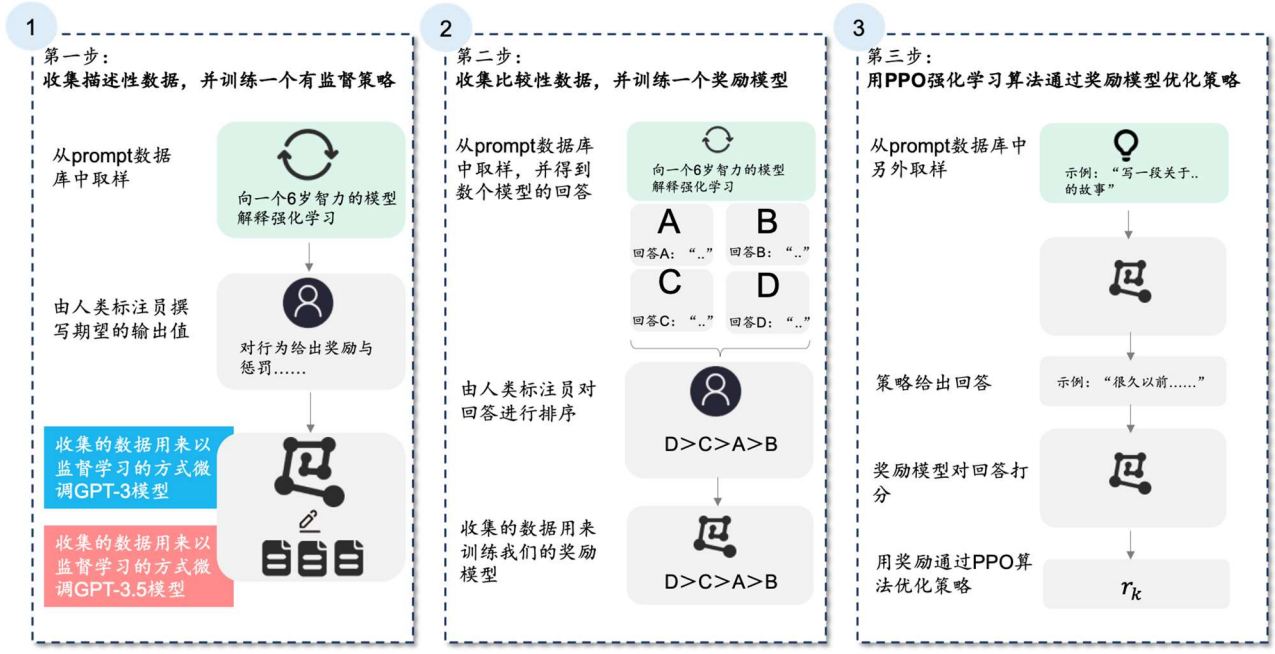
资料来源：《Deep Reinforcement Learning from Human Preferences》，Paul F Christiano（2017）、华泰研究

#2 训练过程：“三步走”实现人工反馈强化学习

ChatGPT 与 InstructGPT 的训练方法基本一致。了解 ChatGPT 与 InstructGPT 的优化目标之后，我们需要进一步梳理模型的具体训练过程。对比论文《Training language models to follow instructions with human feedback》（2022）中的 InstructGPT 训练方法与 OpenAI 官方发布的 ChatGPT 训练方法，我们看到，ChatGPT 与 InstructGPT 的训练方法基本一致，区别在于 InstructGPT、ChatGPT 分别基于 GPT-3、GPT-3.5 进行模型微调。考虑到 ChatGPT 尚未有官方论文发布，我们以 InstructGPT 论文为基准，对 InstructGPT 的训练过程进行了详细梳理，具体可分为有监督微调、奖励模型训练、PPO 强化学习三个步骤。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表165： ChatGPT 与 InstructGPT 的训练方法基本一致



注：图中蓝色部分为 InstructGPT 训练示意、粉色部分为 ChatGPT 训练示意

资料来源：OpenAI 官网、《Training language models to follow instructions with human feedback》，Long Ouyang（2022）、华泰研究

GPT-4：多模态确认，在专业和学术上表现亮眼

#1 GPT-4：支持多模态输入，安全问题或成为 LLM 关注焦点

GPT-4 支持多模态输入，安全问题或成关注焦点。北京时间 3 月 15 日凌晨，OpenAI 召开发布会，正式宣布 GPT 模型家族中最新的大型语言模型（LLM）—GPT-4。GPT-4 可以接受图像和文本输入并产生文本输出。同时，OpenAI 发布了 GPT-4 相关技术文档。从技术文档结构看，OpenAI 并未对 GPT-4 模型本身做过多介绍，而是将大量篇幅留给模型安全性相关讨论。我们认为，在 LLM 走向大规模应用的过程中，随着模型规模和能力的提升，将暴露出更多安全方面的问题，或将成为未来 LLM 的重要关注点。

GPT-4 于 2022 年 8 月完成训练，技术细节未完全公布。据 GPT-4 技术文档，GPT-4 于 2022 年 8 月完成训练，此后，OpenAI 一直在评估、对抗性测试并迭代和改进模型，并提出各种系统级安全缓解措施。此外，OpenAI 在 GPT-4 文档中明确提出，考虑到竞争格局和大型模型（如 GPT-4）的安全影响，本文档没有包含有关架构（包括模型大小）、硬件、训练计算、数据集构造、训练方法或类似内容的进一步细节。文档明确提供的内容包括：1）GPT-4 是一个基于 transformer 的预训练模型，用于预测文字的下一个 token；2）使用公开可用数据（如互联网数据）和第三方提供商授权的数据；3）使用来自人类反馈的强化学习（RLHF）对模型进行微调。

图表166： OpenAI 关于不透露技术细节的说明

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [33] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [34]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

资料来源：《GPT-4 Technical Report》，OpenAI（2023）、华泰研究

免责

更多免费行业报告 并购家 www.ipoipo.cn

GPT-4 API 已开放等待列表（waitlist），API 价格提升明显。当获得访问权限后，即可以对 GPT-4 模型进行纯文本请求（图像输入仍然处于有限的测试阶段），OpenAI 将自动将模型更新为推荐的稳定模型。此外，GPT-4 提供两个版本，GPT-4-8k 和 GPT-4-32k，前者支持的上下文长度为 8192 个 token，后者支持 32768-context（约 50 页文本，提供有限访问）。GPT-4-8k 定价为每 1k 提示 token 0.03 美元，每 1k 完成 token 0.06 美元。默认速率限制为每分钟 40k 个 token，每分钟 200 个请求；GPT-4-32k 定价为每 1k 提示 token 0.06 美元，每 1k 完成 token 0.12 美元。（API 价格网址 https://openai.com/pricing）

图表167：GPT-4 API 和 GPT-3.5-turbo API 价格对比

GPT-4 API 价格

Model	Prompt	Completion
8K context	$0.03 / 1K tokens	$0.06 / 1K tokens
32K context	$0.06 / 1K tokens	$0.12 / 1K tokens

GPT-3.5 API 价格

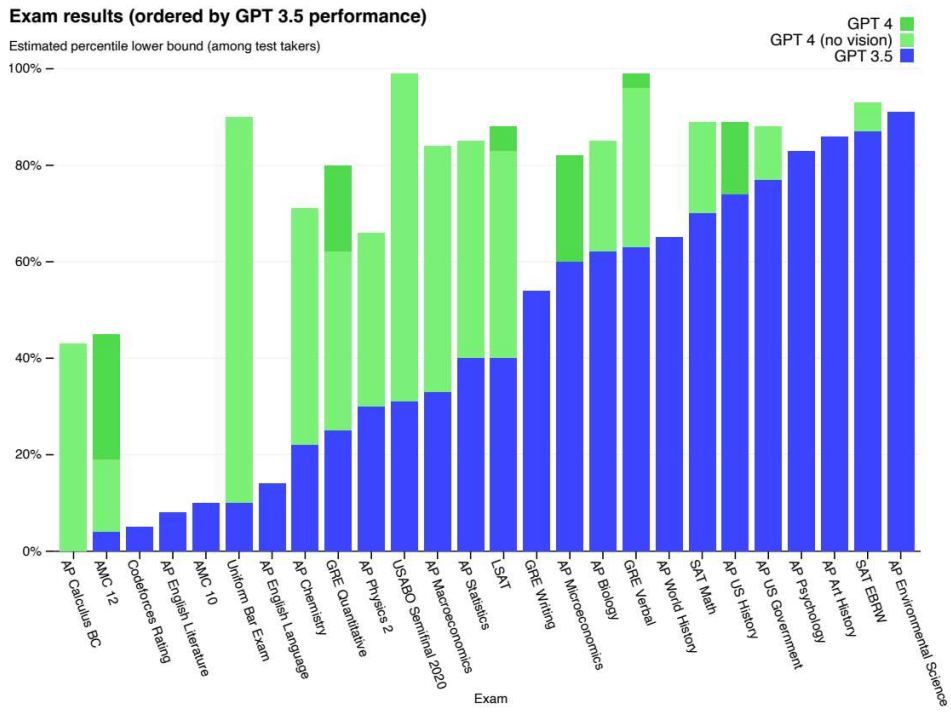
Model	Usage
gpt-3.5-turbo	$0.002 / 1K tokens

资料来源：《GPT-4 Technical Report》，OpenAI（2023）、华泰研究

#2 能力测试：在专业和学术基准上表现出了人类的水平

GPT-4 在各种专业和学术基准上表现出了人类的水平。OpenAI 在一系列不同的基准上测试了 GPT-4，包括最初为人类设计的模拟考试。考试题目包括多项选择题和自由问答题。OpenAI 为每种考试形式设计了单独的提示（prompt），并且在需要的场合增加了图像输入。GPT-4 并未专门针对这些考试科目进行训练，但仍取得了优秀的成绩，例如以大约前 10% 的成绩通过模拟律师资格考试，而 GPT-3.5 在该考试中成绩为后 10%。

图表168：GPT-4 在各种人类考试中表现优异

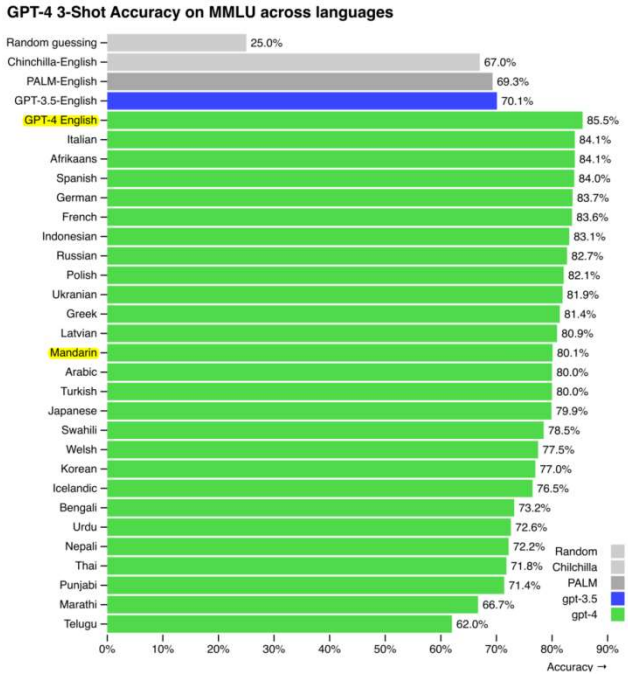


资料来源：《GPT-4 Technical Report》，OpenAI（2023）、华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

GPT-4 在大多数语言上的结果都优于 GPT 3.5 和现有语言模型。现有的多数机器学习测试基准都是基于英语编写的。为了测试 GPT-4 在其他语言中的表现，OpenAI 使用 Azure Translate 将 MMLU 基准测试（涵盖 57 个主题的多项选择问题）翻译成各种语言。结果表明，GPT-4 在大多数语言上的结果都优于 GPT 3.5 和现有语言模型的英语语言性能，包括全球使用人数较低的语言，如拉脱维亚语、威尔士语和斯瓦希里语。

图表169： GPT-4 在大多数语言上的结果都优于 GPT 3.5 和现有语言模型



资料来源：《GPT-4 Technical Report》，OpenAI（2023）、华泰研究

#3 多模态：支持图片输入，未提供具体算法

GPT-4 能够实现图片和文本的多模态输入。GPT-4 接受由图像和文本组成的提示（prompt，与纯文本设置类似），允许用户指定任何视觉或语言任务，同时输出文本信息。此外，语言模型中的 few-shot 提示和思维链等，在 GPT-4 中同样有效。但是，OpenAI 未在技术文档中给出如何实现图片的输入，以及如何对图片进一步处理以注入到模型中的说明。

图表170： GPT-4 视觉输入样例

资料来源：《GPT-4 Technical Report》，OpenAI（2023）、华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

海外落地：BloombergGPT、AutoGPT、SAM

大模型时代到来，模型体系与生态快速扩充，海内外企业坚定发力。受益于 ChatGPT 的成功问世，大模型的理解能力、推理能力、泛化能力得到充分验证，海内外企业纷纷加速大模型相关的产业布局，全面拥抱大模型时代的技术变革。

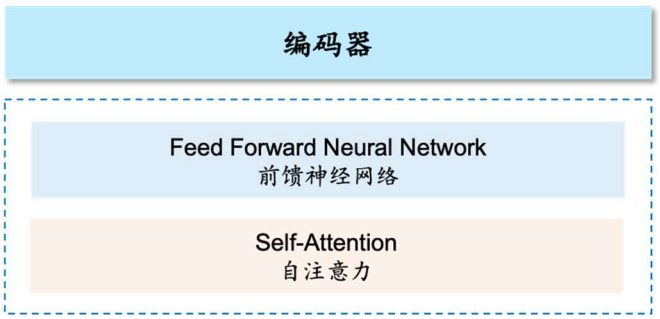
海外大模型持续多维拓展，重点关注大模型垂直落地、工程实现、模态丰富三大趋势。1）垂直落地：得益于通用大模型对于“大模型参数+大训练数据”的有效验证，基于垂直数据训练的行业大模型实现加速发展，BloombergGPT 作为首个金融大模型有效实现了 LLM 与金融垂直领域知识的深度融合，佐证了“开源模型+高质量垂直数据”LLM 搭建思路的有效性；**2）工程实现：**大模型的训练速度、训练成本以及工程化应用受到广泛关注，以 AutoGPT 为代表的工程化思路，充分验证了大模型有望成为 AI 时代的操作系统，改变人机交互范式；**3）模态丰富：**除大语言模型（Large Language Model，LLM）外，大模型逐步向图像、音频、视频等模态拓展，建议关注多模态对齐的大模型技术进展。

#1 BloombergGPT：提供构建垂直行业大模型的有益参考

BloombergGPT 发布，大模型实现金融垂直领域的加速落地。2023 年 3 月 30 日，Bloomberg 官方发布专为金融领域打造的大语言模型（Large Language Model，LLM）BloombergGPT，实现了 LLM 在垂直场景的加速落地。相较 OpenAI 的 GPT 系列（GPT-1、GPT-2、GPT-3、GPT-4）通用大模型，BloombergGPT 在保证 LLM 通用性能的基础上，更加聚焦金融场景，更加适应金融术语的复杂性与独特性，有效实现了 GPT+垂直知识的深度融合。

BloombergGPT 同样采用 decoder-only 路径，模型参数介于 GPT-2 与 GPT-3 之间。根据论文《BloombergGPT: A Large Language Model for Finance》的模型介绍显示，BloombergGPT 同样基于典型的 Transformer 架构，同时与 OpenAI 的 GPT 模型相同，采用了仅有解码器（decoder-only）的技术路径。考虑到 Tranformer 解码器与编码器的构成区别，Bloomberg 沿用了 GPT 模型的 decoder-only 技术路径，训练出的 LLM 在基于前文预测后文的自然语言生成方面具备天然优势（详细分析可参考华泰计算机 2023 年 2 月 14 日发布的报告《GPT 产业复盘（2019）：大容量路线》）。对比来看，BloombergGPT 的模型参数介于 GPT-2 与 GPT-3 之间，GPT-2 模型参数为 1.5 亿，GPT-3 模型参数为 1,750 亿，bloombergGPT 的模型参数为 500 亿。

图表171：Tansformer 的 encoder 编码器结构示意



资料来源：《The Illustrated GPT-2 (Visualizing Transformer Language Models)》，Jay Alammar（2020）、华泰研究

图表172：Tansformer 的 decoder 解码器结构示意

资料来源：《The Illustrated GPT-2 (Visualizing Transformer Language Models)》，Jay Alammar（2020）、华泰研究

大规模金融数据集 FINPILE 是 BloombergGPT 在金融垂直领域成功落地的关键。不同于 OpenAI 的 GPT 模型基于通用文本语料进行模型预训练，Bloomberg 为强化大语言模型对金融垂直领域的专业理解，基于 Bloomberg 自身在金融语料方面的优势，构建了目前最大规模的金融数据集 FINPILE，实现了模型在金融领域的知识强化。总体来看，BloombergGPT 预训练的数据集包括金融数据文本 FINPILE 数据集与通用数据文本 PUBLIC 数据集两部分。

免责
更多免费行业报告 并购家 www.ipoipo.cn

FINPILE + PUBLIC 让模型兼具通用常识与金融知识的双重理解。得益于 Bloomberg 构建的融合通用文本（PUBLIC）与金融知识（FINPILE）的大规模训练数据，BloombergGPT 基于通用场景和金融业务场景进行混合模型训练，使模型在具备通常常识的基础上强化金融领域的知识理解，根据官方 2023 年 3 月发布的论文《BloombergGPT: A Large Language Model for Finance》中的测试结果显示，BloombergGPT 在执行金融任务上的表现超过现有的通用 LLM 模型，在通用场景上的表现与现有通用 LLM 模型能力基本持平。

BloombergGPT 实现在垂直领域模型、训练数据、模型评估、tokenizer、模型构建方法等方面的创新。

1）垂直领域语言模型：过去的大语言模型多为基于通用文本训练的通用模型，垂直领域模型多为仅基于垂直领域数据训练垂直模型，Bloomberg 开创通用+垂直的混合训练方法，让模型兼具通用性与专业性；

2）训练数据：过去的大语言模型的预训练数据很大程度上依赖于网页抓取数据，如 C4、The Pile、Wikipedia 等，Bloomberg 自建了高质量的大规模金融数据集；

3）模型评估：Bloomberg 在对模型进行了公共、金融 NLP 基准测试之外，还对模型进行了一系列基于 Bloomberg 内部任务的性能测试；

4）Token 化（Tokenizer）：将训练文本 Token 化是模型训练的关键步骤，Bloomberg 使用 Unigram 模型取代 greedy merge-based sub-word 模型，实现更智能的 token 化转换；

5）模型构建方法：以 GPT-3、GPT-4 为代表的大语言模型均由大型的专业人工智能团队开发，并且模型训练需要大量算力；受益于开源模型 BLOOM 的项目实践与 Bloomberg 在垂直领域高质量数据的深厚积累，BloombergGPT 成功证明了一个中等规模的团队可以在垂直领域的特定数据上生产同样具有竞争力的大语言模型。

图表173：BloombergGPT 实现垂直领域模型、训练数据、模型评估、tokenizer、模型构建方法等方面的创新

资料来源：《BloombergGPT: A Large Language Model for Finance》，Shijie Wu（2023）、华泰研究

BloombergGPT 开拓了开源模型+垂直数据的 LLM 搭建新思路。通过对比 GPT-3 与 BloombergGPT，我们发现，高质量的预训练数据是提升大语言模型能力的关键，大量的高质量垂直领域知识有望弥补模型在规模上的不足。BloombergGPT 的模型参数为 500 亿，相较于 GPT-3 的 1,750 亿属于较小的模型的规模，不同的是，BloombergGPT 的预训练数据增加了大量的高质量金融数据，同时 BloombergGPT 对于模型的预训练数据进行了一系列的数据清洗和数据标注，进一步提升了数据质量。从测试结果来看，BloombergGPT 在通用能力与 GPT-3 基本持平的情况下，实现了金融垂直能力的大幅增强。

免责
更多免费行业报告 并购家 www.ipoipo.cn

训练数据是 LLM 能力塑造的关键因素。基于 BloombergGPT 的成功案例，我们认为，训练数据是大语言模型能力塑造的关键因素，原因有三，**一是**金融垂直领域数据的输入成功塑造了 BloombergGPT 的金融知识理解能力，具备了更强的专业性；**二是**尽管 BloombergGPT 的模型参数相对而言有所减少，但仍然表现出较强的通用能力与垂直能力，证明了在一定参数规模的前提下，高质量数据才是决定模型能力的关键；**三是** Bloomberg 在论文中明确出于对数据泄露的担忧，BloombergGPT 将采取与 OpenAI 相同的闭源方案，侧面佐证了原始的预训练数据是各家大模型竞争的重要因素。

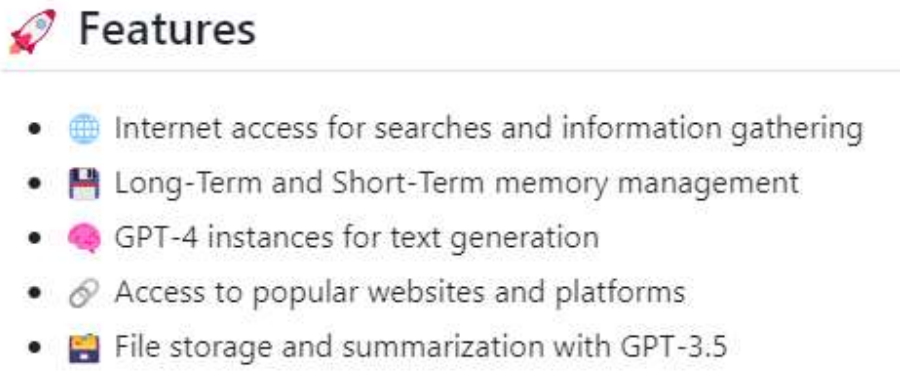
GPT+垂直场景有望全面铺开。垂直知识+通用常识的综合训练方法的有效性已经得到有效验证，大语言模型有望实现与医疗、安全、能源、工业等更多垂直场景的深度融合。以海外的 ProGEN 的模型为例，大语言模型已经实现了在蛋白质结构预测的场景落地。未来，融合垂直领域知识的大语言模型有望加速发展。

#2 AutoGPT：充分展现大模型有望重塑人机交互范式

AutoGPT 充分展现大模型有望重塑人机交互范式。AutoGPT 是一个实验性开源应用程序，展示了 GPT-4 语言模型的功能，标志着自然语言有望成为未来人机交互的全新模式。该程序由 GPT-4 驱动，将 LLM 的“思想”链接在一起，以自主实现用户设定的任何目标。作为 GPT-4 完全自主运行的首批示例之一，AutoGPT 突破了 AI 的可能性界限。

AutoGPT 具备五大典型特征：1）可以从 Internet 搜索和收集信息；2）支持长期和短期内存管理；3）提供文本生成的 GPT-4 实例；4）访问热门网站和平台；5）使用 GPT-3.5 进行文件存储和汇总。

图表174：AutoGPT 五大特征

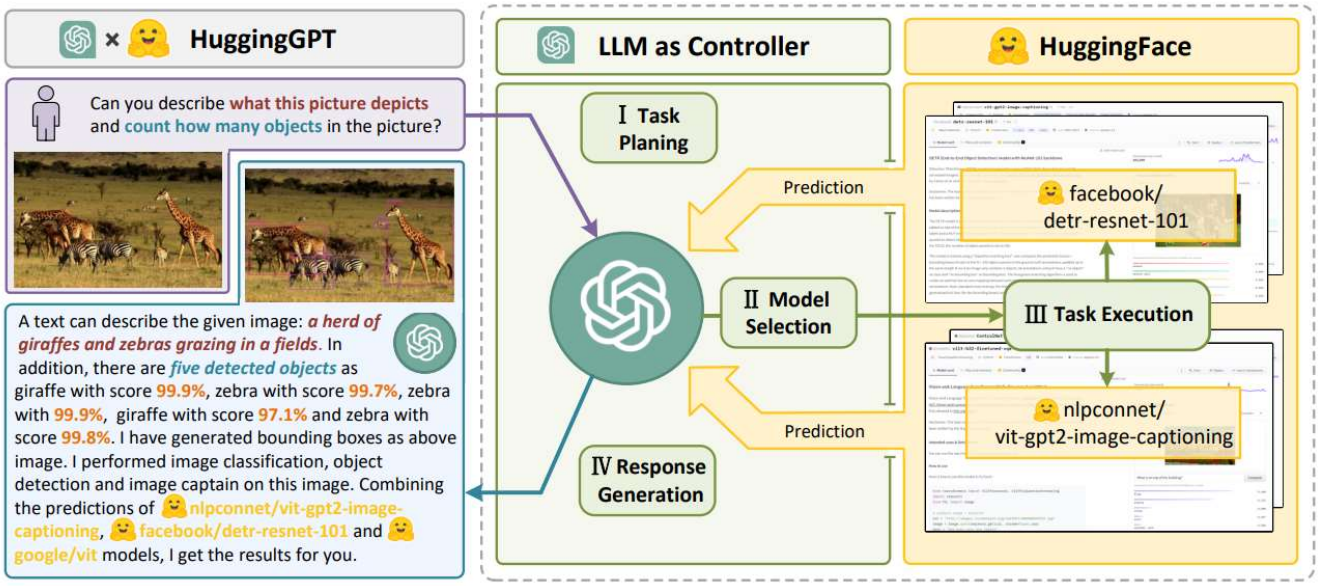


资料来源：Github 官网、华泰研究

AutoGPT 需要人类指导，充分展示了人机交互方式改变的可能性。AutoGPT 并非真的是全自动、全自主，而是需要人类指导。我们认为，AutoGPT 是在 GPT-4 的基础上，衍生出来的创新型大模型使用方法探索。其本质可以认为是 GPT 模型的自动化任务执行工具，能够通过预先设计好的提示来自动生成任务指令，从而快速高效地执行各种任务。基于 GPT-4 模型的自然语言理解能力，AutoGPT 通过 API 接口、现有大模型的创新性结合等方式，实现问题的“自动化解决”。大语言模型 LLM 成为大模型时代的操作系统，可以作为控制器来管理现有的 API 接口、人工智能模型，以解决复杂的人工智能任务。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表175：微软 HuggingGPT 与 AutoGPT 相似，探索 LLM 作为 AI 操作系统的可能性

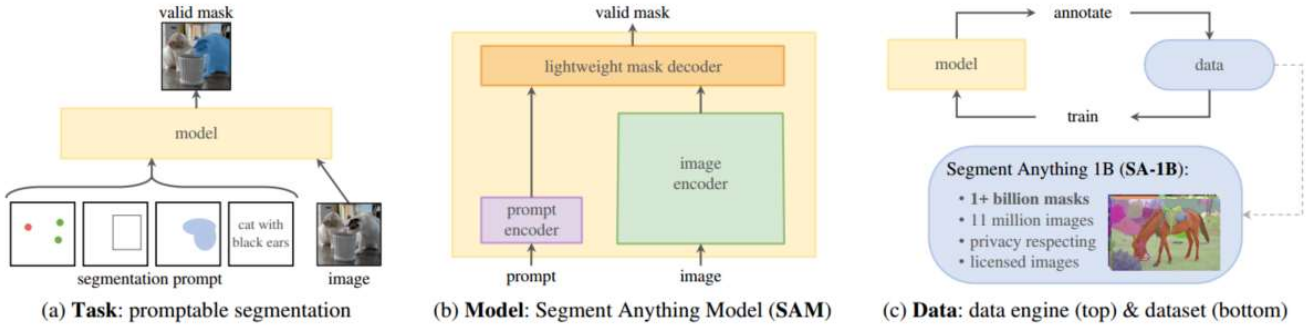


资料来源：HuggingFace 官网、华泰研究

#3 SAM 模型：实现大模型在 CV 领域的模态延展

Segment Anything Model 是 Meta 提出的图像分割模型。Segment Anything Model（SAM）由三部分组成，**1）任务：**为可提示（promptable）的分割任务，描述通用的预训练目标，实现广泛的下游任务；**2）分割模型（SAM）：**支持灵活的提示（prompt）和数据注释，并能通过提示工程（prompt engineering）实时输出分割掩膜，以 zero-shot 方式迁移到一系列任务；**3）数据引擎：**构建了数据引擎，一方面收集新的数据，另一方面使用新数据来迭代模型。Meta 建立了迄今为止最大的分割数据集 SA-1B，在 1100 万张许可和尊重隐私的图像上，有超过 10 亿个掩膜（mask）。结果表明，SAM 的 zero-shot 性能优秀，能够达到甚至优于完全监督模型的结果。SAM 和 SA-1B 已经在 https://segment-anything.com 开源。

图表176：SAM 的三个组成部分

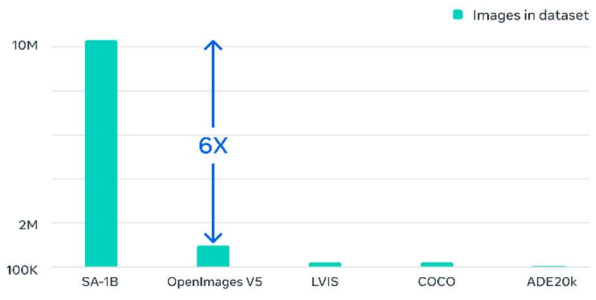


资料来源：《Segment Anything》，Meta（2023）、华泰研究

大规模且多样化的掩膜数据集能够实现 SAM 更好的泛化特性。由数据引擎生成的最终数据集 SA-1B，包含 1100 万张经许可且隐私保护的图像，以及超过 10 亿个掩膜（均为数据引擎最后阶段自动生成的掩膜）。SA-1B 拥有的掩膜数量比任何现有的分割数据集都多 400 倍以上，具有高质量和多样性。除了将其用于训练 SAM 之外，SA-1B 已开源，供其他研究者在此基础上建立新的基础大模型。

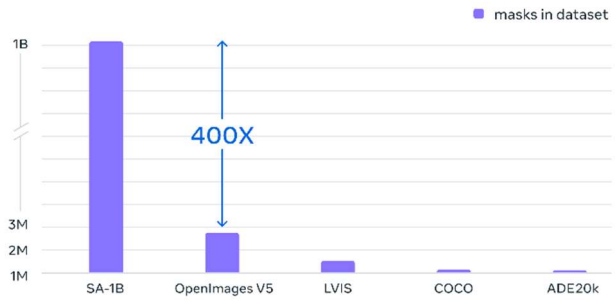
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表177：SA-1B 包含的图像数量比其他数据集高 6 倍以上



资料来源：Meta 官网、华泰研究

图表178：SA-1B 包含的掩膜数量比其他数据集高 400 倍以上



资料来源：Meta 官网、华泰研究

SAM 模型亮点：探索模型组件化应用

以提示工程等技术为基础的可组合系统，能实现更广泛的应用。 Meta 指出，提示和组合是功能强大的工具，使单个模型能够以可扩展的方式使用，有可能完成模型设计时未知的任务。与专门为固定任务集训练的系统相比，以提示工程等技术为基础的可组合系统，能实现更广泛的应用。例如 CLIP 是 DALL-E 图像生成系统的文本图像对齐组件。

SAM 是为通用性和使用广度而设计的，是视觉分割的通才和多面手。Meta 指出，虽然特定领域的工具在各自的领域中会优于 SAM，但 SAM 是为通用性和使用广度而设计的，是视觉分割的通才和多面手。此外，由于 SAM 的强大迁移和泛化能力，以及开箱即用特性，SAM 甚至可以通过可穿戴设备检测到的注视点来进行提示，从而启用新的应用程序。

免责
更多免费行业报告 并购家 www.ipoipo.cn

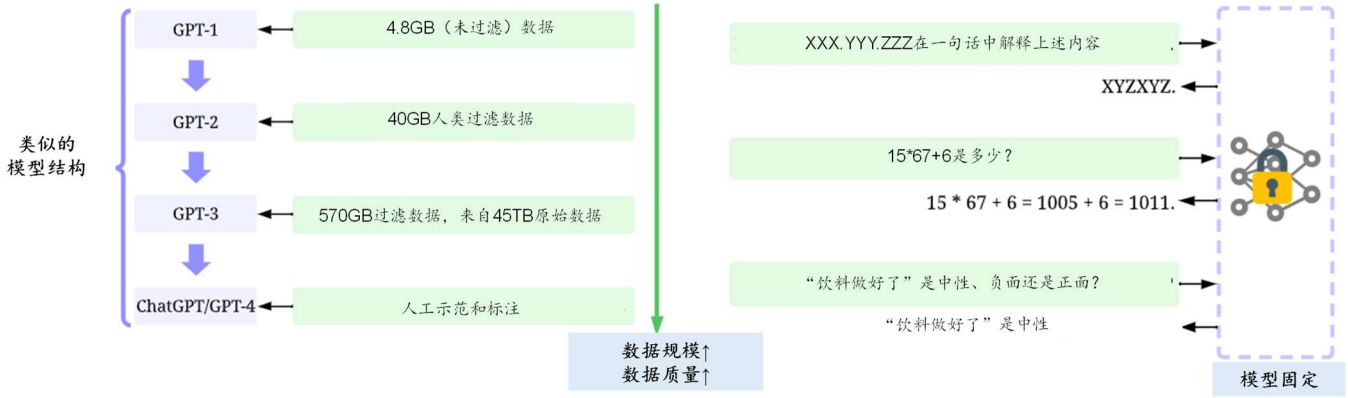
附录：AI 大模型需要什么数据？

AI 的突破得益于高质量数据，我们认为数据是大模型竞争关键要素之一：1）训练大模型需要高质量、大规模、多样性的数据集；2）优质中文数据集稀缺，数字中国战略将促进数据要素市场完善，助力数据集发展。近期欧洲议会议员《人工智能法案》提案、网信办《生成式人工智能服务管理办法（征求意见稿）》对大模型训练数据的版权披露、合法性提出要求，对于数据产业链的投资机会，我们认为：1）数据资产储备公司的商业化进程值得关注；2）行业数据价值高，具有优质数据和一定大模型能力的公司或通过行业大模型赋能业务；3）关注卡位优质客户、技术降低人力成本的数据服务企业。

数据将是未来 AI 大模型竞争的关键要素

人工智能发展的突破得益于高质量数据的发展。例如，大型语言模型的最新进展依赖于更高质量、更丰富的训练数据集：与 GPT-2 相比，GPT-3 对模型架构只进行了微小的修改，但花费精力收集更大的高质量数据集进行训练。ChatGPT 与 GPT-3 的模型架构类似，并使用 RLHF（来自人工反馈过程的强化学习）来生成用于微调的高质量标记数据。

图表179：更高质量、更丰富的训练数据是 GPT 模型成功的驱动力；而除模型权重变化之外，模型架构保持相似



资料来源：Daochen Zha et al. "Data-centric Artificial Intelligence: A Survey" 2023，华泰研究

我们认为 AI 大模型需要高质量、大规模、多样性的数据集。

1）高质量：高质量数据集能够提高模型精度与可解释性，并且减少收敛到最优解的时间，即减少训练时长。

2）大规模：OpenAI 在《Scaling Laws for Neural Language Models》中提出 LLM 模型所遵循的“伸缩法则”（scaling law），即独立增加训练数据量、模型参数规模或者延长模型训练时间，预训练模型的效果会越来越好。

3）丰富性：数据丰富性能够提高模型泛化能力，过于单一的数据会非常容易让模型过于拟合训练数据。

海外主要大模型数据集

我们梳理了海外主要的开源语言和多模态数据集，主要的发布方包括高校、互联网巨头研究部门、非盈利研究组织以及政府机构。我们认为海外积累丰富的开源高质量数据集得益于：1）相对较好的开源互联网生态；2）免费线上书籍、期刊的长期资源积累；3）学术界、互联网巨头研究部门、非盈利研究组织及其背后的赞助基金形成了开放数据集、发表论文-被引用的开源氛围。然而，高质量语言数据或于 2026 年耗尽，AI 合成数据有望缓解数据耗尽的隐忧，Gartner 预测 2030 年大模型使用的绝大部分数据或由 AI 合成。

免责
更多免费行业报告 并购家 www.ipoipo.cn

#1：海外主要大语言模型数据集

参数量和数据量是判断大模型的重要参数。2018 年以来，大语言模型训练使用的数据集规模持续增长。2018 年的 GPT-1 数据集约 4.6GB，2020 年的 GPT-3 数据集达到了 753GB，而到了 2021 年的 Gopher，数据集规模已经达到了 10,550GB。总结来说，从 GPT-1 到 LLaMA 的大语言模型数据集主要包含六类：维基百科、书籍、期刊、Reddit 链接、Common Crawl 和其他数据集。

图表180：大语言模型数据集综合分析

大模型	维基百科	书籍	期刊	Reddit链接	Common Crawl	其他	合计
GPT-1		4.6					4.6
GPT-2				**40**			40
GPT-3	11.4	21	101	50	**570**		753
The Pile v1	**6**	**118**	**244**	**63**	**227**	**167**	825
Megatron-11B	**11.4**	4.6		**38**	**107**		161
MT-NLG	6.4	118	77	63	983	127	1374
Gopher	**12.5**	**2100**	164.4		3450	4823	10550
LLaMA	**83**	**85**	**92**		**4162.2**	**406**	**4828.2**

注：以 GB 为单位，公开的数据以粗体表示，仅原始训练数据集大小
资料来源：Alan D. Thompson “What’s in My AI” 2023，Hugo Touvron et al. “LLaMA: Open and Efficient Foundation Language Models” 2023，华泰研究

数据集#1：维基百科

维基百科是一个免费的多语言协作在线百科全书。维基百科致力于打造包含全世界所有语言的自由的百科全书，由超三十万名志愿者组成的社区编写和维护。截至 2023 年 3 月，维基百科拥有 332 种语言版本，总计 60,814,920 条目。其中，英文版维基百科中有超过 664 万篇文章，拥有超 4,533 万个用户。维基百科中的文本很有价值，因为它被严格引用，以说明性文字形式写成，并且跨越多种语言和领域。一般来说，重点研究实验室会首先选取它的纯英文过滤版作为数据集。

数据集#2：书籍

书籍主要用于训练模型的故事讲述能力和反应能力，包括小说和非小说两大类。数据集包括 Project Gutenberg 和 Smashwords (Toronto BookCorpus/BookCorpus)等。Project Gutenberg 是一个拥有 7 万多本免费电子书的图书馆，包括世界上最伟大的文学作品，尤其是美国版权已经过期的老作品。BookCorpus 以作家未出版的免费书籍为基础，这些书籍来自于世界上最大的独立电子书分销商之一的 Smashwords。

数据集#3：期刊

期刊可以从 ArXiv 和美国国家卫生研究院等官网获取。预印本和已发表期刊中的论文为数据集提供了坚实而严谨的基础，因为学术写作通常来说更有条理、理性和细致。ArXiv 是一个免费的分发服务和开放获取的档案，包含物理、数学、计算机科学、定量生物学、定量金融学、统计学、电气工程和系统科学以及经济学等领域的 2,235,447 篇学术文章。美国国家卫生研究院是美国政府负责生物医学和公共卫生研究的主要机构，支持各种生物医学和行为研究领域的研究，从其官网的“研究&培训”板块能够获取最新的医学研究论文。

数据集#4：WebText（来自 Reddit 链接）

Reddit 链接代表流行内容的风向标。Reddit 是一个娱乐、社交及新闻网站，注册用户可以将文字或链接在网站上发布，使它成为了一个电子布告栏系统。WebText 是一个大型数据集，它的数据是从社交媒体平台 Reddit 所有出站链接网络中爬取的，每个链接至少有三个赞，代表了流行内容的风向标，对输出优质链接和后续文本数据具有指导作用。

Reddit 宣布收取数据使用费。2023 年 4 月，Reddit 宣布将向使用其 API 训练 AI 聊天机器人的公司收取数据使用费，其中便包含微软、谷歌、OpenAI 等，目前具体收费标准暂未公布，但可能会根据不同使用者划分不同等级收费标准。许多公司已经意识到数据的价值，如图片托管服务商 Shutterstock 已把图像数据出售给 OpenAI，推特计划针对 API 使用收取几万到几十万美元不等的费用。

免责
更多免费行业报告 并购家 www.ipoipo.cn

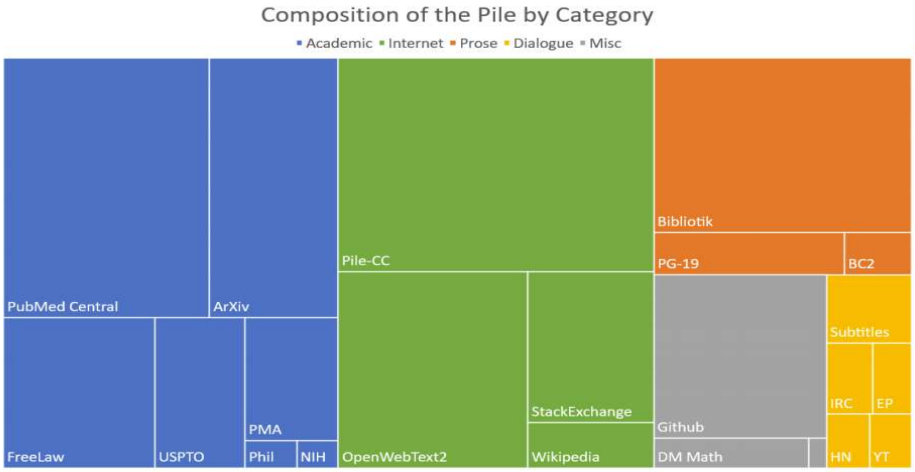
数据集#5：Common crawl/C4

Common crawl 是 2008 年至今的一个网站抓取的大型数据集。Common Crawl 是一家非盈利组织，致力于为互联网研究人员、公司和个人免费提供互联网副本，用于研究和分析，它的数据包含原始网页、元数据和文本提取，文本包含 40 多种语言和不同领域。重点研究实验室一般会首先选取它的纯英文过滤版（C4）作为数据集。

其他数据集

The Pile 数据集：一个 825.18 GB 的英语文本数据集，用于训练大规模语言模型。The Pile 由上文提到的 ArXiv、WebText、Wikipedia 等在内的 22 个不同的高质量数据集组成，包括已经建立的自然语言处理数据集和几个新引入的数据集。除了训练大型语言模型外，The Pile 还可以作为语言模型跨领域知识和泛化能力的广泛覆盖基准。

图表181：按有效尺寸划分的 The Pile 组成树状图



资料来源：Leo Gao et al. “The Pile: An 800GB Dataset of Diverse Text for Language Modeling” 2020，华泰研究

其他数据集包含 GitHub 等代码数据集、StackExchange 等对话论坛和视频字幕数据集等。

图表182：其他常见 NLP 数据集

数据集分类	数据集	简介
代码数据集	Github	一个大型的开源代码库，在多年以前的预训练语言模型例如 BERT、GPT 里几乎没有人用，该代码数据的加入对语言模型的逻辑推理能力有极大的帮助
	CodeSearchNet	一个大型函数数据集，其中包含来自 GitHub 上的开源项目的用 Go、Java、JavaScript、PHP、Python 和 Ruby 编写的相关文档
	StaQC	是迄今为止最大的数据集，大约有 148K Python 和 120K SQL 域问题代码对，它们是使用 Bi-View Hierarchical Neural Network 从 Stack Overflow 中自动挖掘出来的
	CodeExp	其中包含 (1) 2.3 的大分区 百万原始代码-docstring 对，(2) 一个介质 158,000 对的分区从 使用学习的过滤器的原始语料库，以及 (3) 具有严格的人类 13,000 对的分区注释
	ETH Py150 Open	来自 GitHub 的 740 万个 Python 文件的大规模去重语料库
论坛数据集	StackExchange	StackOverflow 的超集，包含有不限于计算机的各种各样不同领域的高质量问答
	Federated Stack Overflow	数据由所有问题和答案的正文组成。Body 被解析成句子，任何少于 100 个句子的用户都会从数据中删除。最少的预处理如下进行：小写文本，对 HTML 符号进行转义，删除非 ASCII 符号，单独的标点符号作为单独的标记（撇号和连字符除外），去除多余的空白，用特殊标记替换 URLS
	QUASAR	一个由 QUASAR-S 和 QUASAR-T 组成的大规模数据集。这些数据集中的每一个都旨在专注于评估旨在理解自然语言查询、大量文本语料库并从语料库中提取问题答案的系统。具体来说，QUASAR-S 包含 37,012 个填空题，这些问题是使用实体标签从流行的网站 Stack Overflow 收集的
	GIF Reply Dataset	发布的 GIF 回复数据集包含 1,562,701 次 Twitter 上的真实文本 - GIF 对话。在这些对话中，使用了 115,586 个独特的 GIF。元数据包括 OCR 提取的文本、带注释的标签和对象名称，也可用于该数据集中的一些 GIF
视频字幕数据集	TVC (TV show Captions)	电视节目 Caption 是一个大规模的多模态字幕数据集，包含 261,490 个字幕描述和 108,965 个短视频片段。TVC 是独一无二的，因为它的字幕也可以描述对话/字幕，而其他数据集中的字幕仅描述视觉内容

资料来源：Hugo Touvron et al. “LLaMA: Open and Efficient Foundation Language Models” 2023，OpenDataLab，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

#2：海外主要多模态数据集

模态是事物的一种表现形式，多模态通常包含两个或者两个以上的模态形式，包括文本、图像、视频、音频等。多模态大模型需要更深层次的网络和更大的数据集进行预训练。过去数年中，多模态大模型参数量及数据量持续提升。例如，2022 年 Stability AI 发布的 Stable Diffusion 数据集包含 58.4 亿图文对/图像，是 2021 年 OpenAI 发布的 DALL-E 数据集的 23 倍。

图表183：多模态大模型数据集介绍

公司	多模态大模型	发布时间	最大参数量（B）	数据集（M_图文对/图像）	数据集类别
OpenAI	DALL-E	2021.1	12	250	Conceptual Captions、YFCC100M、Wikipedia
Meta	Make-a-scene	2022.3	4	35	-
谷歌、Hugging Face	DALL-E mini	2022.4	0.4	15	-
OpenAI	DALL-E 2	2022.4	6.5	650	AVA
谷歌	Imagen	2022.5	7.6	860	内部数据、LAION-400M
谷歌	Parti	2022.6	20	4800	MS-COCO、LAION-400M、FIT400M、JFT-4B
Stability AI	Stable Diffusion	2022.8	na	5840	LAION-5B
谷歌	PaLM-E	2023.3	562	na	Language-Table

资料来源：Aditya Ramesh et al. “Zero-Shot Text-to-Image Generation” 2021，Oran Gafni et al. “Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors” 2022，Aditya Ramesh “Hierarchical Text-Conditional Image Generation with CLIP Latents” 2022，Chitwan Saharia et al. “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding” 2022，Jiahui Yu et al. “Scaling Autoregressive Models for Content-Rich Text-to-Image Generation” 2022，Jay Alammar "The Illustrated Stable Diffusion” 2022，Danny Driess et al. “PaLM-E: An Embodied Multimodal Language Model” 2023，华泰研究

类别#1：语音+文本

SEMAINE 数据集：创建了一个大型视听数据库，作为构建敏感人工侦听器(SAL)代理的迭代方法的一部分，该代理可以使人参与持续的、情绪化的对话。高质量的录音由五台高分辨率、高帧率摄像机和四个同步录制的麦克风提供。录音共有 150 个参与者，总共有 959 个与单个 SAL 角色的对话，每个对话大约持续 5 分钟。固体 SAL 录音被转录和广泛注释：每个剪辑 6-8 个评分者追踪 5 个情感维度和 27 个相关类别。

类别#2：图像+文本

COCO 数据集：MS COCO 的全称是 Microsoft Common Objects in Context，起源于微软于 2014 年出资标注的 Microsoft COCO 数据集，与 ImageNet 竞赛一样，被视为是计算机视觉领域最受关注和最权威的比赛之一。 COCO 数据集是一个大型的、丰富的物体检测，分割和字幕数据集。图像包括 91 类目标，328,000 张图像和 2,500,000 个 label。

Conceptual Captions 数据集：图像标题注释数据集，其中包含的图像比 MS-COCO 数据集多一个数量级，并代表了更广泛的图像和图像标题风格。通过从数十亿个网页中提取和过滤图像标题注释来实现这一点。

ImageNet 数据集：建立在 WordNet 结构主干之上的大规模图像本体。ImageNet 的目标是用平均 5,001,000 张干净的全分辨率图像填充 WordNet 的 80,000 个同义词集中的大多数。这将产生数千万个由 WordNet 语义层次结构组织的注释图像。ImageNet 的当前状态有 12 个子树，5247 个同义词集，总共 320 万张图像。

LAION-400M 数据集：LAION-400M 通过 CommonCrawl 提取出随机抓取 2014-2021 年的网页中的图片、文本内容。通过 OpenAI 的 Clip 计算，去除了原始数据集中文本和图片嵌入之间预先相似度低于 0.3 的内容和文本，提供了 4 亿个初筛后的图像文本对样本。

LAION-5B 数据集：其包含 58.5 亿个 CLIP 过滤的图像-文本对的数据集，比 LAION-400M 大 14 倍，是世界第一大规模、多模态的文本图像数据集，共 80T 数据，并提供了色情图片过滤、水印图片过滤、高分辨率图片、美学图片等子集和模型，供不同方向研究。

Language Table 数据集：Language-Table 是一套人类收集的数据集，是开放词汇视觉运动学习的多任务连续控制基准。

免责
更多免费行业报告 并购家 www.ipoipo.cn

IAPR TC-12 数据集：IAPR TC-12 基准的图像集合包括从世界各地拍摄的 2 万张静态自然图像，包括各种静态自然图像的横截面。这包括不同运动和动作的照片，人物、动物、城市、风景和当代生活的许多其他方面的照片。示例图像可以在第 2 节中找到。每张图片都配有最多三种不同语言（英语、德语和西班牙语）的文本标题。

AVA 数据集：AVA 是美学质量评估的数据库，包括 25 万张照片。每一张照片都有一系列的评分、语义级别的 60 类标签和 14 类照片风格。

OpenViDial 数据集：当人们交谈时，说话者接下来要说什么在很大程度上取决于他看到了什么。OpenViDal 一个用于此目的的大型多模块对话数据集。这些对话回合和视觉环境都是从电影和电视剧中提取出来的，其中每个对话回合都与发生的相应视觉环境相匹配。版本 1 包含 110 万个对话回合以及存储在图像中的 110 万个视觉上下文。版本 2 要大得多，包含 560 万个对话回合以及存储在图像中的 560 万个视觉上下文。

类别#3：视频+图像+文本

YFCC100 数据集：YFCC100M 是一个包含 1 亿媒体对象的数据集，其中大约 9920 万是照片，80 万是视频，所有这些都带有创作共用许可。数据集中的每个媒体对象都由几块元数据表示，例如 Flickr 标识符、所有者名称、相机、标题、标签、地理位置、媒体源。从 2004 年 Flickr 成立到 2014 年初，这些照片和视频是如何被拍摄、描述和分享的，这个集合提供了一个全面的快照。

类别#4：图像+语音+文本

CH-SIMS 数据集：CH-SIMS 是中文单模态和多模态情感分析数据集，包含 2,281 个精细化的野外视频片段，既有多模态注释，也有独立单模态注释。它允许研究人员研究模态之间的相互作用，或使用独立的单模态注释进行单模态情感分析。

类别#5：视频+语音+文本

IEMOCAP 数据集：南加州大学语音分析与解释实验室(SAIL)收集的一种新语料库，名为“交互式情感二元动作捕捉数据库”(IEMOCAP)。该数据库记录了 10 位演员在面部、头部和手上的二元会话，这些标记提供了他们在脚本和自发口语交流场景中面部表情和手部动作的详细信息。语料库包含大约 12 小时的数据。详细的动作捕捉信息、激发真实情绪的交互设置以及数据库的大小使这个语料库成为社区中现有数据库的有价值的补充，用于研究和建模多模态和富有表现力的人类交流。

MELD 数据集：MELD 收录了《老友记》电视剧 1,433 个对话中的 13,708 个话语。MELD 优于其他对话式情绪识别数据集 SEMAINE 和 IEMOCAP，因为它由多方对话组成，并且 MELD 中的话语数量几乎是这两个数据集的两倍。MELD 中的话语是多模态的，包括音频和视觉形式以及文本。

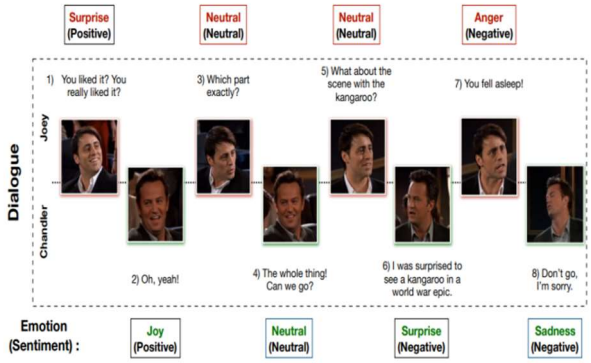
免责

更多免费行业报告 并购家 www.ipoipo.cn

图表184： IEMOCAP——有 8 个摄像头的 VICON 运动捕捉系统

资料来源：Carlos Busso et al. “IEMOCAP: interactive emotional dyadic motion capture database. Lang Resources & Evaluation” 2008，华泰研究

图表185： MELD 数据集——对话中和对话前说话人情绪变化对比



资料来源：Soujanya Poria et al. “MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations” 2018，华泰研究

#3：海外主要大模型数据集由何方发布

海外主要开源大模型数据集发布方主要分为：

1) 非营利组织/开源组织：古腾堡文学档案基金会发布的 Project Gutenberg 截至 2018 年已收录 57,000 部书籍，平均每周新增 50 部。Common Crawl 抓取网络并免费向公众提供其档案和数据集，一般每个月完成一次抓取。艾伦人工智能研究所分别于 2017 年、2018 年和 2019 年发布了基于维基百科的 TriviaQA、QuAC、Quoref。Eleuther AI 发布了 825GB 多样化文本数据集 The Pile。LAION 2021 年发布包含 4 亿图文对的 LAION-400M 数据集，2022 年发布包含 58.5 亿图文对的 LAION-5B 数据集；
2) 学术界：例如多伦多大学和麻省理工学院联合发布了 BookCorpus；
3) 互联网巨头研究部门：例如 Google Research 发布了 C4 文本数据集、AVA 和 Conceptual Captions 等等图像数据集等；
4) 政府机构：政府机构是一些常见的数据集发布方，通常包含关于经济和医学等方面的数据，美国国家卫生研究院发布的 MedQuAD 包括从 12 个 NIH 网站创建的 47,457 个医学问答对；
5) 多种类型机构合作：尤其是学术界与互联网巨头研究部门、开源组织之间的合作。例如 Facebook、伦敦大学学院和 DeepMind 联合发布了 ArxivPaper 数据集。卡内基梅隆大学、雅虎研究院和 International Computer Science Institute 联合发布了 YFCC100M。

我们认为海外积累丰富的开源高质量数据集得益于：1）相对较好的开源互联网生态；2）免费线上书籍、期刊的资源积累；3）学术界、互联网巨头研究部门、非盈利研究组织及其背后的基金形成了开放数据集、发表论文-被引用的开源氛围。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表186：常见大模型数据集发布方总结

类别	类别	名称	数据来源	发布方
大语言模型数据集	维基百科	Identifying Machine-Paraphrased Plagiarism	维基媒体基金会	德国伍珀塔尔大学、布尔诺孟德尔大学
		Benchmark for Neural Paraphrase Detection	维基媒体基金会	德国伍珀塔尔大学
		Quoref	维基媒体基金会	艾伦人工智能研究所、华盛顿大学
		QuAC (Question Answering in Context)	-	艾伦人工智能研究所、华盛顿大学、斯坦福大学、马萨诸塞大学阿默斯特分校
		TriviaQA	维基媒体基金会	华盛顿大学、艾伦人工智能研究所
		WikiQA	维基媒体基金会	微软研究院
	书籍	BookCorpus	Smashwords	多伦多大学、麻省理工学院
		Project Gutenberg	古腾堡文学档案基金会	古腾堡文学档案基金会
	期刊	ArxivPapers	arXiv	Facebook、伦敦大学学院、DeepMind
		MedQuAD	美国国家卫生研究院	美国国家卫生研究院
		Pubmed	PubMed	马里兰大学
		PubMed Paper Reading Dataset	PubMed	伊利诺伊大学厄巴纳香槟分校、滴滴实验室、伦斯勒理工学院、北卡罗来纳大学教堂山分校、华盛顿大学
		PubMed RCT (PubMed 200k RCT)	PubMed	Adobe Research、麻省理工学院
		MedHop	PubMed	伦敦大学学院、Bloomsbury AI
		unarXive	arXiv	Karlsruhe Institute of Technology
		arXiv Summarization Dataset	arXiv	Georgetown University、Adobe Research
		SCICAP	arXiv	宾夕法尼亚州立大学
	Reddit 链接	OpenWebText	Reddit	华盛顿大学、Facebook AI Research
	Commom Crawl	C4 (Colossal Clean Crawled Corpus)	Common Crawl	Google Research
		Common Crawl	Common Crawl	法国国家信息与自动化研究所、索邦大学
	综合	The Pile	-	EleutherAI
多模态数据集		Conceptual Captions	网络	Google Research
		YFCC100M	Flickr	卡内基梅隆大学、雅虎研究院、International Computer Science Institute
		AVA	-	Google Research
		LAION-400M	Common Crawl	慕尼黑工业大学、EleutherAI、LAION
		COCO	微软	微软
		LAION-5B	Common Crawl	LAION
		Language-Table	-	-

资料来源：OpenDataLab，CSDN，华泰研究

中国主要 AI 大模型数据集

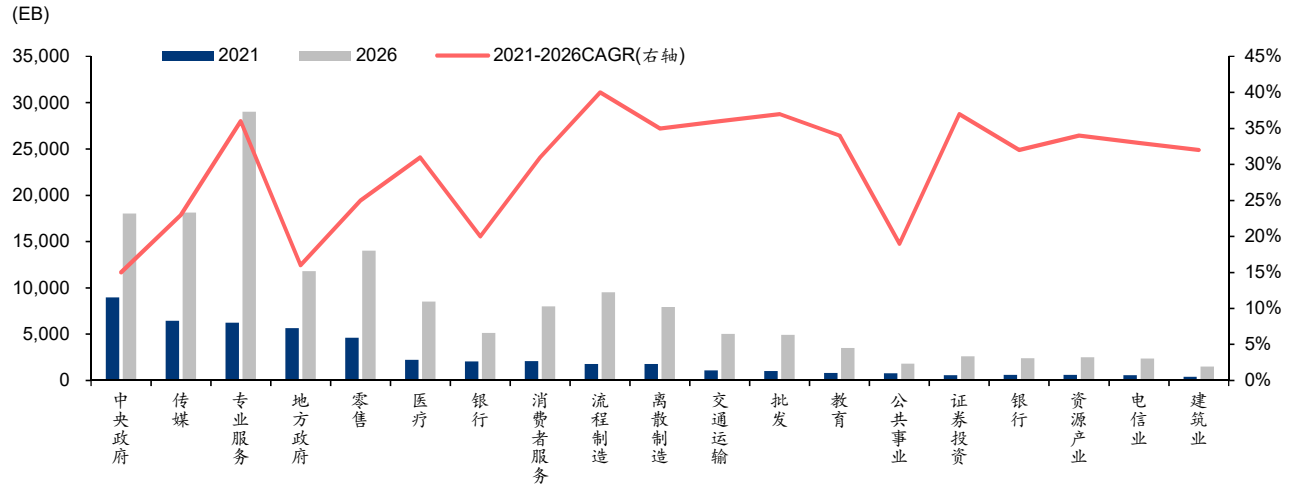
与国外类似，国内大模型的训练数据包括互联网爬取数据、书籍期刊、公司自有数据以及开源数据集等。就开源数据集而言，国内外的发布方都涵盖高校、互联网巨头、非盈利机构等组织。但国内开源数据集数量少、规模小，因此国内大模型训练往往使用多个海外开源数据集。国内缺乏高质量数据集的原因在于：1）高质量数据集需要高资金投入；2）相关公司开源意识较低；3）学术领域中文数据集受重视程度低。看好数字中国战略助力国内数据集发展：1）各地数据交易所设立运营提升数据资源流通；2）数据服务商链接数据要素产业链上下游，激活数据交易流通市场，提供更多样化的数据产品。

中国 AI 大模型数据集从哪里来

国内各行业数据资源丰富，2021-2026 年数据量规模 CAGR 高于全球，数据主要来源于政府/传媒/服务/零售等行业。据 IDC，2021-2026 年中国数据量规模将由 18.51ZB 增长至 56.16ZB，CAGR 达到 24.9%，高于全球平均 CAGR。从数据来源看，国内各行业数据差异化发展，2021 年政府、传媒、专业服务、零售等行业数据量占比较高，分别为 30.4%、13.4%、13.0%、9.6%，其中接近 90%的数据为非结构化数据，这也要求了海量数据采集设备和软件的互联互通以及互动互控。另外随着智能化转型的深入，制造、交通运输、批发、教育等行业数据规模在未来也拥有较大的增长潜力，2021-2026 年数据量增长 CAGR 将分别达到 37.6%、36.1%、37.1%、34.0%。

更多免费行业报告 并购家 www.ipoipo.cn

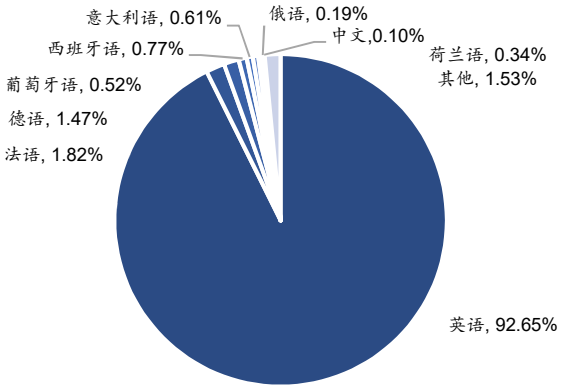
图表187：国内各行业数据量分布及增长预测



资料来源：IDC，华泰研究

尽管国内数据资源丰富，但由于数据挖掘不足，数据无法自由在市场上流通等现状，优质中文优质数据集仍然稀缺。目前中文优质数据仍然稀缺，以 GPT-3 为例，其训练集中中文数据仅占 0.1%，这一方面因为 GPT 为美国开发，以英文语料为主；另一方面也反映了目前可获得数据集中，中文内容仍然相对稀缺。国内大模型训练基于中文原生内容仍是第一选择，翻译数据训练效果劣于中文原生内容。据加利福尼亚大学和 Google 研究机构发现，机器学习和自然语言处理模型使用的数据集 50%由 12 家 Top 机构提供，其中 10 家为美国机构，1 家为德国机构，仅 1 家机构来自中国，为香港中文大学。值得一提的是，数据集与数据机构的基尼系数有升高的趋势，即数据集被少数 Top 机构或特定数据库掌控的集中有所增加。

图表188：GPT3 训练中各国语言占比



资料来源：中国网信网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表189：数据集分布及发展趋势

注：左：截至 2021 年 6 月，每个机构的数据集使用情况图。网点大小表示使用次数。蓝点表示营利机构，橙点表示非营利机构。机构占使用量的 50%以上。右图：机构和数据集在整个 Papers With Code 数据集上使用集中度的基尼系数。圆点大小表示当年的使用次数。

资料来源：Bernard Koch et al. “Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research” 2021，华泰研究

我们认为国内缺乏高质量数据集主要有以下几方面的原因：1）高质量数据集需要大量资金投入，当前国内数据挖掘和数据治理的力度不足；2）国内相关公司不具有开源意识，数据无法在市场上自由流通；3）国内相关公司成立较晚，数据积累相对国外公司更少；4）学术领域中文数据集受重视程度低；5）国产数据集市场影响力及普及度较低等。

中国大模型如何构建数据集#1：LLM

我们选取了在其论文中详细阐述如何构建预训练数据集的三个大语言模型，研究中国大模型预训练数据集的来源。我们发现：1）类似海外大语言模型，中国大语言模型的预训练数据集也主要来自互联网抓取数据（如 Common Crawl、中文公共社交媒体抓取等）、网络百科全书（如百度百科、搜狗百科）、书籍等等；2）充分借助已有的高质量开源数据集，例如对 Common Crawl 等进行二次处理。

图表190：中国大语言模型数据集构成

公司	NLP大模型	发布时间	最大参数量（B）	数据集（TB_文本）	数据来源
百度	Plato-XL	2021.9	11	-	中文：公共领域的社交媒体、英文：Reddit评论
华为	盘古	2021.4	200	10	开源数据集、百科全书、电子书、Common Crawl、新闻
腾讯	WeLM	2022.9	10	80	Common Crawl、新闻、书籍、流行的在线论坛以及学术著作

资料来源：Siqi Bao et al. “PLATO-XL: Exploring the Large-scale Pre-training of Dialogue Generation” 2021，Wei Zeng et al. “PanGu-α: Large-scale Autoregressive Pretrained Chinese Language Models with Auto-parallel Computation” 2021，Hui Su et al. “WeLM: A Well-Read Pre-trained Language Model for Chinese” 2022，华泰研究

百度 Plato-XL 大模型：百度于 2021 年发布 PLATO-XL，包括中英文 2 个对话模型，预训练语料规模达千亿级 token，模型规模高达 110 亿参数。预训练语料库分为：1)英语：会话样本从 Reddit 评论中提取，这些评论由第三方收集，并在 pushshift.io 上公开发布，遵循了 PLATO-2 的精心清洗过程；2）中文：预训练数据来自公共领域的社交媒体，过滤后训练集中有 1.2 亿个样本。

华为盘古大模型：华为于 2021 年发布盘古，是业界首个 2000 亿参数以中文为核心的预训练生成语言模型，目前开源了盘古 α 和盘古 α 增强版两个版本，并支持 NPU 和 GPU 两个版本，支持丰富的场景应用，在知识问答、知识检索、知识推理、阅读理解等文本生成领域表现突出，具备较强的少样本学习的能力。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表191： 华为盘古大模型 1.1TB 中文文本语料库数据组成

数据集	大小（GB）	数据来源	处理步骤
开源数据集	27.9	15 个开源数据集：DuReader、百度 QA、CAIL2018、搜狗 CA 等	格式转换和文本重复删除
百科全书	22	百度百科、搜狗百科等	文本重复删除
电子书	299	各种主题的电子书(如小说、历史、诗歌和古文等)	基于敏感词和模型垃圾邮件过滤
Common Crawl	714.9	来自 Common Crawl 的 2018 年 1 月至 2020 年 12 月的网络数据	所有步骤
新闻	35.5	1992 年至 2011 年的新闻数据	文本重复删除

资料来源：Wei Zeng et al. “PanGu-α: Large-scale Autoregressive Pretrained Chinese Language Models with Auto-parallel Computation” 2021，华泰研究

腾讯 WeLM 大模型：腾讯于 2022 年发布 WeLM，数据来源主要分为三部分：1）Common Crawl：Common Crawl 于 2020.08 至 2022.01 期间的文本内容，使用 langdetect 工具过滤掉非中文的文本；2）特定领域语料库：混合了来自各种来源的数据，包括新闻、书籍、流行在线论坛以及学术著作，仅中文数据。3）英文数据：添加了从上述来源收集到的约 750GB 的英语数据。数据中有大量的噪音如胡言乱语或模板文本、冒犯性语言、占位符文本和源代码等，首先应用一组基于规则的过滤器，再在剩余的数据上手动构建好坏数据分类器提升数据清理泛化能力。

图表192： WeLM 大模型训练语料库统计

来源	%过滤	#剩余 Tokens	预训练比例
Common Crawl	92%	198.5B	50.6%
书籍	40.9%	61.9B	38.7%
新闻	7.5%	1.91B	6.7%
论坛	6.7%	1.0B	3.5%
学术著作	2.5%	0.39B	0.5%

资料来源：Hui Su et al. “WeLM: A Well-Read Pre-trained Language Model for Chinese” 2022，华泰研究

中国大模型如何构建数据集#2：多模态大模型

我们选取了在其论文中详细阐述如何构建预训练数据集的三个多模态模型，研究中国大模型预训练数据集的来源。我们发现网页抓取、自有数据和开源数据集是多模态大模型数据集的重要来源：1）网页抓取图文对：例如阿里 M6 大模型、百度 ERNIE-ViLG 大模型都从网页中抓取文本-图片对，然后经过一定过滤，形成最终数据集的一部分；2）自有数据：例如阿里 M6 大模型有来自电商的图文数据，百度 ERNIE-ViLG 大模型从内部图像搜索引擎中收集查询文本和对应的用户点击图像；3）开源数据集：例如百度 ERNIE-ViLG 大模型的部分图文对数据来自开源的 CC 和 CC12M，并通过百度翻译 API 翻译。

图表193： 中国多模态模型数据集构成

公司	多模态大模型	发布时间	最大参数量（B）	数据集（M_图文对/图像）	数据来源
阿里	M6	2021.3	100	-	百科全书、社区QA、论坛讨论、Common Crawl、抓取网页、电商数据
百度	ERNIE-ViLG	2021.12	10	145	中文网页、内部图像搜索引擎、CC、CC12M
上海人工智能实验室等	InternVideo	2022.12	1.3	-	Kinetics-400、WebVid2M、WebVid10M、HowTo100M、AVA、Something-Something V2、Kinetics-710、自采视频

资料来源：Junyang Lin et al. “M6: A Chinese Multimodal Pretrainer” 2021，Han Zhang et al. “ERNIE-ViLG: Unified Generative Pre-training for Bidirectional Vision-Language Generation” 2021，Yi Wang et al. “InternVideo: General Video Foundation Models via Generative and Discriminative Learning” 2022，华泰研究

阿里 M6 大模型：于 2021 年发布，参数规模达到 1000 亿。阿里构建了最大的中文多模态预训练数据集 M6-Corpus，包含超过 1.9 TB 图像和 292GB 文本，涵盖了百科全书、问答、论坛讨论、产品说明等类型的数据集。研究人员设计了完善的清洁程序：1）文本数据：删除 HTML 标记和重复的标点符号，只保留中文和英文的字符和标点符号。删除短于 5 个字符的标题和短于 15 个字符的文本内容。使用“内部垃圾邮件检测器”筛选包含某些政治问题、色情或脏话等不合适的句子。建立一个语言模型进行评估文本的困惑程度，去掉困惑程度高的句子；2）图片数据：只有超过 5000 像素的图像才有资格被保留用于预训练。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表194：M6 预训练数据集构成

来源	模态	图像（M）	Tokens(B)	段落（M）	平均长度	图像大小（TB）	文本大小(GB)
百科全书	纯文本	-	31.4	34.0	923.5	-	65.1
社区 QA	纯文本	-	13.9	113.0	123.0	-	28.8
论坛讨论	纯文本	-	8.7	39.0	223.1	-	18.0
Common Crawl	纯文本	-	40.3	108.7	370.7	-	83.3
百科全书	图像&文本	6.5	7.9	10.4	759.6	0.1	15.0
抓取网页	图像&文本	46.0	9.1	106.0	85.8	1.5	70.0
电商	图像&文本	8.0	0.5	8.5	62.1	0.3	12.2
总计		60.5	111.8	419.6	266.4	1.9	292.4

注：电商数据包含 260k 来自淘宝的成对产品描述和产品图片

资料来源：Junyang Lin et al. “M6: A Chinese Multimodal Pretrainer” 2021，华泰研究

百度 ERNIE-ViLG 大模型：于 2021 年发布，参数规模达到 100 亿。百度构建了一个由超过 1.45 亿对高质量中文图像-文本对组成的大规模图像-文本数据集，数据来源如下：1）中文网页。从各种中文网页中抓取了 8 亿对原始的中文替代文字描述和图片，进行了几个步骤的过滤，总共收获了 7000 万对文本-图片，过滤规则主要包括文本长度、文本内容和图像-文本相似度；2）图片搜索引擎：从内部图像搜索引擎中收集了大约 6000 万个查询文本和相应的用户点击图像；3）开源图像-文本数据集：从 CC 和 CC12M 中共收集了 1500 万文本图像对，这些数据集中的字幕通过百度翻译 API 翻译成中文。

InternVideo 大模型：由上海人工智能实验室等、南大、港大、复旦、中科院深圳先进技术研究院等于 2022 年发布，使用了 6 个来自各个领域的开源数据集和自采视频片段。

图表195：InternVideo 预训练过程中使用的数据集统计

预训练数据集	域	样本剪辑	帧数×采样率
Kinetics-400	Youtube 视频	240k	16 × 4
WebVid2M	网络视频	250k	16 × 4
WebVid10M	网络视频	10M	16 × 4
HowTo100M	Youtube 视频	1.2M	16 × 4
AVA	电影	21k	16 × 4
Something-Something V2	剧本镜头	169k	16 × 2
自采视频	Youtube, Instagram	250k	16 × 4
Kinetics-710	Youtube 视频	680k	16 × 4

资料来源：Yi Wang et al. “InternVideo: General Video Foundation Models via Generative and Discriminative Learning” 2022，华泰研究

中国开源数据集#1：大语言模型数据集

DuReader 数据集：于 2018 年由百度发布。DuReader 是一个大规模的开放域中文机器阅读理解数据集。该数据集由 200K 问题、420K 答案和 1M 文档组成，是迄今为止最大的中文 MRC 数据集。问题和文档基于百度搜索和百度知道，答案是手动生成的。该数据集还提供了问题类型注释——每个问题都被手动注释为实体、描述或是否以及事实或意见之一。

WuDaoCorpora 数据集：于 2021 年由北京人工智能研究院、清华大学和循环智能联合发布。WuDaoCorpora 是北京智源研究院最新构建的高质量数据集，悟道文本数据集采用 20 多种规则从 100TB 原始网页数据中清洗得出最终数据集，注重隐私数据信息的去除，从源头上避免 GPT-3 存在的隐私泄露风险；包含教育、科技等 50+个行业数据标签，可以支持多领域预训练模型的训练。

CLUECorpus2020 数据集：于 2020 年由 CLUE 发布。CLUECorpus2020 是一个可以直接用于语言模型预训练或语言生成等自监督学习的大型语料库，它有 100G 的原始语料库，包含 350 亿个汉字，这些语料库来自 Common crawl。

免责

更多免费行业报告 并购家 www.ipoipo.cn

CAIL2018 数据集：于 2018 年由清华大学、北京大学、中国科学院软件研究所和中国司法大数据研究院联合发布。CAIL2018 是第一个用于判决预测的大规模中国法律数据集，收录了中国最高人民法院公布 260 万件刑事案件，是现有判决预测工作中其他数据集的数倍。对判断结果的注释也更加详细和丰富。它由适用的法律条款、指控和刑期组成，根据案件的事实描述而定。

Ape210K 数据集：于 2020 年由猿辅导 AI Lab 和西北大学联合发布。Ape210K 是一个新的大规模和模板丰富的数学单词问题数据集，包含 210K 个中国小学水平的数学问题，是 Math23K 的 9 倍。每个问题都包含黄金答案和得出答案所需的方程式，有 56K 个模板，是 Math23K 的 25 倍。

FCGEC 数据集：于 2022 年由浙江大学和华为联合发布。FCGEC 用于检测、识别和纠正语法错误，是一个人工标注的多参考语料库，由 41,340 个句子组成，主要来自公立学校语文考试中的选择题。

E-KAR 数据集：于 2022 年由复旦大学、字节跳动人工智能实验室和 BrainTechnologies,Inc.联合发布。数据集包含来自公务员考试的 1,655 个（中文）和 1,251 个（英文）问题，这些问题需要深入的背景知识才能解决。

Douban Conversation Corpus 数据集：于 2017 年由北京航空航天大学、南开大学和微软研究院联合发布。豆瓣会话语料库包括一个训练数据集、一个开发集和一个基于检索的聊天机器人的测试集，测试数据包含 1000 个对话上下文。

MATINF 数据集：于 2020 年由武汉大学和密歇根大学联合发布。MATINF 是一个联合标注的大规模数据集，用于中文母婴护理领域的分类、问答和总结。数据集中的一个条目包括四个字段：问题、描述、类别和答案。从中国大型母婴护理 QA 网站收集了近 200 万对问答对，其中包含细粒度的人工标记类，数据清洗后，用剩余的 107 万个条目构建。

中国开源数据集#2：多模态模型数据集

WuDaoMM 数据集：于 2022 年由清华大学和北京智源人工智能研究院联合发布。WuDaoMM 是北京智源人工智能研究院 WuDaoCorpora 开源数据集的一部分。WuDaoMM 是图像和文本的多模态预训练数据，完整的数据集包含 6.5 亿对图像和文本，包含几千万对的强相关数据和 6 亿对弱相关数据，包含 19 大类，分别是：能源、表情、产业、医疗、景观、动物、新闻、花卉、教育、艺术、人物、科学、海洋、树木、汽车、社会、科技、体育等。

MUGE 数据集：于 2021 年由清华大学和阿里巴巴联合发布，包括图像描述、图像文本检索和基于文本的图像生成 3 种多模态理解和生成任务。

Noah-Wukong 数据集：于 2022 年由华为诺亚方舟实验室和中山大学联合发布。诺亚悟空数据集是一个大规模的多模态中文数据集，包含 100 万对图文对，数据集中的图像根据大小和宽高比进行过滤，数据集中的文本根据其语言，长度和频率进行过滤。隐私和敏感词也被考虑在内。

Zero 数据集：于 2022 年由北京航空航天大学、清华大学、奇虎 360 人工智能研究所联合发布。Zero 是一种大规模的中文跨模态基准测试，其中包含目前最大的公共预训练数据集 ZERO-Corpus 和五个用于下游任务的人工注释微调数据集。ZERO-Corpus 包含 2.5 亿张图片和 7.5 亿篇文字描述，另外五个微调数据集中的两个也是目前中国跨模式下游任务中最大的数据集。

更多免费行业报告 并购家 www.ipoipo.cn

COCO-CN 数据集：于 2018 年由中国人民大学发布。COCO-CN 是一个双语图像描述数据集，通过手动编写的中文句子和标签丰富了 MS-COCO。新数据集可用于多种任务，包括图像标记、字幕和检索，所有这些都在跨语言环境中完成。COCO-CN 拥有 20,342 张图片，27,218 个中文句子和 70,993 个标签，为跨语言图像标注、字幕和检索提供了一个统一平台。

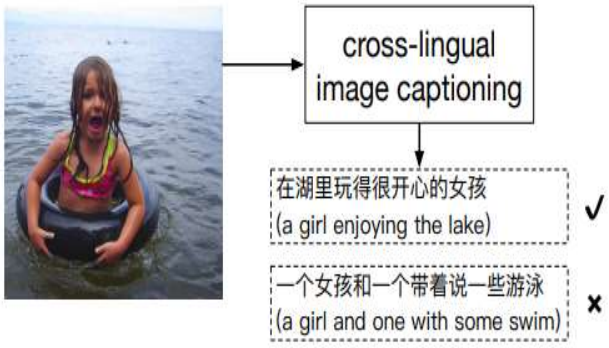
Flickr8k-CN & Flickr30k-CN 数据集：于 2017 年由浙江大学和中国人民大学联合发布。Flickr8k-cn 是公共数据集，每个测试图像与 5 个中文句子相关联，这些句子是通过手动翻译 Flickr8k 中对应的 5 个英文句子获得的。Flickr30k-cn 是 Flickr30k 的双语版本，通过其训练/有效集的英译汉机器翻译和测试集的人工翻译获得。

图表196： COCO-CN 数据集——示例

Image	MS-COCO text	COCO-CN *Sentences*	COCO-CN *Tags*
	man is flying a kite in a field	一个男人在河边的草地上放风筝 (A man flying a kite on the riverside grass)	蓝天 (bule sky) 天空 (sky) 多云 (cloudy) 年轻人 (youth) 青草 (grass) 草地 (lawn)
	young man serving a tennis ball to his opponent	在红土场上，一个身穿蓝色运动服的男人正跳起发球 (On a clay court, a man wearing a blue sportswear is jumping to serve)	红土球场 (clay court) 网球 (tennis) 男人 (man) 球拍 (racket) 网球运动员 (tenni player) 红土 (red clay)
	zebras run on the grass near the trees	一群野生动物在草地上奔跑 (A herd of wild animals are running on the grass)	灌木丛 (bush) 斑马 (zebra) 草地 (grass) 大自然 (nature)

资料来源：Xirong Li et al. “COCO-CN for Cross-Lingual Image Tagging, Captioning and Retrieval” 2019，华泰研究

图表197： Flickr30k-CN 数据集——跨语言图像字幕示例



资料来源：Weiyu Lan et al. “Fluency-Guided Cross-Lingual Image Captioning” 2017，华泰研究

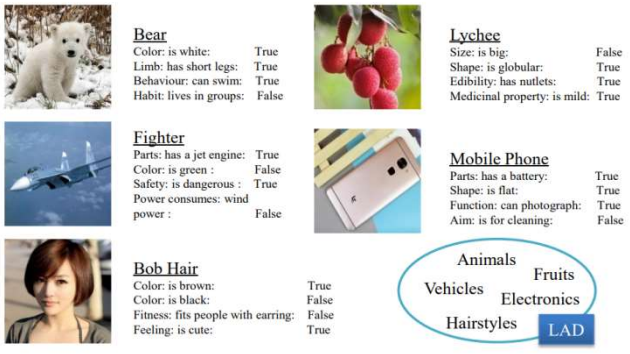
Product1M 数据集：于 2021 年由北京交通大学、阿里巴巴和中山大学联合发布。Product1M 是用于实际实例级检索的最大的多模式化妆品数据集之一，包含超过 100 万个图像对并且由两种样品类型组成，即单产品和多产品样品，其中包括各种化妆品品牌。

AI Challenger 图像中文描述数据集：数据来自 2017 AI Challenger，数据集对给定的每一张图片有五句话的中文描述。数据集包含 30 万张图片，150 万句中文描述。数据集包含人类关键点检测(HKD)、大规模属性数据集(LAD)和图像中文字幕(ICC)三个子数据集。

图表198： Product1M 数据集——多模态实例级检索

资料来源：Xunlin Zhan et al. “Product1M: Towards Weakly Supervised Instance-Level Product Retrieval via Cross-Modal Pretraining” 2021，华泰研究

图表199： AI Challenger 数据集——示例



资料来源：Jiahong Wu et al. “AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding” 2017，华泰研究

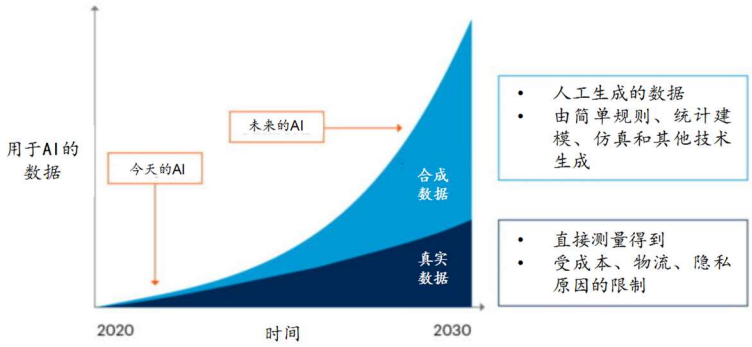
更多免费行业报告 并购家 www.ipoipo.cn

高质量语言数据和图像数据或将耗尽，合成数据有望生成大模型数据

高质量语言数据或将于 2026 年耗尽。数据存量的增长速度远远低于数据集规模的增长速度，如果当前的趋势继续下去，数据集最终将由于数据耗尽而停止增长。在语言模型方面，语言数据的质量有好坏，互联网用户生成的语言数据质量往往低于书籍、科学论文等更专业的语言数据，高质量数据训练出的模型性能更好。根据《Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning》预测，语言数据将于 2030~2040 年耗尽，其中能训练出更高性能模型的高质量语言数据将于 2026 年耗尽。此外，视觉数据将于 2030~2060 年耗尽。

合成数据或将弥补未来数据的不足。合成数据是计算机模拟或算法生成的带有注释的信息，可以替代真实数据。它可以用于模拟实际情况，补充真实数据的不足，提高数据质量和数量，以及降低数据采集和处理的成本。OpenAI 在 GPT-4 的技术文档中重点提到了合成数据的应用，可见其对该领域的重视。根据 Gartner 的预测，2024 年用于训练大模型的数据中有 60%将是合成数据，到 2030 年大模型使用的绝大部分数据将由人工智能合成。

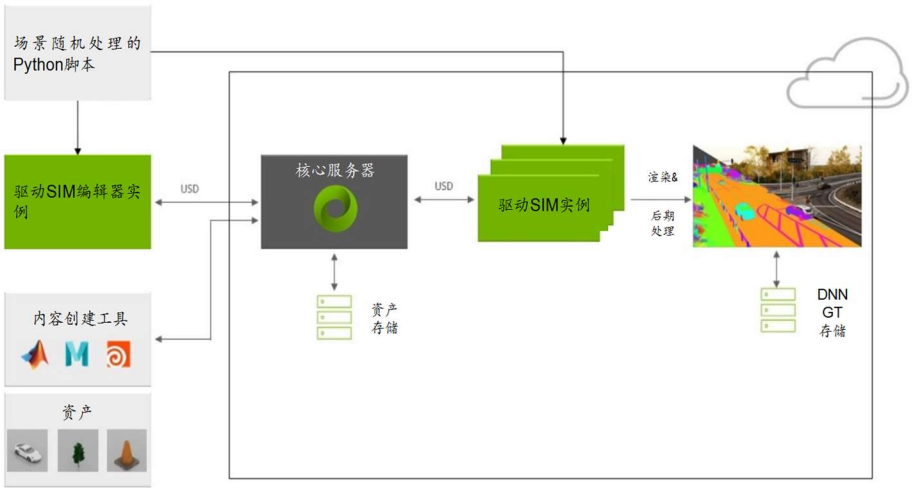
图表200： 到 2030 年 AI 模型中的合成数据将完全盖过真实数据



资料来源：Gartner，华泰研究

合成数据有望首先在金融、医疗和汽车等诸多领域落地。在金融行业，金融机构可以在不提供敏感的历史交易信息前提下，通过合成数据集训练量化交易模型提升获利能力，也可以用来训练客服机器人以改善服务体验；在生物医药行业，可以通过合成数据集，在不提供患者隐私信息的条件下训练相关模型完成药物研发工作；在自动驾驶领域，可以通过合成数据集模拟各种驾驶场景，在保障人员和设备安全的条件下提升自动驾驶能力。

图表201： NVIDIA Omniverse——用户可使用 Python 为自动驾驶车辆生成合成数据



资料来源：英伟达官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

数据产业链投资机会：数据生产环节、数据处理环节

建立数据集的的流程包括数据采集、数据清洗、数据标注、模型训练、模型测试和产品评估。与之相对应，我们认为数据产业链分为数据生产、数据处理、数据使用三大环节。数据使用环节的参与者包括训练、微调大模型的企业，本文不作展开。本节中，我们对数据生产、数据处理环节进行讨论。

数据集如何产生

建立数据集的流程主要分为 1）数据采集；2）数据清洗：由于采集到的数据可能存在缺失值、噪声数据、重复数据等质量问题；3）数据标注：最重要的一个环节；4）模型训练：模型训练人员会利用标注好的数据训练出需要的算法模型；5）模型测试：审核员进行模型测试并将测试结果反馈给模型训练人员，而模型训练人员通过不断地调整参数，以便获得性能更好的算法模型；6）产品评估：产品评估人员使用并进行上线前的最后评估。

流程#1：数据采集。采集的对象包括视频、图片、音频和文本等多种类型和多种格式的数据。数据采集目前常用的有三种方式，分别为：1）系统日志采集方法；2）网络数据采集方法；3）ETL。

图表202：数据采集三种常见方式

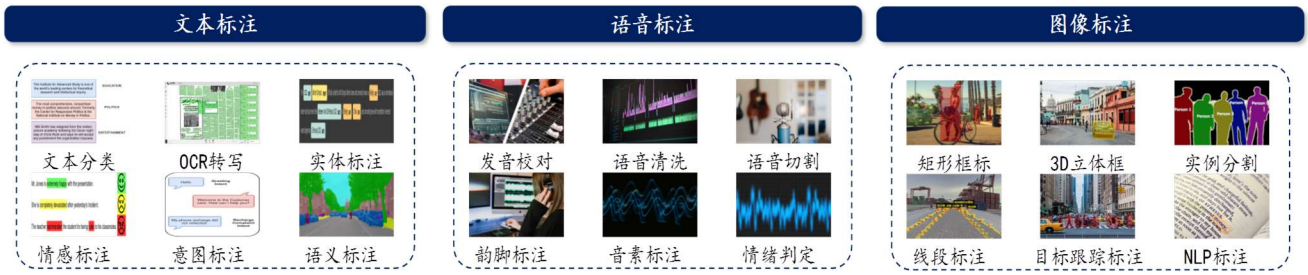
资料来源：CSDN，Apache，Scrible，Python，GitHub，Scrapy，IBM，华泰研究

流程#2：数据清洗是提高数据质量的有效方法。由于采集到的数据可能存在缺失值、噪声数据、重复数据等质量问题，故需要执行数据清洗任务，数据清洗作为数据预处理中至关重要的环节，清洗后数据的质量很大程度上决定了 AI 算法的有效性。

流程#3：数据标注是流程中最重要的一个环节。根据不同的标注需求，将待标注的数据划分为不同的标注任务。每一个标注任务都有不同的规范和标注点要求，一个标注任务将会分配给多个标注员完成。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表203：三大类数据标注



资料来源：Devol Shah “A Step-by-Step Guide to Text Annotation” 2022，CSDN，景联文科技，华泰研究

流程#4：最终通过产品评估环节的数据才算是真正过关。产品评估人员需要反复验证模型的标注效果，并对模型是否满足上线目标进行评估。

数据生产环节投资机会

数据生产环节是数据产业链的上游环节，是数据的源头。环节内的企业或从业务运营中直接产生数据，或作为平台方聚合数据。按照数据的通用程度，我们认为这一环节的公司可以分为通用类型数据及垂直行业数据 2 类。

1） **通用类型数据：**如前文所言，我们认为 AI 大模型需要高质量、大规模、具有多样性的数据。对标海外主要数据集，通用类型数据来自维基百科、书籍期刊、高质量论坛，因此国内的数据或来自文本数据领域的百度百科、中文在线、中国科传、知乎，以及图像视觉领域的视觉中国等公司。

a) 截至 2022 年 6 月，视觉中国拥有超过 2/3 的高水准独家或自有内容，目前提供 4 亿张图片、3,000 万条视频和 35 万首音乐等可销售的各类素材，是全球最大的同类数字版权内容平台之一。

b) 中国科传从事图书出版业务、期刊业务、出版物进出口业务。截至 2022 年底，公司年出版新书超过 3000 种，已累计出版图书超过 5 万种，是国内学科分布最全、出版规模最大的综合性科技出版机构。截至 2022 年底，中国科传出版期刊 554 种，其中中文期刊 254 种，英文期刊 276 种，中英文期刊 5 种，法文期刊 19 种。共有 101 种期刊被 SCI 收录，其中 36 种期刊处于 Q1 区，4 种期刊在国际同学科期刊中排名第一，16 种期刊居国际同学科期刊排名前 10%。

c) 截至 2022 年 6 月，中文在线累积数字内容资源超 510 万种，网络原创驻站作者 440 余万名。

2） **行业数据：**我们认为垂直行业的高价值量数据对于 AI 大模型，尤其是行业大模型的训练和落地至关重要。处于数字化程度领先的行业中的龙头公司在行业数据积累上具有优势，例如：1）计算机视觉领域的海康威视、大华股份；2）城市治理、ToB 行业应用领域的中国电信、中国联通等；3）金融领域的同花顺、东方财富等；4）自动驾驶领域的特斯拉、蔚小理、经纬恒润、德赛西威等。

大模型时代数据价值凸显，国内外数据收费为大势所趋，收费方式尚在摸索中。2023 年 4 月 18 日，美国知名论坛 Reddit 宣布计划向通过其 API 使用数据的公司收费。Reddit 尚未公布具体的收费标准，但表示会分为不同的等级，根据使用者的规模和需求来区分。Reddit 是大模型训练的优质语料库，OpenAI 的 GPT-3 训练使用了来自 Reddit 的数据，Meta 旗下的 Facebook AI Research 与华盛顿大学也联合开源了来自 Reddit 数据的 OpenWebText 数据集。对于通用类型数据和行业数据，我们认为其潜在的变现方式可能存在差异：

免责
更多免费行业报告 并购家 www.ipoipo.cn

1） **通用类型数据：**我们认为通用类型数据所有者可能采用开发自有模型/应用、售卖数据 2 种变现方式。例如，知乎联合清华系 AI 公司面壁智能发布中文大模型“知海图 AI”。中文在线则基于自有数据开发了 AI 辅助文字创作工具，并计划售卖数据：根据中文在线 4 月 19 日回复深交所关注函内容，其收费方式为按照采集数据包的大小及数据类别进行基础包加增量包的收费，目前尚未签署具体合作协议。

2） **行业数据：**我们认为数据是垂直行业企业的护城河之一，结合具体场景和用户充分挖掘数据能更好地赋能业务。因此垂直行业企业或更偏好基于基础模型，使用自有数据来训练自有模型，并且可能会尽量规避售卖数据。建议关注具有丰富行业数据积累的龙头公司。

数据处理环节投资机会

数据处理环节涉及数据设计、数据采集、数据清洗、数据标注和数据质检服务。据艾瑞统计与预测，2022 年我国 AI 基础数据服务市场规模（含数据采集与标注）为 31 亿元。受自动驾驶、对话机器人和消费硬件等 AI 应用，对图像、语音和文本数据集及定制化数据服务的需求上涨影响，2027 年相应规模可达到 79 亿元，2022-2027 年 CAGR 为 20.6%。

数据服务相关公司：百度智能云、慧听、标贝、它思、Appen、Telus international、Scale AI

百度智能云（9888 HK/BIDU US）

百度智能云拥有专业的 AI 基础数据服务平台——百度智能云数据众包，为客户提供专业的 AI 数据标注服务，凭借百度智能云逾 10 年的数据服务经验和标注基地丰富的标注人力，可准确、高效、安全地完成各类型数据标注任务，助力客户算法模型训练。2011 年，公司建立数据采标团队，能力全面，资源丰富。2019 年，公司与政府共建百度山西数据标注基地，拥有 100+数据专家、2000+标准人员和 10 万+外部标注人力，具备百万级数据标注处理能力。

慧听科技（未上市）

北京慧听科技有限公司是一家专业的数据服务提供商。2011 年，公司成立，采用全程质量监控流程，执行完善的标注流程，配合保密管理手段，提供质量上乘的数据服务。团队核心成员拥有十余年的数据制作经验，负责完成过语音识别、语音合成、语音评测、语言文本类和多媒体类等多领域数据制作，并参与过语音合成、语音识别和输入法系统的研发。目前，公司提供语言语音、多媒体两大类几十余种数据服务。

标贝科技（未上市）

标贝（北京）科技有限公司是一家专注于智能语音交互和 AI 数据服务的人工智能公司。2016 年，公司成立，基于 AI+SaaS 开放平台，为客户提供 AI 数据服务、技术能力和智能语音交互方案赋能服务；AI 数据业务涵盖语音合成、语音识别、图像视觉、NLP 等采标服务和平台化自研工具能力。2020 年，公司加入“腾讯 5G 生态计划”。目前，公司已与微软、阿里、腾讯、360、小鹏、中科大和中国银行等国内外百余家企业客户建立合作，服务项目累计超过 1,000 项，涵盖汽车、教育、零售、阅读和智能硬件等多个领域。

它思科技（未上市）

它思科技是专注打造高效学习的智能语音语言技术和产品的领航者。2018 年，公司成立，得到清华大学天津电子信息研究院的扶持并完成千万元融资。公司主要从事语音与语言处理相关 AI 技术及产品的研发，拥有国际一流的语音语言处理技术栈，包括人机对话系统、语音识别与理解、自然语言处理、说话人识别、情感识别等，业务包括智慧商业、智慧医疗、智慧广电及安防等。公司致力于建设第三代人机交互对话系统，是生成式对话技术的先行者，近年来引领发展半监督式 AI，摆脱第二代 AI 技术过度依赖于高成本标注的现实瓶颈，协同利用标注数据和无标数据，发挥数据网络效应，大幅提升数据转化为智能的效率。

免责
更多免费行业报告 并购家 www.ipoipo.cn

Appen（APX AU）

Appen 是全球领先的图像、文本、语音、音频和视频等 AI 训练数据服务提供商。1996 年，公司成立，拥有人工智能辅助数据注释平台，在全球 170 多个国家与 100 多万名专业承包商合作，训练数据涵盖科技、汽车、金融服务、零售和医疗健康和政府等各个领域。目前，公司已与微软、亚马逊及谷歌等大型科技公司、汽车厂商和政府等客户建立合作。2022 年收入 3.9 亿美元。

Telus international（TIXT CN）

Telus international 是数字客户体验创新者，业务涵盖数字化战略、创新、咨询和设计、数字化转型和 IT 生命周期服务、数据注释和智能自动化等。2005 年，公司成立。目前，公司在全球 32 个国家与 650+客户建立合作，涵盖科技和游戏、通信和媒体、电子商务和金融科技、银行、金融服务和保险、医疗保健、旅游和酒店等高增长行业垂直领域。2022 年收入 24.7 亿美元，其中 AI 数据解决方案占比 13%（约 3.2 亿美元）。

Scale AI（未上市）

2016 年，公司成立，专注于标记数据用于训练驱动自动驾驶的机器学习模型。2020 年，公司将重点从交通运输扩展到电子商务、物流、政府和保险。目前，它通过改进底层数据来帮助组织实施 AI，客户涵盖 OpenAI、stability.ai、英伟达、亚马逊和三星等大型科技公司、丰田等汽车厂商、美国和乌克兰等政府。

免责
更多免费行业报告 并购家 www.ipoipo.cn

半导体国产化：供需错配推动半导体制造国产化加速

中国半导体供需错配，国产替代主要集中在设计、封测等环节

经过多年的发展，我国大部分中低端半导体产品实现了国产化替代，但是高端产品有待进一步的发展和提高。目前在消费类电子，如机顶盒芯片、监控器芯片、CIS 芯片以及通信设备芯片，国内厂商能较好地兼顾性能、功耗、成本等因素，被市场广泛认可。根据 Gartner，2021 年国产芯片在分立器件、传感器、无线通讯芯片、应用处理器等细分领域的全球市场份额超过 10%。但半导体材料、设备全球占比仅为 9%、5%。受地缘政治影响，我们认为全球半导体的生产中心未来会从中国台湾走向全球分散。发展设备/零部件/材料等核心环节，避免“卡脖子”是未来重要发展路线。

图表204： A股半导体企业国产化率

		全球市场情况（2021A）				
	类别	市场集中度	全球市场规模（百万美元）	美国	中国大陆	中国公司
产业链环节	EDA软件	★★★	13,000	77%	1%	华大九天、全芯智造、芯华章、国微集团、概伦电子
	设备	★★	102,640	40%	5%	北方华创、盛美上海、中微公司
	材料	★★	64,300	17%	9%	沪硅产业、中环股份、飞凯材料、晶瑞电材
	晶圆代工	★★★	110,100	11%	13%	中芯国际、华虹集团
	封装测试	★	77,700	15%	28%	长电科技、通富微电、华天科技
芯片	存储器	★★★	167,592	28%	1%	兆易创新、长江存储、长鑫存储
	微处理器（CPU/GPU等）	★★★	63,532	93%	5%	龙芯、华为鲲鹏、景嘉微、航锦科技、兆芯
	模拟IC	★★★	42,863	60%	4%	矽力杰、圣邦股份、思瑞浦、纳芯微、艾为电子、晶丰明源
	无线通讯（基带/射频等）	★★★	90,400	72%	5%	卓胜微、华为海思、紫光展锐、翱捷科技
	逻辑IC（显示驱动等）	★	11,826	80%	10%	格科微、韦尔股份、集创北方
	MCU	★	25,704	30%	6%	兆易创新、乐鑫科技、中颖电子、芯海科技
	传感器	★	32,153	19%	20%	豪威、汇顶科技、兆易创新、思特威
	光电子器件	★★	40,318	42%	20%	长光华芯、源杰科技、纵慧芯光
	分立器件	★	30,337	29%	30%	闻泰科技、士兰微、斯达半导

注：美国及中国为各地区本土厂商合计市占率

资料来源：Gartner，SEMI，华泰研究

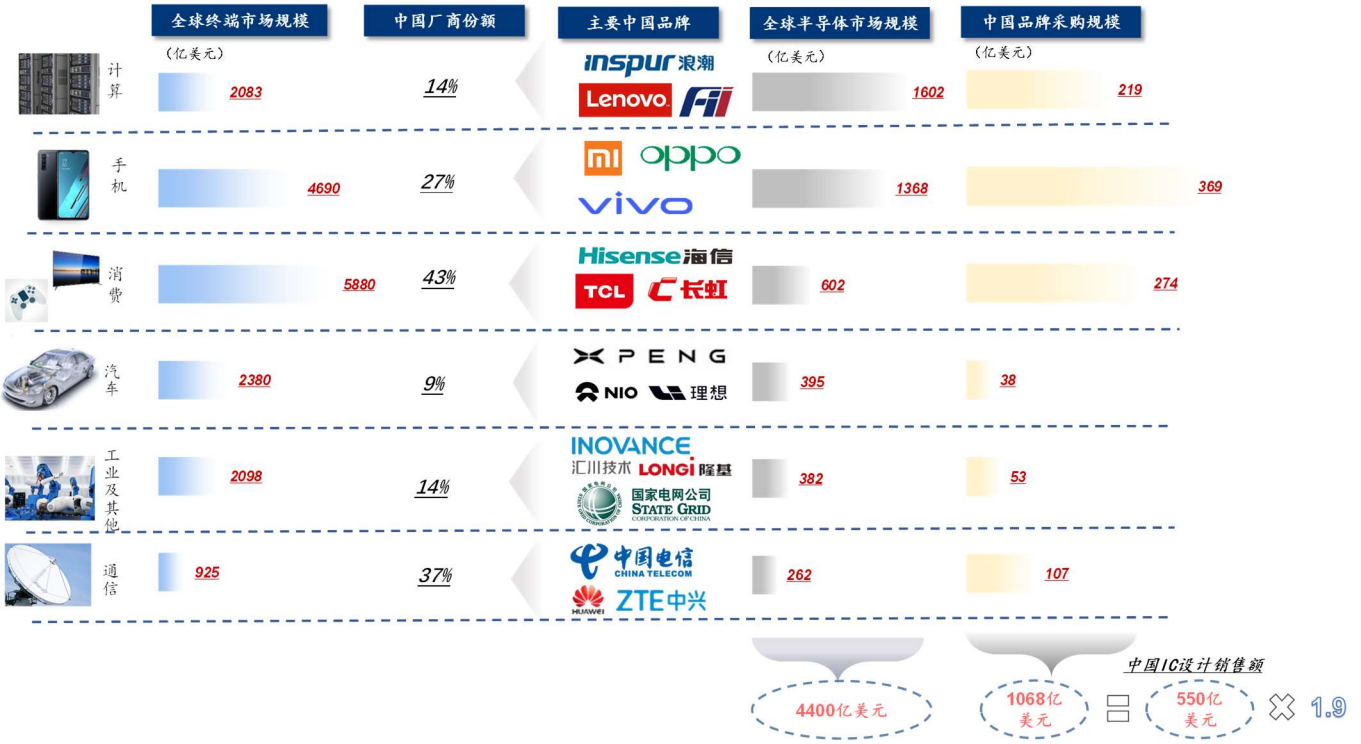
2022 年中国系统厂商半导体消费额约占全球 25%，而代工收入仅为全球 10%

产能紧缺下系统厂商成为推动中国半导体制造国产化的新动力。系统厂商在芯片产能的持续紧张中处于牵一发而动全身的状态，在疫情、地缘政治和需求等多重因素综合影响下，我国手机、汽车、家电和屏厂等整机公司在过去一年多的时间里增加备货，让芯片成为了储备物资；同时，出于供应链安全考虑，系统厂商努力寻求本地化的芯片设计和制造。尤其是在制造方面，即使是采用海外设计公司的芯片，他们也希望海外的公司能够有规划地转向中国制造，这给国内集成电路的设计和制造业带来庞大的需求。

中国企业在全球有明显优势的领域如显示面板/安防/物联网、电动车（IGBT/MCU/电源管理）、碳中和（功率）等领域国产化需求大幅上升，产能的结构性紧张情况仍将持续。经我们测算，2022 年中国计算、手机、消费、汽车、通信等系统厂商半导体消费额超过 1068 亿美元，约占全球半导体消费额的 25%，而中国代工收入仅为全球 10%，仍不能实现全面制造国产化。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表205： 中国系统厂商成为半导体国产化主要推动力（2022A）



资料来源：Yole，IDC，华泰研究

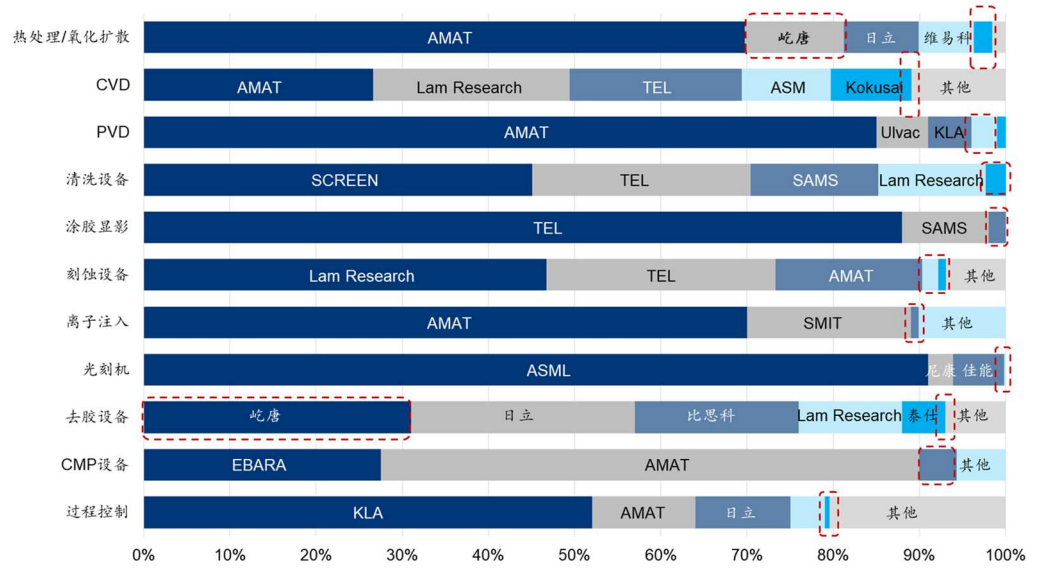
晶圆厂产能从集中走向分布，半导体制造中心向中国转移

全球半导体产业从上世纪四五十年代在美国起源后开始蓬勃发展，在成长过程中历经了从美国到日本，从日本到韩国和中国台湾，以及再到中国大陆的三次产业区域转移。2001年后，随着加入世贸组织，中国大陆逐渐深度参与到全球产业中。半导体下游应用从台式机逐渐拓展到笔记本电脑、手机等各类电子设备，终端产品更加复杂多样。美国、日本、韩国等地区逐渐把劳动力密集型的封装、测试等环节转移到中国大陆。中国大陆半导体产业经历了较长时间周期，逐步完成了技术与资本积累，并在国家战略及产业政策的扶持下，推动半导体全产业链高速发展。

代工产能东移带动半导体设备、材料、零部件需求，2022 年中国大陆前道设备厂商/材料厂商全球市占率约为 5%/9%，国产化需求急切。上游产业主要包括整个产业链需要的硅片、光刻胶、光掩膜、特种气体等等材料以及设备，在这些领域，我国整体国产化水平较低。我们以收入口径计算，国产半导体设备国产化率约 17.3%，本土公司经过多年的技术攻关以及国家项目的支持，目前能满足下游晶圆厂商大部分成熟制程（28nm 及以上的逻辑芯片等）以及少部分先进制程的需求。光刻设备较为特殊，目前国产设备只能应用于 90nm 以上的领域，相对来说与国外的差距更大一些。半导体材料种类繁多，竞争格局分散，目前整体国产化率为 20%左右。具体品类来看，其中电子特气、湿化学品的壁垒主要在于精度/纯度，国产公司已经能在部分产品做到国际标准水平，国产化率约 30%以上；硅片、靶材、CMP 耗材对加工的工艺要求比较高，国产化率约在 20-30%。但在高端领域，如 12 寸硅片国产化率不足 5%；替换成本较高，涉及前后道工序的光掩模版、光刻胶国产化率在 10%以下，尤其 ArF 光刻胶目前仍未实现国产替代量产突破。

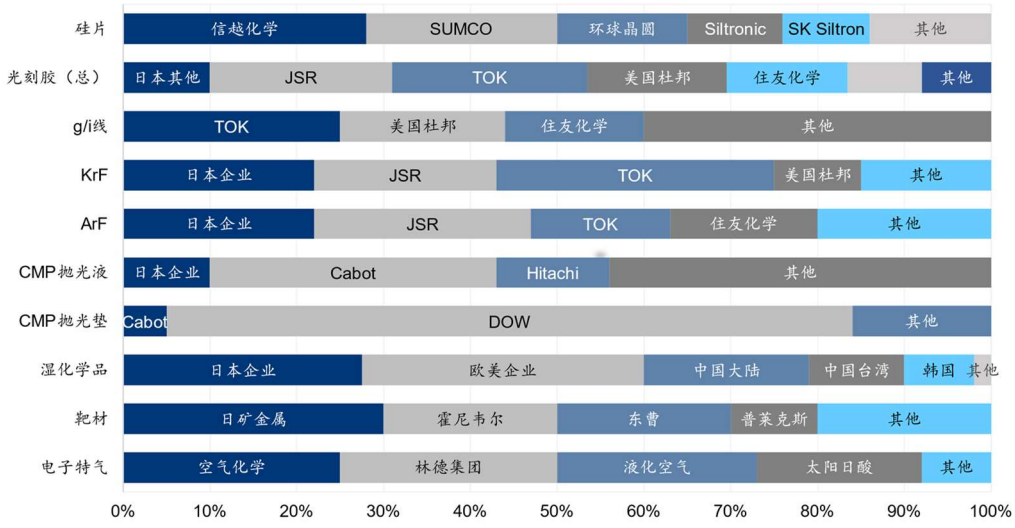
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表206：中国大陆半导体前道制造设备厂商 22 年市占率仅为 5%



注：柱状图表示各公司市占率，红色框为国产公司，以收入口径计算 2022 年国产设备厂商全球市占率约 5%
资料来源：ittbank，Gartner，华泰研究

图表207：中国大陆半导体材料厂商 22 年市占率约 9%



注：柱状图表示各公司市占率，以收入口径计算 2022 年国产半导体厂商全球市占率约 9%
资料来源：ittbank，Gartner，华泰研究

美国、日本进一步扩大对华半导体禁令，催生国产替代需求

美国进一步扩大对华半导体禁令，荷兰、日本共同限制尖端半导体制造设备出口，国产替代迫在眉睫。2022 年，美国出台《芯片和科技法案》，法案限制中国进口可用于加工 14/16nm 及以下逻辑芯片、可用于加工 18nm 及以下的 DRAM 芯片和 128 层及以上的 NAND 芯片的设备，并禁止未获许可的美国公民在中国从事芯片开发或制造工作。除此之外，10 月以来多家半导体、芯片、光电等高科技企业入列实体清单，包括长江存储、合肥万亿芯、寒武纪、深圳鹏芯微、上海微电子等。向该清单上的实体出口、再出口或转让《出口管理条例》（EAR）的管制物项，均需取得美国商务部工业和安全局（BIS）的许可，不适用许可例外。此外，北方华创（002371）间接控股子公司（北方华创磁电科技有限公司）等 31 家实体进入“未经核实清单”（UVL），被禁止通过许可例外的方式获得受 EAR 管辖的物项；即使管辖物项出口无需申请许可证，出口商也须进行备案。紧接着，美国于 2023 年 1 月跟半导体强国日本与荷兰达成共识，将限制向中国出口尖端半导体制造设备。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表208：美国各类清单一览

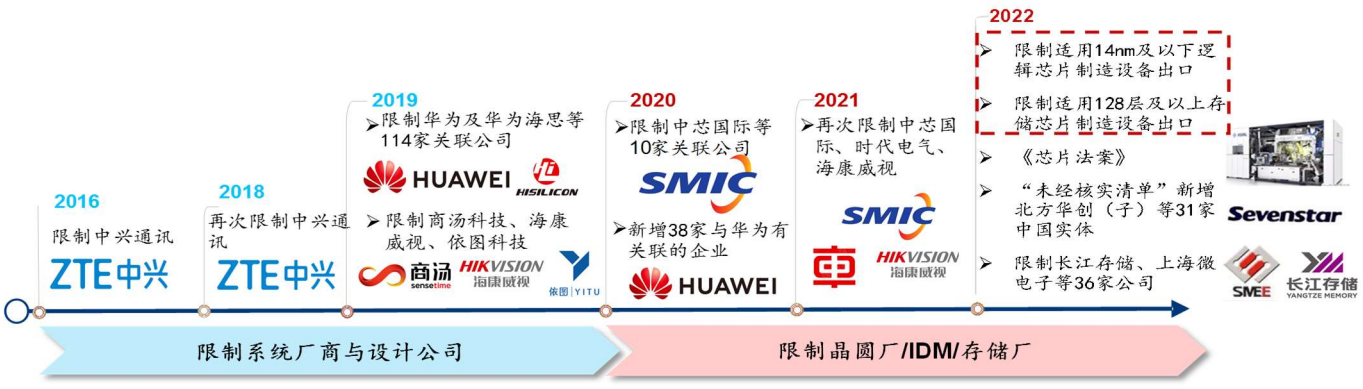
序号	清单名称	发布机构	限制内容	主要影响	部分被列入企业
1	**未经核实清单(UVL)**	商务部BIS	**不适用许可证例外；需履行额外手续**	**涉美受控物项供应链安全受到一定影响**	**北方华创（子）、中捷电子、信维电子、盛维科技、意科股份（子）**
2	**实体清单（EL）**	商务部BIS	**限制获取涉美受控物项(商品、技术及软件)**	**涉美受控物项供应链发生“休克性“中断**	**华为、寒武纪、浪潮、中科曙光、海康威视、大华股份、商汤科技（子）、旷视科技、依图科技、云飞励天、科大讯飞、中国移动、中国电信、大疆、中芯国际、景嘉微、长城军工、格林深瞳、鹏芯微、上海微电子、长江存储**
3	被拒绝人员清单（DPL）	商务部BIS	不得以任何方式参与从美国出口或拟出口的受EAR管辖物项的任何交易	涉美受控物项供应链发生全面中断	/
4	军事最终用户清单（MEU）	商务部BIS	限制获取部分CCL物项	特定行业涉美受控物项供应链安全受到严重影响	国科微、云芯微电子、国盾量子、H3C、杭州中科微电子、嘉兆科技
5	特别指定国民清单(SDN)	财政部OFAC	冻结资产，禁止使用美元；一级制裁及二级制裁	国金融机构及交易方暂停服务和交易导致业务“断崖式”全面停摆	和德宇航、爱立特微电子、旭润科技、灿禾医药
6	中国军工复合体企业清单(NS-CMIC)	财政部OFAC	限制美国人投资公开交易证券；军事最终用户警示红旗	无法使用美国资金；采购涉美物项受到一定影响	华为、浪潮、大疆、中航光电、景嘉微、中电科、中国移动、中国电信、中国联通、云从科技、中科曙光、海康威视、旷视科技、商汤科技、中芯国际、熊猫电子、长城军工

注：“（子）”表示实际列入清单的是其子公司；涉及的子公司分别为北方华创磁电科技有限公司、滁州惠科光电科技有限公司、北京商汤科技开发有限公司
资料来源：BSI，华泰研究

日本与荷兰政府的表态对美国禁令的实施效果至关重要。2021 年，我国 410 亿半导体设备进口总额中，118 亿美元来自日本对华出口，美国和荷兰分别对华出口 68 亿美元和 25 亿美元。我国是日本半导体设备进口大国，日本进口占进口总额超过 1/4；日本半导体设备对中国出口量占其总出口量近 40%。因此，若日本严格限制对华出口，贸易双方的半导体产业均会受到重创。尽管如此，日本于 2023 年 3 月底宣布修改《外汇及外国贸易法》，并于 5 月 23 日公布相关修正法条，限制出口包括尖端极紫外线（EUV）光刻机在内的 23 种芯片制造设备，预计于 7 月 23 日正式实施。由于日本按外汇法对可转用到武器等军事用途的民用产品进行出口管理，虽然该修正法未明确对中国等特定国家提出限制，但中国对其限制的 23 个半导体品类的进口需提前获得经济产业省的单独批准。

荷兰政府议会则在 2022 年 3 月发表公告，将于 6 月公布并实施先进半导体生产设备出口管制措施。此前，荷兰已于 2019 年限制 ASML 出售 EUV 设备给中国大陆。在新声明中，出口监管措施将侧重于先进的芯片制造技术。一旦我国同时被荷、日两国限制进口，我国 14nm 及以下芯片的生产或将受影响。为避免半导体发展掣肘于海外，实现上游产业链自主可控是关键，目前材料虽未受限制，但为分散海外供应链风险，形成产业联动，国产替代紧迫性不断提高，半导体设备作为半导体重要的“卡脖子”环节，国产化推进刻不容缓，而零部件作为半导体设备发展的基石，国产替代也迫在眉睫。

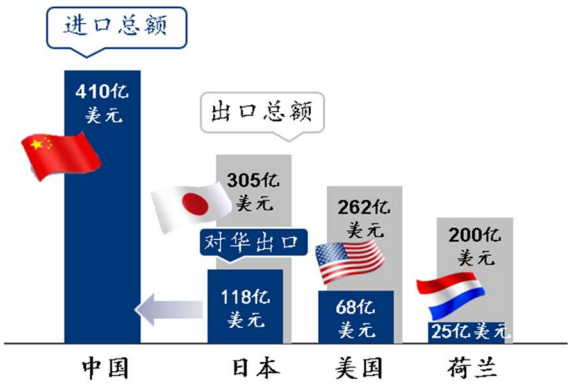
图表209：美国禁令限制中国半导体行业发展



资料来源：各法案公告，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表210：2021 年中国半导体设备进口总额与日本、美国、荷兰半导体设备对华出口对比



资料来源：国际贸易中心，华泰研究

当前国产半导体制造仅在部分环节突破，下一步需走向全面化，高端化

SEMI 公布 2022 年全球半导体设备市场规模约 1076 亿美元，中国半导体设备市场规模达到 283 亿美元。我们估计 2022 年国内半导体设备市场国产化率在 17.3%左右。当前干法刻蚀、清洗、去胶设备等均已实现较高国产化率，CMP、薄膜沉积、量测等设备成熟制程均有产品推出。国内尚未获得突破的设备主要为光刻设备，28nm 及以下薄膜沉积、刻蚀、量测检测、离子注入和涂胶显影等环节仅覆盖部分步骤，14nm 及以下设备仅刻蚀、清洗、氧化发展较快，先进制程为设备国产化的下一个角力点。

根据 SEMI 数据，我们测算 2022 年全球半导体设备厂商所需零部件的市场规模 517 亿美元，同比增长 5.9%。目前我国半导体零部件产业尚处于起步期，核心零部件仍然依赖进口。根据芯谋研究，目前石英、喷淋头、边缘环等零部件国产化率达到 10%以上，射频发生器、MFC、机械臂等零部件的国产化率在 1%-5%，而阀门、静电吸盘、测量仪表等零部件的国产化率不足 1%。此前国产设备厂商为了保障自身设备稳定性，较少尝试国产零部件，新冠疫情造成部分零部件产能紧张，国外厂商零部件交期延迟，为国产替代带来契机，同时美国出口管制加剧化，半导体设备零部件国产化重要性凸显。

SEMI 公布 2022 年全球半导体材料市场规模达到 727 亿美元，中国大陆地区半导体材料规模 130 亿美元，占全球总规模的 17.8%。半导体材料种类繁多，竞争格局分散，目前整体国产化率为 20%左右。具体品类来看，其中电子特气、湿化学品的壁垒主要在于精度/纯度，国产公司已经能在部分产品做到国际标准水平，国产化率约 30%以上；8 寸硅片、靶材、CMP 耗材对加工的工艺要求比较高，国产化率约在 20-30%。但在高端领域，如 12 寸硅片国产化率不足 10%；替换成本较高，涉及前后道工序的光掩模版、光刻胶国产化率在 10%以下，尤其 ArF 光刻胶目前仍未实现国产替代量产突破。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表211：中国半导体设备、材料、零部件国产化情况（2022A）

全球市场规模（亿美元）	国产化率	半导体零部件	全球市场规模（亿美元）	国产化率	半导体材料	全球市场规模（亿美元）	国产化率
42	30%~35%	机械类（钣金件/金属件/腔体/硅电极/喷淋头/石英件/静电吸盘/密封圈）	123	>50%	半导体硅片	133	6寸：>50% 8寸：20%-30% 12寸：<10%
20	20%	电气类（可编程电源/射频电源）	62	<10%	光掩膜	53	I/G线：30% KrF：<10% ArF/EUV：~0%
231	~0%				光刻胶	24	~0%
34	~3%				光刻胶配套试剂	28	~0%
200	20%~25%	机电一体类（废弃处理装置/温控装置/机械手/EFEM/双工台/浸液系统）	82	10-50%	湿化学品	16	30%~40%
7	45%~50%	仪器仪表类（MFC）	10	10-50%	电子特气	57	30%~40%
22	5%	气/液/真空系统类（管路/泵/过滤器/气体输送系统/阀/真空规）	92	10-50%	溅射靶材	12	20%
CVD：135 PVD：36	CVD：5%~10% PVD：10%左右	光学类（光源/物镜系统/曝光系统/光场匀化器）	82	N.A.	CMP耗材	28	20%~30%
28	20%~25%	机电一体类（废弃处理装置/温控装置/机械手/EFEM/双工台/浸液系统）	82	10-50%	湿化学品	16	30%~40%
1076	**~17.3%**	-	**517**	**~10%**	-	**727**	**~20%**

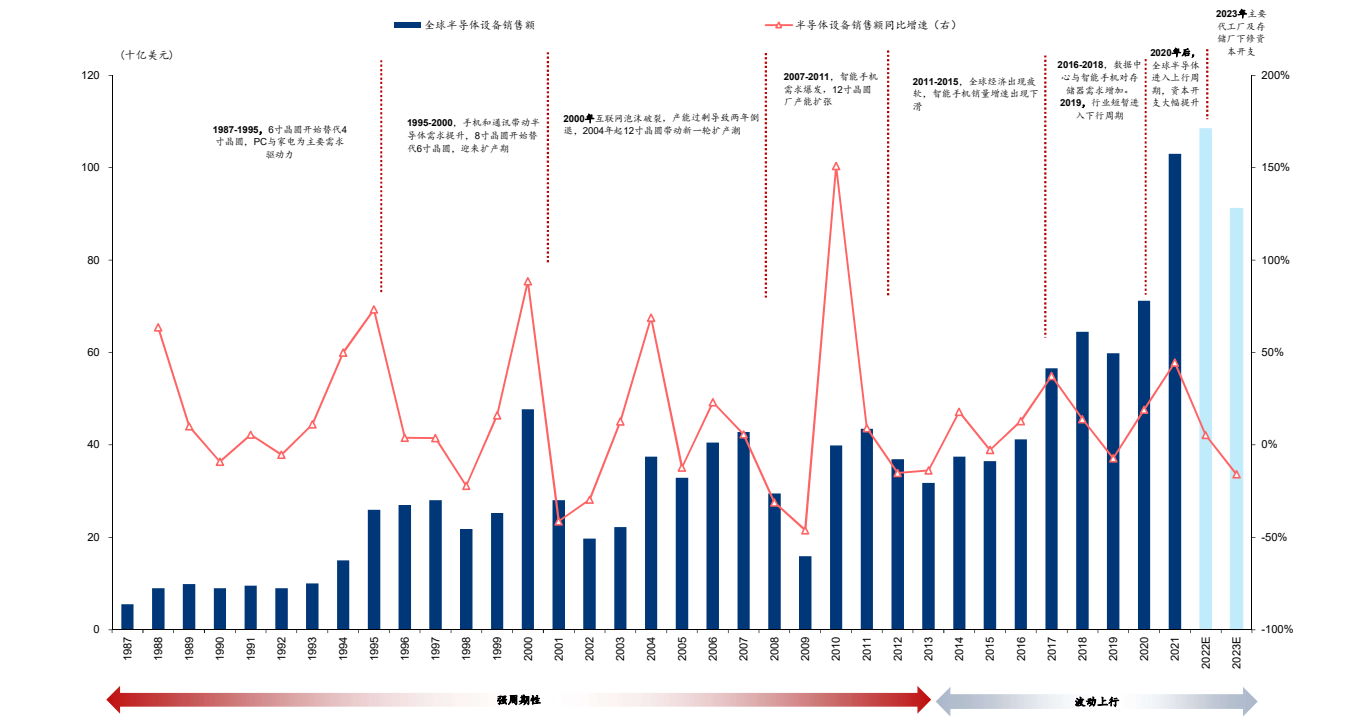
资料来源：Gartner，SEMI，中国国际招标网，富创精密招股说明书，盛美公司招股说明书，华泰研究

机会#1 半导体设备：关注先进制程突破以及国产化率提升

2023 年全球半导体设备市场规模下滑，国产厂商增速优于海外龙头

SEMI 预计 2023 年全球半导体设备市场规模同比下滑 16%至 912 亿美元，而 2024 年恢复成长 17.6%至 1071.6 亿美元。半导体设备行业的核心增长驱动力是下游晶圆厂的资本开支，由于半导体行业的强周期属性，晶圆厂的扩产节奏及资本开支同样也呈现较强的周期性特征，带动半导体设备行业呈现一定的周期性。近年来，随着先进制程投资比例逐步加大，以及本地建厂的趋势兴起，全球半导体设备市场的周期性有所减弱，呈现波动上行的趋势。在经历了 2020-21 年高昂的资本开支后，SEMI 数据表明 2022 年全球半导体设备销售额为 1076.5 亿美元，同比增长 4.9%。其中，中国大陆销售额为 282.7 亿美元。但受下游需求及宏观经济转弱影响，半导体自 2H22 进入下行周期，台积电、美光等半导体制造商宣布削减 2023 年资本开支，SEMI 预计 2023 年全球半导体设备市场规模将同比减少 16.0%至 912 亿美元，而 2024 年恢复成长 17.6%至 1071.6 亿美元。

图表212：全球半导体设备市场规模

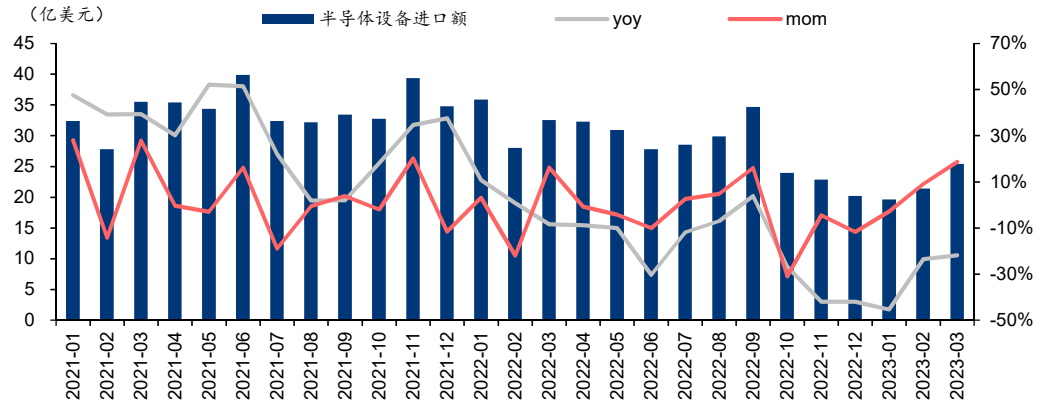


资料来源：Gartner，SEMI,华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

半导体设备厂商在 1Q23 业绩会上预测 2023 全年 WFE 约 750 亿美元。随着景气下行，半导体制造巨头相继削减 23 年资本开支，同时由于美国对华出口管制新规，各设备厂商面临不同程度的收入压力，中国区收入占比高的厂商后续受影响尤其严重。LamResearch 特别提到了其预估 2023 年的 WFE（晶圆厂设备）市场规模将从 2022 年的 950 亿美元衰退至 750 亿美元，约为 21%跌幅，这与 SEMI 预估的-16.9%大致吻合。产业面临经济萧条的冲击，加上美国近期针对华制订的新管制，导致新技术无法输出中国大陆。据 数据，受美国禁令影响，2023 年 3 月中国半导体制造设备进口额显著下滑至 25.4 亿美元（同比-21.9%，环比+18.7%）。

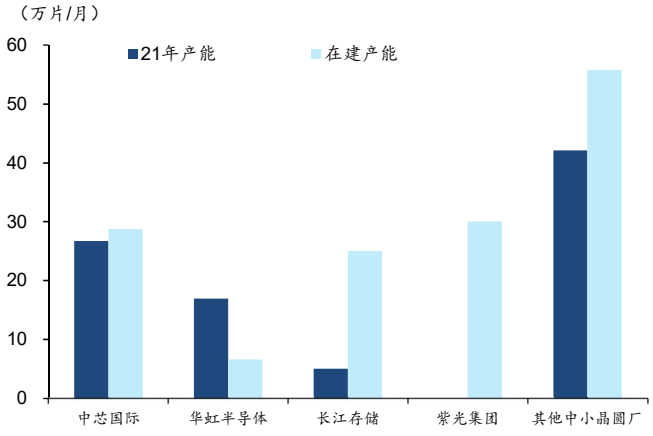
图表213： 中国半导体制造设备进口金额及增长情况



资料来源：，华泰研究

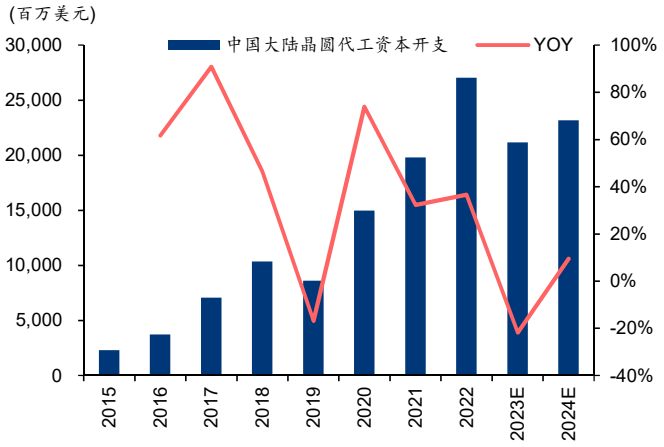
本土晶圆厂持续扩产，国产设备厂商迎来拓宽与加深客户矩阵机遇。不同于全球半导体资本开支放缓趋势，本土晶圆厂产能仍持续扩张，进而拉升对上游设备厂商的需求。根据各公司公开披露数据，我们测算国内内资晶圆厂/IDM/存储厂 2022 年资本开支从 2021 年的 198 亿美元提升至 270 亿美元，同比增长 36.6%，较 2021 年同比增速提升 4.4pct。在 2023 年，虽然中芯国际，华虹等会随着经济周期放慢资本开支步伐，但 2023 年中国大陆地区半导体设备资本开支仍有望超过 200 亿美元，同时在内因外因的推动下，看好设备国产化率持续提升。在国内半导体行业快速发展和国家政策的双重驱动下，国内半导体设备厂商一方面不断扩大产品种类，逐步打破国外厂商垄断；另一方面，稳步提升产品性能，逐步向中高端市场渗透。国内市场表现明显优于全球。

图表214： 22-23 年中国内资晶圆厂整体逆势扩张（等效 12 英寸）



资料来源：ittbank，Gartner，华泰研究

图表215： 中国内资晶圆厂/IDM/存储厂 2020-2024E 资本开支



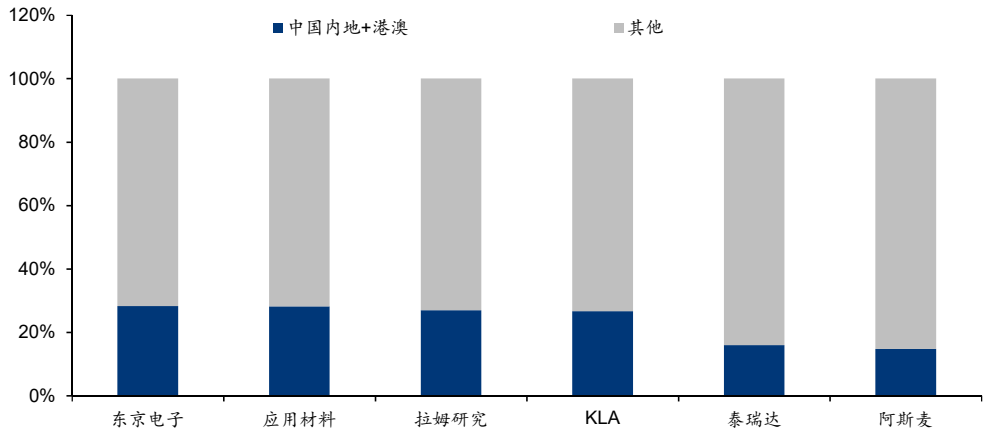
资料来源：各公司公告，华泰研究预测

更多免费行业报告 并购家 www.ipoipo.cn

美对华新规对中国区收入占比较大的海外设备厂商具有较大不利影响

2022 年 10 月 7 日美国商务部宣布修订《出口管理条例》，ASML 和 LAM 对此表态不一，ASML 表示没有直接影响 23 年发货，或间接影响公司 5%积压订单，但这些订单是符合美国出口管制条例的；而 LAM 则表示或因此减少 20～25 亿美元收入，并明确向中国销售设备需购买主体和相应技术均不受限。科磊集团对其大陆地区的业务前景持悲观态度，预估 2023 年全球营收损失 6-9 亿美元。新规影响不同，或因此前 ASML 已不得向中国销售 EUV，此次并未收紧美国对光刻机的出口管制，ASML 可继续向中国晶圆厂销售 DUV，而主要限制了除光刻机以外的其他半导体设备。此外，本次新规主要限制了中国存储厂商，LAM 存储收入占比 52%，受影响较大。

图表216： 半导体设备公司中国内地+港澳收入占比



注：采用各公司 1Q23 数据

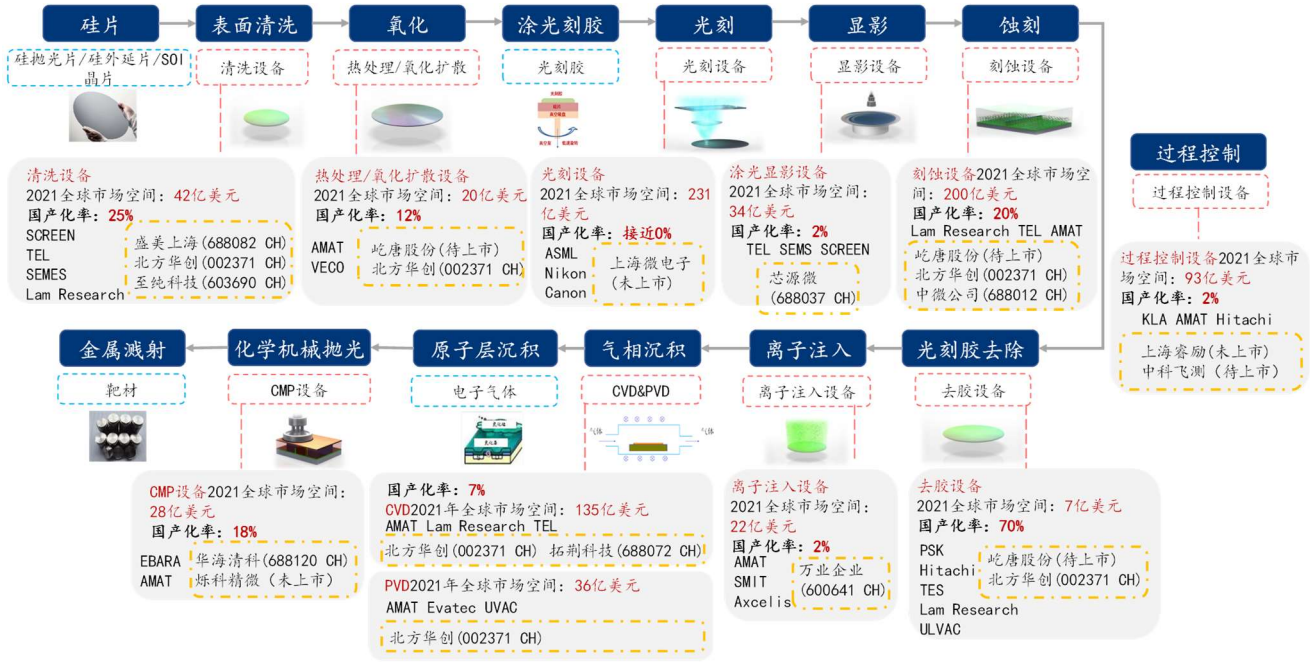
资料来源：各公司财报，DIGTIMES，华泰研究

本土半导体设备公司已进入多个细分领域，但国产替代进程仍处于早期

当前重要的本土半导体设备公司产品已涵盖半导体全产业链，但各环节国产替代进程存在较大差异。2022 年国产设备公司收入约占全球 5%，国产化率约 18.5%。当前中国半导体设备产品主要集中在去胶设备/清洗设备/刻蚀设备/化学机械抛光设备/氧化及热处理设备，根据中国半导体设备招标采购份额计算国产化率，上述环节设备 2021 年国产化率分别为 70%/25%/20%/18%/12%。此外，薄膜沉积设备实现小部分国产化，国产化率达 7%左右。而对于工艺相对复杂的道次所需的半导体设备（如光刻机、过程控制设备、离子注入设备和涂胶显影设备）主要依赖进口，国产化率仍处于低位，2021 年光刻机领域国产化率几乎为 0%，离子注入、涂胶显影和过程控制设备自给率仅为 2%左右。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表217：2021年半导体制造设备产业链地图



资料来源：Gartner，SEMI,华泰研究

根据各个环节所需设备的市场空间及国产化率，我们将目前的半导体设备国产化进程划分为三大环节：国产替代程度较高的环节、正在实现国产替代的环节、易被“卡脖子”的亟待突破的重点环节。

去胶设备及清洗设备国产替代程度较高，屹唐半导体、盛美上海等占据领先地位

去胶设备：根据 Gartner，2021 年去胶设备的全球市场空间为 7 亿美元，国内市场空间为 2 亿美元，**国产化率高达 70%**。目前去胶设备工艺以干法去胶为主，其中屹唐半导体为干法去胶设备全球龙头企业，2021 年全球市占率达 31%，位列第一，比思科、日立高新、LamResearch 为其他主要玩家。根据中国国际招标网数据统计的去胶设备竞争格局，屹唐半导体 2021 年的国内市占率为 81%。根据公司招股说明书，屹唐半导体的干法去胶设备可用于 90nm 到 5nm 逻辑芯片、1y 到 2xnm 系列 DRAM 芯片以及 32 层到 128 层 3D 闪存芯片制造中若干关键步骤的大规模量产。

清洗设备：根据 Gartner，2021 年清洗设备全球市场空间达 42 亿美元，国内市场空间达 14 亿美元，目前我国半导体清洗设备的**国产化率已达 25%**，我们预计到 2030 年有望翻倍。全球半导体清洗设备行业高度集中，龙头企业主要是迪恩士、TEL、韩国 SEMES、LamResearch 等等。近年来盛美上海、至纯科技、北方华创等国产设备公司不断在单片、槽式等清洗设备上有所突破，根据公司年报，盛美上海 14nm 清洗设备正处于客户验证中。至纯科技目前产品可以满足 28nm 全部湿法工艺需求，根据公司年报，2022 年突破 14nm 及以下制程并已交付 4 台设备。根据公司年报，芯源微单片式湿法设备已具备 40nm 制程；根据公司官网，北方华创已具备 28nm 制程技术。盛美上海 2021 年全球市占率达 2.3%。

国产设备有望在刻蚀、CMP、薄膜沉积等领域率先突破

刻蚀设备：根据 Gartner，2021 年刻蚀设备全球市场空间达 200 亿美元，由 LamResearch，TEL，AMAT 三巨头主导市场，国内市场空间为 58 亿美元，主要玩家包括北方华创、中微公司等，目前**国产化率已达 20%**。北方华创以硅刻蚀、金属刻蚀为主，ICP 刻蚀机领域国内领先，14nm 等离子硅刻蚀机在客户端通过多道制程工艺验证，已实现量产应用；用于存储器件的高深宽比刻蚀机已完成研发，进入工艺测试阶段，各项技术指标已基本满足客户要求。中微半导体以介质刻蚀为主，并开始获得台积电及长江存储等公司的刻蚀设备订单。目前国内刻蚀设备虽已实现一定的国产替代，但开发能够处理这些工艺的干式蚀刻技术和原子层刻蚀（ALE）设备成为当务之急。

免责
更多免费行业报告 并购家 www.ipoipo.cn

CMP 设备：CMP 在半导体前道设备中占比较低，仅为 3%。根据 Gartner，2021 年 CMP 设备全球市场空间为 28 亿美元，AMAT 联合日本荏原合计占据了全球 90%以上的市场空间。华海清科为国产 CMP 设备龙头，21 年**国产化率达 18%**，根据公司官网，目前在国内成熟制程领域，华海清科 28nm 及以上制程的主要 CMP 设备与国际龙头已不存在明显技术差距；而在先进制程领域，美国 AMAT 与日本荏原分别已实现 5nm 制程和部分材质 5nm 制程的工艺应用，华海清科的 14nm 制程工艺技术在研。

热处理/氧化扩散设备：根据 Gartner，2021 年全球市场规模达 20 亿美元，AMAT 垄断近 70%的份额，屹唐半导体与北方华创分别占比 12%/2%，**21 年国产化率达 12%**。根据公司招股书，屹唐半导体的快速热处理设备达到国际领先水平，技术突破至 28nm，已覆盖台积电、三星电子、中芯国际、华虹集团、长江存储等知名厂商。

薄膜沉积设备：根据 Gartner，2021 年薄膜沉积设备全球市场空间达 190 亿美元，国内市场空间达 50 亿美元，**21 年国产化率达 7%**。薄膜沉积设备是目前半导体设备国产化的重要突破口，国产替代空间广阔，主要包括 CVD/PVD/ALD 三大类，全球市场中 AMAT/LAM/TEL 占据主导，根据公司招股书，国内企业中拓荆科技的 PECVD 设备已适配 180-14nm 逻辑芯片、19/17nmDRAM 及 64/128 层 FLASH 制造工艺，SACVD 设备适配 12 英寸 40/28nm 及 8 英寸 90nm 以上的逻辑芯片制造工艺，ALD 设备也已适配 55-14nm 逻辑芯片制造工艺需求，基本达到国际厂商设备水准；根据公司官网，北方华创在 PVD 领域实现了国产高端薄膜制备设备零的突破，28nmHardmaskPVD、Al-PadPVD 设备已率先进入国际供应链体系，覆盖 90-28nm 多个制程。

涂胶显影设备：涂胶显影设备是配合光刻机工作的核心工艺设备，根据 Gartner，2021 年全球市场规模达 34 亿美元，长期被 TEL 垄断，**国产化率为 2%处于低位**。目前芯源微是国内唯一可提供前道高端量产涂胶显影设备的厂商，尚处于产业化初期，2021 年全球市占率为 2%。根据公司年报，2022 年前道涂胶显影机 offline、I-line、KrF 机台均实现批量销售；2022 年第四季度，随着公司前道涂胶显影机 ArFi 机台的推出，标志着公司前道涂胶显影设备已完成在晶圆加工环节 28nm 及以上工艺节点全覆盖，并可持续向更高工艺等级迭代，有望逐步实现对海外设备的替代。

离子注入设备：根据 Gartner，2021 年全球市场规模为 22 亿美元，被应用材料、Axcelis 垄断，国产设备仍相对空缺处于快速起步期，**国产化率仅为 2%**。根据公司年报，万业企业旗下的凯世通制造的 IC 离子注入机迎来商业客户及多款设备订单的重大突破，根据公司官网电科装备旗下的烁科中科信的中束流离子注入机已经在 12 英寸和 8 英寸产线量产或 demo 验证超过数十台套，其工艺稳定性获得业内客户高度认可，其大束流离子注入机已实现多台销售，正在陆续进入客户端，2022 年 5 月,烁科中科信宣布特种离子注入机顺利交付。

光刻机及过程控制设备是易被“卡脖子”的亟待突破的重点环节

光刻设备：光刻机是芯片制造与光刻胶研发中的核心工艺设备，根据 Gartner，2021 年全球市场份额高达 231 亿美元，ASML 市占率达 75%，是全球唯一销售高端光刻机-EUV 的厂商。国内市场规模达 67 亿美元，但**国产化率仍接近于零**，极大程度上依赖海外光刻设备。在国产光刻机领域中，根据公司官网，上海微电子产品主要采用 ArF、KrF 和 i-line 光源，目前只能达到 90nm 制程量产，且主要用于 IC 的后道封装和面板领域。

免责

更多免费行业报告 并购家 www.ipoipo.cn

过程控制设备：贯穿于集成电路领域生产过程的质量控制环节可分为前道检测、中道检测和后道测试，检测和量测环节贯穿制造全过程，是保证芯片生产良品率非常关键的环节。根据 Gartner，2021 年全球市场规模为 93 亿美元，竞争格局相对集中，全球范围内主要检测和量测设备企业包括 KLA、AMAT、日立等。国内**设备国产化率较低仅为 2%左右**，目前国内主要企业中，根据中微公司年报，上海睿励自主研发的 12 英寸光学测量设备 TFX3000 系列产品，已应用在 65/55/40/28nm 芯片生产线并进行 14nm 工艺验证，在 3D 存储芯片产线支持 64 层 3DNAND 芯片的生产，并正在验证 96 层 3DNAND 芯片的测量性能。根据公司招股书，中科飞测目前也能够提供 28nm 制程的相关过程控制设备，填补了国内先进封装检测设备在高端市场的空白。此外，根据公司年报，精测电子的膜厚设备、OCD 设备以及电子束设备均已取得多家客户批量性订单，半导体硅片应力测量设备取得客户订单并完成交付，明场光学缺陷检测设备已取得突破性订单，且完成首台套交付。但国内企业目前在过程控制设备产品线上的覆盖广度与深度仍与国际龙头存在较大差距，仍存在较大的国产替代空间。

图表218：半导体设备主要公司及市场空间

公司名称	公司代码	是否上市	成立时间	2022 年营业收入（亿美元）	产品空间（全球，亿美元）	产品种类
盛美上海	688082 CH	科创板	2005	4.1 亿	42 亿	清洗设备、电镀设备、先进封装湿法设备、立式炉设备、化学气相沉积 PECVD 设备、前道涂胶显影设备
屹唐股份	-	未上市	2015	2.2 亿（1H21）	227 亿	干法去胶设备、快速热处理设备、干法刻蚀设备
北方华创	002371 CH	主板	2001	17.4 亿（电子装备业务）	440 亿	等离子刻蚀设备、物理气相沉积设备 PVD、化学气相沉积设备 CVD、原子层沉积设备 ALD、氧化扩散设备、清洗设备、气体量测控制等
上海微电子	-	未上市	2002	-	26.4 亿	产品涵盖半导体光刻机、AOI 晶圆自动光学检测设备等半导体装备
至纯科技	603690 CH	主板	2000	4.4 亿	76 亿	湿法清洗设备、炉管、涂胶显影设备
芯源微	688037 CH	科创板	2002	2.0 亿	283 亿	涂胶显影设备、清洗设备、去胶设备、刻蚀设备
万业企业	600641 CH	主板	1991	0.3 亿（专用设备制造）	22 亿	2018 年收购上海凯世通，进入离子注入机领域；2021 年成立嘉芯半导体，产品覆盖刻蚀、薄膜沉积、快速热处理等多款集成电路核心前道设备
中微公司	688012 CH	科创板	2001	5.5 亿（专用设备）	225 亿	CCP 高能等离子体刻蚀设备、ICP 低能等离子体刻蚀设备、MOCVD 金属有机化学气相沉积设备、环保设备
拓荆科技	688072 CH	科创板	2010	2.4 亿	135 亿	等离子体增强化学气相沉积（PECVD）设备、原子层沉积（ALD）设备、化学气相沉积（SACVD）设备
华海清科	688120 CH	科创板	2013	2.4 亿	28 亿	CMP 设备、减薄设备、金属薄膜厚度测量设备
烁科精微	-	未上市	2019	-	28 亿	聚焦集成电路高端装备 CMP 设备
上海睿励	-	未上市	2005	0.1 亿	93 亿	关键薄膜测量设备、光学关键尺寸测量设备、缺陷检测设备
中科飞测	688361 CH	科创板	2014	0.7 亿	93 亿	无图形晶圆缺陷检测设备、图形晶圆缺陷检测设备、套刻精度量测设备、晶圆介质薄膜量测设备、三维形貌量测设备
精测电子	300567 CH	创业板	2018	0.3 亿（半导体）	93 亿	膜厚量测系统、光学关键尺寸量测（OCD）系统、电子束缺陷检测系统、形貌测量、电子束关键尺寸量测系统
中安半导体	-	未上市	2020	-	93 亿	从事半导体生产线精密检测设备的研发和生产
御微半导体	-	未上市	2019	-	93 亿	已经形成了掩模版检测、晶圆检测、泛半导体检测、晶圆测量四大领域六大类量检测产品
华峰测控	688200 CH	科创板	1993	1.5 亿	78 亿	模拟 IC 测试、模拟/混合 IC 测试、分立器件测试、功率模块测试
深科达	688328 CH	科创板	2004	0.8 亿	78 亿	固晶设备、测试分选设备、影像模组封装设备
深圳矽电	-	未上市	2003	0.6 亿	78 亿	分立器件测试、光电芯片测试、曝光机、AOI 检测
联动科技	301369 CH	创业板	1998	0.5 亿	78 亿	后道封装测试领域半导体自动化测试、激光打标设备
宏泰	-	未上市	2007	-	78 亿	三大品牌覆盖半导体行业集成电路测试系统、分立器件测试系统与转塔式测试分选系统三大品类
悦芯	-	未上市	2017	-	78 亿	从事大规模集成电路自动化生产测试设备（ATE）的研发、生产和销售

资料来源：，各公司官网，各公司年报，华泰研究预测

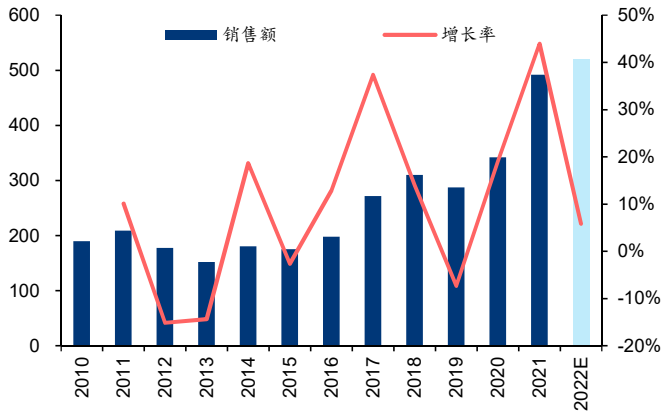
免责
更多免费行业报告 并购家 www.ipoipo.cn

机会#2 半导体设备零部件：国产化地位开始受到重视

半导体精密零部件支撑半导体设备的升级迭代，是半导体产业链中不可或缺的一环。半导体的生产需要经过光刻、刻蚀、离子注入、薄膜沉积、清洗、量测、抛光等一系列步骤，各环节需要不同的制造设备。半导体设备大部分关键核心技术物化于设备精密零部件，因此具备高精密、高洁净、超强耐腐蚀能力、耐击穿电压等特性的精密零部件直接保障了半导体设备的精度和可靠性，其技术突破进而支撑了半导体设备的升级迭代。

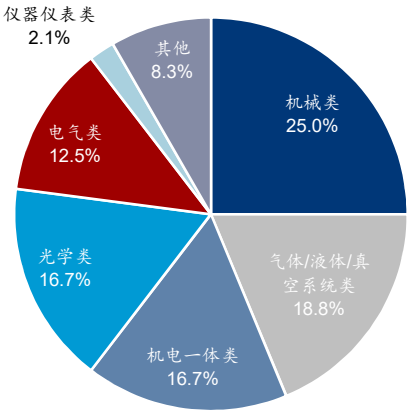
半导体设备零部件市场规模持续上升，我们估算全球 2022 年将达 517 亿美元规模。半导体设备厂商与晶圆厂均对半导体零部件存在一定需求，目前约 90%的需求来自于半导体设备厂商，这是目前半导体零部件的主要市场与最大成长动能。据北方华创、拓荆科技等公司公告，中国半导体设备公司设备业务毛利率在 38%-45%左右，即成本占据设备总销售额的 55%-62%，同时直接材料占比构成了设备总成本中的 88%-94%左右，对应半导体零部件占半导体设备公司总销售额约 48%左右。**SEMI 数据表明 2022 年全球半导体设备销售额为 1076.5 亿美元**，据此我们测算 2022 年全球半导体设备所需精密零部件的市场规模为 517 亿美元。

图表219： 全球半导体设备零部件市场规模



资料来源：SEMI，华泰研究预测

图表220： 2021 年全球市场各类半导体设备零部件占比



资料来源：富创精密招股说明书，华泰研究

半导体零部件种类繁多，机械类占比最高，高端产品技术壁垒高

按产品类别半导体精密零部件具体又分为机械类、电气类、机电一体类、仪器仪表类、气体/液体/真空系统类和光学类六大精密零部件。据富创精密招股书披露，2021 年全球机械类/电气类/机电一体类/仪器仪表类/气、液、真空系统类/光学类零部件市场规模占半导体设备零部件市场比例分别为 25%/13%/17%/2%/19%/17%，市场份额分别为 123/62/82/10/92/82 亿美元，机械类零部件占比最高，市场空间最大。

1）机械类：在设备中起到构建整体框架、基础结构、晶圆反应环境和实现零部件特殊功能的作用，保证反应良率，延长设备使用寿命，具体又分为金属工艺件、金属结构件和非金属机械件。其中金属件技术要求相对较低，国产化程度较高，但应用于高制程设备的产品仍存在较高技术壁垒；而静电卡盘、密封圈等非金属机械件对密封性、真空度等要求较高，国内市场尚处于起步阶段。

2）电气类：在设备中起到控制电力、信号、工艺反应制程的作用，核心模块为射频电源，直接关系到腔体中的等离子体浓度、均匀度和稳定度。目前国产射频电源主要应用于低端工业及泛半导体设备，系电源的波形、频率和功率等参数不够稳定，精度达不到半导体级应用要求。

免责
更多免费行业报告 并购家 www.ipoipo.cn

3）机电一体类：在设备中起到实现晶圆装载、传输、运动控制、温度控制的作用，部分产品包含机械类产品。其中温控系统、废气处理装置等国产化程度较高；EFEM、机械手等产品已实现技术突破，但稳定性及一致性较国外产品存在差距；应用于光刻机的双工台、浸液系统技术壁垒较高，尚未实现突破。

4）仪器仪表类：在设备中起到控制和监控流量、压力、真空度、温度等数值的作用，对响应速度、测量精度及稳定性要求极高。国内厂商通过收购已进入国际半导体设备供应商，但高端产品尚未实现技术突破。

5）气体/液体/真空系统类：在设备中起到传输和控制特种气体、液体和保持真空的作用，对真空度、抗腐蚀性等性能要求较高。其中泵、阀、管路等产品已经实现技术突破，进入国际半导体设备供应商，但气柜模组等高端产品尚未实现技术突破。

6）光学类：在光学设备中起到控制和传输光源的作用，对光学性能要求较高，主要应用于光刻设备和量测设备。国内光刻设备及量测设备均处于发展阶段，技术较国际龙头厂商差距较大。

图表221：各类零部件具体情况

分类	2021 年市场规模占比	具体类别	技术要求	主要应用设备	技术突破难度
机械类	25%	**金属工艺件:**反应腔、传输腔、过渡腔、内衬、匀气盘等 **金属结构件:**托盘、冷却板、底座、铸钢平台等 **非金属机械件:**石英、陶瓷件、硅部件、静电卡盘、橡胶密封件等	满足加工精度、耐腐蚀性、密封性、洁净度、真空度等指标	应用于所有设备	应用最广，市场份额最大，品类繁多，主要产品技术已实现突破和国产替代，应用于高制程设备的产品技术突破难度较高
电气类	13%	射频电源、射频匹配器、远程等离子源、供电系统、工控电脑等	满足输出功率的稳定性、电压质量、波形质量、频率质量等指标	应用于所有设备	控制设备工艺制程的核心部件，技术突破难度较高
机电一体类	17%	EFEM、机械手、加热带、腔体模组、阀体模组、双工机台、浸液系统、温控系统等	满足真空度、洁净度、放气率、SEMI 定制标准等指标，同时保证多次使用后的一致性和稳定性，不同具体产品要求差别较大	应用于所有设备,其中双工机台和浸液系统仅用于光刻设备	品类繁多，部分产品已实现技术突破，但产品稳定性和一致性与国外有差距，技术难度适中
气体/液体/真空系统类	19%	**气体输送系统类:**气柜、气体管路、管路焊接件等	满足真空度、耐腐蚀性、洁净度、SEMI 定制标准等指标	主要应用于薄膜沉积设备、刻蚀设备和离子注入设备等干法设备	品类繁多，部分产品已实现技术突破，但产品稳定性和一致性与国外有差距，技术难度适中
		真空系统类:干泵、分子泵、真空阀门等	满足抽气后的真空指标、可靠性、稳定性、一致性等指标	主要应用于薄膜沉积设备、刻蚀设备和离子注入设备等干法设备	
		气动液压系统类:阀门、接头、过滤器、液体管路等	满足真空度、表面粗糙度、洁净度、使用寿命、耐液体腐蚀等指标	主要应用于化学机械抛光设备、清洗设备等湿法设备	
仪器仪表类	2%	气体流量计、真空压力计等	满足量程时间、流量测量精度、温度测量精度、压力测量精度、温度影响小等指标	应用于所有设备	对测量的精准度要求极高，技术突破难度较高
光学类	17%	光学元件、光栅、激光源、物镜等	满足制造精度、分辨率、曝光能力、光学误差小等指标	主要应用于光刻设备、量测设备等	对光学性能要求极高，国内光刻设备尚在发展，技术突破难度较高
其他	8%	定制装置、耗材等	满足相应设备要求的定制化指标	应用于所有设备	-

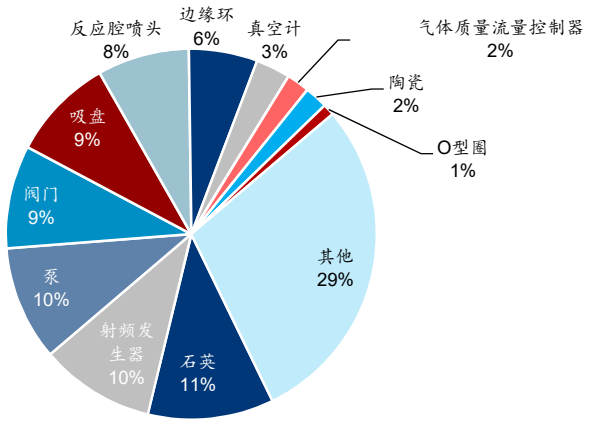
资料来源：富创精密招股说明书，华泰研究

免责

更多免费行业报告 并购家 www.ipoipo.cn

晶圆厂与半导体设备厂商均对半导体零部件存在一定需求。据芯谋研究数据，2020 年中国大陆晶圆厂 8-12 寸晶圆设备零部件直接采购零部件金额超过 10 亿美元，其中中国本土晶圆厂直接采购零部件金额约 4.3 亿美元，晶圆厂具体采购的零部件主要包括石英件、射频发生器、泵、阀、静电吸盘、喷淋头、边缘环、流量计、MFC、陶瓷件、密封圈等。按零部件品类划分，机械类/电气类/机电一体类/仪器仪表类/气、液、真空系统类/光学类零部件分别占比 42%/12%/9%/2%/27%/7%。据 SEMI 数据，中国大陆晶圆厂资本开支占全球份额的 27.4%，我们测算 2021 年全球晶圆厂直接采购零部件金额可达约 37 亿美元，预计晶圆厂持续扩产将进一步催生其对半导体设备零部件的需求，2021 年全球晶圆厂及设备厂合计所需零部件市场规模达 529 亿美元。

图表222：2020 年中国晶圆厂采购 8-12 寸晶圆设备零部件产品结构



资料来源：芯谋研究，华泰研究

半导体零部件市场主要被美日供应商占据，国产厂商主要以机械类为主

国内规模较大的半导体精密零部件厂商以外资控股为主。目前国内规模较大的半导体设备精密零部件厂商主要为台湾地区的京鼎精密和日本 Ferrotec 等外资企业的境内子公司，其主要为国际半导体设备厂商供货。同时，国内设备尤其是先进制程设备零部件采购也主要依赖进口，一方面国内部分零部件在技术上较国际龙头厂商尚有差距，技术突破难度较高；另一方面半导体精密零部件对设计和生产要求较高，国产产品在同等技术水平下的量产稳定性尚有不足，国内设备厂商更愿意选择国际成熟产品。据北京国际咨询对国内主流代工厂调研数据，目前全年日常运营过程中领用的零部件（包括更换和失效更换的零部件）达到 2000 种以上，但国产占有率仅为 8%左右。美国和日本占有率分别为 59.7%和 26.7%。其中高端零部件市场主要被美国、日本、欧洲供应商占有；中低端零部件市场主要被韩国、中国台湾供应商占据。

全球半导体设备零部件龙头均为外企，前十大厂商销售额超 110 亿美元。根据 techinsights 数据，2021 年全球半导体设备零部件十大领军供应商为：Zeiss（光学镜头）、MKS（MFC、射频电源、真空产品）、Edwards（真空泵）、Horiba（MFC）、AdvancedEnergy（射频电源）、VAT（真空阀件）、UltraCleanTech（密封系统）、Ichor（模块化气体输送系统以及其他组件）、Ebara（干泵）和 ASML（光学部件）。以上企业均为外企，且均非纯粹的半导体零部件公司，而是横跨多个行业，将半导体零部件供应作为其业务线之一。其中涉及电气类/仪器仪表类/气、液、真空系统类/光学类零部件的企业分别有 2/2/6/2 家，除 MKS 外均只供应一类零部件，系不同零部件领域壁垒较高，厂商难以全面涉足。对比之下，中国前十大半导体设备零部件供应商零部件收入显著低于全球前十家，且全球市占率总和仅为 1.21%。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表223：全球前十大半导体设备零部件供应商与中国前十大半导体设备零部件供应商对比（2021A）

国外公司	所在国家	主要产品	半导体/泛半导体零部件收入规模（亿美元）	全球占有率	中国公司	零部件收入（亿美元）	全球占有率	中国占有率
Zeiss	德国	物镜系统	~30.7	6.24%	万业企业	1.4	0.28%	1.05%
MKS	美国	MFC、射频电源、真空产品	~19.9	4.05%	富创精密	1.3	0.27%	0.50%
Edwards	英国	真空泵	~18.2	3.70%	新莱应材	0.8	0.17%	0.63%
Horiba	日本	MFC	~8.8	1.79%	神工股份	0.7	0.15%	0.56%
Advanced Energy	美国	射频电源	~8.6	1.75%	华亚智能	0.4	0.08%	0.31%
VAT	瑞士	真空阀件	~7.2	1.46%	江丰电子	0.3	0.06%	0.22%
Ultra Clean Tech	美国	真空阀件	~6.9	1.40%	机器人	0.3	0.05%	0.19%
Ichor	美国	模块化气体输送系统及其他组件	~6	1.22%	凯德石英	0.2	0.04%	0.15%
ASML	荷兰	光学部件及光刻机服务组件	~5	1.01%	伟泰科技	0.3	0.06%	0.21%
Ebara	日本	干泵	~5	1.01%	盛剑环境	0.3	0.05%	0.19%
合计	-	-	~116.3	23.63%	合计	6.0	1.21%	4.01%

注：富创国内收入按照 50%调整，汇率按照 2021 年历史汇率美元对人民币 6.38 计算

资料来源：techinsights，公司年报(中国公司部分)，华泰研究

整体上看，国内设备厂商目前在机械类零部件方面已经进行了一定程度的国产化，但在一些技术壁垒较高的领域国产化率仍然偏低：

图表224：各品类半导体零部件主要玩家及国产化率（2021 年）

资料来源：富创精密招股说明书，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表225：半导体零部件主要公司及市场空间

产品类别	公司名称	公司代码	是否上市	成立时间	2022 年营业收入（亿美元）	产品空间（全球，亿美元）	产品种类
机械类	菲利华	300395 CH	创业板	1999	2.47	123	主要产品包括高性能石英玻璃材料及制品，石英玻璃纤维和以石英玻璃纤维为基材的复合材料及制品。
	江丰电子	688409 CH	科创板	2005	3.34	123	主要产品包括传输腔体、反应腔体、腔体、圆环类组件、腔体遮蔽件、保护盘体、冷却盘体、加热盘体、气体分配盘、气体缓冲盘等
	珂玛科技	A22388 CH	（未上市）	2009	0.33（1H22）	123	产品包括静电卡盘、陶瓷加热器等
	神工股份	688233 CH	科创板	2013	0.77	123	产品包括硅零部件。
	华亚智能	003043 CH	主板	1998	0.88	123	主要产品为高端精密金属零部件。
机械类、气/液/真空系统类	新莱应材	300260 CH	创业板	2000	1.02（电子洁净）	215	主要产品包括高真空和超高真空的法兰、管件、钢瓶、传输阀、铝合金与不锈钢闸阀、角阀、腔室、加热带、视窗、无氧铜垫片等
机械类、机电一体类、气/液/真空系统类	富创精密	688409 CH	科创板	2008	1.86（半导体）	297	主要产品包括工艺零部件、结构零部件、模组产品和气体管路。
电气类	英杰电气	300820 CH	创业板	1996	0.28（半导体等电子材料）	62	主要产品包括系列功率控制器、功率控制电源系统（还原炉电源、单晶炉电源、蓝宝石炉电源、碳化硅炉电源等）、特种电源（直流编程电源、加速器电源、中高频感应电源、高压电源、微波电源、射频电源、固态调制器系统等）和新能源汽车充电桩电源模块及充电桩/站。
	国力股份	688103 CH	科创板	2000	0.98（真空电子）	62	产品型号众多，主要产品包括真空无源器件（继电器、电路开关、电容器）和真空有源器件（电子管）。
电气类、仪器仪表类、气/液/真空系统类	北方华创	002371 CH	主板	2001	3.70（电子元器件）	164	在真空及锂电装备领域，产品包括有晶体生长设备、真空热处理设备、气氛保护热处理设备、连续式热处理设备、磁控溅射镀膜设备、多弧离子镀膜设备；在精密电子元器件领域，产品包括石英晶体器件、石英微机电传感器、高精密电阻器、钽电容器、微波组件、模拟芯片、模块电源等产品。
机电一体类	沈阳新松	300024 CH	创业板	2000	0.35（半导体装备）	82	主要产品包括工业机器人、物流与仓储自动化成套装备、自动化装配与检测生产线、交通自动化系统、半导体装备等。
机电一体类、机械类	华卓精科		（未上市）	2012	0.16（1H22）	205	主要零部件产品包括运动平台、EFEM、静电卡盘、隔振器等。
气/液/真空系统类、仪器仪表类	万业企业(Compart)	600641 CH	主板	1991	0.30（专用设备制造）	102	收购全球领先的零部件公司 CompartSystems，主要产品包括气箱中的高价值气体输送零部件、焊件、密封件、气棒、质量流量控制器（MFC）等。
气/液/真空系统类	中科仪	830852 CH	新三板	2001	1.00	92	产品主要包括干式真空泵和多种真空部件。
	正帆科技	688596 CH	科创板	2009	3.39	92	主要产品包括特气柜、化学品中央供应柜、分流箱、液态源输送设备、GasBox 等。
	汉钟精机	002158 CH	主板	1998	4.69	92	主要产品为 R 系列制冷压缩机、A 系列空气压缩机组和机体、L 系列冷冻冷藏压缩机组及 P 系列真空产品等。
光学	炬光科技	688167 CH	科创板	2007	0.79	82	收入主要来源于上游产业，产品包括上游半导体激光元器件（有源器件、光纤耦合模块及无源器件、专业医疗健康应用元器件）、激光光学元器件（光束准直转换系列：单（非）球面柱面透镜、光束转换器、光束准直器、光纤耦合器；光场匀化器、光束扩散器、微光学晶圆等）和中游汽车应用模块（激光雷达面光源发射模组、激光雷达线光源发射模组、激光雷达光源光学组件等）。
	国科精密	-	（未上市）	2014	-	82	主要产品包括复杂光学系统（曝光光学系统）、日盲紫外探测模组、高端光学检测产品（标准镜头）

资料来源：SEMI，各公司年报，各公司官网，华泰研究预测

免责

更多免费行业报告 并购家 www.ipoipo.cn

机会#3 材料：国产基础较好，掩模版、光刻胶国产化率较低

半导体材料是半导体发展的基石，贯穿生产的全过程。在晶圆制造过程中，需要使用硅片、电子气体、光刻胶及其配套试剂、掩膜版、湿电子化学品、CMP 耗材（抛光垫、抛光液等）以及靶材。硅片是晶圆制造的第一步，是晶圆的基底材料；电子气体几乎贯穿晶圆制造的全流程，在清洗、氧化、光刻、刻蚀、离子注入、沉积中均发挥着重要作用；掩膜版是生产的母版，发挥着模具的作用；光刻胶则用于将掩膜版的图案转移到硅片上；湿电子化学品又称超净高纯试剂或化学工艺品，使用于清洗、显影、刻蚀、去胶等多个环节中；CMP 耗材是集成电路制造过程中实现晶圆全局均匀平坦化的关键；靶材是利用溅射法制备电子薄膜的关键原材料，在形成具有特定功能的金属层（如导电层、阻挡层）中发挥重要作用。在封装测试过程中，需要使用封装基板、引线框架、键合丝、包封材料、陶瓷基板、芯片粘接材料等。

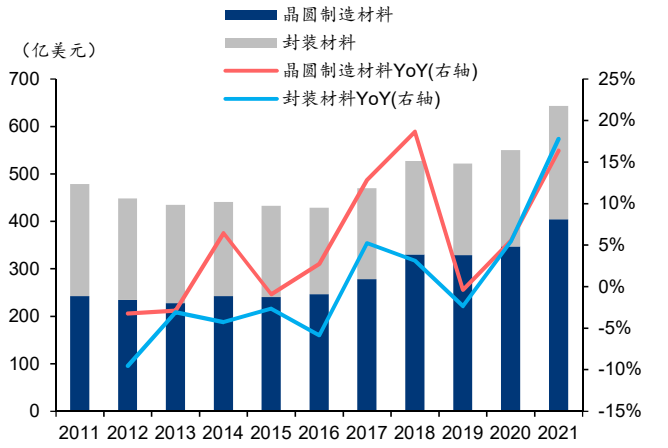
图表226： 晶圆制造流程和封装测试流程中需要的半导体材料

资料来源：SEMI，三星官网，华泰研究

半导体材料市场规模超 600 亿元，晶圆制造材料占比 63%。根据 SEMI 数据，2022 年全球半导体材料市场规模达到 698 亿美元。中国大陆地区半导体材料规模 119 亿美元，占全球总规模的 19.2%。其中晶圆制造材料市场规模在半导体材料中的占比逐年提高，从 2011 年的 51%逐步提升并稳定在 63%左右，主要系制程越先进、存储容量要求越高，需要的晶圆制造每步骤的重复次数就越多，所需材料也就越多，其中，硅片、湿化学品、CMP 和光掩模领域在晶圆制造材料市场中表现出最强劲的增长，而封装材料市场的增长主要受有机基板、引线框架和键合线领域的推动。从晶圆制造材料占比来看，2020 年占比最高的为硅片（33%），其次是电子特气（14%），光刻胶及其配套试剂、光掩膜版占比为 13%，CMP 耗材、湿化学品以及靶材占比不足 10%。

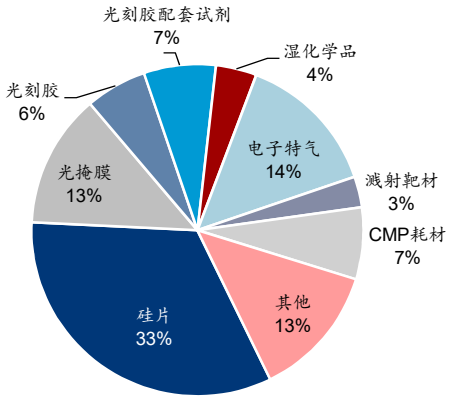
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表227：2011-2021 全球半导体材料市场规模及增长率



资料来源：SEMI，华泰研究

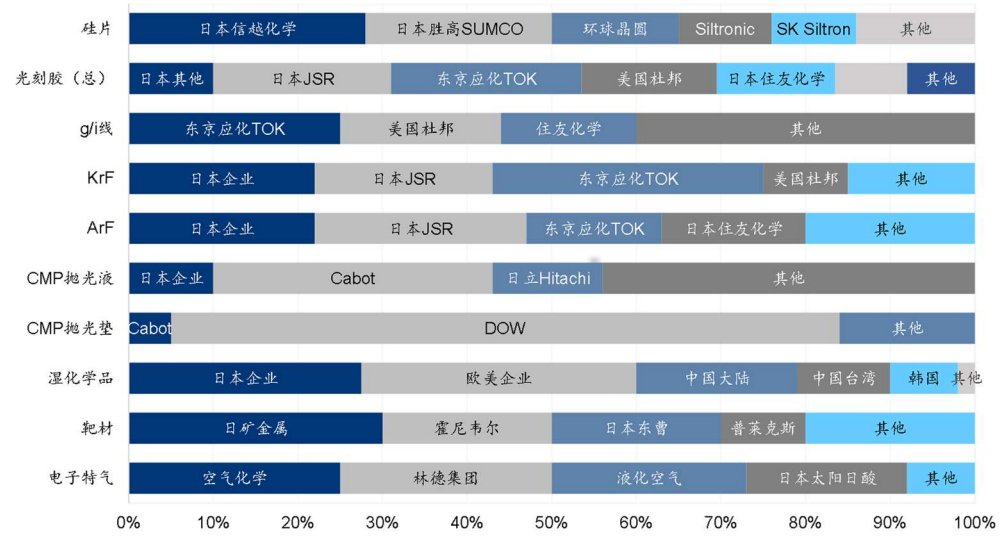
图表228：晶圆制造材料细分占比（2020A）



资料来源：SEMI，华泰研究

日本在光刻胶、CMP 粉浆、各类高纯度溶液等产品中市占率高，日本企业在化学领域（而非物理）优势较大。日本主要在液体（或者气体、流体）等相关的材料以及加热后凝固的材料的占比较高，其在半导体材料领域的强势离不开日本的工匠精神。这些材料最初的形状不是固定的，质地柔软，优化的标准非常多且复杂，需要基于多年的经验和直觉得出最合适的见解，在这些产品中有很多无法实现标准化的隐性知识和技巧，需要在车间里不断地改善和改良，以自下而上（BottomUp）的方式进行研发。

图表229：日本在半导体材料领域具有领先地位（2021A）



注：图中柱状图表明各公司市占率
资料来源：EETimes，华泰研究

低端半导体材料国产化率相对较高，部分高端产品出现突破

6 寸硅片已实现国产替代，8 寸国产替代正在进行中，12 寸进口依赖度依旧较高；光刻胶各品类国产化进程不一，其中在 PCB 光刻胶中的湿膜光刻胶与阻焊油墨，面板光刻胶中的 TFT 正性光刻胶与触控用光刻胶，半导体光刻胶中的 g/i 线光刻胶领域已实现国产替代；电子气体国产化率约为 20%，其中在 2021 年电子特种气体全球市场规模排名前十的电子特种气体（三氟化氮、六氟化钨、六氟丁二烯、氨气、氙气、硅烷、一氧化二氮、磷烷、激光气及三氟化氯）中，国内公司除三氟化氯外均具有量产能力；掩膜版国产化率不足 15%，高精度产品依旧依赖进口；湿电子化学品整体国产化率较高，为 30%-40%，但 8 英寸及以上产品国产化率低；CMP 耗材中，抛光垫国产化率较低，不足 15%；靶材整体国产化率不足 15%，但部分厂商技术已进入国际前列。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表230：半导体材料国产化进程

材料种类	国产化进度	代表厂商	主要情况
硅片	6 寸：>50%	沪硅产业	涵盖 12 英寸抛光片及外延片、8 英寸及以下抛光片/外延片/SOI 硅片，其中 12 英寸大硅片已实现 14nm 及以上逻辑工艺与 3D 存储工艺的全覆盖
	8 寸：20%~30%		
	12 寸：<10%	立昂微	已实现 4/6/8/12 寸轻重掺硅片的量产
		TCL 中环	实现 8 英寸及以下主流产品全覆盖，12 英寸已完成 28nm 以下产品的量产
		有研硅	实现 6-8 溢出抛光片产业化
		神工股份	8 英寸硅片项目顺利开支，已进入客户认证阶段
		中晶科技	立足小尺寸硅片，进一步拓展 8 英寸抛光片业务
半导体光刻胶	G/I 线：30%	晶瑞电材	i 线光刻胶量产，KrF 光刻胶通过中试
	KrF：<10%	上海新阳	KrF 光刻胶已实现销售
	ArF：~0	南大光电	ArF 光刻胶在一家存储芯片制造企业的 50nm 闪存平台上通过认证；在逻辑芯片制造企业 55nm 技术节点的产品上取得了认证突破
	ArFi：~0		
	EUV：~0	彤程新材	i 线、KrF 光刻胶批量供货给国内 18 家 12 寸以及 16 家 8 寸 IC 制造企业
		徐州博康	上游原材料全面布局，KrF、ArF 光刻胶均有产品形成销售
面板光刻胶	彩色光刻胶：5%	雅克科技	收购科特美和 LG 化学旗下的彩色光刻胶资产成功进入 LCD-TFT 光刻胶与 LCD 彩色光刻胶市场
	黑色光刻胶：5%		
	触控屏光刻胶：30%-40%		
	TFT-LCD 胶：国内可大批量供应		
电子气体	20%左右	雅克科技	主要气体产品为六氟化硫、四氟化碳，已成功进入 SK 海力士、美光、三星电子、英特尔、LG、友达光电、中芯国际、长江存储等供应链
		华特气体	在国内 8 寸、12 寸集成电路制造厂商实现超过 80%的客户覆盖率，是国内唯一通过 ASML 认证的气体公司
		金宏气体	主要产品为超纯氨（6N）、高纯氨、高纯一氧化二氮等，已成功进入中芯国际、海力士等厂商供应链
		南大光电	主要产品为砷烷（6N）、磷烷（6N）、三氟化氮、六氟化硫
		正帆科技	主要产品有砷烷（6N）、磷烷（6N）、硅烷和电子混合气等高纯气体等
		昊华科技	核心产品有三氟化氮、四氟化碳、六氟化硫、六氟化钨、磷烷、砷烷、三氟化硼、绿色四氟化二氮、高纯硒化氢、高纯硫化氢、高纯氦气（6N）、VOCs 标气、标准混合气体等
		派瑞特气	主要产品包括高纯三氟化氮、高纯六氟化钨（6N）、高纯氯化氢、高纯氟化氢、高纯四氟化硅、高纯氘气、高纯六氟丁二烯、高纯八氟环丁烷、高纯电子混合气（6N）等
掩膜版	<15%	清溢光电	250nm 工艺节点的 6 英寸和 8 英寸半导体芯片用掩模版量产
		路维光电	实现 250nmIC 掩模版量产，掌握 180nm/150nmIC 掩模版制造核心技术
湿电子化学品	整体 30%-40%	江化微	硫酸、氨水、盐酸达 G5，硝酸达 G4，具备 G2-G5 产品的规模化生产能力
	6 英寸及以下：83%左右	晶瑞电材	高纯双氧水、高纯氨水及高淳硫酸等产品品质已达到 SEMI 最高等级 G5 水准
	8 英寸及以上:<20%	格林达	核心产品 TMAH 显影目前已达到 SEMIG5 标准要求
		达诺尔	超净高纯氨水、超净高纯硫酸技术突破了国外国际技术垄断
CMP 耗材	整体 20%-30%，抛光垫<15%	安集科技	全品类 CMP 抛光液布局，核心原材料自主可控
		鼎龙股份	抛光液、抛光垫均有布局，抛光垫产品实现全制程、全技术节点覆盖
靶材	<20%	江丰电子	高纯金属溅射靶材在技术门槛最高的半导体领域已具备一定的国际竞争力，部分产品成功打入 5nm
		有研新材	铜系列高端靶材产品全面实现技术突破，12 英寸高纯铜及铜合金靶材、高纯镍铂靶材和高纯钴靶材的多款产品
		阿石创	高纯钼靶、ITO 靶材量产线投产，2020 年半导体芯片用靶材开始投产
		隆华科技	研发生产高纯钽及钽合金靶材、ITO 靶材、银合金靶材等产品

资料来源：各公司公告，华泰研究

硅片：国内厂商逐步实现技术突破，国产化进程呈加速趋势

半导体硅片市场长期处于垄断格局，据 SEMI 数据，2020 年全球硅片市场份额前五大公司为信越化学（27.5%）、环球晶圆（14.8%）、德国世创（11.5%）、SUMCO（21.5%）和 SKSiltron（11.3%），合计超过 85%，大陆硅片厂商沪硅产业市占率仅 2.2%。但随着国内厂商的技术突破与产能爬升，硅片国产化进程呈加速趋势。根据公司年报，目前沪硅产业产品涵盖 12 英寸抛光片及外延片、8 英寸及以下抛光片/外延片/SOI 硅片，其中 12 英寸大硅片已实现 14nm 及以上逻辑工艺与 3D 存储工艺的全覆盖；根据公司年报，立昂微已实现 6/8/12 寸轻重掺硅片的量产，其中特色 6/8 英寸特殊规格重掺硅外延片更是供不应求。

免责

更多免费行业报告 并购家 www.ipoipo.cn

光刻胶及配套材料：国内厂商在高端 ArF/EUV 光刻胶领域尚未突破

全球光刻胶市场被日本和美国厂商垄断，其中日企总占比约 80%，占据绝对优势地位，主要厂商包括 JSR、东京应化、住友化学、信越化学、美国杜邦、东进世美肯等。国内光刻胶厂商主要集中在中低端市场，根据各公司官网及年报，仅有北京科华、徐州博康、上海新阳三家公司能量产 KrF 光刻胶，ArF/EUV 光刻胶尚无涉足。此外，南大光电（ArF 光刻胶“02”专项）、上海新阳、徐州博康正在突破 ArF 光刻胶研发。

其他半导体材料：低端产品国产化程度相对较高，高端产品国产替代亟待推进

掩膜版：据 Omdia 统计，2022 年全球各大掩膜版厂商平板显示掩膜版的销售金额情况前五名分别为福尼克斯、SKE、HOYA、LG-IT 和清溢光电，CR5 超 86%，且仅有一家国产厂商入围，在 AMOLED/LTPS 高精度掩膜版上国产化率不足，仍严重依赖进口，国产替代的空间巨大。国内厂商中，根据公司年报，清溢光电已实现 250nm 工艺节点的 6 英寸和 8 英寸半导体芯片用掩模版量产；根据公司年报，路维光电已实现 180nmIC 掩模版量产，掌握 150nmIC 掩模版制造核心技术。

湿电子化学品：全球湿化学品主要厂商包括巴斯夫，霍尼韦尔、亚什兰、汉高、住友化学、三菱化学等，具备 G1-G5 级技术，而国内大部分企业已达 G3 标准，根据各公司年报，江化微、晶瑞电材、格林达等头部企业已逐步实现 G4-G5 级产品布局。2022 年湿化学品领域 6 英寸及以下国产化率为 83%左右，8 英寸及以上国产化率不足 20%，有望持续提升。

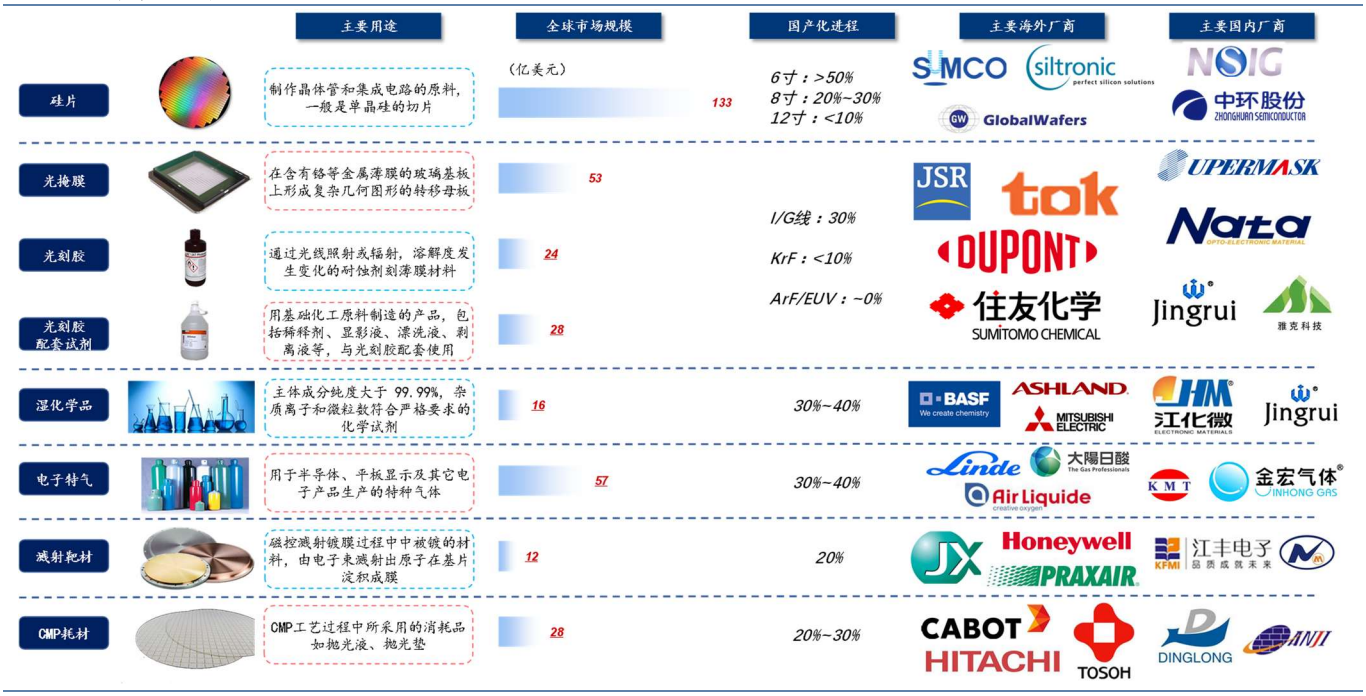
电子气体：电子气体分为电子大宗气体和电子特种气体，是重要的晶圆制造材料。电子气体市场集中度较高，海外头部厂商德国林德、法国液化空气、美国空气化工、日本大阳日酸等，CR4 为 77%。2022 年电子特气国产化率约为 20%，目前国内主要玩家为雅克科技、华特气体、金宏气体、南大光电、正帆科技、昊华科技、派瑞特气等。

CMP 耗材：CMP 耗材包括抛光液、抛光垫，主要海外厂商包括 Cabot、Hitachi、Fujimi、Versum、Dow、ThomasWest 等。国内 CMP 耗材厂商已实现从零到一突破，其中抛光液厂商安集科技产品线齐全，率先打破国外厂商垄断，约占全球抛光液份额的 2.4%；抛光垫厂商鼎龙股份产品型号覆盖率已接近 100%，并已成为部分客户的第一供应商。

靶材：溅射靶材同样由美日企业垄断，2020 年日矿金属、霍尼韦尔、东曹及普莱克斯占据了约 80%的市场份额。根据公司年报，江丰电子产品主要包括铝靶、钛靶/环、钽靶/环、钨钛靶，同样也生产铜靶、镍靶、钴靶、铬靶等，是国内唯一一家能向台积电供应的半导体靶材公司，根据公司官网，目前已在 7 纳米技术节点实现批量供货,应用于 5 纳米技术节点的部分产品评价通过并量产，部分产品进入验证阶段。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表231：半导体材料分类别情况（2020A）



资料来源：SEMI，华泰研究

图表232：半导体材料主要公司及市场空间

材料种类	公司名称	公司代码	是否上市	成立时间	2022 营业收入（亿美元）	2022 半导体收入（亿美元）	产品空间（全球，亿美元）	产品种类
硅片	沪硅产业	688126 CH	科创板	2015	5.17	5.17	133	半导体硅片企业，产品类型涵盖 300mm 抛光片及外延片、200mm 及以下抛光片、外延片及 SOI 硅片
	立昂微	605358 CH	主板	2002	4.18	4.18	133	半导体硅片、半导体分立器件芯片及分立器件成品等，已实现 4/6/8/12 寸轻重掺硅片的量产
	TCL 中环	002129 CH	主板	1988	96.22	4.69	133	聚焦半导体节能产业和新能源产业，实现 8 英寸及以下主流产品全覆盖，12 英寸已完成 28nm 以下产品的量产
	有研硅	688432 CH	科创板	2001	1.69	1.61	133	半导体硅材料研究，实现 6 英寸、8 英寸硅片的产业化及 12 英寸硅片的技术突破
	神工股份	688233 CH	科创板	2013	0.77	0.72	133	半导体级单晶硅材料供应商，生产的半导体级单晶硅材料纯度达到 11 个 9，量产尺寸最大可达 19 英寸，可满足 7nm 先进制程芯片制造刻蚀环节对硅材料的工艺要求
	中晶科技	003026 CH	主板	2010	0.49	0.47	133	8 英寸硅片项目顺利开支，已进入客户认证阶段
半导体光刻胶	晶瑞电材	300655 CH	创业板	2001	2.51	1.11	40	微电子业用超纯化学材料和其他精细化工产品，种包括氢氟酸、过氧化氢、氨水、盐酸、硫酸、硝酸、异丙醇、冰醋酸、混合酸(硅腐蚀液、铝腐蚀液、铬腐蚀液、BOE、金蚀刻液)氢氧化钾、氢氧化钠、配套试剂等
	上海新阳	300236 CH	创业板	2004	1.72	0.92	24	主要产品包括半导体传统封装、半导体制造及先进封装、航空航天飞行器电子元器；KrF 光刻胶已实现销售
	南大光电	300346 CH	创业板	2000	2.27	2.14	81	在 MO 源的合成制备、纯化技术、分析检测、封装容器等方面已全面达到国际先进水平，主要产品有光刻胶、三甲基镓、三甲基铟、三乙基镓、三甲基铝等，主要气体产品为砷烷（6N）、磷烷（6N）、三氟化氮、六氟化硫，可以实现 MO 源产品的全系列配套供应
	彤程新材	603650 CH	主板	2008	3.59	0.55	24	化工材料制造和服务商，主要产品如橡胶用酚醛树脂，包括增粘树脂、补强树脂、粘合树脂等，广泛应用于轮胎及其他橡胶制品制造行业。
	徐州博康	-	未上市	2010	-	-	24	光刻胶（ArF、KrF、I-LINE、电子束、封装光刻胶）、电子材料（KrF 单体材料、ArF 单体材料、聚酰亚胺单体、OLED 材料）及功能性化学品

免责
更多免费行业报告 并购家 www.ipoipo.cn

材料种类	公司名称	公司代码	是否上市	成立时间	2022 营业收入（亿美元）	2022 半导体收入（亿美元）	产品空间（全球，亿美元）	产品种类
面板光刻胶	雅克科技	002409 CH	主板	1997	6.12	5.80	81	主要产品为电子半导体材料、深冷复合材料以及塑料助剂材料，主要气体产品为六氟化硫、四氟化碳；收购科特美和 LG 化学旗下的彩色光刻胶资产成功进入 LCD-TFT 光刻胶与 LCD 彩色光刻胶市场
电子气体	华特气体	688268 CH	科创板	1999	2.59	1.90	57	主要气体产品包括氖气、氦气、六氟化硫、乙烯、一氧化碳、消毒气、甲烷、氦气等
	金宏气体	688106 CH	科创板	1999	2.82	2.82	57	特种气体、大宗气体和天然气三大类 100 多个气体品种，包括天然气、甲烷、三氯化硼、三氟化氮、六氟化氪、氯化氢、一氧化碳、笑气、硅烷等
	正帆科技	688596 CH	科创板	2009	3.88	0.35	57	主要产品有砷烷（6N）、磷烷（6N）、硅烷和电子混合气等高纯气体等
	昊华科技	600378 CH	主板	1999	13.02	12.49	57	核心产品有三氟化氪、四氟化碳、六氟化硫、六氟化钨、磷烷、砷烷、三氟化硼、绿色四氧化二氮、高纯硒化氢、高纯硫化氢、高纯氦气（6N）、VOCs 标气、标准混合气体等
	中船（邯郸）派瑞特气	688146 CH	科创板	2016	2.81	-	57	主要产品包括高纯三氟化氪、高纯六氟化钨（6N）、高纯氯化氢、高纯氟化氢、高纯四氟化硅、高纯氖气、高纯六氟丁二烯、高纯八氟环丁烷、高纯电子混合气（6N）等
掩膜板	清溢光电	688138 CH	科创板	1997	1.09	1.07	53	250nm 工艺节点的 6 英寸和 8 英寸半导体芯片用掩模版量产
	路维光电	688401 CH	科创板	2012	0.92	0.92	53	具有 G2.5-G11 全世代掩膜版生产能力，可以配套平板显示厂商所有世代产线；实现了 250nm 制程节点半导体掩膜版量产，并掌握了 180nm/150nm 节点半导体掩膜版制造核心技术，满足先进半导体芯片封装和器件等应用需求
湿电子化学品	江化微	603078 CH	主板	2001	1.35	0.68	16	主要产品包括硫酸、氨水、盐酸达 G5，硝酸达 G4，具备 G2-G5 产品的规模化生产能力
	格林达	603931 CH	主板	2001	1.22	0.30	16	主要产品包括显影液、蚀刻液、稀释液、清洗液等，其中高纯双氧水、高纯氨水及高淳硫酸等产品品质已达到 SEMI 最高等级 G5 水准
	达诺尔	833189.NQ	新三板	2004	0.19	0.19	16	为多家电子芯片生产厂家提供超纯氨水和超纯异丙醇产品，超净高纯氨水、超净高纯硫酸技术突破了国外国际技术垄断
CMP 耗材	安集科技	688019 CH	科创板	2006	1.55	1.55	28	全品类 CMP 抛光液布局，核心产品包括化学机械抛光液、功能性湿电子产品等
	鼎龙股份	300054 CH	创业板	2000	3.91	0.75	28	公司主营业务是半导体 CMP 制程工艺材料、半导体显示材料、半导体先进封装材料；主要产品包括 CMP 抛光垫、CMP 抛光液、CMP 清洗液、黄色聚酰亚胺浆料 YPI、光敏聚酰亚胺 PSPI、面板封装材料 INK、临时键合胶 TBA、封装光刻胶 PSPI、底部填充胶 Underfill
靶材	江丰电子	300666 CH	创业板	2005	3.34	2.82	12	从事高纯溅射靶材的研发、生产和销售，主要产品包括铝靶、钛靶及钛环、钽靶及钽环等
	有研新材	600266 CH	主板	1999	21.90	16.60	12	主要产品有高纯/超高纯金属材料、稀土材料、光电材料、红外光学、光纤材料、医疗器械材料
	阿石创	300706 CH	创业板	2002	1.00	0.96	12	主要产品包括光学光通讯靶材（化合物材料、单质与混合物材料、光学光通讯靶材、镀膜配件）、太阳能光伏、半导体靶材、平板显示、Low-E 镀膜等
	隆华科技	300263 CH	创业板	1995	3.30	0.68	12	主要产品包括电子信息靶材及超高温材料、军民融合高分子复合材料及节能环保产业板块产品

资料来源：各公司官网，各公司年报，华泰研究预测

免责
更多免费行业报告 并购家 www.ipoipo.cn

半导体产业链未上市公司：悦芯科技、华封科技、先锋半导体、御微半导体、世禹精密、汉骅半导体、佑伦真空、富乐华半导体、泰科天润、锐杰微科技

悦芯科技：国内半导体中高端制造装备提供商，致力于打破高端设备市场日美垄断格局。公司成立于 2017 年，主要从事大规模集成电路自动化生产测试设备（ATE）的研发、生产和销售，客户包含集成电路晶圆制造企业、晶圆测试厂、封装测试厂、芯片设计公司及科研院所等企事业单位。

公司深耕中国，面向全球，在中国合肥、西安、韩国水原等地设有研发部门，与国内院校设立产学研合作基地。公司员工中研发技术人员占比近 60%，累计申报知识产权 100 余项。公司自主研发的具有完全自主知识产权的 SoC 芯片生产测试设备-T800 系列、功率半导体测试设备-TP1200 系列及存储器芯片生产测试设备-TM8000 系列填补了国产高端 ATE 设备领域的空白，整体性能指标已接近国际主流高端测试平台；市场累计装机量已超过 300 台，服务近 100 家国内外客户。未来，产品规划方面公司会持续加强 ATE 细分市场的技术整合，覆盖 Flash、CIS、RF 等细分市场领域。在规划下一代 ATE 系统构架研究的同时，也将加大投入设备专用 ASIC 芯片开发，逐步提升自研 ATE 测试系统核心技术竞争力。

华封科技：全球领先半导体先进封装解决方案供应商

华封科技成立于 2014 年，在新加坡、菲律宾、北京、中国台湾、苏州等地设有分支机构。公司是聚焦先进封装设备领域的高端装备制造商，致力于为客户提供半导体先进封装的产品技术和解决方案，是现阶段能与国际头部设备供应商进行直接竞争的非上市公司。在全球范围内，华封科技是近年来唯一打破现有竞争格局的新兴企业，尤其在晶圆级先进封装 Die bonder 领域站稳头部席位。华封科技成立 3 年就进入 ASE 全球供应链，并成为其核心供应商，目前华封科技已经服务了近 30 家客户，覆盖国内前十的国内客户，和国际排名前十中的七家客户，包括日月光、矽品、长电科技、通富微电、华天科技等。

华封科技在技术上处于国际领先地位。创始团队来自 ESEC、K&S、ASM 等企业，拥有近 30 年的半导体行业经验，拥有自主核心技术的电控系统和机械设计，最高精度的 Hybrid bond 设备可以达到亚微米级，独创的双动梁多头设计，实现了设备精度和稳定性媲美甚至超过国际同类产品，同时速度提升 1-2 倍，打破高精度、高端速度和高稳定性相矛盾的技术屏障。华封科技在晶圆级封装（Fan-out）领域处于国际领先地位，并成为日月光 M-series 工艺 Die Bond 环节的独家设备供应商。

公司产品定位于半导体先进封装领域，对先进封装贴片工艺实现了全面覆盖，技术全球领先。公司有面向晶圆级封装、倒装贴片机、系统级封装的主力设备，工艺涵盖 FOWLP（Face Up/Down）、POP、MCM、EMCP、SIP、2.5D/3D、FCCSP/FCBGA。针对未来先进封装的需求，公司推出了全球首个大尺寸玻璃面板级封装设备，后续将相继推出面向 Chiplet 封装的 Hybrid Bonder 设备，面向 IGBT、汽车电子大功率器件的 Soft Solder Die Bonder，面向层叠封装的 Stack Die 设备，面向 Micro LED 封装的巨量转移设备等。

先锋半导体：中国半导体设备精密零部件供应商，实现蚀刻机核心部件国产化

靖江先锋半导体成立于 2008 年，主要从事精密金属零部件生产制造、以数控加工中心为主体的精密加工和针对铝、不锈钢等金属材料表处理，为国内外半导体设备厂商提供金属零部件。公司专注技术，在刻蚀机、薄膜沉积设备金属精密零部件如腔体、内衬、加热器、匀气盘等领域已公开专利申请 72 件。公司产品质量优异，已实现规模出货，主要客户包括北方华创、中微半导体、中芯国际、华海清科、屹唐股份等行业龙头企业。其中，公司腔体等关键零部件已用于中微半导体公司生产的 5nm 刻蚀机，并应用于台积电全球首条 5nm 制程生产线，助力蚀刻机核心部件国产化。

更多免费行业报告 并购家 www.ipoipo.cn

御微半导体：半导体量检测设备提供商

御微半导体成立于 2019 年，是一家为集成电路制造提供先进装备的高新技术企业。公司聚焦于集成电路光学量测及检测装备的系统设计与系统集成，已经形成了掩模版检测、晶圆检测、晶圆测量、泛半导体检测 4 大领域 6 大类量检测产品，具备高检测率、高灵敏度、高分辨率等特征。公司致力于成为半导体设备行业第一梯队的技术提供者。其中，公司推出的全国首台集成电路掩模缺陷检测产品的关键性能已经达到行业领先水平，订单表现良好；集成电路前道晶圆缺陷检测产品已经成功通过知名集成电路制造商认证并获得全面推广；集成电路晶圆套刻测量设备，已发运至国内重要客户，并获得良好验证。2023 年上半年，公司顺利完成由合肥产投集团领投的 B 轮融资，融资资金将用于公司的研发投入，帮助公司构建集成电路前道量检测领域的核心竞争力。

世禹精密：半导体自动化设备供应商

世禹精密成立于 2017 年 10 月，主营业务为半导体自动化设备研发和制造。公司主要产品有植球类设备，芯片外观检查分选机，精密芯片贴装机，激光打标机，基板传送类设备，AGV+Robot 自动化系统，客户定制各类自动化系统等。公司拥有丰富的半导体自动化设备的制作经验，硬件和软件全部自主开发，模块化设计，除了提供标准的设备，还可以根据客户要求进行硬件和软件的多种定制；此外，公司还擅长高精度，多种工艺结合。公司开发的多种机型达到业界先进水平，主要客户有 Sandisk，NXP，Amkor，AT&S，通富微电，长电科技，星科金朋等。2023 年 4 月，公司完成新一轮数亿元融资，投资方包括 IDG 资本、中电科研投基金、厚雪资本、物产中大投资、尚颀资本、泓生资本、天马股份、东证资本、山高弘金等，其中 IDG 资本、东证资本二次跟投。

汉骅半导体：半导体材料研发商，打造国际领先高端半导体研发生产基地

苏州汉骅半导体成立于 2017 年，属于国家扶持的战略新兴产业，主要致力于第三代半导体关键材料的研发与生产，是国内领先的具备规模化研发、生产化合物半导体材料与芯片能力的公司。2023 年 6 月，汉骅于 ISPSD 会议展示多种功率氮化镓高端外延片，主要产品可以应用于电力电子等诸多关键应用场景。目前公司在苏州工业园区核心区域已建成约 20000 ㎡厂区、5000 ㎡洁净室，数十台 MOCVD，6/8 英寸光电芯片-MEMS 量产产线，是具备国际领先水准的高端半导体闭环研发与生产基地。公司秉承“创新、开放、合作、共赢”为核心的经营理念，以科技创新为本，为高端芯片产业链内上下游企业提供全方位的解决方案。

佑伦真空：专注高真空薄膜沉积设备供应

佑伦真空的前身苏州佑伦电子成立于 2012 年，佑伦真空成立于 2016 年，成立以来公司专注于高真空薄膜沉积设备领域，主营业务覆盖高端定制型薄膜沉积设备、零配件业务、为行业提供设备所需耗材和精密机加工四个业务板块，已形成独立自主可控的核心技术和一定的成本优势。

公司产品包括金属镀膜机、ITO 镀膜机、大腔体单泵系列、大腔体双泵系列等 8 个机种 80 多个系列的薄膜沉积设备，应用领域覆盖 LED 芯片（Micro LED、Micro OLED 等）、半导体（分立器件、功率器件、集成电路）、光学膜（光刻机光栅镀膜和激光芯片镀膜）、太阳能、科研等众多领域。公司产品具有高性价比、售后服务完备、交货周期短（1.5-2 月）等优势，公司核心人员均有多年在国际先进设备制造大厂中从业经验，具备国内领先的工艺理解能力。公司产品已成功进入了华为海思、中电科、士兰微、中芯绍兴、中芯宁波、泰科天润、海康微影、维信诺、深天马、康佳、长春光机所、西安光机所等国内头部客户和科研院所。

截至目前，公司获得授权专利超过百余项，涉及整机设计、关键结构、系统设计、材料和工艺整合等技术，并且已经成功开发出高精度薄膜沉积设备，实现了蒸镀机高均匀性、高重复性、高稳定性的性能保障。公司将始终以“打造真空产业链，成为薄膜沉积设备领域的 NO.1”的目标不断前进。

免责

更多免费行业报告 并购家 www.ipoipo.cn

富乐华半导体：全球领先的功率模块封装用覆铜陶瓷载板生产企业

富乐华半导体成立于 2018 年，公司主要产品为直接覆铜陶瓷载板（DCB）、活性金属钎焊覆铜陶瓷载板（AMB）及直接镀铜陶瓷载板（DPC），为功率半导体厂商提供从材料、制造、封装、测试的完整解决方案，产品广泛应用于新能源汽车、新能源发电、工业变频控制、消费电子、光通信、激光器等领域。公司拥有全球领先的陶瓷覆铜技术，部分核心技术为全球唯一，公司坚持“以研发设计为核心、以市场需求为导向”的经营理念。截止目前，公司已累计申请专利 172 项，授权专利总计 104 项，其中发明专利 56 项，实用新型 48 项，受理中专利 68 项。专利领域不仅涵盖 DCB 制造技术，还有针对未来市场的 AMB 制造技术、DPC 制造技术、DBA 制造技术、高分子材料埋阻制造技术、半导体材料制造技术等，公司通过对技术的不断创新，与时俱进时刻把握住市场脉搏。

泰科天润：国内碳化硅功率器件产业化领军企业

泰科天润半导体科技（北京）有限公司成立于 2011 年，是国内碳化硅功率器件产业化领军企业，专业从事碳化硅器件研发与制造，并提供应用解决方案。总部坐落于北京，拥有两条碳化硅半导体工艺晶圆生产线，并在上海、深圳设有办事处，海外均有代理渠道提供服务支持。公司专注于 IDM 模式，拥有 10 年碳化硅器件量产经验，集设计、研发、制造、应用为一体，已通过 IATF16949 体系认证。公司的碳化硅产品已通过 RoHS、REACH、UL、DNV·GL、USCG、AEC-Q101 等标准认证，产品质量比肩行业先进水平。产品规格范围覆盖 650V-3300V（1A-100A）等多款规格，可以提供 TO220、TO220 全包封、TO220 内绝缘、TO247-2L、TO247-3L、TO247-4L、TO252、TO263、TO268、DFN8*8、DFN5*6、SOD-123FL、SMA、SMD 等多款塑封以及高温封装形式，并可按客户需求提供其他封装形式。产品已批量应用于光伏逆变器、充电桩、OBC、车载 DC-DC、通信电源、高端服务器电源、工业电源、PC 电源等领域，并多次获得行业优秀产品奖。

锐杰微科技：高端芯片封测方案服务商

锐杰微科技成立于 2011 年，是一家提供全流程 Chiplet&高端芯片封测制造方案商。聚焦复杂芯片封装设计& 仿真、规模化封装加工制造和成品测试，配套提供 Bumping 和 CP、先进材料采购、流片和板级电路开发服务。公司已为服务器级 GPU、国内首款商用 RISC-V 架构高端 AI 处理器、国产主流商用计算处理器、国产高性能复杂 SiP、Chiplet 架构的国产高性能网络处理器、国内首款基于 RISC-V 架构工业级二层交换芯片等众多创新产品提供封测服务。公司获得苏州市“独角兽”培育企业、郑州市电子信息 50 强企业、芯榜“2022 中国先进封测新锐企业”等资质和称号。 团队具备 25 年以上的先进封装设计、量产经验和 600+SiP\国产高端处理器项目积累，同时还可为客户提供定制化服务。2022 年公司完成近 3 亿 B 轮融资，截至当前公司已服务超过 200 家科研院所及高端商业客户。

更多免费行业报告 并购家 www.ipoipo.cn

汽车智能化：红海竞争的破局点

在汽车电动化渗透率不断提高的背景下，汽车智能化较大可能成为下一个红海竞争的破局点，具体表现在：**智能驾驶领域：**领航辅助驾驶（NOA）等高级别辅助驾驶功能模型成熟度不断提高，市场领先者的测试版逐渐落地。我们认为今年或是城市 NOA 普及的元年并且成本有望进一步下降，国产供应商有望受益。**智能座舱领域：**现阶段智能座舱已经处于全面向 4.0 阶段发展的时期，硬件的装载是实现软件层交互的基础，座舱硬件已率先迎来大规模上车机遇；其次在智能座舱大模型生态发展下，我们认为拥有大模型和机器视觉技术布局的软件层企业同样有望受益。

智能驾驶：高级别辅助驾驶蓄势待发，国产供应商受益

目前 LCC 等 L2 级别的辅助驾驶功能逐渐标配化，NOA 等高级别辅助驾驶功能模型成熟度不断提高，市场领先者的测试版逐渐落地。我们认为今年或是城市 NOA 普及的元年。我们测算当前旗舰车型高级别辅助驾驶 BOM 成本为 1.4 万元。我们预测 2023-2030 年高级别辅助驾驶 BOM 成本的平均年降幅度为 11%，国产供应商有望受益。

高级别辅助驾驶的普及趋势

领航辅助驾驶（NOA），即车辆在部分高速公路或高架等封闭路段行驶时，结合车载导航路线让车辆进行自动变道、自动进入和驶出匝道口的技术功能，可实现一定道路场景范围内的点到点智能驾驶。根据场景的不同，领航辅助可进一步分为高速领航和城区领航。高速领航普遍限制在特定高速公路和城区高架路开启，包含自动调节车速、自动进出匝道、自动变道超车等功能，目前已在国内落地。城区领航则可以在复杂的城市场景中实现点到点的“导航辅助驾驶”功能，车主在导航上设定好目的地，车辆可以实现全程辅助驾驶到达终点，并在路途中实现变道、超车、过红绿灯等行为动作，其难度也远远大于高速 NOA，有望于今年开始导入。

2022 年，中国乘用车市场加速导入高速领航。在中国，高速领航功能的落地始于 2019 年 6 月，特斯拉官方正式向中国市场内全部选配了 FSD 完全自动驾驶的车型，推送最新版本 NOA，能够令车辆自动驶入和驶出高速公路匝道或立交桥岔路口，并超过行驶缓慢的车辆。随后小鹏、蔚来、理想也纷纷入局，进入 2022 年，长城、上汽等自主品牌也开始在旗下的部分车型上推出该功能。

小鹏、华为领衔，城区领航 2023 年开始落地。目前城区领航仍较为依赖高精地图，因此在初始落地时，可运行区域范围仍较小，普遍按城市进行开放。小鹏和华为的方案已于 2022 年 9 月开始落地，并在逐步拓宽开放区域。除此之外，蔚来、理想、集度、长城等车企也计划在 2023 年推出各自的城区领航功能。

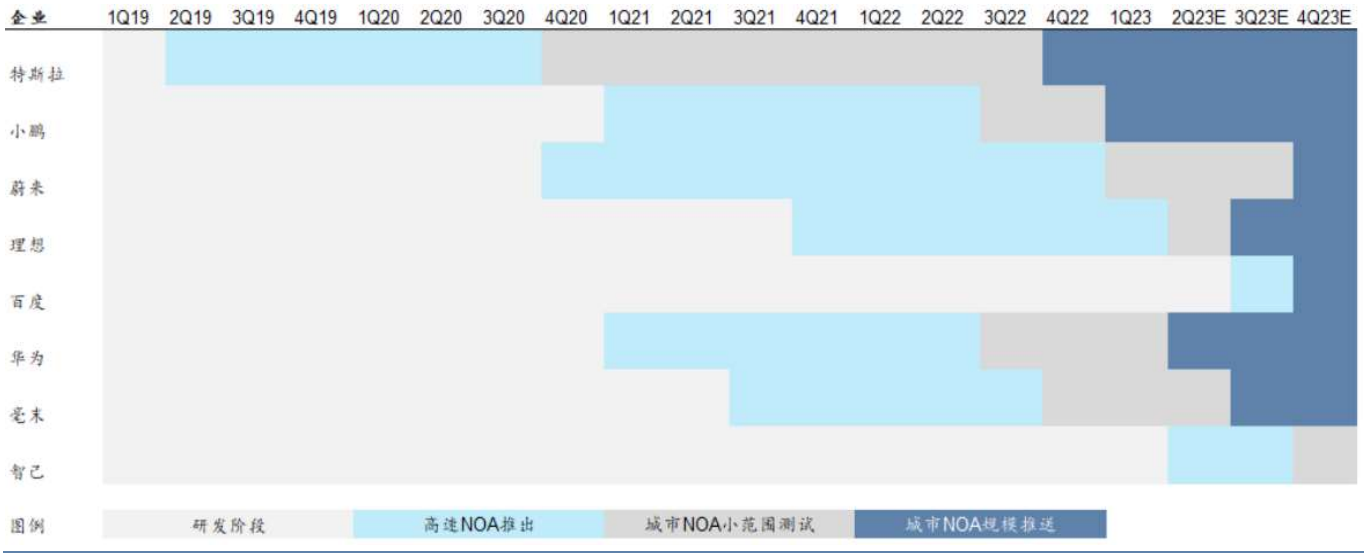
1）3 月 31 日，小鹏汽车启动推送 Xmart OS 4.2.0 的同时，释放了 XNGP 第一阶段能力：G9 及 P7i Max 版在上海、深圳、广州，小鹏 P5 P 系列在上海的高精地图覆盖区域开放点到点的城市 NGP（智能导航辅助驾驶）；在无高精地图覆盖区域，开放具备跨线绕行，识别红绿灯并直行通过路口能力的 LCC 增强版。
2）4 月 11 日，在第八届 AI DAY 上，毫末智行发布消息称，城市 NOH 将依次搭载在魏牌摩卡 DHT-PHEV 和蓝山上，目前已经在保定和北京做大规模泛化测试，将于今年三季度实现城市 NOH 功能，并于 2024 年开拓 100 个城市。
3）4 月 12 日，智己汽车宣布了智己城市 NOA 领航辅助以及替代高精地图的数据驱动道路环境感知模型，预计将于 2023 年内开启公测。
4）4 月 16 日，华为发布 HUAWEI ADS 2.0 系统，同时表示，城市 NCA 已经在深圳、上海、广州落地，重庆、杭州也将在今年二季度解锁。目前这些城市还是基于高精地图技术，且广州、重庆、杭州为部分区域覆盖。今年三季度，华为城区 NCA 将实现 15 个无图城市的落地，Q4 将新增 30 个无图城市落地，至 45 城。
5）4 月 16 日，百度 Apollo 推出城市智驾产品 Apollo City Driving Max，城市智驾产品今年会量产。

更多免费行业报告 并购家 www.ipoipo.cn

6）4 月 18 日，理想 AD Max 3.0 的城市 NOA 将于今年第二季度开始推送内测用户，到年底将推送 100 座国内城市。

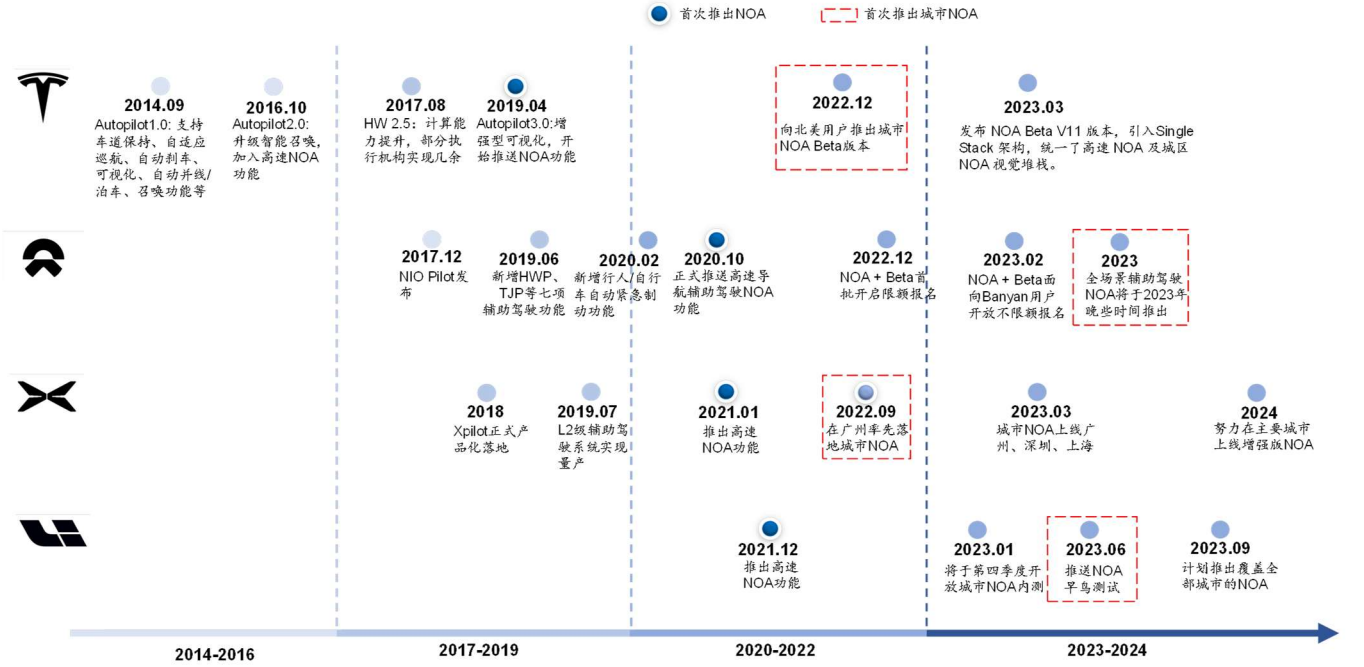
7）4 月 18 日，蔚来宣布于今年初开启内测的 NOP+（增强型领航辅助），要在今年 7 月 1 日正式开启商用，NOP+之后的 NAD（自动驾驶）系统有望可以于今年 6 月份在上海开启 Beta 版本。

图表233： 国内头部智能驾驶企业高速及城市 NOA 进展



资料来源：公司公告，佐思汽研，华泰研究

图表234： 特斯拉、蔚小理领航辅助功能演进

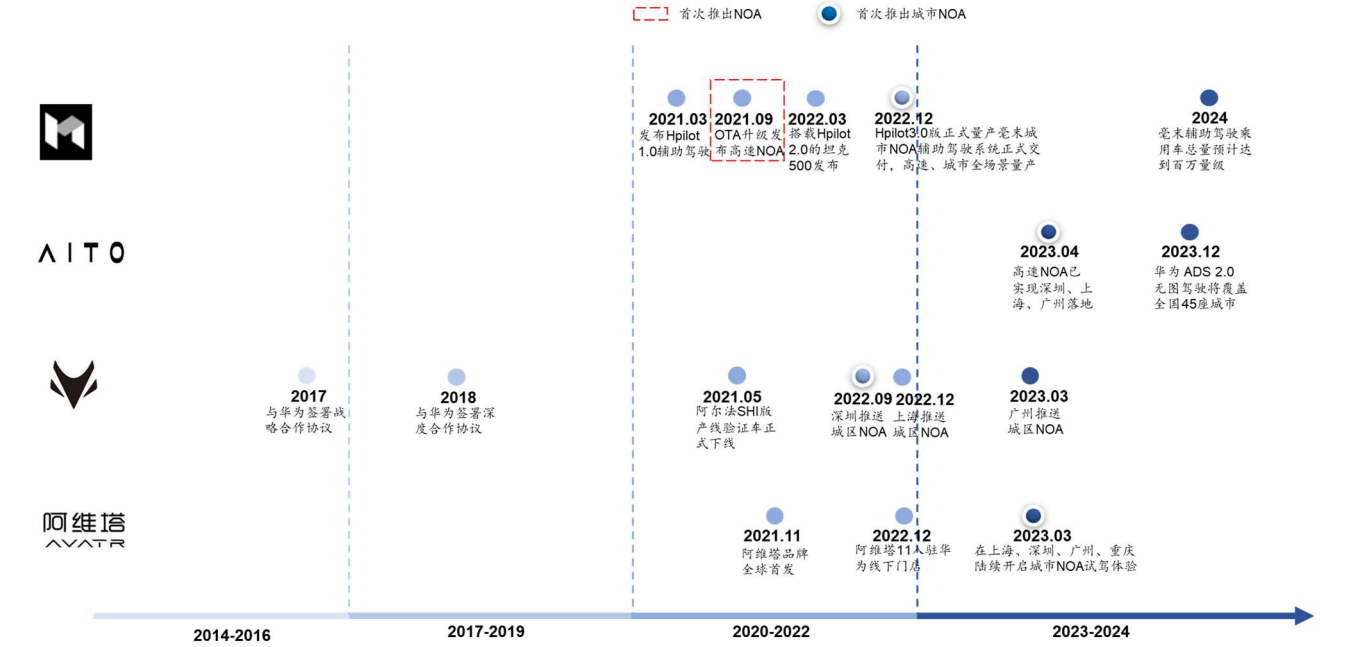


注：特斯拉 NOA——FSD；蔚来 NOA——NOP；小鹏 NOA——NGP；理想 NOA-NOA。

资料来源：公司公告，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表235：毫末智行、问界、极狐、阿维塔领航辅助功能演进



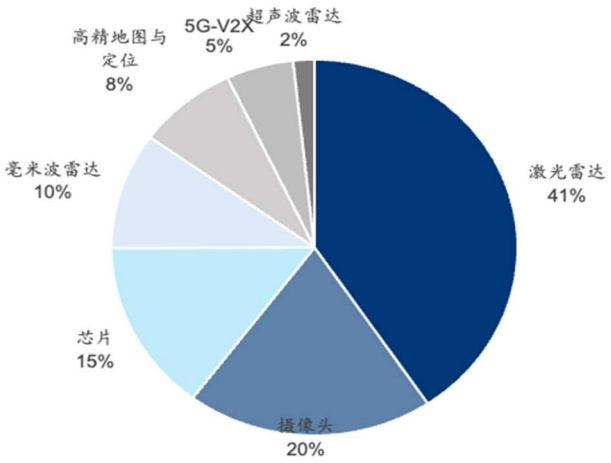
注：毫末智行 NOA——NOH；问界 NOA——HUAWEI ADS；极狐 NOA——NCA；阿维塔 NOA——NCA。
资料来源：公司公告，华泰研究

高级别辅助驾驶的降本趋势

当前旗舰车型高级别辅助驾驶 BOM 成本为 1.4 万元。我们基于市面上主流高级别辅助驾驶的硬件配置成本测算[3]。当前各品牌旗舰车型智能驾驶硬件的平均总成本为 14,181 元；其中激光雷达占 41%、摄像头占 20%、芯片占 15%、毫米波雷达占 10%、高精度地图与定位占 8%、V2X 占 5%、超声波雷达占 2%。

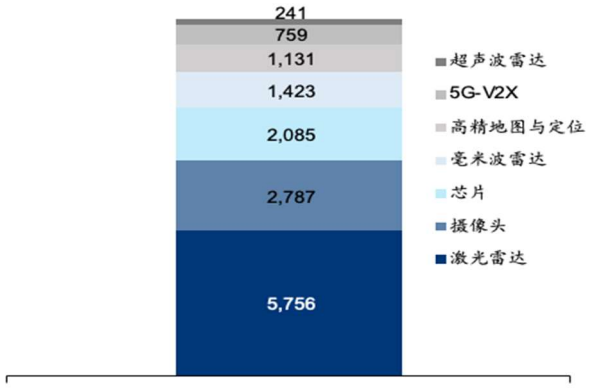
我们认为，未来几年，高级别智能驾驶功能的落地可分为，1）软件端，主要由大模型升级主导；2）硬件端，由过去量的堆叠与冗余转向性能提升与成本下降。具体而言，我们预测 2023-2030 年高级别辅助驾驶 BOM 成本的平均年降幅度为 11%。

图表236：主流高级别辅助驾驶汽车 BOM 成本-占比



资料来源：公司公告，华泰研究

图表237：主流高级别辅助驾驶汽车 BOM 成本-价值量（元）

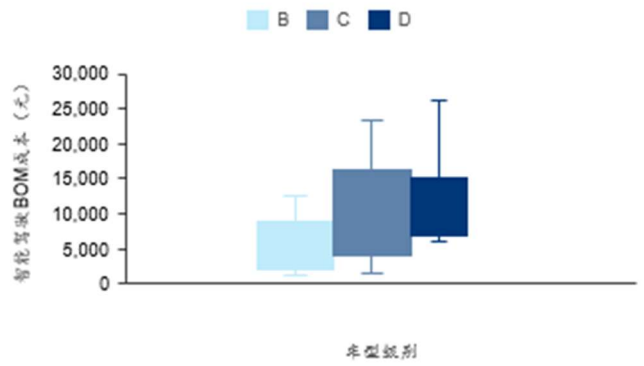


资料来源：公司公告，华泰研究

[3]注：样本为蔚来、理想、小鹏、问界、极氪、智己、飞凡、广汽埃安、瑞虎、红旗、岚图、极狐、丰田、大众、现代、宝马、福特、凯迪拉克 L2/L2+级别车型

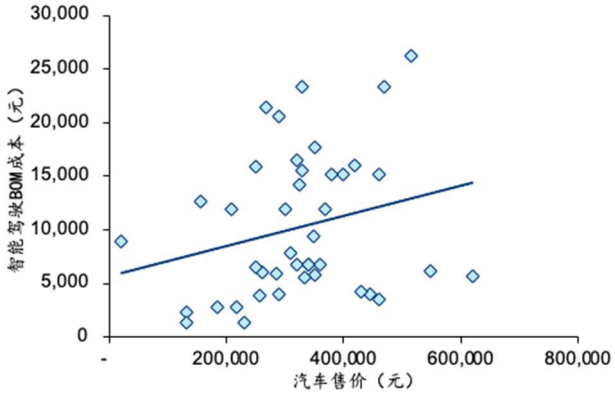
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表238： 主流高级别辅助驾驶汽车 BOM 成本-分车型级别



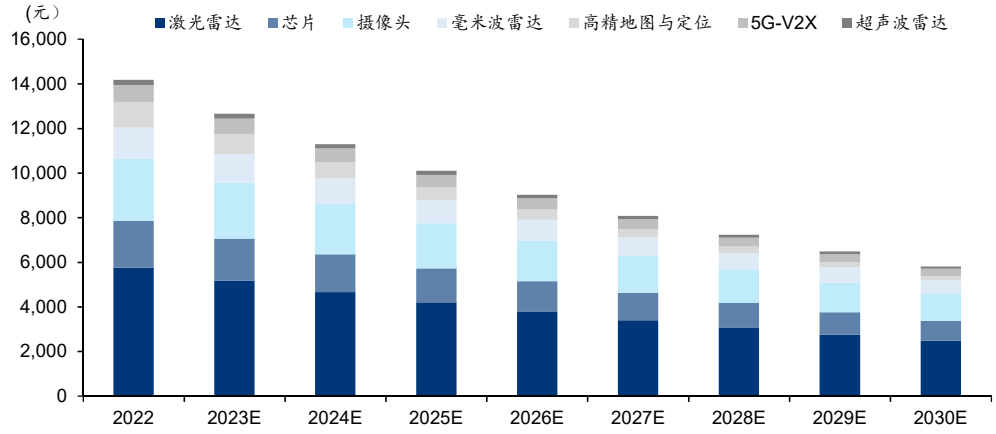
资料来源：公司公告，华泰研究

图表239： 主流高级别辅助驾驶汽车 BOM 成本-分整车售价



资料来源：公司公告，华泰研究

图表240： 主流高级别辅助驾驶汽车 BOM 成本-降本趋势展望



资料来源：公司公告，华泰研究预测

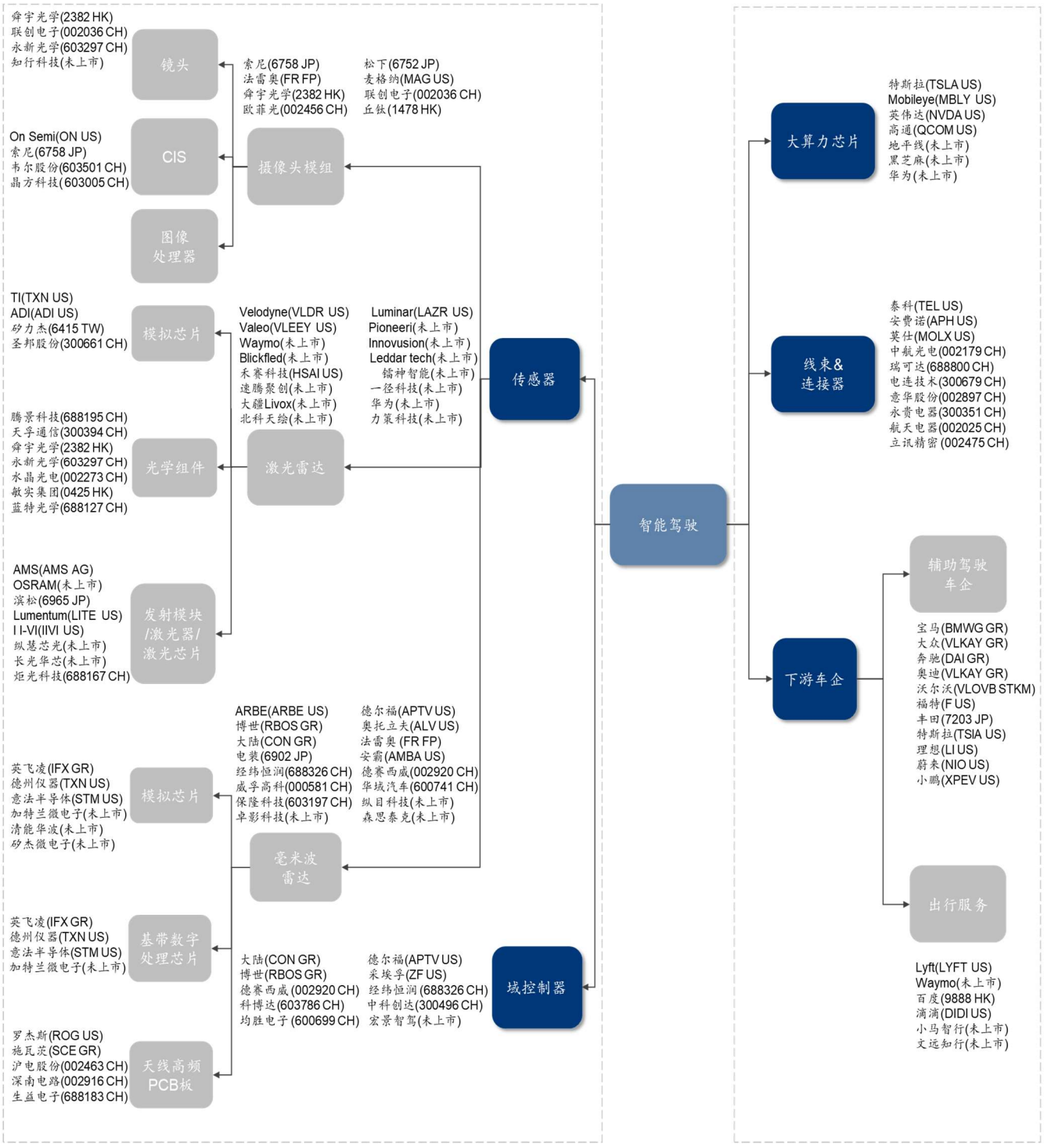
供应链的国产替代趋势

我们自上而下地梳理了自动驾驶产业链及代表公司。我们认为，受益于产业链各环节国内玩家产品性能的提升与下游自主品牌的崛起，零部件国产替代趋势显著。1）芯片：英伟达引领智能驾驶芯片迭代，Mobileye、TI 技术成熟，国内玩家地平线、黑芝麻等与先行者距离逐步缩短。2）域控制器及解决方案：国内玩家德赛西威、经纬恒润、纵目科技等均已规模化上车，技术成熟度不断提升。3）激光雷达：国产供应商禾赛、图达通、速腾等在量产节奏上更快。4）4D 毫米波：国内玩家有加特兰（MMIC 芯片）、森思泰克等；5）高速连接器：罗森伯格技术积淀深厚，电连技术、瑞可达等加速追赶。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表241： 智能驾驶产业链主要公司梳理



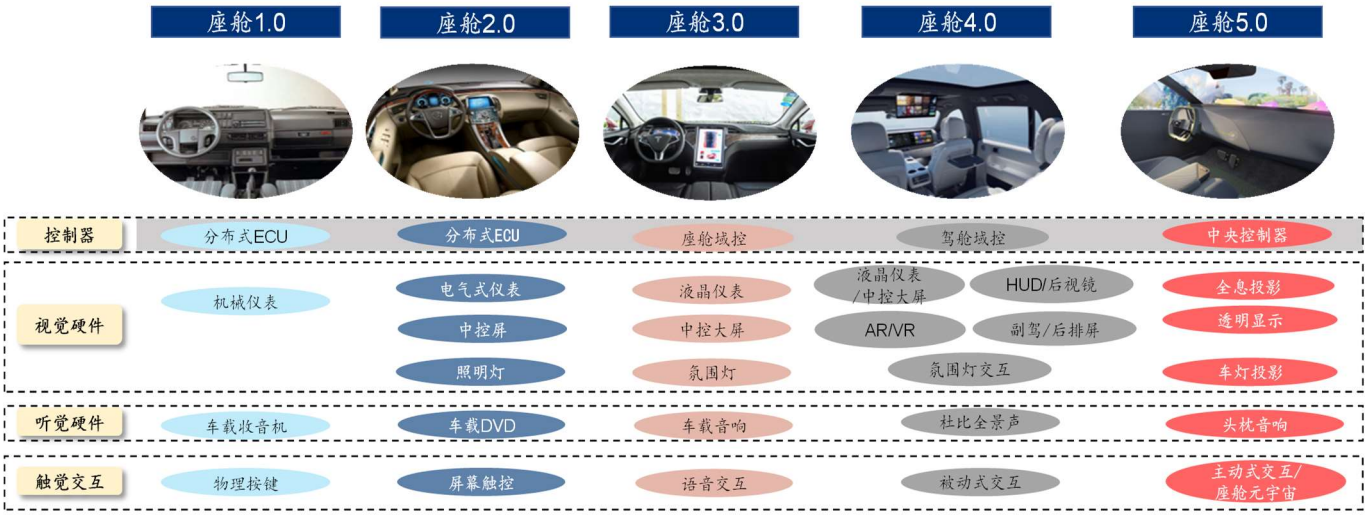
资料来源：Bloomberg，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

智能座舱：从 4.0 多模态到 5.0 大模型，座舱智能化全面升级

如果我们将汽车座舱从机械时代到智能时代的发展划分为 5 个阶段，我们认为现阶段下座舱正处于从 3.0（以特斯拉 Model3/Y 的简配为代表）到 4.0（以理想 L9 的高配为代表）的全面演进阶段。在 2023 年上海车展中，我们观察到除了新势力依旧维持激进的座舱智能化配置外，燃油车也正在开启座舱智能化方案迭代。**4.0 时代加速到来，座舱视听触多模态硬件加速升级，丰富汽车座舱智能生态；大模型赋能软件，打造全车 AI 智能管家。**

图表242：汽车座舱发展阶段

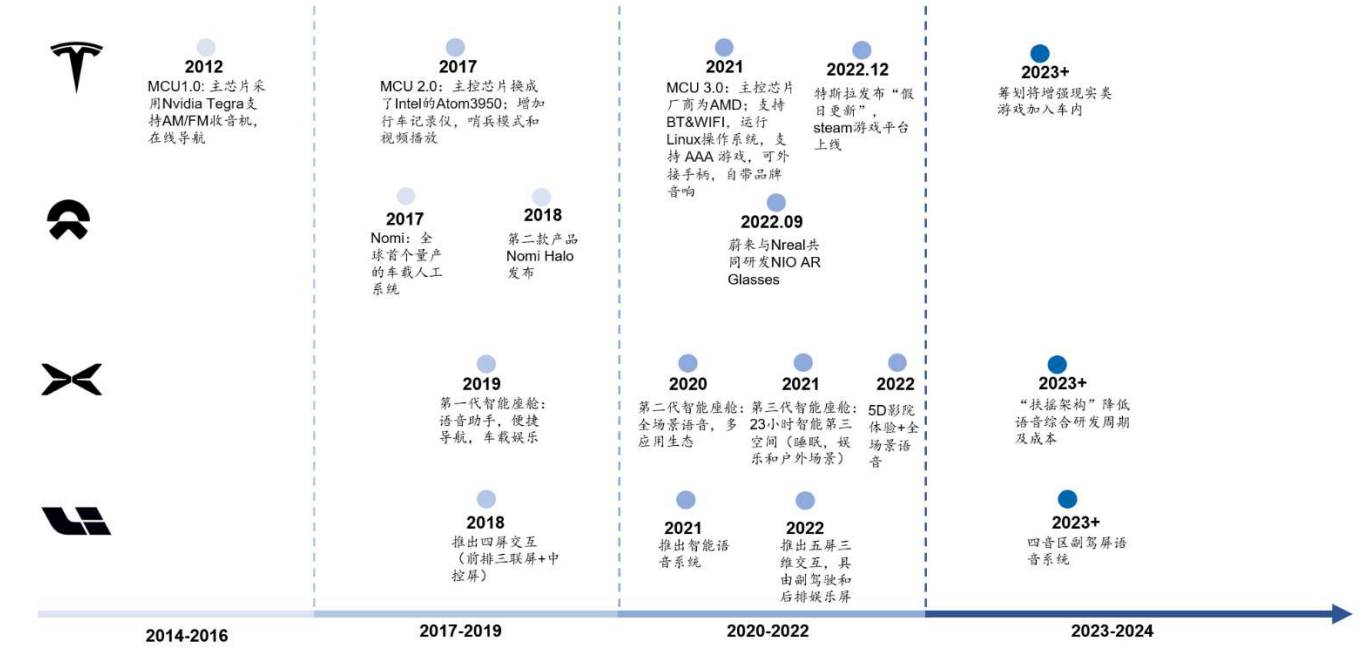


资料来源：CES2023 大会，电子发烧友，有驾，华泰研究

从新势力到传统车企全面普及

新势力引领目前智能座舱演进。从 2012 年特斯拉引领了座舱中控大屏开始，汽车座舱由机械时代朝着智能时代演变，中控屏、液晶仪表盘、音响等基础视听硬件，以及车灯、座椅等均得到了全方位升级。到 2017 年蔚小理等新势力开始逐步从特斯拉手中接过智能座舱创新的交接棒，推动座舱生态由原来的行车空间向“第三生活空间”演变。目前我们看到理想（五屏三维交互、主卧大床、HUD 等）、小鹏（5D 影院、全场景语音等）、蔚来（AR 眼镜、NOMI 等）均走在座舱创新前列。

图表243：特斯拉、蔚小理智能座舱功能演进



资料来源：公司公告，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

燃油车逐步跟进座舱智能化方案。从车展新车迭代中，智能座舱的多模态硬件升级已经在23款车型开始向全燃油车，尤其是国外的传统车企演进，传统车企对于舱内屏幕、音响升级的意愿显著加强。包括保时捷Cayenne、宝马X1、本田雅阁、福特锐界L、北汽现代MUFASA、林肯航海家等均升级成大尺寸中控/仪表屏，部分车型甚至加装了副驾驶屏。音响方面，部分车型增加了扬声器数量；智能语音系统同样在快速迭代。

图表244：上海车展展出部分车型智能化情况

类型	车企	动力	车型	外观	视觉	听觉	系统/语音助手
传统车企	奔驰	燃油	2022款 GLC SUV		10.25英寸触控+控制器中控屏，无AR实景导航	选装Burmester柏林之声，最多13个扬声器	MBUX系统
		燃油	全新长轴距GLC SUV		11.9英寸全触摸中控屏，AR实景穿越导航，HUD	15个扬声器，Burmester柏林之声3D环绕立体声音响系统	第二代MBUX智能人机交互系统
	宝马	燃油	2022款X1		10.25英寸中控屏，无全液晶仪表盘，没有HUD	哈曼卡顿高端音响系统	iDrive8
		燃油	新一代X1		10.25英寸液晶仪表和10.7英寸中央显示屏的一体悬浮式曲面屏，HUD	哈曼卡顿高端音响系统	iDrive9
	本田	燃油	雅阁 2022款(2021年上市)		10.25英寸中控屏，没有HUD，没有液晶仪表盘	非品牌音响，8个扬声器	Honda CONNECT 3.0智导互联
		燃油	雅阁 2023款		豪华12.3英寸彩色智能屏互联系统，仪表盘10.2英寸彩色TFT多功能信息显示屏，HUD	BOSE豪华音响系统（12扬声器）	Honda CONNECT 4.0智导互联，AI语音助手
	大众	燃油	2022款 T-ROC 探歌		10.3英寸全液晶仪表盘；9.2英寸悬浮式中控屏；	Beats音响系统，6扬声器+1备胎池低音炮+功放	--
		燃油	2023款 T-ROC 探歌		10.3英寸全液晶仪表盘；9.2英寸悬浮式中控屏；后视影像	Beats音响系统，6扬声器+1备胎池低音炮+功放	ID.OS 2.0系统
	保时捷	燃油	Cayenne 2022年上市		12.3英寸中控屏，无副驾屏，无全液晶仪表盘	BOSE/Burmester，10个扬声器	--
		燃油	Cayenne 最新款		12.3英寸中控屏，选装10.9英寸副驾触显示屏，12.65英寸曲面显示屏仪表	BOSE/Burmester，10个扬声器	Active Safe Plus智能安全系统，Porsche Connect智能服务平台
	福特	燃油	锐界L 2022款		13.2英寸中控屏，8英寸液晶仪表盘	6个扬声器	AI智能语音交互系统，SYNC+智行信息娱乐系统
		燃油	锐界L 2023款		27英寸4K高清中控屏，12.3英寸全液晶数字仪表盘	8个高保真扬声器	AI智能语音交互系统，SYNC+2.0智行互联系统
	林肯	燃油	2022款航海家		3英寸全数字仪表盘，12.8英寸高清触控屏，5.8英寸二排中控	林肯高级音响系统(9个扬声器+1外置功放)	SYNC+3.0智行互联，智能语音助理
		燃油	全新一代航海家		双23.6英寸环抱式4K超清屏，11.1英寸触控屏	林肯高级音响系统(9个扬声器+1外置功放)	SYNC+3.0智行互联，智能语音助理

注：燃油车部分我们对比了展出的2023款新车（蓝色行）和上一代2022款旧车型

资料来源：公司官网，华泰研究

座舱硬件向多模态持续升级

#1 车载屏幕：

车载显示系统主要包括传统的中控屏、仪表盘、副驾驶/后排屏等，以及新兴的HUD抬头显示、E-Mirror/CMS电子后视镜。我们认为在座舱智能化中，车载显示是目前最核心受益环节之一。（1）从需求侧，车载显示是辅助行车、体验娱乐、车身控制的必不可少的视觉接口；（2）从供给侧，和机分离趋势下，高算力芯片、OTA给多屏化发展提供技术基础。我们认为座舱屏幕朝着多屏化、大屏化、高清化发展的趋势带来车载显示系统价值量提升。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表245：汽车显示屏分布及作用

资料来源：奔驰官网，蔚来官网，小鹏官网，现代官网，佐思汽研，华泰研究

传统车载显示的迎大屏、多屏、高清化趋势

1）中控屏和仪表盘：据高工智能汽车研究院， 2022 年 10 月国内乘用车中控屏渗透率已经高达 89.7%。展望 2030 年，我们认为受前装及后装市场对中控屏/仪表盘需求稳步推进的影响，中控屏/仪表盘仍有望实现持续增长。

2）副驾/后排等其他屏幕：随着座舱娱乐属性增强，中控屏一定情况下难以满足乘客需求。因此理想 L9/L8/L7 以及岚图梦想家、极氪 009 等部分车型推出副驾驶屏、后排屏，并利用多屏联动打造高科技座舱生态。

新型应用：HUD/电子 E-Mirror，从 0-1 阶段高速放量

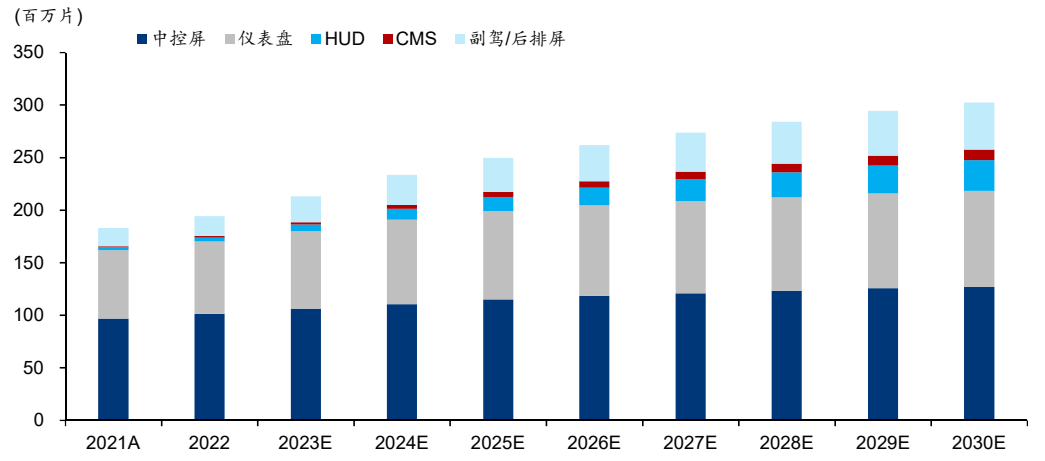
1）HUD（抬头显示）：通过将时速、导航等行车信息投影到驾驶员前的挡风玻璃上或驾驶员和挡风玻璃之间的透明表面上，让驾驶员尽量做到在不低头、不转头的情况下就能兼顾重要驾驶信息。目前 HUD 逐渐演化成了 C HUD、W HUD 和 AR HUD 三种路线。

2）E-Mirror/CMS（后视镜）：替代传统玻璃镜，驾驶员坐在驾驶室座位上直接通过显示屏获取汽车后方、侧方和下方等摄像头合成画面信息，减少恶劣天气干扰。乘用车上电子后视镜主要应用于：1）提供内后视野（也称流媒体后视镜）；2）提供外后视野（也称电子外后视镜）。2022 年 12 月 29 日于全国标准信息公众服务平台发布的《机动车辆间接视野装置性能和安装要求》允许电子外后视镜于 2023 年 07 月 01 日上市，法规有望助推电子后视镜渗透率实现从 0 至 1 提升。

整体来看，我们测算中控/仪表/HUD/CMS/副驾和后排屏 23-30E 市场规模 CAGR 将分别为 7.8/8.6/25.7/19.3/16.6%。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表246：2021-2030E 车载显示器出货量及预测

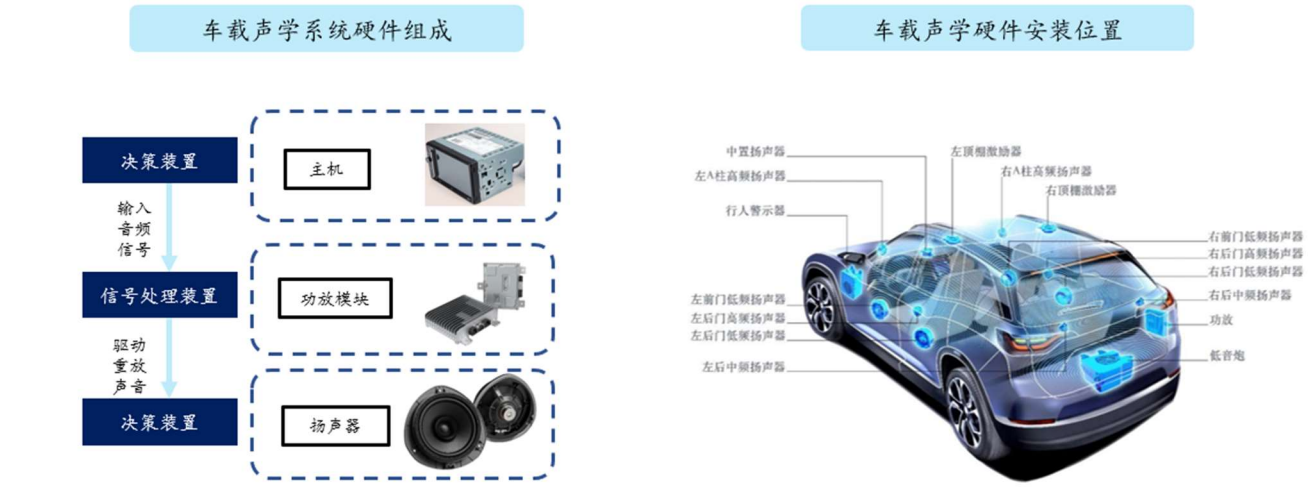


资料来源：Digitimes，华泰研究预测

#2 车载音响：

车载声学是座舱人机交互的重要方式。车载声学硬件系统主要由主机、扬声器、功放组成。1）主机作为声学系统的控制核心，主要负责音源的读取、输出以及开关静音等功能；2）功放是声学信号处理装置，负责将输入信号进行预处理后驱动扬声器重放声音。3）扬声器负责将电能转换为声能，是声学系统的执行装置。

图表247：车载声学硬件示意图

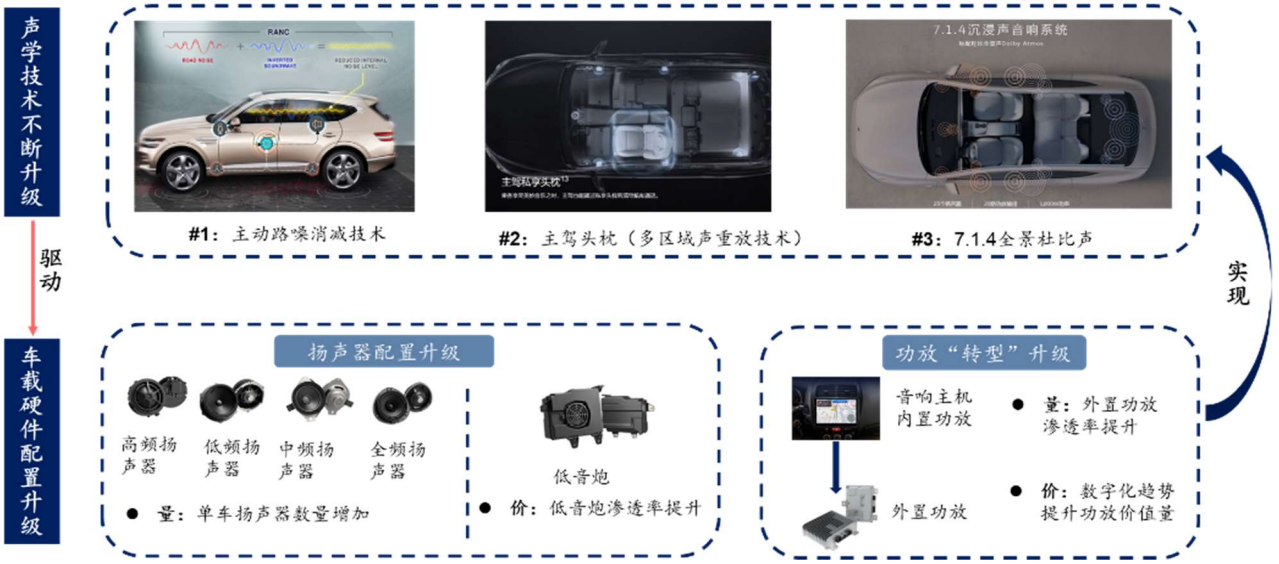


资料来源：上声电子招股说明书，华泰研究

声学技术升级驱动硬件配置升级。声学系统作为座舱娱乐体验的重要组成部分，整体呈现高端化发展趋势。未来 RNC、杜比全景声、多区域声重放等技术驱动扬声器和功放等执行元件升级，低音炮/数字功放有望加速上车。

更多免费行业报告 并购家 www.ipoipo.cn

图表248： 车载声学硬件升级



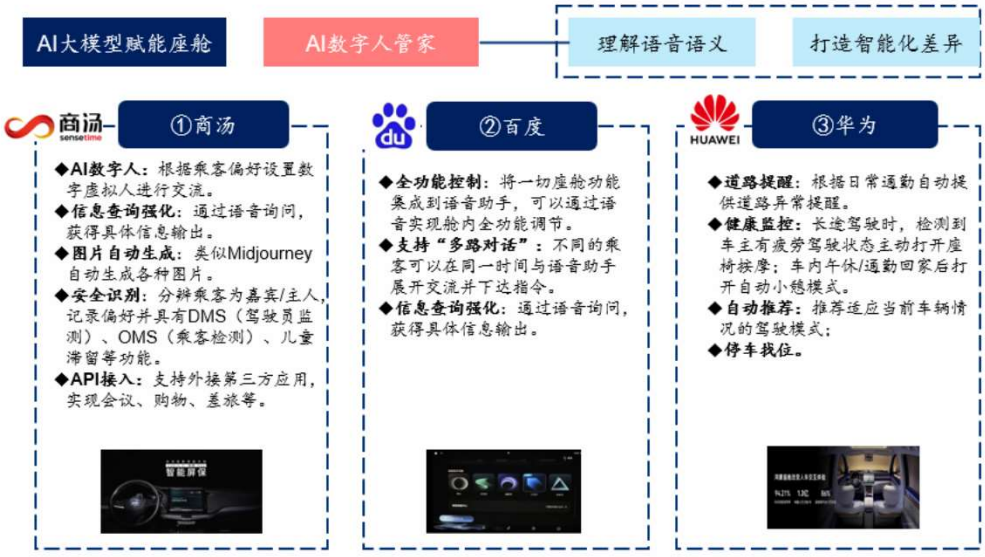
资料来源：上声电子招股说明书，华泰研究

座舱软件受益大模型赋能

硬件逐步上车，座舱交互性和操控便捷性的重要程度逐步凸显，车企开始探索舱内智能化交互，“语音助手”或成为核心交互载体。我们看到目前新势力在座舱交互上已经有所突破。例如智能化领先的小鹏语音助手小 P 可以支持多指令识别、离线响应。蔚来人车交互系统 NOMI 具有较强的学习功能和语音指令识别功能，此外还能根据不同状况表现出相对应的情感，与车主进行简单对话。而理想的“理想同学”同样属于可以由关键词触发,能够执行语音/手势/触控命令的初级智能 AI。

往后看随着 2023 上海车展上，商汤、百度、华为相继推出 AI 座舱大模型，或将加速原本的语音软件能力升级，推动座舱 5.0 人机主动式交互时代的到来。大模型有望：（1）通过赋能语音助手对于乘客的语音语义理解能力，提升全功能操控力。打通其在视觉、听觉、触觉等多模态应用上的操控力，形成深度的乘驾人机主动式互动体验。（2）打造智能生态差异化，摆脱座舱同质化。通过大模型赋能语音助手或将成为“AI 智能管家”，因算法不同而具有不同的特色和功能，进而形成车企自身独特的智能化差异。

图表249： 商汤、百度、华为宣布将 AI 大模型应用于智能座舱



资料来源：公司官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

部分环节国产供应链已经占优

我们梳理了智能座舱产业链代表公司，我们认为受益于国内智能座舱的率先兴起，部分产业链国内玩家在全球市占率中已经占据领先地位。

1）芯片：高通引领智能座舱芯片，TI、瑞萨、恩智浦成熟，国内玩家地平线、芯驰、芯擎、瑞芯微正奋起直追。

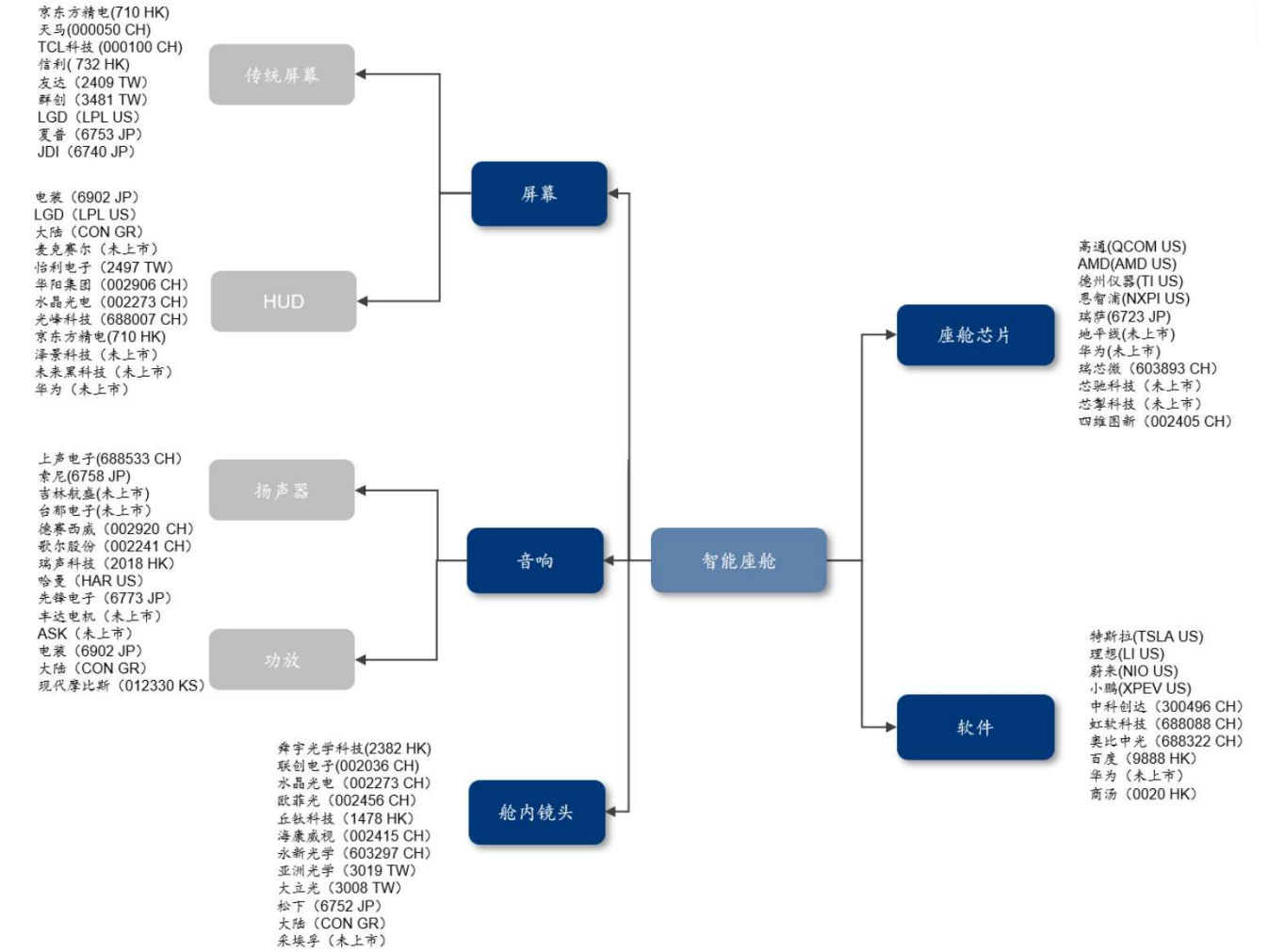
2）屏幕：受益于面板产业优势，国内车载显示玩家：京东方精电、天马占据了全球车载显示出货量前 2 名，其余主要对手包括华星、友达、群创、LGD、夏普等。

3）音响：上声电子居全球扬声器出货量首位，其余国内玩家有台郁电子、航盛，主要系与海外 ASK、丰达电机、哈曼、电装等竞争。

4）舱内摄像头：国内龙头舜宇稳居全球车载摄像头市场第一，随着舱内 TOF、DMS/CMS 镜头需求提升，国内光学玩家舜宇、联创、欧菲、水晶等具有领先技术优势。

5）软件：国内百度、商汤、华为、虹软、奥比中光在机器视觉、座舱 AI 模型上有先发优势。

图表250： 智能座舱产业链公司



资料来源：Bloomberg，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

车载芯片相关公司：地平线、黑芝麻智能、芯驰科技、芯擎科技，旗芯微

地平线（未上市）：行业领先的高效能智能驾驶计算方案提供商

地平线成立于 2015 年 7 月，是行业领先的高效能智能驾驶计算方案提供商。作为推动智能驾驶在中国乘用车领域商业化应用的先行者，地平线致力于通过软硬结合的前瞻性技术理念，研发极致效能的硬件计算平台以及开放易用的软件开发工具，为智能汽车产业变革提供核心技术基础设施和开放繁荣的软件开发生态，为用户带来无与伦比的智能驾驶体验。2020 年，地平线正式开启中国汽车智能芯片的前装量产元年，实现从 0 到 1 的突破。据地平线官网，目前地平线征程芯片累计出货量已突破 200 万片，与超过 20 家车企签下了超过 70 款车型前装量产项目定点，携手合作伙伴实现从 1 到 N 的价值共探。

黑芝麻智能（未上市）：车规级自动驾驶计算芯片和平台研发企业

黑芝麻智能成立于 2016 年，是行业领先的智能汽车计算芯片和平台研发企业，专注于车规级高性能计算芯片与平台技术领域的高科技研发。基于核心 IP 技术和芯片产品，黑芝麻智能提供完整的智能汽车、车路协同解决方案，支撑智能汽车产业链相关产品方案的快速产业化落地。公司和客户在 L2、L3 级 ADAS 和自动驾驶感知系统解决方案上开展了一系列商业合作，算法和图像处理等技术已在智能手机、智能汽车、智能家居等消费电子领域布局和商业落地。黑芝麻智能分别在武汉、硅谷、上海、成都、深圳、重庆以及新加坡成立研发及销售中心，目前已有超过 1000 名员工，核心团队均来自业内顶尖公司，平均拥有 15+年行业经验。

芯驰科技（未上市）：未来出行的探索者和创新者

芯驰科技成立于 2018 年，在上海、北京、南京、深圳、大连拥有研发中心，同时在长春、武汉设有办事处。芯驰专注于提供高性能、高可靠的车规芯片，也是全球首家“全场景、平台化”的芯片产品与技术解决方案提供者。产品覆盖智能座舱、智能驾驶、网关和 MCU，涵盖了未来汽车电子电气架构最核心的芯片类别，从而实现“四芯合一 赋车以魂”。芯驰拥有近 20 年车规级量产经验的国际水平团队，是国内为数不多的具有车规核心芯片产品定义、技术研发及大规模量产落地的整建制团队。目前芯驰已完成 4 个系列芯片的流片、最高规格车规认证及大规模量产上车，服务超过 260 家客户，覆盖了中国 90%以上车厂。

芯擎科技（未上市）：国内汽车智能座舱芯片领军企业

湖北芯擎科技有限公司于 2018 年在武汉经济技术开发区成立，在北京、上海和武汉均设有研发中心，是国内汽车智能座舱芯片领军企业。芯擎科技专注于设计、开发并销售先进的汽车电子芯片，以“让每个人都能享受驾驶的乐趣”为发展使命，致力于成为世界领先的汽车电子芯片整体方案提供商。芯擎科技具备多核异构超大规模 SoC 设计、自主设计多种创新核心 IP。其自研的国内首款 7nm 先进制程车规级智能座舱芯片，在设计、工艺和性能等方面对标海外厂商目前最先进的产品，实现了国产高端汽车芯片领域的技术突破。目前，该款座舱芯片“龙鹰一号”已通过全部车规测试，正式量产出货，所搭载的首款车型全新新能源中型 SUV 领克 08 于上海国际车展亮相，强大的 CPU、GPU、NPU 算力可流畅支持智能车机互娱需求，提供沉浸式座舱体验；同时，芯擎科技与全球汽车零部件巨头伟世通达成合作，搭载“龙鹰一号”的 SmartCore™ 座舱域控制器于上海车展对外展示，这也是伟世通首次推出采用“中国芯”的智能座舱域控制器解决方案。目前芯擎科技的第二款 7nm 高等级自动驾驶 SoC 也已开发超一年，年底即将流片；后续伴随整车电气架构演进所布局的车载中央计算芯片，也可与现有产品形成完整生态组合，对汽车智能化领域核心高算力芯片做到全面覆盖。

免责
更多免费行业报告 并购家 www.ipoipo.cn

旗芯微（未上市）：国内汽车与工业控制器领域领先厂商

苏州旗芯微半导体有限公司成立于 2020 年 10 月，基于 ARM Cortex M4、M7 等系列架构构建面向汽车不同应用场景的高性能、高可靠性的片上系统，开发智能汽车高端控制器芯片。目标填补国内新一代智能网联汽车控制器芯片领域空白，致力于发展成为中国汽车与工业控制器领域领导级厂商。公司通过采用自研 IP，多核锁步等技术以及应用于车规芯片的六西格玛模拟电路设计流程，设计出覆盖安全标准 ISO26262 ASIL-B 至 ASIL-D 的全系列产品家族。公司产品均满足车规 AEC-Q100、功能安全标准 ISO26262 以及各项车规可靠性测试，可广泛应用于车身、汽车仪表、安全、动力、电池管理等领域。公司核心团队研发人员拥有平均超过 18 年的车规芯片设计经验，是国内唯一完整开发过车规级 8/16/32 位控制器的顶级研发团队。目前公司在苏州、上海、北京等地设有研发中心及办公室。

域控及数据标注相关公司：德赛西威、经纬恒润、纵目科技、魔视智能、宏景智驾、柏川数据

德赛西威（00290 CH）：大算力域控制器龙头

德赛西威成立于 1986 年，是国际领先的移动出行科技公司之一，致力于成为出行变革的首选伙伴。德赛西威深度聚焦于智能座舱、智能驾驶和网联服务三大领域的高效融合，持续开发高度集成的智能硬件和领先的软件算法，为全球客户提供安全、舒适、高效的移动出行整体解决方案和服务。2022 年，公司营业总收入 149.3 亿元，同比增长 56.0%，归母净利润 11.8 亿元，同比增长 42.1%。

经纬恒润（688326 CH）：汽车电子平台型公司

经纬恒润成立于 2003 年，专注于为汽车、无人运输等领域的客户提供电子产品、研发服务和高级别智能驾驶整体解决方案。总部位于北京，并在天津、南通建立了现代化的生产工厂，形成了完善的研发、生产、营销、服务体系。本着“价值创新、服务客户”的理念，公司坚持 “专业聚焦”、“技术领先”和“平台化发展”的战略，致力于成为国际一流综合型的电子系统科技服务商、智能网联汽车全栈式解决方案供应商和高级别智能驾驶 MaaS 解决方案领导者。经纬恒润长期供应国内外知名整车制造商和一级供应商，产品类型包括智能驾驶电子产品、智能网联电子产品、车身和舒适域电子产品、底盘控制电子产品、新能源和动力系统电子产品。2022 年，公司营业总收入 40.2 亿元，同比增长 23.3%，归母净利润 2.3 亿元，同比增长 60.5%。

纵目科技（未上市）：领先的智能驾驶产品及技术供应商

纵目科技成立于 2013 年，是一家基于中国面向全球的世界一流汽车先进驾驶辅助系统（ADAS）的软硬件方案提供商。总部位于上海张江国际科创中心，在上海、北京、厦门、深圳、重庆、美国密西根 Novi 市以及德国斯图加特都设有研发中心，生产制造中心位于厦门、湖州、东阳（在建）。纵目科技通过领先的系统能力和首屈一指的量产实力，成为国内率先获得整车厂 L4 级别量产项目定点合同的自动驾驶企业之一，并与一汽红旗、长安汽车等多家国内主流主机厂商建立了量产合作关系。

魔视智能（未上市）：嵌入式人工智能自动驾驶赋能者

魔视智能成立于 2015 年，是一家专注于自动驾驶的创新科技公司。魔视智能以领先的人工智能算法，赋能嵌入式芯片平台，实现自动驾驶和高级辅助驾驶产品的大规模量产。魔视智能拥有自主知识产权的全栈式自动驾驶核心技术，包括环境感知、多传感器融合、高精度的车辆定位、路径规划、车辆控制、驾驶决策等所有核心算法，支持 L1-L4 级自动驾驶。魔视智能独有的深度学习框架，全面支持国际及国内六大主流嵌入式芯片平台，实现高度优化及精准的人工智能引擎。魔视智能自主研发的自动驾驶和高级辅助驾驶产品，涵盖乘用车及商用车、行车及泊车、舱内及舱外、前装及后装等主流市场，以百万套的量产规模成为当之无愧的行业引领者。魔视智能是率先实现国内一线乘用车主机厂量产的自动驾驶企业之一，业已与国内几乎所有头部乘用车及商用车主机厂建立量产及项目合作关系，并在主动安全市场占据主要市场份额。魔视智能公司总部位于上海张江，在澳大利亚设有人工智能研究院，在深圳、武汉和苏州设有研发中心，在南通设有制造基地。

免责

更多免费行业报告 并购家 www.ipoipo.cn

宏景智驾（未上市）：全栈式自动驾驶解决方案服务商

宏景智驾成立于 2018 年，总部位于杭州，在北美、上海和北京分别设有技术中心。公司致力于成为全球智能驾驶科技领航者，是一家全栈式自动驾驶解决方案服务商，具备全栈自动驾驶软件算法和完整的系统集成能力，可针对不同客户需求提供定制化的高性能智能驾驶解决方案，全周期赋能 L1-L4 级别智能驾驶。主要产品是软硬一体自动驾驶计算平台(ADCU – Autonomous Driving Computing Unit)、智能摄像头产品，产品覆盖硬件、感知、定位、融合、规划与控制等自动驾驶全研发链。目前，宏景智驾已在上汽、长城、奇瑞、江淮、比亚迪、合众等 OEM 共计 30 个车型上实现量产和定点。同时，宏景智驾还获得了顶级风险投资机构包括华登、高瓴、线性、蓝驰、沙特阿美 Prosperity 7、中信金石等投资。

柏川数据（未上市）：人工智能数据服务领域头部服务商

柏川数据成立于 2021 年，是人工智能数据服务领域头部服务商之一。专注于为包括主机厂、科技公司、AI 公司等数十家产业链相关方，提供集数据采集、数据清洗、数据标注、数据管理、数据存储于一体的一站式服务。在国内 6 个动态区域管理超千人的标准化采标管存团队，并获得乙级测绘资质以及 ISO9001 和 ISO27001 双重认证。柏川数据自研数据处理平台具备高精度 AI 预标注功能，配置丰富多样的标注工具，支持工作进程的可视化、灵活性、自定义管理，全面覆盖多种数据标注场景，在大幅降低标注成本的同时，为客户带来数智升级的全新服务体验。同时，柏川数据正加速布局自动化标注、数据集、仿真数据库等前沿技术，致力于实现标注数据的仿真场景还原和仿真测试，以及端到端算法的回灌测试，最终打通自动驾驶数据闭环，加速自动驾驶规模化量产落地。客户方面，柏川数据持续为百度、地平线、MOMENTA、挚途科技、MAXIEYE、天翼交通、天瞳威视、主线科技、清智科技、交控科技等 30 家+企业提供规模数据支撑。立足自身强大的数据服务实力，坐拥相城区高铁新城赋予的产业集群优势，柏川数据旨在成为国内最大的自动驾驶数据服务商，并积极拓展海外业务，为更多的主机厂及自动驾驶算法公司提供更前沿、更深度、更全面的智能驾驶数据解决方案，助力未来智慧出行进入发展“快车道”。

连接器相关公司：电连技术、瑞可达

电连技术（300679 CH）：深耕连接器行业，领军汽车高速连接

电连技术成立于 2006 年，是一家专业从事连接器、连接线，天线以及电磁屏蔽产品研发和制造的供应商，同时为电子设备提供一站式射频解决方案，产品广泛应用于消费电子、智能物联、通信设备及基础设施、移动终端和汽车电子等领域。电连技术拥有从组件到系统级别的大容量射频产品设计和交付能力。电连技术的微型射频连接器和汽车连接器分别在中国的智能手机市场和车载电子市场占有领先的份额，在全球市场占有重要地位。2022 年，公司营业总收入 29.7 亿元，同比增长 8.5%，归母净利润 4.4 亿元，同比增长 19.3%。

瑞可达（688800 CH）：高压连接器龙头，配套客户切入高速连接器

瑞可达成立于 2006 年，专注于连接系统产品的设计开发和制造，系国内知名连接器生产制造商，行业地位名列前茅。瑞可达与全球的客户和伙伴开展广泛业务合作，在苏州、无锡、绵阳、成都、武汉、西安、墨西哥、美国等地设有工厂或研发中心，在美国、德国、瑞典等多地设立海外销售办事处。主要产品分为连接器类、线缆组件类、系统模块类等，广泛应用于数据通信、新能源汽车、储能/清洁能源、工业控制、医疗设备、轨道交通装备等领域。公司产业链完整，从模具设计和制造、连续冲压、注塑成型、机械加工、精密压铸等生产，拥有强大的在线检测能力。2022 年，公司营业总收入 16.3 亿元，同比增长 80.2%，归母净利润 2.5 亿元，同比增长 121.9%。

更多免费行业报告 并购家 www.ipoipo.cn

传感器相关公司：加特兰、森思泰克、行易道、禾赛科技、速腾聚创、图达通、探维科技、灵明光子

加特兰（未上市）：毫米波雷达 MMIC 芯片开发与设计的领导者

加特兰微电子科技创立于 2014 年，是 CMOS 工艺毫米波雷达芯片开发与设计的领导者。公司汇聚了射频毫米波电路设计、雷达系统算法研发、大规模数字电路实现、高频天线设计以及汽车级芯片量产运营等领域的顶尖人才，并于 2017 年成功量产了全球首个汽车级 CMOS 工艺 77/79GHz 毫米波雷达射频前端芯片，率先实现了在汽车前装市场的突破。随着汽车自动驾驶和驾驶辅助技术的快速发展，高性能、易开发、小型化成为毫米波雷达发展的热点和趋势。加特兰率先推出了集成雷达基带处理的 SoC 芯片，为毫米波雷达传感器的开发实现带来了全新的变革。公司又进一步推出了 AiP 技术，通过在芯片封装内部集成天线阵列，减少用户天线设计和高频板材投入，并大幅缩短模块研发和生产周期，加速毫米波雷达在汽车和行业市场的普及。

公司坚持汽车级可靠性和安全性的设计理念，通过了 ISO 9001 质量体系认证和 ISO 26262 功能安全管理体系认证。产品满足 AEC-Q100 的可靠性规范，包含计算单元的 SoC 产品达到 ASIL-B 的功能安全等级。

2022 年 12 月 20 日，加特兰举办“Next Wave”Calterah Day 活动，发布了下一代全新毫米波雷达 SoC 芯片系列——Andes，代表了目前毫米波雷达技术的前沿。Andes 系列芯片可实现 4D 高端雷达以及成像雷达功能。加特兰 CEO 陈嘉澍博士在新品发布会上提到，车载毫米波雷达有三大发展趋势：小尺寸低功耗雷达、高性价比雷达和 4D 成像雷达。加特兰拥有全面的毫米波雷达芯片产品组合，可满足三大雷达发展趋势的需求。截至 2022 年 12 月 20 日，加特兰已服务超 400 家客户，与 20 余家 OEM 车企达成合作，赋能 70 余款乘用车，累计汽车芯片出货量超过 300 万片。

森思泰克（未上市）：国产毫米波雷达品牌“小巨人”企业

森思泰克注册于 2015 年，是一家专业从事毫米波与激光雷达传感器智能产品研发、生产与销售的高科技企业，现已形成石家庄、北京、西安、芜湖、杭州、韩国等多地区多位一体的发展布局。公司具备 ISO9001 和 IATF16949 双体系认证，拥有从算法到高频硬件、软件完整的系统设计能力及量产工艺核心技术，获得多项自主知识产权并成功实现科技成果转化，先后研发出多系列高性能雷达产品，主要面向车载安全与自动驾驶、智能交通与智慧停车、安防监控、无人机等领域。产品涵盖 24GHz、77GHz、79GHz 所有频段，并融合相控阵、DBF、MIMO、稀疏阵、超分辨等科技前沿的雷达技术。凭借丰富的产品线与出色的研发能力，森思泰克已成为多家国内外一线车厂的体系内供应商。

行易道（未上市）：聚焦高端车载成像雷达，自研技术助 4D 雷达实现行业领先

行易道由中科院赵捷博士创立于 2014 年，9 年来深耕毫米波车载前向 AEB 雷达，在毫米波雷达新体制、雷达信号处理新算法、车载合成孔径雷达（SAR）技术等领域已获得丰富的项目经验和技术积累，先后通过 ISO14001 和 IATF16949 体系认证，并获得北京市“高新技术企业”和“专精特新”中小企业等资质认定，并通过知识产权管理体系认证。

产品线目前涉及车载前向雷达、角雷达、交通雷达、物联网监测雷达等系列产品。公司在北京、上海、深圳、韩国等地设有分公司，并在北京、盐城两地拥有获得 IATF16949 体系认证的高品质产线，年产能达数百万套。

2020 年，商业上实现韩国、东南亚跨区域的 AEB 功能前装定点和交付，并通过 EN-CAP 测试，韩国 KC 认证和越南 MIC 认证，ISO17387 标准测试，成为率先出口的前装车载 AEB 雷达企业。

免责
更多免费行业报告 并购家 www.ipoipo.cn

2022 年，推出第三代 4D 成像雷达产品：ALRR300，仅用双片级联实现角分辨率均小于 1°（同期可量产 4D 雷达在 1.5°*2°以上），该产品基于稀疏阵列天线设计、压缩感知超分辨算法等创新技术，产品力领先行业，同时尺寸、功耗、成本方面均得到优化，完美解决了当下智能驾驶技术发展中传感器应用的瓶颈问题，并同年实现韩国车企定点，2023Q2 实现 SOP。

原创技术方面，2019 年车载 SAR（合成孔径雷达）技术获评“全球新能源汽车前沿技术”；2020 年，获得中国雷达协会科技进步二等奖。2020 年，获得北京市科委资助项目“支持高级别自动驾驶的成像毫米波雷达研制”，于 2022 年底验收通过；至今，行易道已实现了 17 项发明专利、100+项实用新型和外观设计专利，11 项软著，发表 3 篇 IEEE 论文，获得良好的业内口碑和影响力。

禾赛科技（HSAI US）：全球领先的激光雷达制造商

禾赛科技成立于 2014 年 10 月，是一家全球化的激光雷达研发与制造企业。公司产品广泛应用于支持高级辅助驾驶系统（ADAS）的乘用车和商用车，以及自动驾驶汽车。禾赛的激光雷达技术也致力赋能各类机器人应用，例如无人配送车和封闭区域内的物流机器人等。禾赛在光学、机械、电子、软件等激光雷达核心领域有着卓越的研发能力和深厚的技术积累，其激光雷达产品已成功经过市场验证，截至 2022 年底累计交付量超过 10 万台。禾赛将激光雷达的制造工艺融入到研发设计流程中，在促进产品快速迭代的同时，保证高性能、高可靠性与高性价比。禾赛的客户包括全球主流汽车厂商、自动驾驶和机器人公司，遍及全球 40 个国家、90 多个城市。

速腾聚创（未上市）：激光雷达硬件、感知软件与芯片三大核心技术闭环

速腾聚创(RoboSense)成立于 2014 年 8 月，是全球领先的智能激光雷达系统（Smart LiDAR Sensor System）科技企业。RoboSense 通过激光雷达硬件、感知软件与芯片三大核心技术闭环，为市场提供具有信息理解能力的智能激光雷达系统，颠覆传统激光雷达硬件纯信息收集的定义，赋予机器人和车辆超越人类眼睛的感知能力，守护智能驾驶的安全。RoboSense 总部位于深圳，企业员工超过 1200 人，拥有来自全球顶尖企业和科研机构的人才团队，为 RoboSense 提供源源不断的创新能力，截止 2022 年底，全球布局激光雷达相关专利超过 1000 项。RoboSense 产品技术的领先建立在多学科多层级的技术积累之上。公司以市场为导向，为客户提供多种的智能激光雷达系统解决方案，产品技术包括：MEMS 与机械式激光雷达硬件，硬件融合技术，感知软件等。

图达通（未上市）：深耕图像级超远距激光雷达领域

图达通(Innovusion)成立于 2016 年，是全球图像级激光雷达提供商。在硅谷、苏州和上海设有研发中心，在宁波和苏州拥有高度工业化的车规级激光雷达制造基地。Innovusion 潜心致力于激光雷达的探索与创新，不断提供高性能的激光雷达产品及多样化的应用解决方案。其图像级超远距激光雷达猎鹰 (Falcon) 已于 2022 年作为蔚来 ET7、ES7 及 ET5 的 Aquila 超感系统标配量产交付。依托产品技术的先进性和成本的可控性，Innovusion 在国内外还与多家车联网、智慧高速、轨道交通以及工业自动化等行业龙头企业开展着积极合作，用先进的硬件及软件解决方案赋能安全与智能生活。

探维科技（未上市）：新一代固态扫描激光雷达技术方案和 ALS 平台技术

探维科技成立于 2017 年，核心团队起源于清华大学精仪系国家重点实验室，早在 2008 年即开始从事激光雷达系统方向的技术研发。探维科技开发了行业唯一的硬件级图像前融合产品 Tanway Fusion，解决了多传感器融合的可靠性问题，深入挖掘激光雷达在车端的应用价值与用户价值，开启激光雷达 3.0 时代。为了实现车规级稳定性，探维科技基于全自研的 ALS 平台技术，打造了高性能、低成本、易量产的车规级固态激光雷达，并于 2019 年率先完成量产交付。探维科技总部和研发中心位于北京，在成都落地电子研发中心，2021 年底在苏州落地车规级产线。2022 年探维科技获得 IATF16949 质量管理体系认证，与合创汽车联合发布全球首款搭载激光雷达的量产 MPV-V09，今年内计划落地 3 个乘用车定点项目。

免责
更多免费行业报告 并购家 www.ipoipo.cn

灵明光子（未上市）:dToF 传感芯片和系统解决方案提供商

灵明光子是由顶尖海归博士团队于 2018 年 5 月创立于深圳南山。灵明光子设计的单光子雪崩二极管（SPAD），是帮助现代电子设备实现 3D 感知的核心器件，广泛赋能汽车、智能手机、机器人、自动控制、人机交互、智慧家居等领域。公司提供一系列的 SPAD dToF 传感芯片产品，包括:SiPM、3D 堆叠 dToF 模组、有限点 dToF 传感器等，其产品拥有行业领先的精准度、能效比和测距范围。灵明光子使用了最前沿的 BSI 3D 堆叠技术——将背照式的传感芯片晶圆和数字逻辑电路晶圆混合键合以达到行业内最高的性能效果和集成度，并提供以 SPAD 为基础的 dToF 系统解决方案。

AR-HUD 相关公司：泽景科技、未来黑科技

泽景科技（未上市）：国内领先的车内视觉解决方案供应商

泽景成立于 2015 年，是一家集科研、生产、服务于一体的国家高新技术企业，主要产品包括 W-HUD、AR-HUD、CMS、透明 A 柱、透明窗口显示等智能座舱相关产品。在全国已经布局仪征（总部和制造中心）、西安（研发中心）、上海（管理及客户服务支持中心）、长春（北方服务中心）、重庆（西南服务中心）5 个区域。泽景始终致力于用科技手段提高智能汽车的数字化体验，已然成为目前国内领先的车内视觉解决方案供应商。同时泽景是 HUD 产业本土化的引领者，建立了中国本土第一条 WHUD 自动化生产线，取得了第一款中国本土供应商开发的 WHUD 的产品量产等多项成就。公司已获得多家头部主机厂的认可，在技术、工程、量产经验和客户粘性等方面壁垒明显，拥有蔚来、北汽、奇瑞等量产客户。

未来黑科技（未上市）：搭建虚拟与现实世界的沟通桥梁

未来（北京）黑科技成立于 2016 年，是一家集光学成像、增强现实、人机交互设计（HMI）、全车窗显示和计算机算法研发于一体的创新型高科技公司，产品覆盖经典 HUD、光场 AR HUD、全景显示器和全息影像等。FUTURUS 未来实验室结合基础科学研究和应用技术创新，发明了全息光场技术和光场增强现实显示技术，创造了人与现实和虚拟世界同步沟通的全新窗口。截止目前，FUTURUS 产品已累计荣获 4 座 CES/CES Asia 创新奖奖杯，主要为宝马、上汽、理想汽车等品牌提供产品支持与技术供应。

免责

更多免费行业报告 并购家 www.ipoipo.cn

商业航天：大年开启，关注行业机遇

我们认为“商业航天”是中国企业有机会在全球占据重要一席的重要硬科技赛道之一。近一年来，我们观察到我国商业航天产业链上下游发生明显变化。在技术端，国内民营液体火箭公司积极探索发动机技术与可回收技术突破，国内卫星制造在国家统筹规划、民企合力组建下步入规模化发展。在需求端，我国多项星座计划有序开展部署，中国星网预计明年上半年开始发射；“卫星直连”成为行业热点，手机、汽车、模组厂商积极开展尝试；工信部提出加快 6G 研发，为卫星互联网提供重要机遇。在政策面，国内多省市利好政策频出，助力商业航天产业成长。建议关注我国商业航天产业链发展机遇。

技术端：火箭、卫星制造技术均有突破

过去一年中，海外公司引领火箭技术前沿，开展多项意义重大的技术试验。海外方面，美国 SpaceX 星舰虽然首飞失败，但实现了推力结构、发动机技术、控制辅助系统设计等多种技术创新，在加快全球卫星互联网建设等方面呈现出广阔想象空间；人族一号则首次验证了 3D 打印制造的火箭发动机及其他相关组件的适用性。

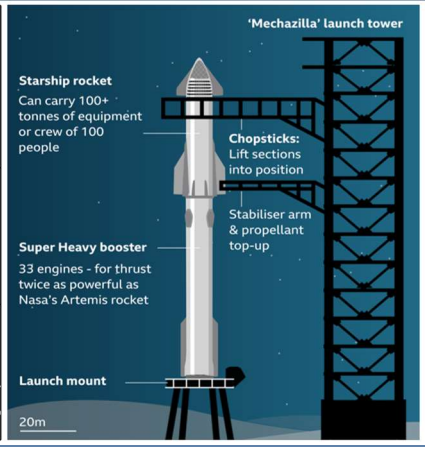
以 SpaceX 为代表的海外企业引领商业火箭技术前沿

“星舰”首飞失败但意义重大，技术创新与应用前景值得关注。2023 年 4 月 20 日，SpaceX 在位于得克萨斯州博卡奇卡的星基地对“超重（Super Heavy）B7+星舰（Starship）S24”组合体进行了轨道试飞。火箭在起飞后约 4 分钟意外解体，试飞所收集的火箭、发动机、计算机和地面系统性能等数据，将用于“星舰”设计模型改进与后续研发迭代。此次试飞意义重大，“星舰”不仅在此次飞行中实现了推力结构、发动机技术、控制辅助系统设计等多种技术创新，还呈现出广阔的想象空间，未来有望大大促进商业航天应用。

“星舰”是“人类历史上体积最大、推力最强”的超重型运载火箭，在推力结构、发动机技术、控制辅助系统设计等关键技术上均有突破。首飞的“星舰”采用“猛禽”二代发动机，推力由一代 185 t 提升至 230 t，采用小贮箱推进剂管理模式，实现推进剂灵活调控。为增强系统可回收性能，“星舰”还采用了新型起降导航与控制技术、返回热防护、原位发射与回收技术，更新了外形设计与增压、航电、热交换等辅助系统。此次飞行中，“星舰”实现了 33 台液氧甲烷发动机的并联，通过协同调整每台发动机的推力，在最多有六台发动机失效的情况下，仍在相当长时间内保持了火箭的正常飞行姿态。

图表251：Starship 星舰基本参数与结构示意图

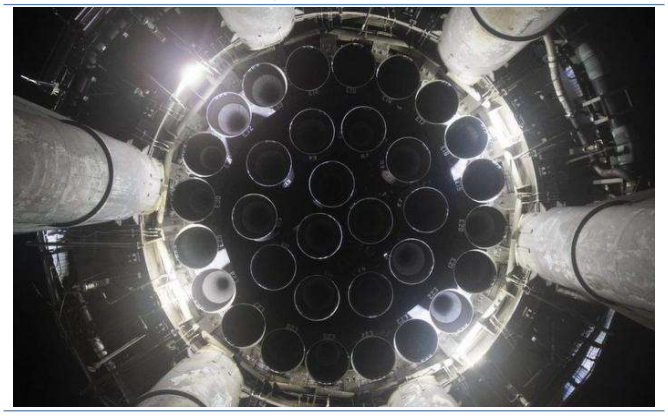
火箭总体	高度	119m
	直径	9m
	重量	5000t
	近地轨道有效载荷重量	150t
	近地轨道有效载荷体积	1000m³
	推进剂	液氧、液态甲烷
二级星舰飞船 Starship	高度	50m
	直径	9m
	空重|总重|推进剂重量	100t|1300t|1200t
	发动机	猛禽发动机3台，真空发动机3台
	最大推力	14700000N
一级超重助推器 Super Heavy	高度	69m
	直径	9m
	空重|总重|推进剂重量	200t|3600t|3400t
	发动机	猛禽发动机33台
	最大推力	74500000N



资料来源：SpaceX 官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表252：33 台猛禽二代液氧甲烷发动机并联



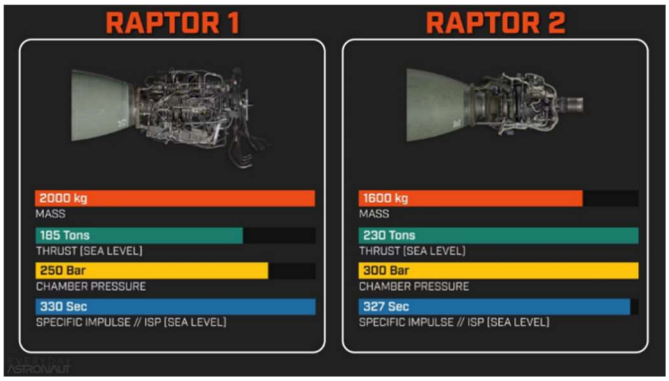
资料来源：BBC，华泰研究

图表253：直播画面中最多有 6 台发动机同时失效

资料来源：SpaceX 官网，华泰研究

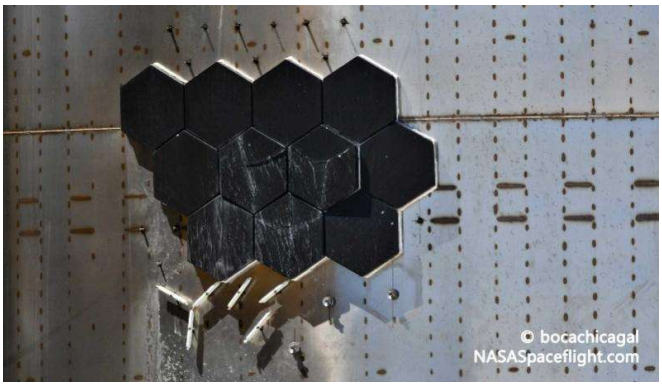
作为 SpaceX 最新一代的重型运载火箭，“星舰”强大的运载能力和低廉的发射成本有望为第二代“星链”部署提供强大支持。随着卫星互联网的迅速发展，SpaceX 正积极扩大其第二代“星链”星座规模，以满足偏远地区和军方用户的需求。“星舰”设计最大运载能力达 150 吨，相较于猎鹰九（最大运载能力 22.8 吨）和猎鹰重型火箭（最大运载能力 64 吨），运载能力实现数倍提升。同时，“星舰”采用全新可重复使用技术，相较于传统的一次性火箭发射成本显著降低，这将进一步加快“星链”部署，进而加速全球卫星互联网建设。

图表254：猛禽一代与猛禽二代发动机性能对比



资料来源：航空学报 CJA，华泰研究

图表255：“星舰”抗氧化陶瓷隔热瓦



资料来源：NASA Space Flight 官网，华泰研究

近年来受军事用途和太空经济等因素驱动，超重型火箭研制得到全球航天大国重视，“星舰”的成功或将加速全球航天竞赛。作为 NASA 月球着陆器唯一中标方案，“星舰”有望服务美国“阿尔特弥斯计划”，承担在月球运输任务，延展人类深空探索能力。“星盾”推出后，SpaceX 计划为美军提供全球覆盖的实时卫星图像数据与各类军事载荷平台，“星舰”的成功将大大降低美国 SDA 军用低轨卫星部署成本。此外，其强大的运输能力对于打造全球 1 小时“点对点”直达的火箭运输系统，提高美军快速响应运输、应对突发事件、长期在轨驻留等皆具有广阔军事价值。美军航天能力的提升势必对全球其他航天强国产生压力，促使各国在航天领域投入更多关注和资源。

相比于传统航空运输，星舰火箭运输系统具有载荷量大、系统可回收与载荷成本低廉等优势。星舰设计运载能力达到 100 吨，可一次运送两辆 M1 坦克，具备快速战略运输的能力，超过 C-17 的 77 吨，达到与 C-5 相当的运载能力。星舰采用垂直降落回收，相比 C-5 和 C-17 省去了跑道，可在非常规地面降落。据马斯克在 2022 年博卡奇卡 SpaceX 星舰发布会介绍，星舰每次飞行只需几百万美元，同时飞船可以将 100 到 150 吨有效载荷送上轨道，并可能携带 200 吨推进剂到轨道上。以每次发射成本 1000 万美元，载荷 100-200 吨计算，SpaceX 星舰的入轨载荷成本为 50-100 美元/千克。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表256：Starship 载人着陆系统电梯与气闸原型照片

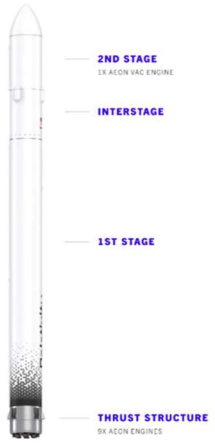
资料来源：NASA，华泰研究

人族一号 85%的零件均为 3D 打印，增材制造技术可能重塑火箭供应链。人族一号（Terran 1）是相对论空间自 2017 年以来开发的一种一次性两级小型运载火箭，于 2023 年 3 月 23 日首次发射升空，最后未进入飞行轨道。2023 年 4 月 12 日，相对论空间宣布人族一号退役。关于人族一号的故障，目前的调查显示 Aeon 真空发动机阀门打开的速度低于预期，导致燃料未能在点火时及时抵达燃烧室和燃气发生器。同时该发动机的氧气泵启动时未能成功增压，从而导致二级未能达到满推力。

此次发射，首次向全世界展示了全 3D 打印的火箭可以承受最恶劣的轨道发射条件。对 3D 打印技术应用与增材制造行业而言是一个重要的里程碑。此前民营航天公司 Launcher、Rocket Lab 等均已采用基于激光的粉末床熔融 3D 打印技术制造火箭发动机与零部件，但像人族一号这样大规模使用 3D 打印技术则是首次。人族一号火箭占质量 85%以上的零件均由 3D 打印制造，包括火箭整个箭体和几乎所有发动机部件。火箭成功通过 Max-Q 阶段，意味着 3D 打印的箭体结构在飞行最大动压力下的安全性得到了验证；其次，火箭一级飞行正常，一二级分离正常，3D 打印制造的火箭发动机及其他相关组件的适用性得到了验证。

图表257：人族一号火箭基本参数与结构示意图

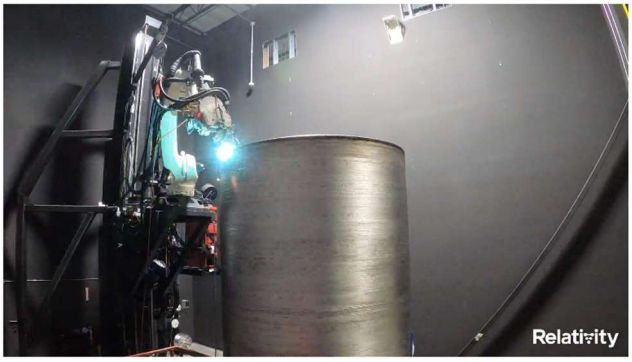
火箭总体	高度	35.2m
	直径	2.3m
	近地轨道有效载荷重量	1479kg
	太阳同步轨道有效载荷重量	898kg
	推进剂	液氧、液态甲烷
二级	高度	8.1m
	直径	2.3m
	发动机	Aeon一号真空发动机1台
	推力	126kN
一级 Block2	发动机	Aeon一号发动机1台
	推力	1300kN
一级	高度	24.3m
	直径	2.3m
	发动机	Aeon一号发动机9台
	推力	920kN



资料来源：Relativity Space 官网，华泰研究

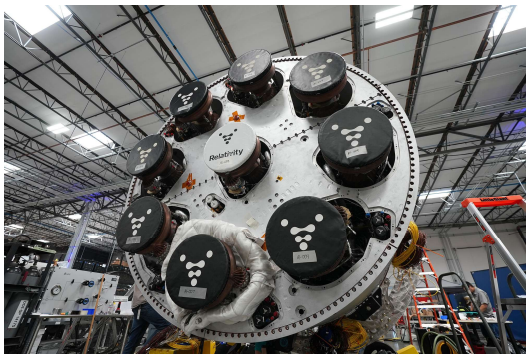
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表258：人族一号 85%的零件使用 3D 打印技术打印



资料来源：Relativity Space 官网，华泰研究

图表259：人族一号一级 9 台 Aeon 一号发动机



资料来源：Bloomberg，华泰研究

对于火箭制造，3D 打印技术的价值在于：大型结构一体化快速制造，复杂结构集成化快速制造和轻量化结构制造。其中，大尺寸构件直接制造则无需模具，能带来制造环节、人员参与、成本投入和材料浪费的减少；复杂结构集成化快速制造，除带来上述的优势之外，还具备安全性提升、部件性能提升以及迭代效率提升的优势；而轻量化制造的显著特点是大幅降低发射成本，从而使火箭能携带更多载荷。以火箭燃烧室的制造为例，根据 NASA 测算，使用 3D 打印激光粉末床熔融技术（L-PBF）能将工期从传统制造的 18 个月降低至 5-8 个月，相应成本也能降低 35%-60%。**因此，3D 打印技术能为火箭制造带来明显的成本优势。**

图表260：增材制造工艺的成本节约——以火箭燃烧室为例

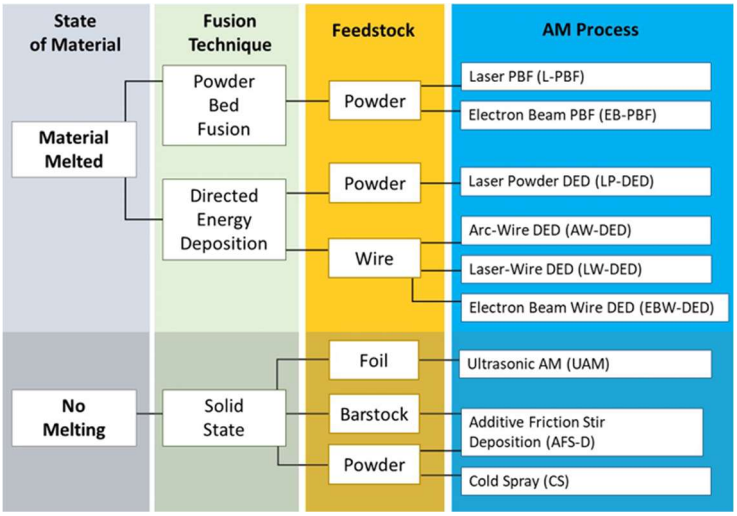
Category	Traditional Manufacturing	Initial AM Development	Evolving AM Development
Design and Manufacturing Approach	Multiple forgings, machining, slotting, and joining operations to complete a final multi-alloy chamber assembly	Four-piece assembly using multiple AM processes; limited by AM machine size. Two-piece L-PBF GRCop-84 liner and EBW-DED Inconel 625 jacket	Three-piece assembly with AM machine size restrictions reduced and industrialized. Multi-alloy processing; one-piece L-PBF GRCop-42 liner and Inconel 625 LP-DED jacket
Schedule (Reduction)	18 months	8 months (56%)	5 months (72%)
Cost (Reduction)	$310,000	$200,000 (35%)	$125,000 (60%)

资料来源：NASA 官网，华泰研究

目前，NASA 已在航天领域 3D 打印技术应用方面进行整体布局，对不同尺寸和结构的火箭零部件进行全面 3D 打印工艺探索。目前新材料和现有合金在不断开发优化中，很多合金仅处于开发阶段，不完全符合使用特定增材制造工艺的航空航天应用要求。**我们认为，未来 3D 打印为火箭制造带来的优势很可能发生连锁反应，革新火箭制造方式，重塑供应链。**

更多免费行业报告 并购家 www.ipoipo.cn

图表261：NASA 布局各种金属增材制造工艺



资料来源：NASA 官网，华泰研究

国内企业有望在火箭发动机技术与可回收技术等方面实现突破

2022 年以来，我国国内液体火箭技术路径进行了以蓝箭航天、天兵科技为代表的技术探索，2023 年有望实现进一步突破。我国民营火箭公司技术路径分固体、液体两种，其中液体发动机能够实现可回收。固体燃料发动机的优势在于技术难度较低、制造成本较低、可靠性高，但运载能力低，且无法实现可回收。液体燃料火箭研发的技术难度较大，但在比冲、可回收性上有固体火箭不可比拟的优势，能够在中大型固体火箭研制短期难以突破的情况下，满足卫星组网快速增加的大运力需求。

蓝箭航天“朱雀二号”执行首次发射，国内液氧甲烷发动机技术未来可期。2022 年 12 月 10 日，蓝箭航天朱雀二号（ZQ-2）运载火箭在酒泉卫星发射中心执行首次飞行试验任务。12 月 15 日，蓝箭航天在官微发布消息称，火箭一级、二级主机飞行正常，二级游机工作异常，发射任务失利。**朱雀二号火箭是目前国内运力最大的民营运载火箭，其首飞标志着液氧甲烷发动机技术在中国的首次应用。**

图表262：固体燃料火箭与液体燃料火箭对比

	固体燃料火箭	液体燃料火箭
动力装置	固体燃料火箭发动机	液体燃料火箭发动机
火箭发动机组成	由固体燃料火箭推进剂装药、燃烧室、喷管和点火装置等部件组成	由推力室、涡轮泵、燃气发生器、阀门、总装管路等部件组成
火箭发动机特点	比冲较低，为 2000-3000 秒/kg，能量密度不如液体燃料，无法实现流量控制，重复起动困难	比冲较高，可高达 2500-4600 秒/kg，能量密度高，可实现流量控制，能反复起动
储存周期	固体药柱，不挥发、无腐蚀性，可保存数年之久	常温推进剂为四氧化二氮和偏二甲肼等，加注后存储周期 7 天左右；低温推进剂为液氢、液氧等，加注后存储周期 1 天左右
技术难度	研制难度小，制造成本较低，可靠性高，实现喷管摆动技术难度大	研制难度大，制造成本高，可靠性相对较低，低温火箭技术尚未成熟
箭体结构	结构简单，推进剂贮存在发动机燃烧室内，无需贮箱和输送系统	结构复杂，推进剂分别贮存于火箭的氧化剂箱和燃料箱内，工作时由输送系统送入发动机燃烧室
发射周期	固体燃料火箭可以长时间储存，发射周期最少可达 24 小时，使用维护方便，可快速响应	液体燃料火箭发射前需要测试、加注推进剂，发射周期较长，加注完成后，无法长期储存
运载能力	相对低，小火箭居多	高，大推力火箭基本都是液体燃料火箭

资料来源：头豹研究院，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

朱雀二号为两级低温液体运载火箭，采用自主研发的液体火箭发动机，箭体直径为 3.35 米，全箭总长 49.5 米，近地轨道运载能力 6 吨，太阳同步轨道运载能力 4 吨，虽然与 SpaceX 星舰、蓝色起源 New Glenn、火箭实验室 Neutron 等国际在研液氧甲烷发动机火箭相比运载能力较弱，但这也标志着我国商业航天民营企业的一大进步。朱雀二号一级火箭采用 4 台“天鹊-12”液氧甲烷发动机，推力达 216 吨；二级发动机采用 1 台“天鹊-12” 液氧甲烷发动机和 1 台“天鹊-11”液氧甲烷发动机的组合。

图表263： 朱雀二号火箭

资料来源：蓝箭航天官网，华泰研究

液氧液氢、液氧甲烷、液氧煤油 3 种低温推进剂均满足发动机重复使用的基本需求，其中，朱雀二号所采用的液氧甲烷路径最具商业趋势。 这主要是由于液氧甲烷发动机成本低、性能适中、结焦少、与液氧沸点接近。液氧液氢发动机由于性能高、可检测维修、推进剂洁净无结焦、返航后易处理、推重比高、燃烧稳定性好等，易于实现重复使用，但推进剂制造、贮存成本高。液氧煤油发动机成本低、推重比高，但由于煤油易结焦，自蒸发能力最差，返航后处理复杂、复用困难。

更多免费行业报告 并购家 www.ipoipo.cn

图表264：液氧甲烷发动机综合优势突出，或为发展方向

技术对比	液氧甲烷	液氧液氢	液氧煤油
推力室比冲(m/s)	3481	4363	3367
燃效稳定性	稳定	最不稳定	适中
使用安全性	中等	最不稳定	最好
满足回收技术需要	满足	满足	不满足
重复使用次数	可达 100 次	可达 100 次	10 次以上
造价	中	高	中
技术门槛	高	高	中
代表性发动机	SpaceX 猛禽、蓝色起源 BE-4	蓝色起源 BE-3	SpaceX 梅林

资料来源：《航天制造技术》2011 年 4 月第 2 期，华泰研究

此外，科学家们已经在诸如火星、土卫六等星球上发现存在液体甲烷“海洋”，如果未来开展星际航行，或可直接从目标星球获取液体甲烷作为燃料。美国 SpaceX 和蓝色起源也积极开展液氧甲烷发动机的研制工作，其发动机型号分别为“猛禽”和 BE-4。我们认为，液氧甲烷火箭推进剂综合优势突出，液氧甲烷发动机或为民营火箭公司未来一致选择。

天兵科技“天龙二号”液体火箭首发成功，将探索大型液体运载火箭重复使用路径。2023 年 4 月 2 日，天兵科技自主研发的天龙二号遥一运载火箭在酒泉卫星发射中心点火起飞，成功将“爱太空科学号”遥感卫星送入 500 公里高度太阳同步轨道，发射任务取得圆满成功。天龙二号为中型运载火箭、三级液体火箭，采用液氧和煤基航天煤油推进剂，火箭直径 3.35 米，长 32.8 米，起飞质量 153 吨，起飞推力 193 吨，500 公里高度太阳同步轨道运载能力为 1.5 吨，LEO 运力为 2 吨，改进后可达 4 吨。

图表265：天龙二号火箭首发即成功

资料来源：天兵科技官网，华泰研究

在此之前的近 20 年时间里，全球共有 8 款液体火箭首飞失败，包括国外的 SpaceX(天空探索)、Rocket Lab(火箭试验室)、Virgin Orbit(维珍银河)、Astra Space(群星公司)、ABL、Firefly(萤火虫航天)、Relativity(相对论航天)，以及中国蓝箭航天(Landspace)的朱雀二号。**天龙二号成功入轨，打败了世界范围内液体火箭首发失败的魔咒，也开辟了我国商业航天从固体向液体运载火箭跨越的新时代。**

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表266：全球液体火箭首发对比

名称	SpaceX	Rocket Lab	Vrigin Orbit	Astra Space	Firefly	ABL	Landspace	Relativity	Space Pioneer
国家	美国	美国	美国	美国	美国	美国	中国	美国	中国
公司成立	2002	2006	2004	2016	2014	2017	2015	2015	2019
火箭型号	猎鹰1号	电子号	运载器1号	火箭3	阿尔法	RS1	朱雀2号	人族1号	天龙2号
首飞时间	2006	2018	2021	2021	2021	2022	2022	2023	2023
首飞情况	试验失败	试验失败	试验失败	试验失败	试验失败	试验失败	试验失败	试验失败	试验成功
发动机	液氧煤油	液氧煤油	液氧煤油	液氧煤油	液氧煤油	液氧煤油	液氧甲烷	液氧甲烷	液氧煤油
起飞重量	27.2t	10.5t	25.9t	11t	54t	45t	219t	95t	153t
箭体直径	1.7m	1.2m	1.8m	1.3m	2.2m	1.8m	3.4m	2.2m	3.4m
失利原因	海雾腐蚀致煤油泵入口压力传感器铝制螺母失效，煤油泄露着火。	第三方遥测错误导致数据异常，触发火箭自毁。	一级发动机液氧泵后管路破裂。	一级飞行过程中导航系统故障，火箭偏离线路触发发动机关机坠地。	一级的二号引擎异常关闭，地面工作人员控制自毁。	九个一级发动机出现异常提前关闭，火箭落回发射场。	二级游机氧泵低压壳体破裂，异常关机。	第二级出现异常，未能进入预定轨道。	

资料来源：各公司官网，华泰研究

据天兵科技官网介绍，天龙二号的成功飞行为煤基航天煤油后续交付我国长征五号、长征七号等新一代运载火箭的飞行使用，降低发射成本、实现重复使用，奠定了坚实基础。此外，天兵科技还计划于 2024 年上半年完成天龙三号基础级火箭首飞。天龙三号大型运载火箭是天兵科技自主研制的大型液体运载火箭。火箭直径 3.8 米，总长 71 米，起飞质量 590t，起飞推力 770t，近地轨道运力 17t，太阳同步轨道运力 14t，一子级具备重复使用功能，可自主返回。未来几年，天兵科技将在大型液体运载火箭重复使用、重型液体运载火箭以及空天往返载人飞船等主营产品上做出产品布局。

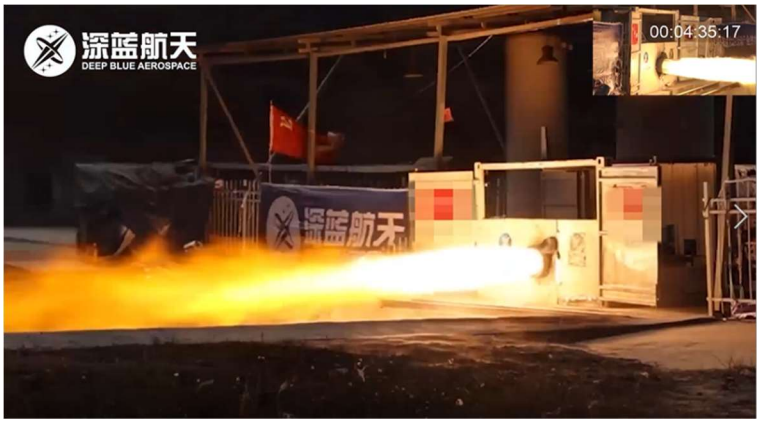
深蓝航天“雷霆-R1”突破发动机多次点火技术，2024 年有望完成首次发射-回收任务。发动机二次起动技术是指火箭在回收飞行过程中，为保证发动机在空中再次正常点火起动而采取的各种措施的总和。随着航天科技的渗透，航天空间资源抢夺战愈演愈烈，我国的火箭发射需求呈现井喷式增长。提高发射频率、降低发射成本、快速技术迭代的可回收复用液体运载火箭是构建中国未来大型卫星星座的必然需求，其中**发动机二次起动点火技术是可回收复用液体运载火箭的核心技术之一。**

2022 年 9 月 29 日，深蓝航天液氧煤油发动机“雷霆-R1”首次“二次起动”试车取得圆满成功，一次起动发动机工作 20s，二次起动发动机工作 5s。10 月 8 日，同一台发动机在不下台状态下再次进行了二次起动试车，一次起动发动机工作 20s，二次起动发动机工作 20s。“雷霆-R1”液氧煤油发动机在设计之初就具备多次点火的能力，此次试车验证了多次点火设计方案的正确性和合理性，也**标志着我国首次突破掌握燃气发生器循环液氧煤油发动机的多次起动技术。**

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表267：深蓝航天“雷霆-R1”试车成功



资料来源：深蓝航天官网，华泰研究

深蓝航天主打产品“星云”系列包括“星云-1”、“星云-1H”可回收复用液体火箭，与之配套的“雷霆”系列液氧煤油液体火箭发动机型号名为“雷霆-R1”和“雷霆-R2”。“雷霆-R1”发动机的二次起动技术的突破，为深蓝航天首枚入轨级可回收运载火箭“星云 1 号”实现垂直降落奠定了扎实基础。未来，雷霆-R1 推力将达到 35-40 吨量级，深蓝航天还将快速进行发动机的常态化试车验证和技术迭代，保障 2023 年度“星云 1 号”火箭一子级高空垂直起降实验的顺利实施。同时，深蓝航天还将按计划执行“星云-1”运载火箭的试样箭总装，并于**2024 年年底之前完成“星云-1”运载火箭的首次入轨发射、回收任务。**

图表268：2022 年 1 月-至今全球商业运载火箭发展大事记

时间	国家	事件
2022 年 5 月 7 日	中国	深蓝航天研发的“星云-M”1 号试验箭完成 1 公里级垂直起飞及降落（VTVL）飞行试验。
2022 年 5 月 13 日	中国	星际荣耀研发的双曲线一号第 4 发火箭由于飞行异常，发射任务失败。
2022 年 7 月 27 日	中国	“力箭一号”（ZK-1A）运载火箭在酒泉卫星发射中心成功发射，采用“一箭六星”的方式，成功将六颗卫星送入预定轨道。
2022 年 8 月 26 日	中国	升力式亚轨道运载器重复使用飞行试验获得圆满成功。
2022 年 11 月 3 日	中国	梦天实验舱顺利完成转位操作，中国空间站“T”字基本构型在轨组装完成。
2022 年 12 月 9 日	中国	捷龙三号运载火箭在黄海海域成功首飞，将 14 颗卫星准确送入预定轨道。
2022 年 12 月 10 日	中国	蓝箭航天朱雀二号（ZQ-2）运载火箭在酒泉卫星发射中心执行首次飞行试验任务
2023 年 3 月 23 日	美国	相对论空间人族一号（Terran 1）一次性两级小型运载火箭首次发射升空
2023 年 3 月 26 日	印度	萨迪什·达万航天中心通过 Mark 3 运载火箭（Launch Vehicle Mark III, LVM3-M3）成功发射并部署了 36 颗一网（OneWeb）互联网卫星。
2023 年 4 月 2 日	中国	天兵科技自主研发的天龙二号遥一运载火箭在酒泉卫星发射中心点火起飞
2023 年 4 月 7 日	中国	双曲线一号固体商业运载火箭在中国酒泉卫星发射中心成功发射升空，火箭按照预定程序安全顺利完成飞行，发射任务获得圆满成功。
2023 年 4 月 20 日	美国	星舰 SN24 首飞，于发射升空后 3 分钟爆炸，发射任务失败。

资料来源：SpaceX 官网，星际荣耀官网，深蓝航天官网，新华社，澎湃新闻，界面新闻，央广网，华泰研究

国内卫星制造技术成熟，步入规模化发展

国家统筹规划，民企合力组建有利于控制卫星互联网建设成本。组建互联网星座难度远远大于发射试验星，需要大量的前期建设投入，后续才能通过推动终端应用市场获得需求和收益。此前各项目与公司单打独斗的策略无法发挥我国新兴举国体制的优势，组建由国家统筹协调、各方力量充分参与的卫星网络公司有利于让中国星座突出重围。中国卫星网络集团有限公司应运而生，由“星网”统筹布局，可以集中一切资源和力量，助力项目有序推进。而国家队企业生产进度受制于体制内的科研生产计划，目前而言成本较高。在由国家总体把控的前提下，充分发挥民营企业效率优势、打造“卫星互联网+”生态，赋能产业集群，促进相关产业协同发展有利于控制卫星互联网建设成本。

免责
更多免费行业报告 并购家 www.ipoipo.cn

以银河航天“小蜘蛛网”计划为例，民营企业星座计划对国家统筹的“GW”星座形成支持与补充。2022 年 3 月，银河航天批量研制的六颗低轨宽带通信卫星成功发射，这批卫星与银河航天首发星组成了低轨宽带通信试验星座，并构建星地融合的 5G 试验网络。

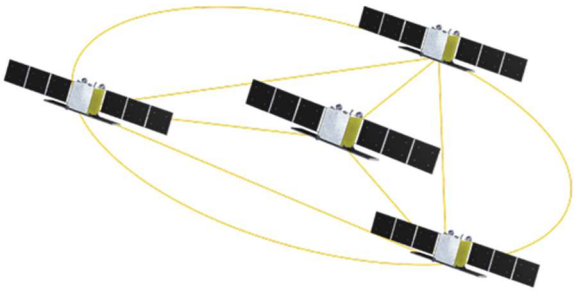
图表269：国内卫星互联网公司及星座统计

国内卫星互联网公司							
国家队				商业卫星公司			
中国星网、航天科技五院/八院、航天科工集团、中国电科				银河航天、时空道宇、国电高科、微纳星空、九天微星、欧科微			
国内卫星互联网星座统计							
序号	**星座名称**	**星座用途**	**建设单位**	**规模数量**	**卫星轨道高度**	**单星重量**	**启动时间**
1	星网工程	宽带通信	中国卫星网络集团有限公司	12992	508km/590km/600km/1145km	300kg/1000kg	2020
2	虹云工程	宽带通信	航天科工	156	500km	300kg	2018
3	银河星座	宽带通信	银河航天	650	500km-1200km	300kg	2019
4	LaserFleet	激光通信	航星光网/上海光机所	288	550km	150kg	2019
合计				**14086**			

资料来源：《卫星与网络》，华泰研究

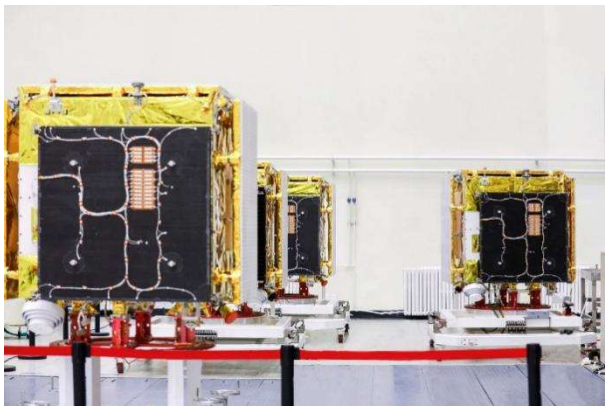
民企合力组建，进一步降低卫星制造成本。引入民营企业形成规模制造是降低卫星制造成本的重要方向。以银河航天为例，2023 年 3 月 30 日太原卫星发射中心使用长征二号丁运载火箭，以“一箭四星”的方式，成功将银河航天承担研制的四颗干涉合成孔径雷达（InSAR）卫星（宏图一号 01 组卫星）发射升空。据银河航天报道，该组卫星在轨构成国际上首个车轮式卫星编队，主星位于中部，三颗辅星均匀分布在“车轮”轮毂上，采用星间通信链路和相位同步链路，在辅星与主星相距仅几百米的情况下，通过精密的轨道控制，保证卫星编队在轨构型的稳定性和空间安全性，相对于传统的干涉卫星系统，车轮式编队具有编队构型相对稳定、干涉基线多、测绘效率高的优势。

图表270：银河航天 INSAR 编队星座示意图



资料来源：银河航天官网，华泰研究

图表271：银河航天研制的四颗干涉合成孔径雷达（InSAR）卫星



资料来源：银河航天官网，华泰研究

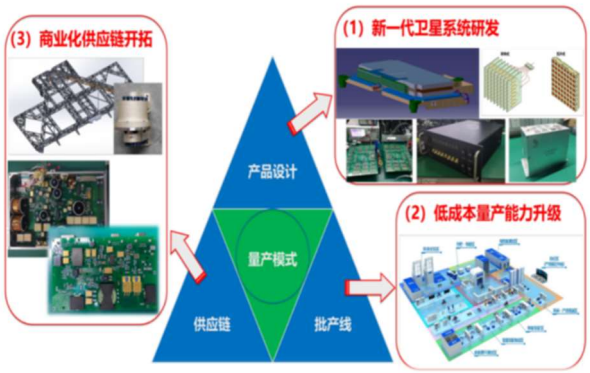
本次车轮式卫星编队的研制也验证了银河航天在产业链垂直整合方面的能力。银河航天通过核心元器件自研降低卫星成本，同时构建了基于民用工业体系的商业化供应链，探索卫星量产模式，为卫星低成本批量研制提供了宝贵的实践经验。针对民用 SAR 卫星小批量研制，银河航天在卫星方案设计、整星集成、发射场测试等阶段，通过数字化仿真、自动化测试、流程优化等途径，极大地提升了研制效率，相比常规研制时间缩短了 60%以上。未来，银河航天等民企的加速技术创新，有望进一步推动卫星低成本批产，支持更多卫星更高频次地发射入轨。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表272：银河航天批量研制的低轨宽带通信卫星

资料来源：银河航天官网，华泰研究

图表273：银河航天“卫星设计—生产线—供应链”的量产铁三角模式



资料来源：银河航天官网，华泰研究

“中星 26 号”卫星成功发射，我国通信卫星单星容量进入百 Gbps 时代。2023 年 2 月 23 日，我国首个超百 Gbps 容量的高通量卫星“中星 26 号”发射成功，我国通信卫星单星容量进入百 Gbps 时代。“中星”26 号采用我国自主研发的东方红四号增强卫星发射平台，星上设计的 Ka 频段高通量通信载荷主要覆盖中国全境、东亚、东南亚、南亚部分地区、太平洋及印度洋部分区域。该星配置的 94 个用户波束及 11 个信关波束将为我国及周边地区的固定终端、车载终端、船载终端、机载终端提供高速、可靠、优质的宽带数据服务，在航空、航海、远程教育与医疗、自然灾害应急通信、自然保护区监测等领域发挥重要作用。

图表274：中国卫通星队

资料来源：中国卫通官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表275：2022 年 1 月-至今全球商业卫星发展大事记

时间	国家	事件
2022 年 2 月 27 日	中国	“安溪铁观音一号”卫星发射成功，是福建省第一颗亚米级地面分辨率的商业光学遥感卫星，也是我国第一颗以茶叶冠名的遥感卫星。
2022 年 2 月 27 日	中国	“启明星一号”卫星发射成功，是我国首颗微纳遥感卫星。
2022 年 3 月 5 日	中国	银河航天 02 批卫星（6 颗）发射成功，是中国首次批量研制的低轨宽带通信卫星。
2022 年 3 月 5 日	中国	我国首颗商业组网 SAR 卫星“巢湖一号”成功获取首批图像。
2022 年 7 月 24 日	中国	问天实验舱与火箭成功分离并进入预定轨道，发射取得圆满成功。
2022 年 10 月 22 日	俄罗斯	代号为 Kosmos-2561 和 Kosmos-2562 的两颗军用卫星成功发射升空。
2022 年 10 月 28 日	美国	“星链”v1.5G4-31，将 53 颗 V1.5 型“星链”卫星发射成功。
2022 年 10 月 13 日	中国	5 米 S-SAR 01 星（环境减灾二号 05 卫星）发射升空，卫星顺利进入预定轨道。
2023 年 2 月 23 日	中国	我国首颗超百 Gbps 容量高通量卫星成功发射。
2023 年 2 月 27 日	美国	首批 21 颗星链 V2 Mini 卫星发射
2023 年 3 月 30 日	中国	成功将宏图一号 01 组卫星发射升空，发射入轨后将组成国际首创的车轮式编队，可以全天候对地观测，提供商业遥感数据服务。
2023 年 3 月 22 日	中国	以“一箭四星”方式，将天目一号气象星座 03 星、04 星、05 星、06 星送入预定轨道。
2023 年 4 月 16 日	中国	风云三号 07 星发射升空，卫星顺利进入预定轨道，发射任务获得圆满成功。
2023 年 4 月 19 日	美国	第二批 21 颗星链 V2 Mini 卫星发射
2023 年 5 月 1 日	美国	Viasat-3 超高通量通信卫星发射成功。

资料来源：环球网，中国日报网，福建日报，中国新闻网，腾讯网，中国网科学，新民晚报，华泰研究

需求侧：行业需求加快落地

星际争霸，starlink 快速推进，新星座不断涌现

2018 年~2022 年全球发星数量年均增速达 54%，Starlink、OneWeb 按计划继续推进星座建设。截止 2023 年 5 月 14 日，“星链”发射总数达 4447 颗，在轨卫星 4095 颗。按 SpaceX 的计划，到 2024 年将部署约 12000 颗“星链”卫星，到 2027 年将部署约 42000 颗“星链”卫星。2023 年 3 月 29 日，Oneweb 完成一代卫星组网，即将提供全球卫星互联网服务。在全球卫星星座快速组网的背景下，卫星互联网布局得到政府和资本重视，欧盟 IRIS2 与亚马逊 Project Kuiper 计划展开部署。

图表276：全球发星数量及星链、oneweb 发星占比（截至 2023 年 5 月 14 日）

	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年(截止 5.14)
全球发星数量	446	492	1278	1816	2505	1115
年增速	-	10.31%	159.76%	42.10%	37.94%	-
Starlink 发星数量	2	120	833	976	1722	781
Starlink 发星占比	0.45%	24.39%	65.18%	53.74%	68.74%	67.62%
OneWeb 发星数量	-	6	104	284	146	116
OneWeb 发星占比	-	1.22%	8.14%	15.64%	5.83%	10.04%

注：2023 年全球发星数量为不完全统计

资料来源：SpaceX 官网、OneWeb 官网、《卫星与网络》，华泰研究

我国卫星需求持续落地，商业星座计划加速部署。根据未来宇航《2022 年中国商业航天产业投资报告》，2022 年中国研制发射航天器共 185 颗，同增 66 颗，其中商业卫星发射 100 颗，同增 59 颗，覆盖商业通信、遥感、导航、技术验证等多个领域。“天府”星座、“小蜘蛛网”、“吉林一号”、“天启卫星互联网”等商业星座计划有序部署。2023 年 5 月 13 日，据海南商发公司官网报道，中国星网工程首次公开披露预计发射时间，预计于 2024 年上半年开始执行发射。航天宏图高分遥感“女娲星座”实现首发四星，一期工程 38 颗业务星计划于 2025 年完成发射。

“星链”全球用户数量已突破 150 万人，服务范围覆盖七大洲 53 个国家和地区。SpaceX 是目前全球唯一一家星座建设与应用推广同步开展的公司；2023 年初，SpaceX 高层宣布“星链”公司预计 2023 年能够扭亏为盈，实现盈利，标志着“星链”即将取得商业化成功。用户数量方面，2023 年 5 月，SpaceX 公司宣布“星链”卫星互联网全球用户数量已超过 150 万人。“星链”用户地区分布方面，“星链”卫星互联网服务已覆盖七大洲 53 个国家和地区并在美国 48 个州被激活。并且，2022 年之后发射的星链卫星均配置了星间激光链路，有效帮助“星链”实现南极洲等全球无人区的覆盖。

更多免费行业报告 并购家 www.ipoipo.cn

星链商业化版图扩大，应用场景不断拓展。“星链”卫星互联网服务场景覆盖：1）家用，为全球偏远和农村地区的高速、低延迟宽带互联网，包含 599 美元硬件费+120 美元月费；2）商用，面向企业提供宽带互联网服务，包含 2500 美元硬件费+250 美元月费；3）漫游，允许用户携带天线旅行，包含 599 美元硬件费+150 美元月费；4）移动服务，终端在车辆固定安装，提供全球动态链接，包含 2500 美元硬件费+250 美元月费；5）海事，提供全球海运覆盖，海上下载速度高达 220Mbps，包含 2500 美元硬件费+250 美元月费；6）航空，提供全球机上互联网连接，包括硬件费 15 万美元+1.25~2.5 万美元月费；7）物联网，整合蜂群技术公司星座与物联网频谱资源，为物联网设备提供低带宽全球连接，月费 5 美元。

图表277：星链漫游服务

资料来源：SpaceX 官网，华泰研究

图表278：星链移动服务

资料来源：SpaceX 官网，华泰研究

SpaceX 发布“星盾”新服务，向军事化应用迈出关键一步。2022 年 12 月 2 日，SpaceX 发布了“星盾（Starshield）”卫星网络，目的是利用现有的“星链”卫星平台技术满足美国国防和情报机构日益增长的需求。“星链”一直以来用于消费者和商业用途，而“星盾”的设计则是针对政府服务。

星盾初期应用主要集中在三个领域：1）地球观测：发射带有遥感载荷的卫星，并将处理后的数据直接发送给用户；2）安全通信：通过星盾用户终端为政府用户提供可靠的全球通信；3）载荷托管，建造卫星总线以支持最苛刻的客户载荷任务。其中，据 SpaceNews 报道，“星盾”的卫星通信服务将借鉴 SpaceX 在俄乌冲突中的经验，因为“星链”在乌克兰的应用证明了其可以在战区良好运行并能抵挡对手的电磁和网络攻击，目前美国空军已购买了“星链”服务以支持欧洲和非洲的部队。

为满足政府使用，“星盾”主要有以下特点：1）安全性：尽管“星链”已经提供了无与伦比的端到端用户数据加密，为满足政府最苛刻的要求，“星盾”使用了额外的高度加密功能来托管机密有效负载并安全地处理数据。2）模块化设计：满足不同任务要求，能集成各种有效载荷，为用户提供多种功能选择；3）互操作性：成熟的星载激光通讯链路可以集成到合作伙伴卫星上，以便将其纳入星盾网络，任何军用卫星只要搭载了星链的激光通讯链路，就可以使用星链进行中继或者遥控；4）快速开发和部署：凭借 SpaceX 快速迭代的能力，以及从运载火箭到用户终端的全栈开发能力，能够实现最快速度的大规模部署；5）弹性和扩展能力：“星盾”星座的近地轨道架构提供固有弹性，同时 SpaceX 能够快速补充卫星。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表279：“星盾”卫星图片

资料来源：SpaceX 官网，华泰研究

Project Kuiper 将于 2023 年初首次发射卫星。Kuiper Systems 是成立于 2019 年的亚马逊子公司，其柯伊柏项目旨在建设大型宽带卫星互联网星座，计划部署 3236 颗卫星。柯伊柏项目的两颗初始原型卫星“KuiperSat-1”和“KuiperSat-2”将在 2023 年初联合发射联盟(ULA)的新型 Vulcan Centaur 火箭首次部署这两颗卫星，目前首发日期尚未确定。2023 年 3 月 14 日，亚马逊发布了 Kuiper 的卫星互联网用户终端工程模型，由亚马逊自研基带芯片提供通信支持，制造成本将低于 400 美元。亚马逊表示，计划于 2023 年底前开始大规模量产卫星，在 2024 年上半年发射第一批卫星，并开始向其首批客户提供卫星互联网服务。

图表280：Kuiper 卫星互联网用户终端

资料来源：Amazon 官网，华泰研究

图表281：Kuiper 星座部署计划

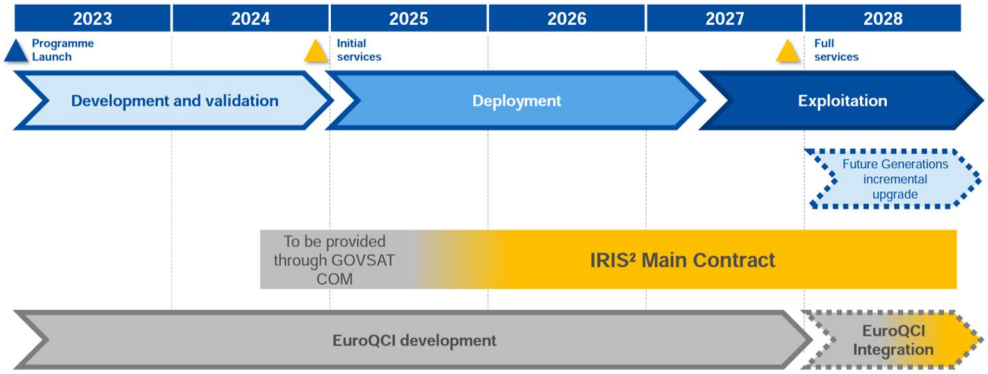
规划	
轨道高度	卫星规划总数
590km	784颗
610km	1296颗
630km	1156颗
总计	**3236颗（2020.7获批）**
进度	
2023年	首发**2**颗原型星
2024年	发射第一批，**578**颗卫星在轨后提供服务
2026年前	共部署**1618**颗
2029年前	共部署**3236**颗

资料来源：Amazon 官网，华泰研究

欧洲通过 IRIS2 提案，将建设欧盟宽带卫星网络。在全球卫星星座快速组网的背景和需求下，卫星制造能力和火箭发射能力是各国布局卫星互联网的必要前提，更是综合实力的体现。2023 年 2 月 14 日，欧洲议会全体会议通过了欧洲议会和理事会制定的关于安全连接计划（IRIS2, Infrastructure for Resilience, Interconnectivity and Security by Satellite）提案，旨在到 2027 年部署一个欧盟拥有的通信卫星群，为欧盟国家提供自己的宽带卫星网络，用于军事、政府和公众用途。项目总投资 60 亿欧元，其中欧盟预算 24 亿欧元、欧洲航天局 6.85 亿欧元，其余资金将由私营部门承担。欧盟预计，此次建设的卫星互联网系统可创造 170 亿～240 亿欧元的收益。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表282：IRIS2 卫星网络项目建设规划



资料来源：欧盟委员会官网，华泰研究

中国星网低轨卫星星座预计于 2024 年上半年开始执行发射。据央视新闻 5 月 12 日报道，中国低轨卫星互联网星座（中国星网）预计明年上半年开始执行发射。这是中国星网工程首次公开披露预计发射时间，被外界期待 2 年之久的“中国版星链”有望开始落地实现。据央视新闻报道，据航天科技集团一院海南商业航天发射场建设项目专项总设计师吴义田介绍，“年底确保按期完成 1 号工位所有的建设工作，等联调联试结束以后，我们会在 2024 年上半年的 6 月份之前择机完成星网工程的发射任务，星网工程也是我们海南商业航天发射场首次执行的发射任务，也是我们国家的重大工程任务。”

图表283：我国 GW 卫星星座计划

星座	子星座	轨道高度（km）	轨道倾角（度）	轨道面数	单轨星数	卫星数量
1	1	590	85	16	30	480
	2	600	50	40	50	2000
	3	508	55	60	60	3600
			小计			6080
2	1	1145	30	48	36	1728
	2	1145	40	48	36	1728
	3	1145	50	48	36	1728
	4	1145	60	48	36	1728
			小计			6912
卫星总数量						12992

资料来源：Circle ID，华泰研究

航天宏图高分遥感卫星星座“女娲星座”首发四星，2025 年前计划完成 38 颗卫星发射。2022 年 12 月 21 日，根据航天宏图官网，航天宏图将建设高分遥感卫星星座“女娲星座”，一期工程计划发射 38 颗业务星，包括 28 颗卫星组成的雷达遥感星座，10 颗分辨率优于 0.5 米的高分和 2 米八谱段的光学卫星星座；“女娲星座”二期工程也在筹划中。2023 年 3 月 30 日，航天宏图 1 号星座 4 颗高分辨率 X 波段合成孔径雷达卫星搭载长征二号丁运载火箭顺利入轨。据公司官网介绍，这是国际上首个“1+3”车轮编队构型的分布式干涉 SAR 卫星星座，由“一主加三辅”四颗卫星组成，主星发射雷达信号，主辅四星接收散射信号，可实现多基线干涉成像，在城市地质灾害监测、海洋监测、洪涝灾害监测、地表沉降监测等方面具备极大优势。

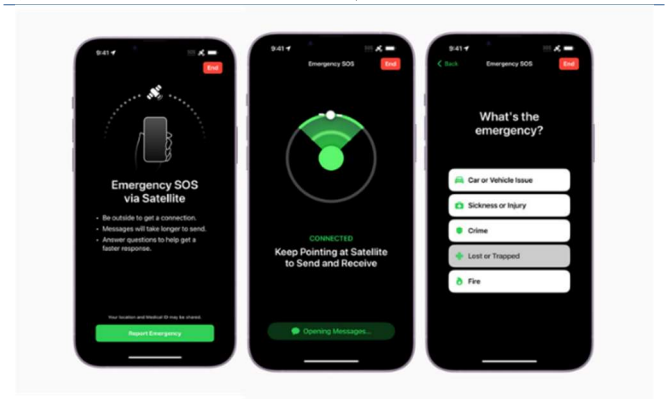
航天科工计划构建超低轨卫星星座，提供遥感公众服务。在 2023 年中国航天日国际论坛中，航天科工集团指出，遥感产业化的实现要与公众结合。为了提供公众服务，需要部署超低轨卫星星座，并实现感传算融合。超低轨飞行能够有效降低卫星的研制发射成本，提高信息获取和传输的效率，特别适用于卫星批量化生产和高密度发射。在规划上，航天科工集团计划年底发射技术验证卫星、明年进行 9 颗卫星星座的业务试验；规模化组网阶段中，2026 年起聚焦形成全球半小时重访能力，2031 年进行换代升级，聚焦通过千颗规模卫星，实现全球任意范围内分钟级重访的新型产业生态构建。

免责
更多免费行业报告 并购家 www.ipoipo.cn

卫星直连进展亮眼，终端侧、卫星侧、监管侧齐发力

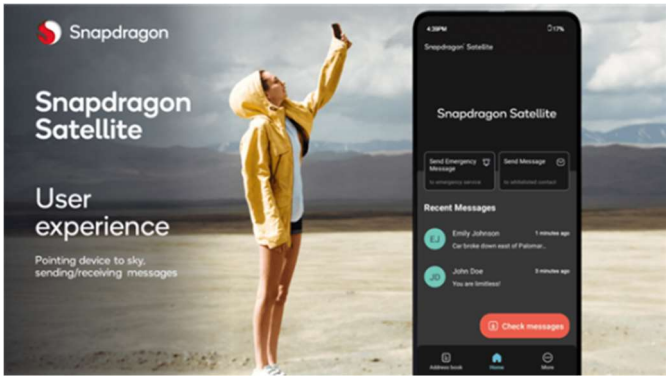
在终端侧，过去一年内卫星直连有了多项喜人的进展。在手机产业链，高通、苹果、华为等在推出了手机端卫星直连服务，解决方案为改变手机硬件以适应卫星星座，其中苹果和高通采用骁龙基带芯片和低轨通信卫星星座，华为采用北斗短报文芯片和高轨道的北斗卫星。在 MWC 2023 上，高通宣布与小米、vivo、OPPO、荣耀、摩托罗拉、Nothing 等厂商合作，未来两年卫星通信有望成为安卓旗舰机标配。大众手机搭载卫星通信技术，标志着卫星通信应用场景已由军用需求逐步拓展至民用市场，我们认为卫星智能手机时代将至，未来智能手机有望替代专门终端，成为卫星网络的主要接入方式。

图表284： iPhone 14 卫星通信功能界面



资料来源：Apple 官网，华泰研究

图表285： 高通 Snapdragon Satellite 双向消息通信解决方案



资料来源：高通官网，华泰研究

图表286： 高通、苹果、华为卫星通信功能对比

厂商	高通	苹果	华为
品牌型号	安卓旗舰机型	iPhone 14 系列	Huawei Mate50
上线时间	2023 年下半年	2022 年 11 月	2022 年 9 月
直连方案	卫星频段+智能终端升级	卫星频段+智能终端升级	卫星频段+智能终端升级
使用星座	铱星二代	全球星	北斗卫星
星座特点	低轨道（LEO）	低轨道（LEO）	高轨道（GEO）
芯片类型	骁龙 X70 基带芯片	骁龙 X65 基带芯片	移动通信基带芯片+北斗短报文芯片
服务种类			
位置	支持	支持	支持
发短信	支持（单次 140 字符）	支持	支持（单次 19 个字符）
发送时间	3s~10s	15s~1min	10s~1min
收短信	支持	支持	不支持
话音	不支持	不支持	不支持
服务费用	未公布	两年内免费，后恢复收费	每月 30 条免费，超出付费
支持地区	全球覆盖	美国、加拿大、法国、德国、爱尔兰、英国	中国大陆（不含港澳台）

资料来源：苹果官网、华为官网、高通官网，华泰研究

在模组厂商方面，移远通信发布了卫通模组。2023 年 3 月，移远通信推出面向物联网行业应用的 CC200A-LB 卫星通信模组，根据移远通信官网介绍，该模组搭载全球领先的物联网通信和解决方案提供商 ORBCOMM 的卫星物联网连接平台，旨在为蜂窝网络无法覆盖的偏远区域提供可靠的全球无线网络覆盖和连接，并通过更具成本效益、超低时延等优势，为诸多海事、运输、重型设备、农业、采矿以及石油和天然气监测等场景打造无线解决方案。

在卫星侧，SpaceX、AST、Lynk 等公司加速验证卫星直连手机。SpaceX 方面，继去年 9 月发布 Starlink V2.0 与 T-Mobile 的手机直连卫星计划后，SpaceX 于去年 12 月向美国联邦通信委员会（FCC）申请移动卫星服务频谱授权，使得普通智能手机以手机蜂窝网的形式接入卫星服务，并于今年 2 月和 4 月发射了两批 Starlink V2.0 mini。

更多免费行业报告 并购家 www.ipoipo.cn

AST 公司方面，继 22 年 9 月完成 Bluewalker3 试验卫星发射后，今年 4 月 AST 公司宣布打通世界第一个天基蜂窝语音通话，测试使用了三星 Galaxy S22 智能手机，利用 AT&T Band5（850MHz）频段，该方案验证了天基 3GPP 技术向地面超过 80 亿部存量手机提供应急通信的服务能力。除了呼叫测试外，AST 还在各种智能手机和设备上进行了初步兼容性测试，这些手机成功将 SIM 卡信息和网络信息直接交换到 BW3，对于从太空向任何设备提供宽带连接非常重要。后续，AST 计划于 2024 年一季度发射前 5 颗 BlueBird Block 1 卫星，将与移动网络运营商合作，为蜂窝覆盖之外的智能手机提供连接。

Lynk 公司方面，Lynk 主打太空双向短信功能，在 22 年 9 月宣布获得了由美国联邦通信委员会（FCC）颁布的“手机直连卫星”商业许可，成为全球第一家获此批准的公司。2023 年 1 月，Lynk 发射了第 2 个与第 3 个太空蜂窝塔，并发射了 5G 载荷。目前，Lynk 已与多个移动运营商签署商业协议，并联合开展测试，计划提供卫星通信商用服务。

图表287： Starlink V2.0 mini 卫星

资料来源：SpaceX 官网，华泰研究

图表288： Bluewalker3 试验卫星

资料来源：AST 官网，华泰研究

在监管侧，FCC 发布了 SCS（Supplemental Coverage from Space）新规则，允许卫星运营商使用地面移动通信频谱运营卫星，促进卫星与地面网络融合。3 月 16 日，FCC 宣布通过太空补充覆盖（SCS）新规则提案，旨在通过构建清晰透明的监管流程，促进卫星运营商和无线通信公司之间的合作（手机直连卫星）。根据 SCS 新规则规定，卫星运营商与地面移动运营商合作后，将能够获得 FCC 授权，在获得许可后分配给地面服务的频谱上运营卫星。FCC 建议，允许经授权的非地球同步卫星运营商申请访问地面频谱，前提是满足某些先决条件，包括在指定地理区域内获得地面许可持有人的租赁；然后，卫星运营商可以为无线服务提供商的客户提供在偏远地区连接的服务。

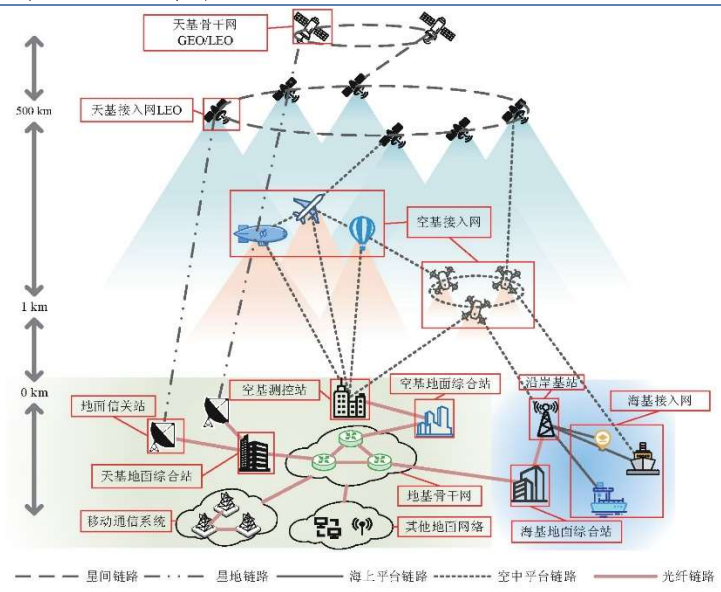
6G 全域覆盖倍受期待，为卫星互联网发展提供重要机遇

从 1G 到 5G 移动通信网络覆盖主要依靠基站增加实现，卫星互联网是实现 6G 全域覆盖的关键。传统移动通信技术主要依靠地面固定的蜂窝基站实现网络覆盖范围的扩大。根据 2022 年世界 5G 大会上中国信科集团副总经理陈山枝表示，全球现有的移动通信网络只覆盖了陆地表面的 20%、地球总面积的 6%。对于未实现信号覆盖的地域，使用地面基站的方案并不可行，一方面由于人口稀少缺乏经济性，另一方面在海洋、森林、沙漠等地区受地形限制无法建设。因此实现移动网络全覆盖，必须使用卫星通信融合地面移动通信的方案。

免责
更多免费行业报告 并购家 www.ipoipo.cn

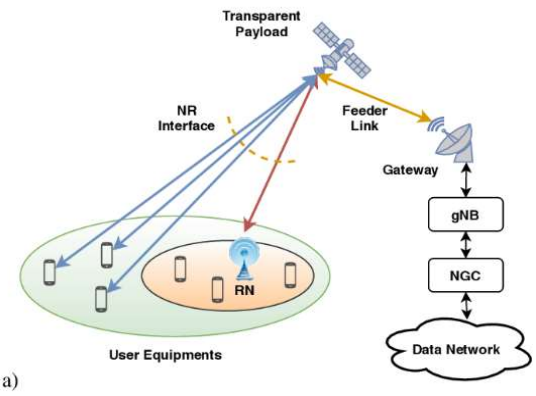
空天地一体化网络是 6G 网络的核心趋势，卫星互联网在 6G 研发中将获重点发展。星地融合网络以地面网络为基础、以卫星网络为延伸，覆盖太空、空中、陆地、海洋等自然空间，为天基、空基、陆基等各类用户的活动提供信息保障。为突破 4G、5G 移动通信受地形限制的缺点，构建全球无缝覆盖的星地融合立体网络已成为当前国际讨论的热点和未来趋势，包括 3GPP 和 ITU 在内的国际组织已经成立了相应的工作组开展星地融合的标准化研究，我国通信标准化协会也于 2019 年成立了航天通信技术工作委员会开展星地一体化的研究工作。其中，3GPP 立项的非地面网络（NTN）致力于将卫星通信与 5G 网络融合，以实现更广阔的覆盖满足用户接入服务需求。5G NTN 标准化持续演进，ITU 6G 标准规划已初步成型，这些均为面向 6G 的星地融合奠定了技术基础。

图表289： 空天地海一体化网络场景图



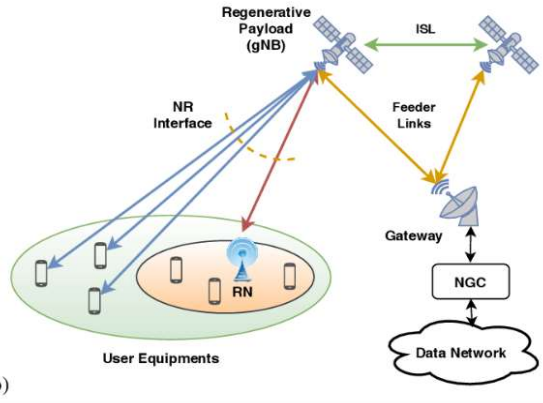
资料来源：《空天地海一体化网络体系架构与网络切片技术》付书航等（2021），华泰研究

图表290： Transparent Payload NTN 架构



资料来源： Satellite Communications in the New Space Era: A Survey and Future Challenges，华泰研究

图表291： Regenerative Payload NTN 架构



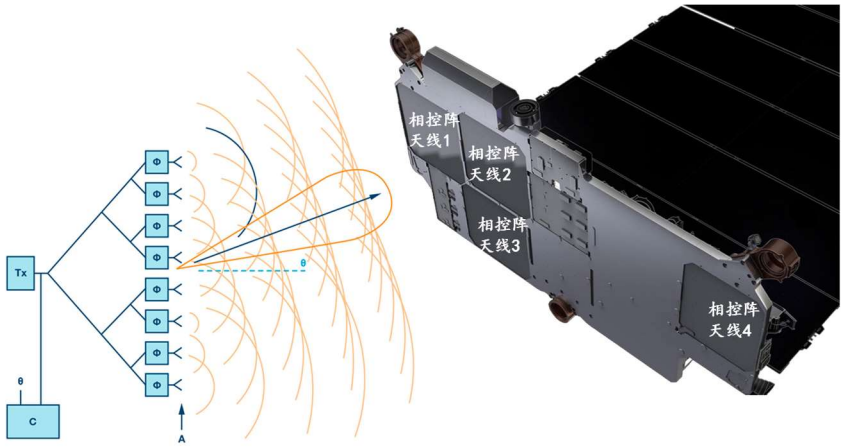
资料来源：Satellite Communications in the New Space Era: A Survey and Future Challenges，华泰研究

低轨卫星互联网重点关注相控阵天线技术。电磁波或声波频率因馈元本身或/和目标物相对运动所引起的频率改变称为多普勒效应。通信卫星低轨化后，卫星移动速度快，多普勒效应明显，需采用相控阵技术。根据《卫星通信中相控阵天线的应用及展望》（李靖等，2019），相控阵天线通过控制阵列天线中辐射单元的馈电相位来改变方向图形状，控制相位可以改变天线方向图最大值的指向，以达到波束扫描的目的，也可以通过加权优化控制副瓣电平、最小值位置等参数；卫星通信空间段主要是利用相控阵天线的同时多点波束、敏捷波束和空域滤波能力。

免责
更多免费行业报告 并购家 www.ipoipo.cn

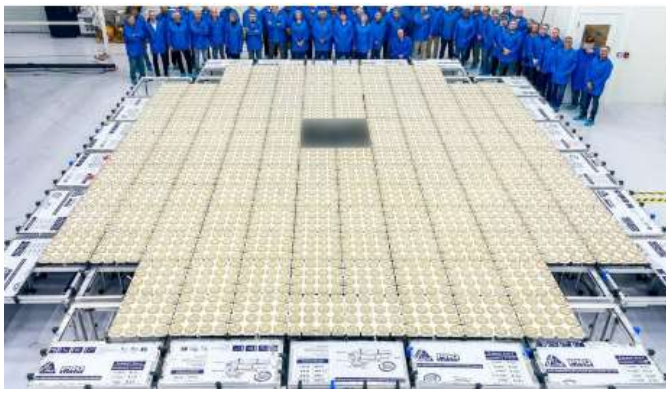
应用与技术的发展对低轨通信卫星的数据传输容量提出了更高的要求，低增益天线已不能满足卫星性能，机械扫描天线不能改变波束的形状及实现多移动波束模式，相控阵天线成为低轨卫星的主流方案。星链 1.0 卫星上采用 4 副工作在 Ku/Ka 频段的相控阵天线，负责向地球上的地面天线发送和接收信号。星链 2.0 卫星的相控阵天线规格将进一步升级，根据 2022 年 8 月马斯克在德州与 T-mobile 的合作声明会上的表示，Starlink V2.0 卫星长约 7 米，并将配备 5 到 6 米宽，面积约 25 平方米的相控阵天线。2022 年 9 月由 AST 公司发射的 Bluewalker3 的相控阵天线为 64 平方米，为目前部署到近地轨道卫星中最大规模。

图表292： 相控阵单元框图和 Starlink 1.0 卫星相控阵天线示意图



资料来源：SpaceX，华泰研究

图表293： Bluewalker 3 巨型相阵控天线



资料来源：AST Spacemobile 官网，华泰研究

图表294： Bluewalker 3 单个阵列天线面板折叠状态

资料来源：AST Spacemobile 官网，华泰研究

政策面：利好政策频出，投融资稳健增长

国内多省市相继发布利好政策，推动商业航天快速发展

济南制定空天信息产业发展五年规划，加快打造卫星总装和火箭发动机制造两大基地

2023 年 2 月 10 日，济南市委市政府召开“凝心聚力抓项目蓄势赋能促发展”主题系列新闻发布会，今年，济南将制定空天信息产业发展五年规划，加快打造卫星总装和火箭发动机制造两大基地；完成济南低运量轨道交通一期建设规划报批，启动首条云巴线路建设；推动中国融通、中国能建、中国电建等企业设立区域总部。

重庆创建国家级卫星互联网产业创新中心

2023 年 3 月 6 日，重庆市政府发布了《关于加快推进以卫星互联网为引领的空天信息产业高质量发展的意见》（下称《意见》），提出到 2025 年，构建空天地一体化、通导遥深度融合的空天信息服务体系，创建国家级卫星互联网产业创新中心。到 2030 年，重庆市要形成千亿级空天信息产业集群。

免责
更多免费行业报告 并购家 www.ipoipo.cn

河南省推进卫星及应用产业发展做大郑州北斗产业集群

2023 年 4 月 4 日，河南省印发《2023 年河南省卫星及应用产业发展工作方案》，将采取系列措施，让“河南星”闪耀太空。《方案》提到，河南省要聚力建设郑州市和鹤壁市全链条卫星产业集群，鼓励各省辖市打造一批卫星应用示范区和产业配套基地。支持郑州市建设高新区北斗产业孵化园，打造航空港区北斗+智能制造示范基地，培育金水区地理信息导航产业，做大郑州市北斗产业集群。

重庆布局空天领域，“解放碑一号”卫星将发射

2023 年 4 月 7 日，重庆市渝中区召开重庆空天信息产业发展论坛暨遥感产业应用高峰论坛，为加快拓展卫星遥感技术应用，近期，渝中区政府将与航天宏图公司共同发射以“解放碑一号”命名的遥感卫星，为重庆卫星矩阵再添一颗新星。该颗卫星整体技术达到国际领先水平，具备高精度地形测绘、高精度形变检测、高分辨率宽幅成像以及三维立体成像等能力，在自然资源、应急管理、水利林业等领域具有极高的应用价值。

河南：推动郑州和鹤壁市建设全链条卫星产业集群

2022 年 9 月 20 日，河南省发展改革委印发《河南省卫星产业发展规划》。《规划》提出：到 2025 年，河南省卫星产业规模达到 200 亿元，初步形成涵盖卫星生产制造、应用服务、研发创新的产业体系。推动郑州和鹤壁市建设全链条卫星产业集群，支持洛阳、新乡、信阳等市做强卫星制造配套产业，鼓励各地差异化发展卫星应用产业，形成“2+N”卫星产业发展格局。

广州南沙发布“探天九条”，商业航天企业最高可获亿元补贴

2022 年 8 月 30 日，广州市发布《广州南沙新区(自贸片区)促进商业航天产业发展的扶持办法》(征求意见稿)。据南沙区政府介绍，“探天九条”在研发创新支持方面，聚焦火箭、卫星产业创新领域，对解决国家战略需求，专注航天关键产品研发、产业化的商业航天企业，按其上一年度在南沙实际产生的研发费用的 20%予以补贴，每家企业累计补贴金额最高不超过 1 亿元，补贴期限 3 年。

北京大兴发布“航天十条”助力商业航天发展

2022 年 4 月 9 日，北京市大兴区发布《支持商业航天产业发展暂行办法》(以下简称“航天十条”)，将支持企业创新发展，聚焦火箭、卫星、地面终端等产业化项目，新购置研发和生产设备的企业每年最高可获得 2000 万元资金支持。据了解，“航天十条”主要支持商业航天领域相关企业、民办非企业单位，促进产业高质量发展。政策的制定“突出优势产业主体、覆盖产业关键环节”，形成了“以点带面、点面结合”的 10 条措施。根据“航天十条”，为鼓励企业落地发展，大兴区对首次纳入规模以上统计范围的企业，给予连续三年、最高 1000 万元的资金支持。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表295：2014年以来国家政策支持

时间	发布单位	文件名称	相关内容
2014年11月	国务院	《国务院关于创新重点领域投融资机制鼓励社会投资的指导意见》	鼓励民间资本参与国家民用空间基础设施建设。完善民用遥感卫星数据政策，加强政府采购服务，鼓励民间资本研制、发射和运营商业遥感卫星，提供市场化、专业化服务。引导民间资本参与卫星导航地面应用系统建设。
2015年10月	国家发改委 财政部 国防科工局	《关于印发国家民用空间基础设施中长期发展规划（2015-2025)的通知》	分阶段逐步建成技术先进、自主可控、布局合理、全球覆盖，由卫星遥感、卫星通信广播、卫星导航定位三大系统构成的国家民用空间基础设施。
2016年11月	国务院	《"十三五"国家战略性新兴产业发展规划》	推动“互联网+天基信息应用”深入发展，打造空间信息消费全新产业链和商业模式。推进商业卫星发展和卫星商业化应用。
2016年12月	国务院	《2016中国的航天》	鼓励引导民间资本和社会力量有序参与航天科研生产、空间基础设施建设、空间信息产品服务、卫星运营等航天活动，大力发展商业航天。
2017年12月	国务院	《国务院办公厅关于推动国防科技工业军民融合深度发展的意见》	加快空间基础设施统筹建设。加快论证实施重型运载火箭、空间核动力装置、深空探测及空间飞行器在轨服务与维护系统等一批军民融合重大工程和重大项目。以遥感卫星为突破口，制定国家卫星遥感数据政策，促进军民卫星资源和卫星数据共享。
2018年4月	国家发展改革委 财政部	《关于降低部分无线电频率占用费标准等有关问题的通知》	调整网络化运营的高通量卫星系统频率占用费收费方式，降低开展空间科学研究的卫星系统频率占用费标准。
2019年4月	工信部	《遥感和空间科学卫星无线电频率和轨道资源使用规划（2019-2025年）》	规范和引导商业遥感卫星发展：发挥市场作用，鼓励有条件的单位推进自主可控、布局合理的商业航天测控网建设，作为国家空间基础设施的重要补充和组成部分。
2019年6月	国防科工局 中央军委装备发展部	《关于促进商业运载火箭规范有序发展的通知》	引导商业航天规范有序发展，促进商业运载火箭技术创新。
2020年5月	国家发改委	《关于2019年国民经济和社会发展计划执行情况与2020年国民经济和社会发展计划草案的报告》	支持商业航天发展，扩展通信、导航、遥感等卫星应用。
2021年1月	中国载人航天办公室	《关于征集“面向空间站运营的低成本货物运输”方案设想的公告》	拟探索发展低成本货物运输手段，构建起灵活高效、形式多样、成本低廉的货物运输体系，火箭运输成本对标国际先进水平。
2021年4月	国家发展改革委 商务部	《关于支持海南自由贸易港建设放宽市场准入若干特别措施的意见》	推动卫星遥感、北斗导航、卫星通信、量子卫星、芯片设计、运载火箭、测控等商业航天产业链落地海南。
2022年1月	国务院	《2021中国的航天》	研究制定商业航天发展指导意见，促进商业航天快速发展。优化商业航天在产业链中布局，鼓励引导商业航天企业从事卫星应用和航天技术转移转化。制定吸引国际商业航天领域高端人才与创新团队落户的特别优惠政策。
2022年1月	国家发改委	《关于深圳建设中国特色社会主义先行示范区放宽市场准入若干特别措施的意见》	充分理解和用好用活《特别措施》，在互联网+医药健康、高端装备产业化、农业金融、艺术品交易、特色职业教育、商业航天、智慧驾驶等领域参照特别措施大胆试验和勇于突破，以自身企业发展战略主动对接国家战略，加大科技创新投入增强新动能。
2022年1月	国务院	《“十四五”数字经济发展规划》	加快建设信息网络基础设施演进升级，加快布局卫星通信网络等，推动卫星互联网建设。
2022年2月	国务院	《“十四五”国家应急体系规划的通知》	稳步推进卫星遥感网建设，开发应急减灾卫星综合应用系统和自主运行管理平台，推动空基卫星遥感网在防灾减灾救灾、应急救援管理中的应用。
2023年2月	工业和信息化部	《关于电信设备进网许可制度若干改革举措的通告》	将卫星互联网设备、功能虚拟化设备纳入进网许可管理，既符合信息通信技术发展方向，也是保障电信安全、互联互通和广大用户权益的必然要求。
2023年3月	工信部	《对地静止轨道卫星动中通地球站管理办法》	推动了国内对地静止轨道卫星动中通地球站（以下简称动中通地球站）在船舶、航空器、铁路机车、车辆等移动平台上的应用发展。

资料来源：工信部官网、中国政府网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

商业航天领域投融资活跃，促进产业发展

中美投资总额领先，带动商业航天市场发展。据太空资本（Space Capital）数据统计，2013-2022 年以来，全球 1791 家太空技术领域的企业，一二级市场累计募集 2724 亿美元。2021 年全球对太空领域的各项技术投资达 474 亿美元，为 10 年来新高。2022 年投资总额 201 亿美元，下降了 58%，SpaceX 在 2022 年仍筹集到 20 亿美元。2013-2022 年，美国共投资 1260 亿美元，占全球总投资的 46%；中国共投资 790 亿美元，占全球总投资的 29%。全球 77%的投资聚焦在导航定位和授时的相关应用领域，总额达 2084 亿美元；太空基础设施投资额约为 547 亿美元（占 20%），主要由火箭发射和卫星通信投资推动；涉及信息分发的分配领域仅有 93 亿美元投资，投向连接和处理大量地面用例数据的软硬件开发公司。

国内商业航天企业融资逐渐回暖

睿兽分析数据显示，2016-2022 年商业航天一级市场累计公开披露融资事件 366 个，披露融资总额 262.7 亿元。从 2016 到 2020 年，融资事件稳步增长，2020 年披露融资金额最高（70.4 亿元）；2021 年融资热度稍降（披露融资金额 48.91 亿元）；2022 年呈现回温上升趋势，已公开披露融资事件 54 个，披露融资金额 66.65 亿元。

星辰空间完成数千万元 Pre-A 轮融资

2022 年 8 月 17 日，北京星辰空间科技有限公司（以下简称“星辰空间”）已经完成数千万元的 Pre-A 轮融资，本轮融资由普华资本和启赋资本投资，北拓资本担任本轮独家财务顾问。星辰空间创始人潘海林表示，本轮融资资金将用于产品研发和批量生产等业务方面。星辰空间成立两年时间，是一家以空间推进技术为核心能力的商业航天高科技创业公司，产品研发能力覆盖了霍尔推进、离子推进、化学推进、空间流体管理等，其致力于成为国际一流空间在轨服务提供商。

航天科工火箭技术有限公司公司完成 B 轮融资

2022 年 6 月 27 日，航天科工火箭技术有限公司 B 轮增资扩股项目签约仪式在武汉举行，募集资金总额达 15.855 亿元。火箭公司于 2016 年 2 月在武汉注册成立，是我国首家专业提供商业航天发射服务的公司。截至目前，火箭公司成功完成 13 次商业发射服务，将 23 颗卫星送入预定轨道。此次募资主要用于固体运载火箭研发及性能提升、液体动力关键技术攻关及发动机研制、研发保障条件建设，提高火箭公司研发和生产能力，助力快舟火箭型谱系列化、商业发射常态化，进一步提升公司核心竞争力。

“中科天塔”宣布完成近亿元 Pre-A 轮融资

2022 年 11 月 29 日，商业航天硬科技企业“中科天塔”宣布完成近亿元 Pre-A 轮融资。本轮股权融资投资方为常青资本、中科创星以及唯尔投资。本轮融资主要用于中科天塔加快布局卫星应用的研发能力和产业化能力，建设一流的商业航天中心。

星启宇航完成数千万元天使轮融资

2023 年 2 月 16 日，北京星启空间宇航科技有限公司（以下简称“星启宇航”）近期宣布完成数千万元的天使轮融资，本轮融资由启赋资本领投，星启宇航创始人曾小金表示，本轮融资资金将用于产品研发和市场拓展等业务方面。星启宇航成立于 2018 年，是一家以各类飞行器天线、相控阵天线、大型展开天线、特种天线等产品研发生产为主的通讯设备硬件供应商。目前，星启宇航的主要业务方向为星载天线及各类相控阵天线，能够有效服务通信、导航、遥感卫星的数据传输环节，助力我国航天事业发展。

免责

更多免费行业报告 并购家 www.ipoipo.cn

极光星通获数千万元 A 轮融资，商业化落地加速，助力卫星星座组网建设

2023 年 3 月 10 日，北京极光星通科技有限公司（以下简称“极光星通”）完成 A 轮融资。本轮融资由方正和生领投，顺为资本联合投资，用于后续持续进行的研发投入，并开展卫星激光通信终端及组件批量交付的产能建设。极光星通成立于 2020 年，是聚焦于空间激光通信技术研发及相关产品制造的国家高新技术企业，核心技术自主可控。公司核心产品涵盖星间同轨、星间异轨、星地和临近空间等应用领域，致力于为用户打造多场景的空间激光通信组网解决方案。

火箭相关公司：星际荣耀、星河动力、蓝箭航天、深蓝航天、中科宇航、天兵科技、凌空天行、东方空间

星际荣耀（未上市）

北京星际荣耀空间科技股份有限公司成立于 2016 年 10 月，致力于研发优秀的商业运载火箭并提供系统性的发射解决方案，为全球商业航天客户提供更高效、更优质、更具性价比的发射服务，以大幅提升人类自由进出空间的能力。公司专注于智能运载火箭产品研发，为全球卫星及星座客户提供一体化的商业发射服务。

星河动力（未上市）

星河动力航天公司是国内领先的商业航天技术公司，公司核心产品包括“智神星”系列中型重复使用液体运载火箭、“谷神星”系列轻小型固体运载火箭等。在航天发射业务方面，公司创造了国内首家实现连续发射成功、首家将商业组网卫星送入 500km 太阳同步轨道、首家掌握一箭多星发射能力等多个行业第一。

蓝箭航天（未上市）

蓝箭航天空间科技股份有限公司是国内领先的航天运输系统创建及运营的商业公司，成立于 2015 年，国家高新技术企业，国际宇航联合会（IAF）会员。公司以国家战略为指引，以市场需求为导向，致力于研制以液氧甲烷作为推进剂的中大型运载火箭系列产品，为全球市场提供高性价比、高可靠性的发射服务。蓝箭航天是全国首家取得全部准入资质的民营运载火箭企业，并于 2018 年 10 月 27 日完成中国首次民营运载火箭发射。2022 年 12 月，蓝箭航天朱雀二号（ZQ-2）运载火箭在酒泉卫星发射中心执行首次飞行试验任务，标志着液氧甲烷发动机技术在中国的首次应用。

深蓝航天（未上市）

江苏深蓝航天有限公司聚焦在液体回收复用运载火箭方向，是一家自研自产可回收复用火箭并向市场提供发射服务的商业航天高新企业。深蓝航天总部位于江苏省南通市，在北京亦庄设有火箭总体、液体发动机研发中心；在陕西铜川建有火箭动力系统测试基地。目前，公司先后成功完成十米级、百米级和公里级 VTVL 垂直回收飞行试验。

中科宇航（未上市）

广州中科宇航探索技术有限公司由中科宇航正式宣布迁入广州南沙后更名而来，成为广州首家商业航天“独角兽”企业。该公司由中国科学院以及国内航天领域领军团队共同发起设立，致力于构建宇航科技探索和宇航运载飞行器研发集成平台，打造中国一流、国际领先的军民融合商业航天企业。核心团队均来自中国运载火箭国家队，团队平均拥有超过 20 年的运载火箭研制经验。

天兵科技（未上市）

北京天兵科技有限公司，成立于 2015 年 6 月，是一家先进宇航推进系统供应商及航天飞行器提供商。产品布局上，除了即将发射入轨的国内首枚民营中型液体运载火箭，未来几年，天兵科技还在重复使用大型液体运载火箭、重型液体运载火箭以及空天往返载人飞船等主营产品上做出了战略布局。

更多免费行业报告 并购家 www.ipoipo.cn

凌空天行（未上市）

北京凌空天行科技有限责任公司成立于 2018 年 10 月 8 日公司，研发总部位于北京市亦庄经济技术开发区。公司是国内从事高超音速飞行技术与应用服务的商业航天企业，具有全系统覆盖的高超音速飞行器总体设计能力，具备完整的火箭总装队伍，可独立完成火箭总装、测试、发射实施全流程工作。

东方空间（未上市）

东方空间科技有限公司成立于 2020 年 6 月，主营业务为太空运输，是一家运载火箭和新型空天运输工具的设计制造公司。公司通过“引力”系列运载火箭的研发，打造一次性使用、可回收重复使用和载人飞行等系列化、多样化的空天运输产品，快速形成低成本、规模化、高效快捷的发射服务能力，覆盖 LEO、SSO、GTO、GEO 等各类轨道。

卫星制造相关公司：银河航天、长光卫星、时空道宇、工大卫星、国博电子、铖昌科技、海格通信、天银机电

银河航天（未上市）

银河航天成立于 2018 年，致力于通过敏捷开发、快速迭代模式，规模化研制低成本、高性能小卫星，打造全球领先的低轨宽带通信卫星星座，建立一个覆盖全球的天地融合通信网络。2022 年 3 月 5 日，银河航天 02 批卫星发射成功，与首发星共同组成我国首个低轨宽带通信试验星座，构建了星地融合 5G 试验网络“小蜘蛛网”，具备单次 30 分钟左右的不间断、低时延宽带通信服务能力，可用于低轨卫星互联网、天地一体网络等技术验证。

长光卫星（未上市）

长光卫星成立于 2014 年，是我国商业航天领域第一家集卫星研发制造、运营管理和遥感信息服务于一体的商业遥感卫星公司。公司自主建设并运营管理着目前全球最大的亚米级商业遥感卫星星座“吉林一号”，能够为客户提供高时间分辨率、高空间分辨率、高光谱分辨率、快速广域覆盖的卫星遥感数据以及以卫星遥感数据为基础的空间信息综合应用服务。同时，公司能够为客户提供定制化的卫星制造及相关服务，包括卫星整星及部组件、试验与测试服务、搭载服务、冠名服务等。

时空道宇（未上市）

时空道宇在商业卫星制造、天地一体化高精度位置服务、天基通信服务领域，为用户提供低成本、高可靠的卫星及供应链产品，提供极具竞争力的、安全可信赖的解决方案与服务，深度融合航天技术与汽车制造、未来出行、人工智能，与生态伙伴开放合作，持续为客户创造价值，构建天地一体化高精时空信息系统，打造新一代航天数字经济业态。

工大卫星（未上市）

哈尔滨工大卫星技术有限公司于 2020 年成立，是哈尔滨工业大学卫星制造科技成果转化平台，是由哈尔滨工业大学资产经营有限公司与哈工大卫星技术研究所技术团队共同发起设立的商业航天公司。公司具备航天系统内完整研制体系，拥有 SATware 柔性化卫星技术平台体系及核心关键技术，具备通、导、遥微小卫星整星批量化设计、制造及在轨交付能力。公司致力于提供国内领先的整星在轨交付服务。

国博电子（688375 CH）

国博电子成立于 2000 年，专注于有源相控阵 T/R 组件和射频集成电路相关产品的研发、生产和销售。在民品领域，公司砷化镓基站射频集成电路主要应用于移动通信基站领域，并逐步拓展到移动通信终端和无线局域网领域。军品领域，公司于 2019 年整合中电科 55 所（未上市）微系统事业部，夯实了从芯片到组件的完整设计平台，并发展成为能够批量提供有源相控阵 T/R 组件的领先企业。公司 T/R 组件产品主要应用于弹载和机载领域。在星载领域，公司保持研发跟进和技术储备，已根据市场需求开发相关产品。

免责
更多免费行业报告 并购家 www.ipoipo.cn

铖昌科技（001270 CH）

铖昌科技成立于 2010 年，主营业务为微波毫米波模拟相控阵 TR 芯片的研发、生产、销售和技术服务，主要向市场提供基于 GaN、GaAs 和硅基工艺的系列化产品以及相关的技术解决方案。公司产品主要包含功率放大器芯片、低噪声放大器芯片、模拟波束赋形芯片及相控阵用无源器件等，频率可覆盖 L 波段至 W 波段。产品已应用于探测、遥感、通信、导航、电子对抗等领域，在星载、机载、舰载、车载和地面相控阵雷达中列装，亦可应用至卫星互联网、5G 毫米波通信、安防雷达等场景。

海格通信（002465 CH）

海格通信成立于 2007 年，前身是国家第四机械工业部（原国家电子工业部）国营第七五〇厂，目前是广州无线电集团的主要成员，业务覆盖“无线通信、北斗导航、航空航天、软件与信息服务”四大领域，是全频段覆盖的无线通信与全产业链布局的北斗导航装备研制专家、电子信息系统解决方案提供商，是行业内用户覆盖最广、频段覆盖最宽、产品系列最全、最具竞争力的重点电子信息企业之一。2023 年 3 月 16 日，公司公告定增不超过 20 亿元，定增项目覆盖“北斗+5G”通导融合、无人系统产业、卫星互联网；我们认为，此次定增有利于公司增强在北斗行业的龙头地位，并拓展卫星互联网的业务边界，抓住市场机遇。

天银机电（300342 CH）

天银机电控股子公司天银星际是一家专注于星敏感器研发的企业。其核心技术来自于清华大学 20 年空间技术积累。公司自主研发生产了皮型、纳型两大系列星敏感器，约 200 台+产品已无故障在轨运营，在我国探月工程、高分专项等国家重大航天任务实践中发挥了重要作用。公司产能达 2000 台套每年，适应多种航天器应用要求，为不同领域的客户提供专业、先进、经济的产品。公司在国内商业卫星市场占据优势地位，应用的卫星包括吉林一号卫星、高分卫星、月球探测卫星、全球多媒体通信卫星、鸿雁卫星、科学探测卫星、立方星等。

地面设备相关公司：震有科技、航天驭星、信科移动

震有科技（688418 CH）

震有科技成立于 2005 年，是一家专注于提供通信设备和解决方案的国家级高新技术企业。公司主营业务为核心网络系统、集中式局端系统、指挥调度系统和技术与维保服务等。在核心网领域，公司是国内为数不多的可以提供卫星 5G 核心网的公司之一，包括高轨卫星和低轨卫星核心网。公司于 2022 年已发布基于国产化软硬件平台的 5G 核心网商用版本、发布支持卫星通导一体增强、卫星物联网、多种体制卫星接入网、卫星安全加密等丰富卫星业务的 5G 核心网商用版本。公司持续参与卫星通信的标准体系建设和方案论证设计，助力卫星通信系统建设。

航天驭星（未上市）

航天驭星成立于 2016 年，是国内领先的全球航天基础设施建设和运营服务商。航天驭星核心团队均来自传统航天系统核心单位，2 人现为航天法编写组成员。航天驭星以“让卫星更好用，让卫星更易用”为使命，致力于为全球商业卫星提供低成本、高可靠、更便捷的卫星在轨综合管理和航天数据应用一站式解决方案。航天驭星的业务主要围绕天地通信和航天器在轨管理开展，可以提供包括卫星频率协调与发射许可申请、卫星测运控数传一体化解决方案、卫星寿命周期空间态势与遥感卫星定标服务。目前提供的产品主要有多功能基带、星上通信机、应答机、测运控软件；测控数传一体化多功能地面站、便携站等。

免责

更多免费行业报告 并购家 www.ipoipo.cn

信科移动（688387 CH）

信科移动成立于 1998 年，是从事移动通信国际标准制定、核心技术研发和产业化的唯一一家央企控股的高新技术企业。公司在星地融合与卫星互联网方向进行战略布局，推动 5G NTN 的国内外标准制定，牵头多个标准立项项目。公司 6G 技术预研的前沿技术方向 10 项，在新型多天线技术、网络分布式智能化、以用户为中心的接入网架构、星地融合卫星互联网、高精度定位等多个研究方向上保持业界领先或先进水平，完成 2022 年国内 IMT-2030（6G）推进组首次关键技术测试，在无线空口传输和网络的多个方向开展关键技术原理验证的测试，在业界首次实现基于 RIS 的毫米波基站与商用终端的多波束赋形技术验证。

卫星运营服务相关公司：中国星网、中国卫通、航天宏图、中科星图

中国星网（未上市）

中国星网成立于 2021 年，是中央直接管理的唯一一家从事卫星互联网设计建设运营的国有重要骨干企业。中国星网致力于打造卫星互联网产业发展的核心力量和组织平台，为用户提供先进、优质、安全、经济的空间网络信息服务，成为具有全球竞争力的世界一流卫星互联网公司。

中国卫通（601698 CH）

中国卫通成立于 2001 年，是中国航天科技集团有限公司从事卫星运营服务业的核心专业子公司。公司运营管理各类通信广播卫星，为用户提供广播电视、通信、视频、数据等传输服务。公司是我国唯一拥有自主可控通信广播卫星资源的基础电信运营企业，被工信部列为国家一类应急通信专业保障队伍，是国家行业主管部门直接指挥调度的保障力量。公司广播通信服务覆盖中国全疆域、周边及“一带一路”主要区域。“海星通”全球网服务覆盖范围全球超过 95%的海上航线，Ka 高通量卫星覆盖中国全境及周边区域，能够为我国及“一带一路”沿线航空、航海、应急、能源、林草等行业及普遍服务用户提供高速的专网通信和卫星互联网接入等服务，为边远地区提供安全可靠、无缝覆盖的信息传输手段。

航天宏图（688066 CH）

航天宏图成立于 2008 年，是国内领先的全球航天基础设施建设和运营服务商。公司研发了具有完全自主知识产权的遥感与地理信息一体化软件 PIE（Pixel Information Expert），拥有国内首个遥感与地理信息云服务平台 PIE-Engine，实现遥感基础软件的国产化替代，为政府、企业、高校以及其他有关部门提供基础软件产品、系统设计开发、遥感云服务等空间信息应用整体解决方案。

中科星图（688568 CH）

中科星图创立于 2006 年，是中国科学院空天信息创新研究院投资的国有控股高新技术企业。公司长期专注数字地球行业，持续研发数字地球相关产品和核心技术，陆续推出 GEOVIS 数字地球基础软件系列产品，并在此基础上，形成以特种领域、智慧政府、气象生态、航天测运控、企业能源、线上业务六大板块业务为核心的 GEOVIS 数字地球应用软件系列产品，为政府、企业、特种领域及大众用户提供软件销售与数据服务、技术开发与服务、专用设备及系统集成等业务。

免责

更多免费行业报告 并购家 www.ipoipo.cn

元宇宙 2023: 苹果这次有什么不同？

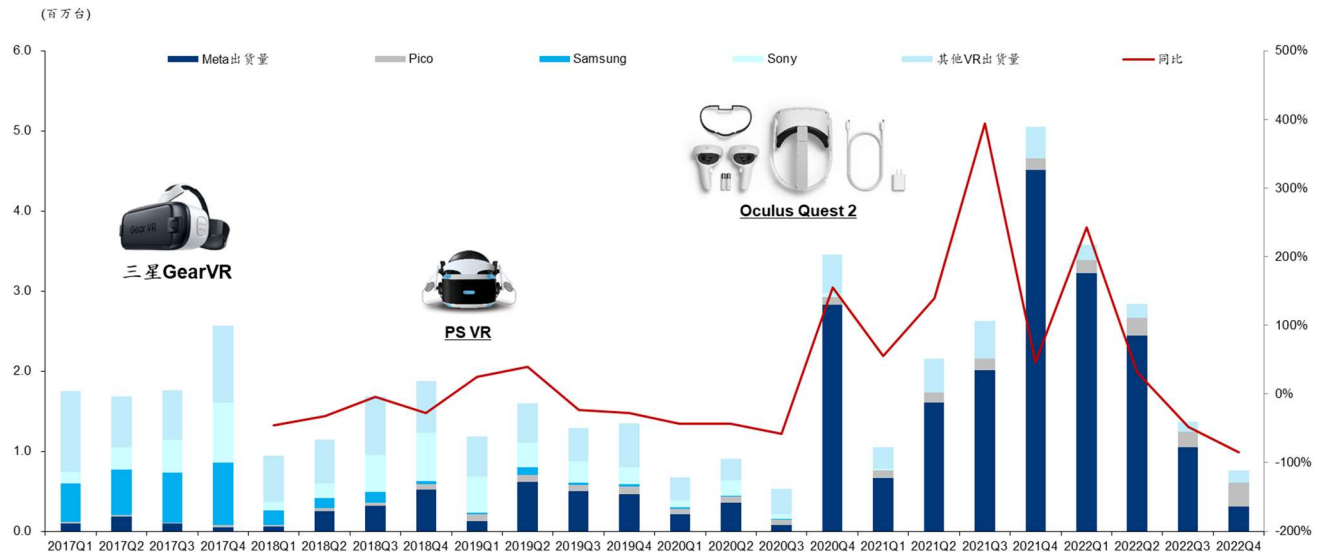
Meta Oculus 的失败教训

北京时间 2023 年 6 月 6 日，苹果在 WWDC 2023 大会上，正式发布其第一款头显硬件苹果 MR。该产品定价 3499 美元起，预计生命周期出货量或在 100 万台（每年 50 万台），主要面对苹果应用的开发者，同时，彭博社报道，苹果目前正在研发第二代量产版的 MR，或将于 2024/2025 年推出。第二代产品出货量或比第一代高一个数量级。

此次苹果 MR 是科技巨头第四次尝试 MR 这个新硬件品类。前三次（2015-2017 三星 VR 盒子，2016-2018 索尼 PS VR，2020-2022 Meta Oculus 2）上市时市场抱有很大期待，但最后都销量不达预期。

以 Meta（Meta US）为例，早在 2014 年，Facebook 就开始切入了 VR/AR 市场，2014 年 3 月，Facebook 以 20 亿美元收购了 Oculus VR，切入元宇宙硬件市场，2019 年 6 月，Facebook 发布 Libra 虚拟货币，并收购 VR 游戏发行商 Unit 2 Games，不断丰富元宇宙内容端。2021 年 12 月，Facebook 集团正式改名为 Meta。

图表296：三星、索尼和 Meta 在 MR 上的探索



资料来源：IDC，华泰研究

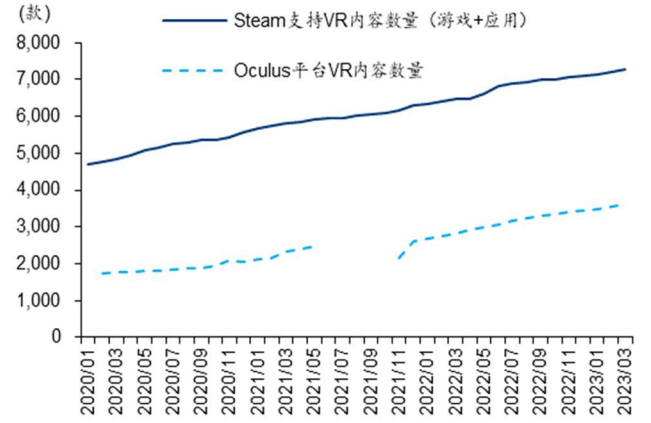
#1：社交健身等轻量级应用发展缓慢，Oculus 依然是游戏终端

我们认为上 Quest 2 未能达到 Meta 目标的主要原因是轻量化应用没有成功。当前大部分用户使用 Quest 2 仍然仅局限于游戏，而办公、社交等轻量化应用并没有成功，Horizon 平台忠实用户始终限制于某些游戏极客。

截至 2023 年 3 月，Steam 平台支持 VR 的内容为 7,290 款，同比增长 12.6%，Oculus 平台支持内容数为 3627 款，同比增长 28.4%。但平台用户活跃度持续疲弱，根据青亭网估算，2023 年 2 月 SteamVR 活跃用户约 182.16 万人，同比下滑 28.3%，而 Meta Horizon 自 2022 年春季以来，用户群数量一直呈下降趋势，到 22 年中月活总用户不到 20 万。

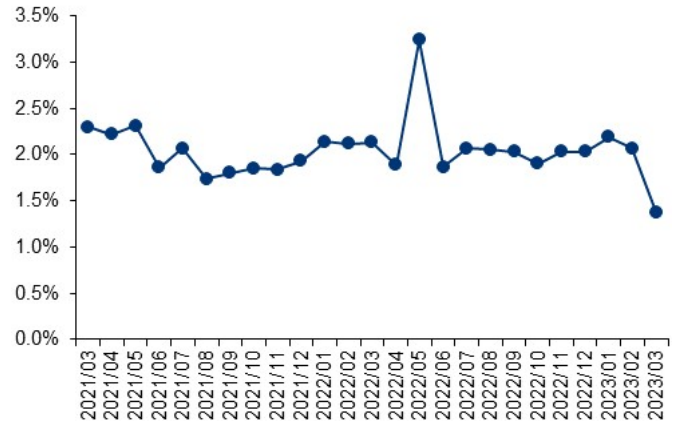
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表297： Steam 与 Oculus 平台 VR 数量



资料来源：Steam 官网，Oculus 官网，华泰研究

图表298： 2021.03-2023.03 SteamVR 占 Steam 总玩家份额统计



资料来源：Steam 官网，Oculus 官网，华泰研究

2021 年 12 月，Meta 正式开放 VR 社交平台 Horizon，用户可以在平台上创建虚拟形象，并与其他用户交互，但 Horizon 中用户的化身只有上半截身体。Horizon 实际上由 Horizon Worlds、Horizon Venues、Horizon Workrooms 和仍在开发中的 Horizon Home 组成。

Horizon Worlds 由用户生成的内容构成，是一个全新的社交虚拟世界。创作者可以通过使用 VR 应用程序中的工具来自行设计世界。用户可以在 Horizon Worlds 中冒险、娱乐，获取与日常生活完全不同的新奇体验。

图表299： Horizon 应用#1：Oculus 重度游戏

资料来源：Meta 官网，华泰研究

图表300： Horizon 应用#2：社交游戏#

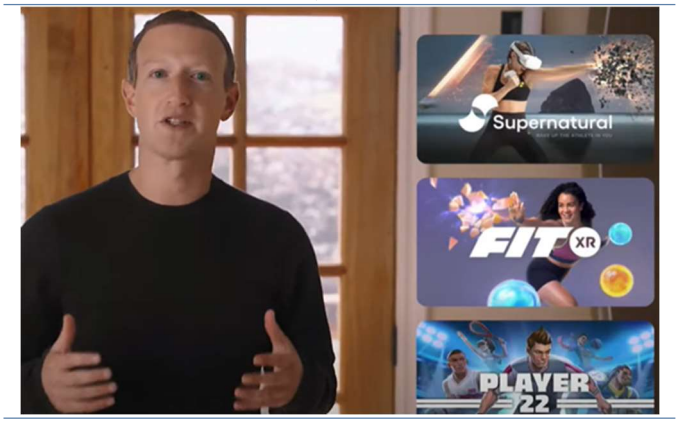
资料来源：Meta 官网，华泰研究

Horizon Venues 则为在 VR 中举办大型活动提供了必要的条件，2021 年 11 月，Meta 官方指出，用户可以通过 Horizon Venues 在 Quest 中观看 NBA 比赛。2021 年 2 月超级碗表演结束之后，美国老牌摇滚乐队 Foo Fighter 在 Horizon Venues 中举办一场 VR 演唱会，Quest 用户将可以坐在前排体验。

而 Horizon Workrooms 让 VR 办公更易用、更好用，为企业和员工提供了更多在疫情期间居家办公提高沟通效率的方式。然而 Horizon Workrooms 并不是 Meta 在打造元宇宙社交平台时的第一次尝试。

更多免费行业报告 并购家 www.ipoipo.cn

图表301： Horizon 应用#3：健身游戏



资料来源：Meta 官网，华泰研究

图表302： Horizon 应用#4：电话会议

资料来源：Meta 官网，华泰研究

#2：产品没有成熟之前以亏损为代价大规模强推产品不可取

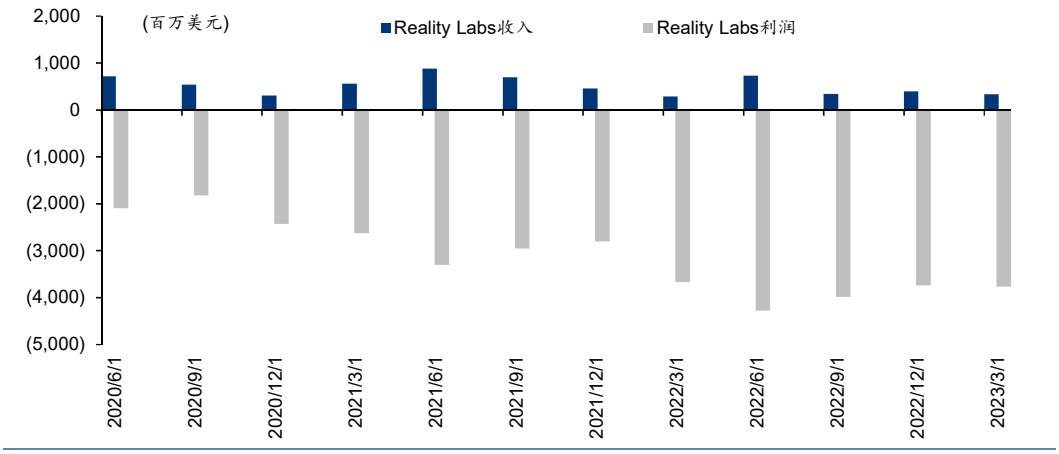
从 Meta 的商业角度来看，Meta Quest 2 尝试以失败告终。20 年到 22 年期间公司通过补贴进行亏损销售，期望能够迅速提高 VR 渗透率，推广自己 Horizon 平台。但大部分 Horizon Worlds 的用户在使用该 App 一个月后便不再使用，自 2022 年春季以来，用户群数量一直呈下降趋势，到 22 年中月活总用户不到 20 万，最后公司降低补贴止损退出，这次商业尝试以失败告终。

Meta Quest 2 产品只是一个及格产品。在很多方面并没有达到成熟产品的水平。重量方面，我们认为一款成熟产品重量最高不能超过 70-80g，而 Quest 2 产品重量达到 500g，佩戴体验并不舒适。续航方面，Quest 2 产品只能支持 2 个小时的使用，不能够支持全天候的佩戴。显示方面，产品只能提供双眼 4K 画质，且不能根据用户注视点的变化而调整，沉浸感一般，且仍有部分使用者感到眩晕。交互方面，产品仍然只能通过手柄进行交互，使用并不方便。

产品不成熟时，通过亏损大规模急推产品不可取。根据 Digitimes 测算，Quest 2 产品成本超过 300 元，Meta 以近乎亏损的方式进行销售。我们认为，一方面 Meta Quest 2 硬件并没有能够为使用者提供有吸引力的使用体验，另一方面，软件方面，办公、社交等轻量化应用没有成功，大部分用户使用 quest 2 仍然仅局限于游戏，Horizon 平台忠实用户始终限制于某些游戏极客。

大规模裁员，以节省成本和资源。2022 年 11 月份 Meta 已经宣布裁员 1.1 万人；2023 年 3 月份，扎克伯格又宣布将在未来的几个月裁员 1 万人。Meta 今年已经累计裁掉了 2.1 万名员工，相当于去年 9 月报告的雇员总数（87314 人）的 24%。

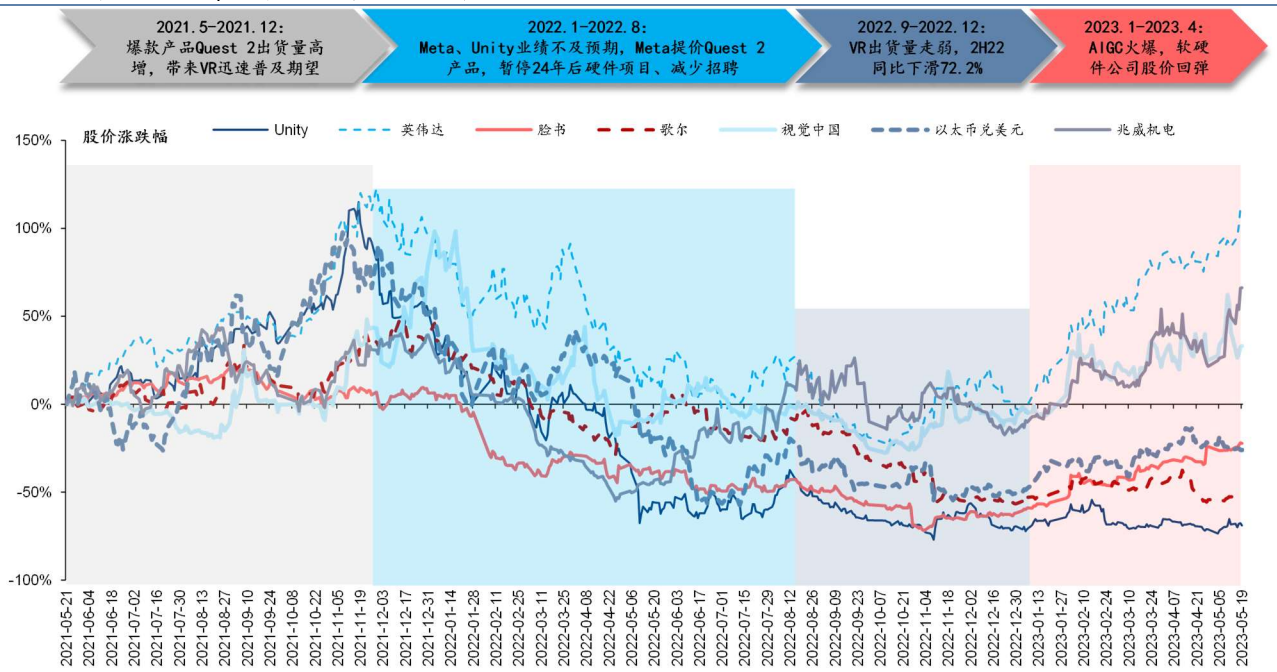
图表303： Meta Reality Lab 的季度收入和亏损



资料来源：彭博，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表304：AI 需求快速崛起，元宇宙二级市场热度回升



资料来源：，华泰研究

苹果带来的变化

#1：用户体验更好的硬件

在 WWDC 2023 大会的 One More Thing 环节中，库克正式发布了公司的第一款 VR/MR 设备 Vision Pro，定价 3499 美元起，计划 2024 年初在美国正式发售，然后在其他国家和地区逐步开售。同时，Apple 也在 WWDC 上同步发布了为空间计算打造的全新操作系统 VisionOS，可支持 iOS 以及 iPadOS 上的各类应用。

图表305：Vision Pro 2024 年开始发售

资料来源：苹果官网，华泰研究

图表306：苹果首款 MR Vision Pro

资料来源：智东西，华泰研究

硬件方面，专属感知芯片、无手柄交互是亮点，整体规格与彭博社等媒体预测差别不大。1）芯片：配备两块芯片，其中一颗为专属感知计算芯片 R1，苹果表示，R1 需处理来自 12 部相机、5 个传感器和 6 个麦克风的信息，以确保内容毫无延迟地出现在用户眼前。R1 可在 12 毫秒内将新影像串流至显示器；**2）显示：**采用 Pancake+MicroOLED 方案，搭载了两块 1.4 英寸、4K 级高分辨率的 Micro OLED 屏幕和一块 AMOLED 屏幕用于双向透视；**3）交互：** Vision Pro 头显支持交互方式以眼动交互、语音交互和手势交互。这也是目前业内罕见地采用无手柄交互的头显产品。**4）续航：**配备一个比 MagSafe 电池包稍大一些的外接电池，单次充电的续航大概为两个小时。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表307：苹果 MR Vision Pro 与其他器件头显对比

	Pico 4Pro	Meta Quest Pro	苹果Vision Pro	HTC Vive XR Elite	PlayStation VR 2	Meta Quest 3
	￥3,799	约￥7,098	约￥24,850	约￥7,809	￥4,499	约￥3,546
	597克	722克	未公布	625克	560克	未公布
	20W充电功率 3小时续航	45W充电功率 2.5小时续航	外置电池 续航2小时	30W充电功率 2小时续航	外接PS5使用	未公布
	高通骁龙XR2	高通骁龙XR2+	Apple M2+Apple R1	高通骁龙XR2	凭借PS5驱动	新一代高通骁龙XR芯片
	LCD Pancake光学透镜	Mini LED Pancake 光学透镜	定制 Micro-OLED	LCD Pancake 光学透镜	OLED 菲涅尔光学透镜	未公布
	105°	106°	未公布	110°	110°	未公布
	双眼4K+	双眼2K+	双眼8K	双眼4K+	双眼4K+	未公布
	面部追踪 眼动追踪手柄	面部追踪 眼动追踪 手势控制手柄	眼动追踪 面部追踪 手势控制 键鼠操作	面部追踪 眼动追踪手柄	眼动追踪手柄	未公布
	Type-C	Type-C+磁吸充电站	外置电池充电	Type-C	Type-C	未公布

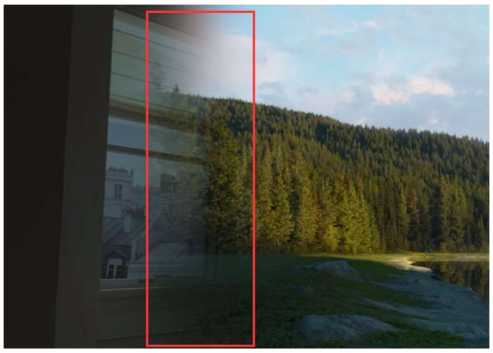
资料来源：爱范儿，华泰研究

外观设计上，分离式电池设计仍有改进空间。产品正面是以三维方式构造的层压玻璃，作为 EyeSight 的显示和一系列环境感知摄像头和传感器的镜片，外框有个按钮可拍摄空间照片和视频，数码旋钮点按时可调出主视图，旋转时可调节空间环境的沉浸度，外框由特质的铝合金精制而成。一整片 3D 编制的头带营造出独特的弧形结构，可缩放。和蔡司一起打造定制的光学插片，通过磁力吸附在镜片上。插电可全天使用，接到外置高性能电池最长可用两小时，分离式的电池让头部不用承受过多重量。

图表308：分离式电池设计仍有改进空间

资料来源：苹果官网，华泰研究

图表309：通过调整 Vision Pro 设备上方的数码旋钮无缝切换



资料来源：智东西，华泰研究

芯片方面，配备一颗为专属感知计算芯片 R1。搭载 M2 和全新的 R1 芯片（处理器芯片），R1 需处理来自 12 部相机、5 个传感器和 6 个麦克风的信息，以确保内容毫无延迟地出现在用户眼前。R1 可在 12 毫秒内将新影像串流至显示器。

空间音频。集成的双驱动单元音频组件可带来个人化的音效，除了个人音效，Vision Pro 会分析空间的特征和材质，以使声音适配用户的房间，完成音频射线追踪。

丰富传感器配置。头显配备 12 个摄像头，5 个传感器和 6 个麦克风，实现精准的头部和手部追踪，进行实时 3D 映射等，同时通过高速相机和一圈 LED，在眼球上投射非可见光阵来实现灵敏又直观的采集。同时通过传感器实现 Optic ID，以眼球虹膜解锁 Vision Pro。

更多免费行业报告 并购家 www.ipoipo.cn

显示方面，Micro-OLED 提供更优秀沉浸式体验。Micro-OLED 技术，显示背板以一颗 iphone 像素的空间容纳了 64 颗像素，宽度却只有 7.5μm，两块背板加一起有 2300 万颗像素，超过 4K 分辨率，设计用了定制的三镜式镜片。

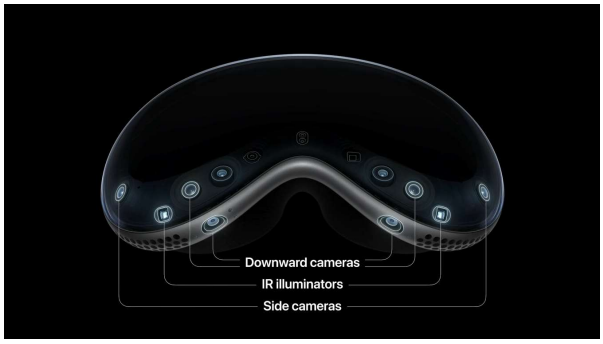
图表310：配备一颗专属感知计算芯片 R1

资料来源：苹果官网，华泰研究

图表311：苹果 Micro-OLED 显示屏提供

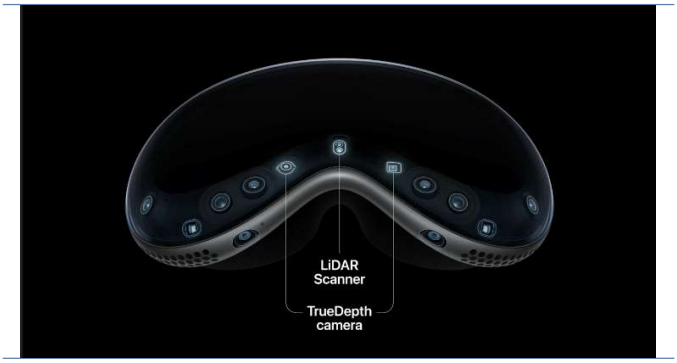
资料来源：智东西，华泰研究

图表312：追踪摄像头



资料来源：苹果官网，华泰研究

图表313：LiDAR 与深度摄像头



资料来源：智东西，华泰研究

双向透视技术 EyeSight 效果大大超出上面 Meta 的研究效果。通过一片前置 OLED 显示屏，当有人出现在周围时，Vision Pro 会自动显示用户的双眼。也让他人知道用户在干什么，是在使用 App 还是体验完全沉浸。即使在完全沉浸状态中，也能和他人保持交流，当有人经过时，会进入使用者的视野，此时 Vision Pro 会自动对他人显示用户的双眼。

图表314：Vision Pro 会自动对他人显示用户的双眼

资料来源：苹果官网，华泰研究

图表315：有人经过时，会进入使用者的视野

资料来源：苹果官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表316： Meta 双向透视效果较差

资料来源：映维网，华泰研究

图表317： LED 投射光阵帮助实现眼动追踪

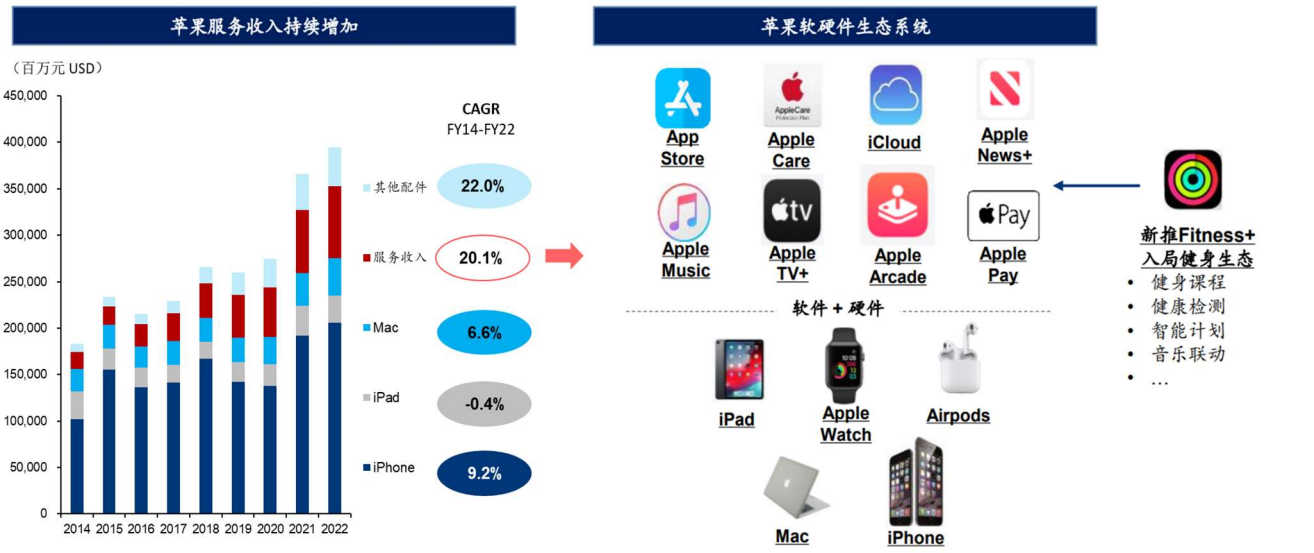
资料来源：苹果官网，华泰研究

#2：全球超过 3400 万的苹果开发者生态和丰富的 App 矩阵

经过 15 年的发展，苹果已经开发了涵盖云存储、音乐、流媒体、支付和健康在内丰富的应用矩阵。根据 4QFY22 财报，FY2022 年度苹果营业收入 3943.28 亿美元，根据苹果 1Q23 披露，苹果全球目前有 20 亿活跃用户，软件付费应用用户数到超过 9 亿。根据彭博社，苹果首款头显中，用户或能够通过 3D 交互方式访问其中部分应用。我们认为这可以大规模拓宽优秀的多维感知交互体验。

此外，根据苹果 WWDC2023 公布的数据，截至 2022 年 6 月 7 日，该公司目前有超 3400 万注册开发者，大中华区的注册开发者数量现已超过 500 万。我们认为通过苹果良好的软件生态，MR 能够快速实现与 iPad 应用程序的兼容，并进一步开发办公协作、健身冥想和游戏娱乐等主打领域的新应用。

图表318： 苹果的主要应用矩阵



资料来源：Apple 官网，，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

#3：大力发展影视娱乐和其他轻度交互娱乐软件及应用

注重影视娱乐和其他轻度交互娱乐。在《元宇宙 2023：苹果这次有什么不同？》(6/4)中，我们指出，Meta 此前商业尝试以失败告终主要因为办公、社交等轻量化应用并没有成功，Horizon 平台忠实用户始终局限于重度游戏玩家。此次苹果更多专注于轻量化应用的开发。

推出 visionOS，引入空间计算新时代。visionOS 具有全新的三维界面，使数字内容的外观和感觉存在于用户的物理世界中。通过动态响应自然光和投射阴影，它可以帮助用户了解比例和距离。为了实现用户导航和与空间内容的交互，Apple Vision Pro 引入了一个全新的输入系统，由人的眼睛、手和声音控制。用户只需查看应用程序、点击手指进行选择、轻弹手腕滚动或使用语音听写即可浏览应用程序。

图表319：VisionOS 是为空间计算打造的操作系统

资料来源：苹果官网，华泰研究

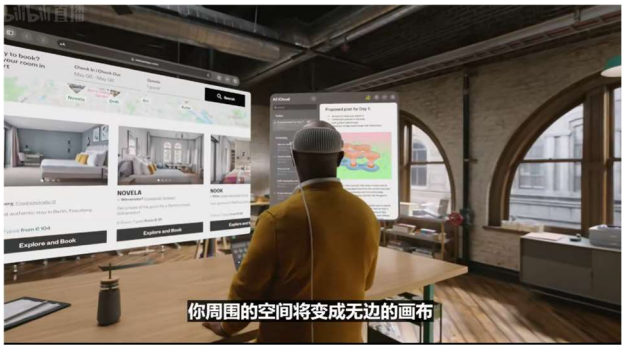
图表320：Vision Pro 登录界面

资料来源：苹果官网，华泰研究

所有 App 都摆在用户面前。可以随处摆放，调整大小来适应当前的任务。你所看到所处的空间和里面的一切，主视图会出现在正前方，所有 app 也出现在正前方，体验不受显示框限制，可以轻松随意缩放 app 界面，也可将 app 放置空间的任何一个地方。打开新的 app 不会出现挤占之前 app 的空间，而是以使用者为中心出现在新的空间。ipad 和 iphone 上的成千上万款 app 也能在 Vision Pro 上使用。用户可以将 APP 拉出来从各种角度看。

家居：实现 3D 视频拍摄和观影功能。首先是空间照片和视频功能，Vision Pro 是苹果的首台 3D 相机，它也能让你用 3D 和空间音频来播放视频。其次是娱乐功能，空间电影可以改变任何空间，Vision Pro 能够自动调暗周围光线，形成优美的光晕同时还有空间音频，为了营造更好的体验，还可将屏幕放大到超过你房间的维度，任何地方都能成为私人影院，配合 airpods 还能带来更好的降噪体验。最后是游戏功能，100 多款 Arcade 游戏在产品发售当天即可享用。

图表321：各类 APP 同时呈现在空间中



资料来源：苹果官网，华泰研究

图表322：应用：巨幕电影

资料来源：苹果官网，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表323： 应用：3D 视频拍摄

资料来源：苹果官网，华泰研究

图表324： 应用：3D 观影

资料来源：苹果官网，华泰研究

办公凸显多任务处理能力。Safari 界面更大，文字也非常清晰、Safari 能在更大的界面展开所有打开的标签页；打开的 app 置于三维空间内，例如可以在信息 app 中拖出三维物体，同时戴着 Vision Pro 也能看到周围，不影响周围人的互动。Vision Pro 还能与蓝牙配件（比如妙控板和妙控键盘）和 Mac 协作，可以将 Mac 的内容也提取到三维空间内。远程办公时，FaceTime 带有空间感，在 FaceTime 中还可借助同播共享与大家分享并使用各种 app，Vision Pro 上的 FaceTime 也能与 iphone、ipad、mac 等流畅配合。

图表325： 应用：办公视频会议与演讲

资料来源：苹果官网，华泰研究

图表326： 应用：办公协作

资料来源：苹果官网，华泰研究

与 Unity、微软、迪士尼等合作打造生态。苹果和 Unity 合作，将更多 3D app 带到 Vision Pro 上，常用的 Unity 游戏和 app 可获得完整权限来显示 Vision Pro 功能；与迪士尼合作引入更多娱乐内容；与微软合作打造生态，目前 Excel、Word 和 Teams 等 Microsoft 应用程序支持 Vision Pro。Zoom、Teams、Webex 等流行的视频会议应用程序可以支持虚拟渲染。数十万款 iPhone 和 iPad 应用程序能在 Vision Pro 发售时就绪。

图表327： 应用：赛车工业设计

资料来源：苹果官网，华泰研究

图表328： 应用：生产线设计

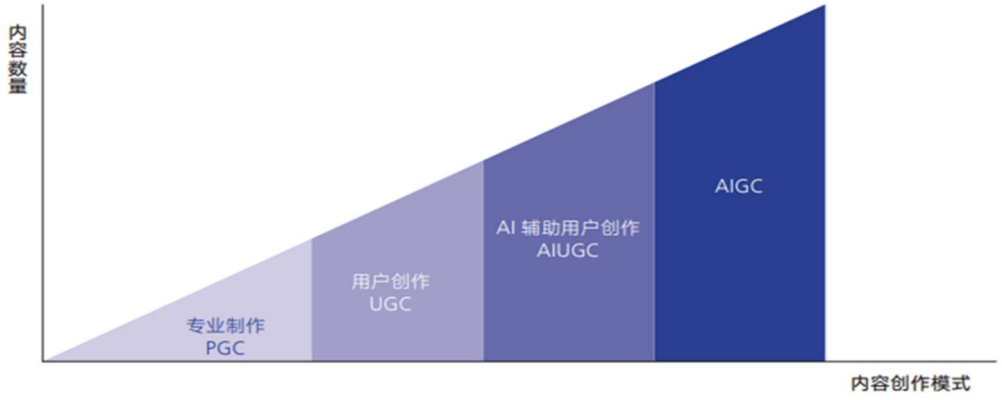
资料来源：苹果官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

生成式 AI 可能带来哪些变化

AIGC 技术的突破性进展引发内容生产方式变革。内容生产由 PGC（专业制作）和 UGC（用户创作）时代逐渐步入 AIGC 时代。AIGC 顺应了内容行业发展的内在需求，一方面内容消费量增加，急需降低生产门槛，提升生产效率；另一方面用户端表达意愿明显上升，消费者对内容形态要求更高，内容生成个性化和开放化趋势明显。AIGC 通过其强大的生成能力广泛服务于内容生产的各类场景和内容生产者，在内容行业的应用场景不断增加和拓展，将在内容生产中产生变革性影响。

图表329：AIGC 技术的突破性进展引发内容生产方式变革



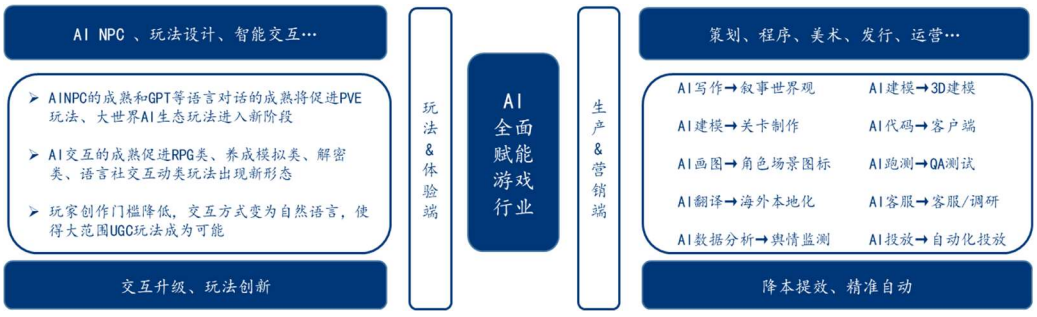
资料来源：华泰研究

AIGC 将为游戏玩家带来玩法体验端的创新升级，以及生产营销端的降本提效。

1）玩法体验端：目前看 AIGC 有望在 NPC 智能交互及玩法设计创新方面加速应用，AINPC 的不断成熟将改变游戏玩家与角色的交互方式，进一步提升游戏沉浸感；同时，AIGC 带来的 UGC 等玩法创新也有望带来全新的游戏类型，进一步提升游戏内容的丰富度及自由度。

2）生产营销端：以 ChatGPT 等大语言模型为基础的文本类 AI 生成工具、Stable Diffusion 及 Midjourney 等 AI 图像生成工具、Gen2 等 AI 视频生成工具等已在游戏行业各环节有所应用，同时英伟达、微软、Epic Games、Unity 等公司的工具产品也在探索 AIGC 相关功能的应用与结合。随着工具化 AIGC 产品和服务快速迭代，行业“卖铲人”将加速游戏行业享受 AIGC 赋能，带来进一步的降本提效空间。

图表330：AI 赋能游戏行业



资料来源：华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

#1：玩法&体验：AI NPC 改变玩家交互体验，助力玩法创新

游戏中加入 AI 驱动型虚拟 NPC，增强游戏人格化与陪伴感。2023 年 2 月，网易《逆水寒》手游官宣采用网易伏羲 AI 技术实装国内首个“游戏 GPT”，实现 NPC 高智能交互。游戏内的智能 NPC 系统完全由 AI 驱动，可以基于对话语义进行无限延展，与玩家进行更有场景感、真实感的交流。NPC 被赋予性格特质，与玩家之间的差异进一步缩小；NPC 之间可形成社会关系网，产生蝴蝶效应，提高游戏可玩性与多样性。

2023 年 4 月 28 日，MOD 开发者结合 ChatGPT、xVASynth 和 Whisper，为《上古卷轴 5》游戏中的 NPC 进行了一次人性化的改造。开发者使用这些 AIGC 工具为每个 NPC 赋予了独立思考，独立行为，记忆系统，并为每个 NPC 设定了丰富的基本背景，使之更好的符合游戏角色设定，将 NPC 从简单那的虚拟角色转化成具有独立人格的角色。此外，开发者使用 Whisper 将玩家的声音转换为文本，并通过 ChatGPT 生成聊天内容，依次实现玩家与 NPC 之间的互动对话，增强了游戏的真实感和玩家沉浸感。

通过自然语音与游戏角色交互，改变传统交互方式。基于 ACE for Games，在 GTC2023 上，NVIDIA 与初创公司 Convai 合作演示开发人员如何使用 ACE 构建游戏 NPC。在名为 Kairos 的游戏 demo 中，一个赛博朋克风格的拉面店场景中，玩家使用自己的声音与 NPC 拉面店的供应商 Jin 互动，Jin 根据玩家语音实时生成回答。示例中，Jin 对自然语言问题的回答真实，且与叙事背景故事一致。NPC 角色具备智能和反应能力，进一步增强游戏的沉浸感。

AIGC 驱动实时互动剧情引擎为游戏创新带来全新可能。2023 年 1 月，游戏公司 Cyber Manufacture Co.发布了最新技术预览的 AIGC 引擎 Quantum Engine。在其网站上，用户可以通过英语、法语、中文等多种语言，以实时对话形式与 AI 进行游戏剧情互动，AI 依据用户的表达生成相应剧情。在下一版本更新中，用户可以自主上传故事剧本，AI 通过学习上传剧本中故事背景、人物关系、表达逻辑和角色人物特点等，依据用户选择扮演剧本中角色并展开互动。

Open AI 助力生成实时游戏剧情，提升用户实时交互体验。Quantum Engine 利用低级 Open AI APIs 和开源模型，由 Cyber Manufacture Co.进行训练和微调，实现了上下文高度相关、基于文本的多模态（语音+文本+3D 模型）和多语言（例如英语、法语、中文、日语、西班牙语）理解场景中的用户意图、人物与人物的互动和环境反应。在公司发布的演示中，玩家可以扮演《黑客帝国》中的角色，与 NPC 进行交互，AI 根据剧本设定的世界观进行相应回答并推动剧情的发展。在更新的下一个版本中，用户还可以自主上传剧本，AI 通过学习上传剧本的人物关系、表达逻辑和角色人物特点等，领悟角色并与用户展开互动，从而在剧本设定内拓展全新互动内容，带来“千人千面”的个性化游玩体验。

以文本内容为基础，AIGC 助力玩法创新。根据 Quantum Engine 团队的规划，该游戏引擎发展规划分为三个阶段：（1）实现用户与 AI 的实时语音互动；（2）实现 AI 实时互动剧情与游戏画面相结合；（3）将该技术封装为一款可操作的游戏，最终成为 AIGC 方向的杀手级应用。目前公司发布的 Quantum Engine 技术预览，属于其整体规划的第一阶段成果。未来公司将重点推进 AI 生成剧情与游戏画面的结合，不断提升游戏可玩性。

#2：生产&营销：全流程赋能，AI 赋予行业降本增效广阔空间

AIGC 对游戏各环节应用正快速落地，游戏脚本创作、角色场景设计、代码编写、营销投放等多领域应用有望实现降本增效。目前我们看到在美术绘图、营销投放层面，AI 落地有望较快实现。游戏行业由于其强交互性，非常注重产品、内容的丰富程度，而在拓展游戏“开放度”过程中，剧情、美工等均需要大量的投入。AIGC 赋能内容生产，可以通过 AI 辅助美术绘画、程序化人物及场景生成等多领域应用，在技术层面打通了游戏厂家之间的壁垒，打破了传统游戏开发领域成本、质量、速度只能三选其二的三角模型。

免责

更多免费行业报告 并购家 www.ipoipo.cn

Stable Diffusion 模型成为 AI 图像领域的里程碑，由 Midjourney 生成的《太空歌剧院》出圈。2022 年 8 月，Stability AI 推出 Stable Diffusion 模型，随着算法和模型的持续优化，Stable Diffusion 的运行速度快、消费资源及内容较少，使用消费级显卡即可迅速生成高质量的图像，且该模型完全免费开源，所有的代码均在 GitHub 上公开，因此也为后续整个图像模型的更迭打下基础。据 TechCrunch，截至 2022 年 10 月，已有超 20 万开发者下载和获得 Stable Diffusion 的授权，各渠道的累计日活已经超过 1,000 万；基于 Stable Diffusion，面向消费者的 AI 智能绘图软件 Dream Studio 用户数超 150 万，已生成 1.7 亿张图片。2022 年 8 月，由 Midjourney 生成的《太空歌剧院》在美国科罗拉多州艺术博览会上获得“数字艺术”类别的冠军，使得 AI 绘画引发了广泛的关注，围绕 AI 能够替代艺术创作的讨论热度较高。

图表331：《太空歌剧院》

资料来源：The Information 官网，华泰研究

2023 年 2 月初，基于 Stable Diffusion 通过多个模型融合的 ChilloutMix 在模型分享社区 CivitAI 出现，两周内下载数量超过 5 万次。主要组件包括两大模型，首先是 Basilmix，专攻逼真的纹理和亚洲脸型；随后合并了用照片集数据优化的 Dreamlike 的衍生模型，让最终生成结果更接近 3D 写实风格。在此基础上，针对一些特定微调需求，LoRA 技术同样被应用在图像生成模型中。LoRA 是最初由微软研究团队早在 2021 年提出的一项大语言模型微调的技术，用于提高微调训练速度。2022 年底韩国科学技术院 KAIST 一位研究员 @cloneofsimo 与 HuggingFace 合作，在 Stable Diffusion 上拓展此技术应用，可以让模型固定生成一个物件或人物。

AI 也能够支持从草图到完整图像、多张图片组合生成新图像等功能，在游戏设计初期、宣传图像领域也有望实现应用。谷歌推出的 Chimera Painter 可以将粗略草图生成 3D 怪物图像，垂直类公司包括 VanceAI、Deepdream Generator、Rosebud.ai 等，其中 VanceAI 旗下的 VansPortrait，可在 5 秒内将图片变成绘画、素描或动画，截至 2023 年 1 月全球有 100 万用户使用。

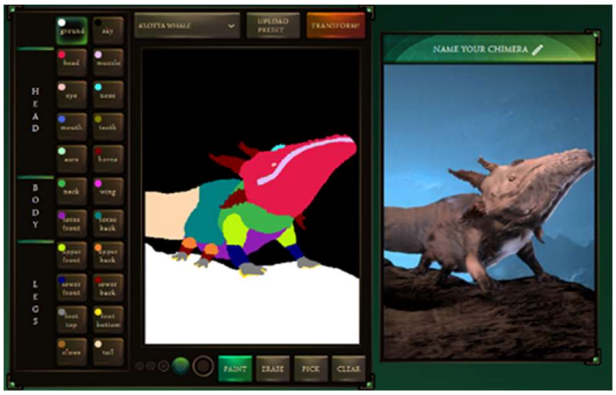
美术绘图环节应用有望最快落地并实现大幅成本及时间优化。现阶段已有工作室通过应用 AI 创作角色，节省成本及时长超 80%。据 Gamelook，俄罗斯游戏工作室 Lost Lore 通过在角色设计阶段应用 AI 相关技术后，将角色创作的时间及开发成本大幅缩短。制作 100 名角色由 5 万美元的开发成本及 6 个月的工时缩短至 1 万美元及 1 个月。该工作室选用 Midjourney，采用 AI 文字输入调整配合人工调整的方法，利用 AI 快速生成角色造型。在加载了已经画好的部分角色作为参考的基础上，使用详细的文字指令（Prompt）来规定新创作角色的特征，包括主体颜色、道具、角色姿势、背景元素等等。比如生成左下图的指令为“一只行军之中的邪恶灰熊，穿着铁制的盔甲和头骨装饰，动态的攻击姿势，古老的头盔和防毒面具，后启示录风格”。随后，美术画师基于 AI 生成结果对画面细节进行进一步监修，大幅节省前期创意时间。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表332：基于已画好角色及详细文字指令生成的新角色形象

资料来源：GameLook 官网，华泰研究

图表333：Chimera Painter 可将草图变成怪物



资料来源：Chimera Painter 官网,华泰研究

AIGC 辅助素材创意生产及精准性提升，不断推动营销内容生产效率。在产品推广阶段，AIGC 通过辅助素材创意生产和内容精准性提升，不断推动营销内容的生产效率，能够极大解放素材制作者的创意工作，为游戏产品营销和出海提供更多可能性。（1）**创意尝试方面**：在同等时间内，通过 AI 赋能能够将创意尝试数量从原本 10 组转变为 100 组及以上，更多的创意尝试能够带来更高转化率，极大解放开发者的创意工作；（2）**素材转换方面**：多模态 AI 模型的发展使得 AIGC 应用更具通用性，如可应用于文本素材生成的 ChatGPT 及文心一言、将文本转化为图片的文心一格、将现实生活中 3D 物品转化为游戏建模的 Scenario、将文本到图像模型由 2D 转化为 3D 网络模型的 Magic3D 等，其发展将简化素材转换过程，增加素材内容多样性；（3）**内容生产及合规审查方面**：《犬与少年》采用 rinna 开发的 AI 辅助背景进行制作，简化了从导演分镜表到 Layout“设计图”的工序环节，极大地提升了内容生产效率。除此之外，行者 AI 不断打磨 2D、3D、音乐、音效的资产生成技术，尝试实现一键生成可编辑任意步骤的营销视频，并探索开发 AI 对内容的合规性审查功能。

海外公司：英伟达，Epic games，Unity

英伟达

英伟达（NVDA US）的元宇宙产品矩阵主要包括：(1) 应用（Audio2Face、CreateXR 等近 10 款应用），（2）云服务，(3)核心技术（MDL、Physx、RTX、USD），(4)硬件（GeForce 显卡、RTX 显卡、NVIDIA 认证系统、NVIDIA studio、NVIDIA OVX 等）。

本届 GTC 上，英伟达对主要产品线做了以下升级。

云服务方面，黄仁勋先后宣布了将在微软 Azure 托管 AI 超级计算服务 DGX Cloud 以及 Omniverse Cloud 的计划。此外，微软和英伟达正在试图把 Teams、OneDrive、SharePoint 等 Microsoft 365 应用与 NVIDIA Omniverse 相结合，以提升平台的协作能力。

硬件方面，英伟达发布了 6 款面向专业消费者的新显卡：NVIDIA RTX 5000、RTX 4000、RTX 3500、RTX 3000、RTX 2000 以及 NVIDIA RTX 4000 SFF。其中前 5 款为笔记本显卡，最后一款为紧凑型桌面级显卡。官网指出，新款笔记本显卡的性能与能效比上一代同类型产品提高了一倍。新显卡配备了英伟达最新的 Max-Q、RTX 等技术，可实现更好的能效和图形效果。这六款显卡均配备了采用 Ada 架构的最新技术，比如 CUDA 核、第三代 RT Core、第四代 RT Core、DLSS 3、更大的 GPU 显存以及更强的扩展现实能力。

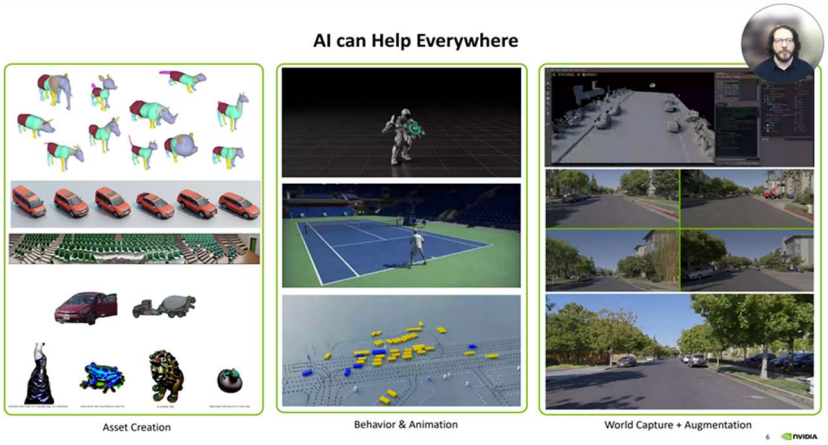
免责
更多免费行业报告 并购家 www.ipoipo.cn

应用方面，英伟达正在把众多 AI 能力集成到 Omniverse ACE（虚拟人），如语音 AI 应用 Riva、用于自然语言理解的 NeMo 服务、AI 驱动动画工具 Audio2Face 以及 Live Portrait 等。此外，Omniverse ACE 的最新功能改进还可以与 Tokkio 或者 NVIDIA Maxine SDK 工作流相结合，为网络会议以及客户服务部署更为高级的虚拟形象。

图表334：Omniverse 中集成了众多 AI 能力

资料来源：NVIDIA 官网，华泰研究

图表335：AI 对各领域的效率提升



资料来源：NVIDIA 官网，华泰研究

Epic Games：新品 Unreal Engine 5.2 支持优质图像与虚拟人生成

在 2023 年 3 月 23 日 Epic Games State of Unreal 2023 的主题演讲上，Epic Games（EPOR US）展示了 Unreal Engine 5.2 的新增强功能、全新 MetaHuman 技术、一个统一的 3D 资产市场以及 Fortnite Unreal Editor 工具。

优质图像生成：Unreal Engine 5.2 推出的 Substrate 是一种新的材质框架，可以让使用 Unreal Engine 5 的用户更好地控制实时应用程序中使用的对象的外观和感觉，而 Procedural Content Generation Framework 允许艺术家更快、更高效地定义规则和参数以快速填充广阔、高度详细的空间。将预制峭壁结构放入虚拟环境中，随后整个事物便会通过 AI 自动重新调整自身，为上部区域添加木桥，并让下面看起来更加自然，移动预制件还能使虚幻引擎用逼真的水坑、小石头和浅色树叶填充间隙。这种技术与生成式 AI 相结合，可以真正加速游戏开发。

免责
更多免费行业报告 并购家 www.ipoipo.cn

虚拟人技术：Metahuman Animator 可以通过将人物形象录制在 iPhone 上短时间内生成高度逼真的人体模型或元宇宙虚拟人。用户通过手机录制面部视频，上传至 Live Link Face 应用程序捕获面部动态，MetaHuman Animator 就可以使用视频和 Dev 数据将其转换为高保真度的动画。在完成面部捕捉之后，仅需要 3 帧视频就可以完成 3D 面部建模，且能在短暂的几分钟内完成全流程。拥有虚拟人之后，MetaHuman Animator 可以让用户能够使用 iPhone 或立体头盔相机 (HMC) 在 MetaHuman 角色上以高保真动画的形式重现任何面部表演。

除 AIGC 生成场景与虚拟人外，Epic Games 结合 UGC，给提供用户原创内容机会。Epic Games 发布了 Fortnite 虚幻编辑器 (UEFN)，允许玩家在 Fortnite Creative 模式下免费创建自己的游戏和体验。UEFN 是适用于 PC 版 Fortnite 的新附加组件，与 Fortnite 现有的创意工具集共同工作，并将用于直接在 Fortnite 中设计、开发和发布游戏和体验。

此外，在今年晚些时候，Epic Games 还计划将 Unreal Engine Marketplace、Quixel Bridge、ArtStation Marketplace 和 Sketchfab 将全部合并为一个市场 Fab。Fab 是 Epic Games 的一个新市场，让所有数字内容创作者能够发现、分享、购买和出售数字资产。理想情况下，MetaHuman Animator 能够帮助提高面部动画水平，并简化从一个模型到另一个模型进行相似程度的面部动画转换所需的工作量。该款动画工具将于今年夏天推出。

图表336： Metahuman 生成虚拟人

资料来源：Unreal Engine 官网，华泰研究

Unity：布局生成式创造能力

在 2023 年 2 月 25 至 26 日的 2023 全球人工智能开发者先锋大会(GAIDC)上，Unity（US）展示了 Unity+AI 在汽车、数字人、元宇宙等领域已呈现的成熟解决方案。此前，Unity 已经通过 AI 技术大幅度简化了 3D 资产的生产流程，比如由 AI 赋能的 Ziva 工具集可以让虚拟人的脸部创作周期从原来的数周降低为单次按钮点击，美术辅助工具 World Generation 能够帮助用户快速生成地形，并实时模拟出真实的光影效果，RestAR 借助计算机视觉和深度学习实现在移动设备中对产品进行高质量的 3D 扫描和渲染。

在 GDC 2023（游戏开发者大会）期间，Unity 宣布了他们即将开发一个 AI 工具并加入到引擎中。在预告视频中，Unity 正在研发文生产品，提供输入指令得到角色、NPC、形象等。目前该项目还在测试中。

Unity 探索游戏引擎+ChatGPT 制作方式。在 UnityAI 的功能下，开发者可以输入生成/移动箱子、随机增加光源、随机为光源配色等文字指令，在既定空间内由 AI 生成相应的 3D 场景。目前它暂时不能被认为是一个 100%实用的解决方案，无法完成所有的工作或一些复杂的任务。

更多免费行业报告 并购家 www.ipoipo.cn

图表337：引擎编辑器生成场景

资料来源：Unity 官网，华泰研究

Unity 中国技术副总裁左瑞文在全球人工智能开发者先锋大会（GAIDC）期间（2023 年 2 月 25 日）接受澎拜科技采访时提出，“在 AIGC 加持下 Unity 的工具能更高效地产生更高质量的场景。如 Speetree 支持 AI 赋能的程序化建模，利用 ArtEngien 美术内容创作工具，可以自动生成 PBR 材质，一键实现风格转换。”他表示 Unity 认为实现生成式创作的答案就是 U3Dcopilot，其打通了 Unity 与用户资产库、数据索引工具以及插件全家桶，就提供给开发者对应资产相关的生成式创造能力，是接下来重点布局的一套系统。

Unity 未来将推出一个标准化 AI 工具交易平台，帮助用户更轻松地获取 AI 工具，也为 AI 工具的创作者提供新的创收渠道。目前 Unity 中国已经建立了本土 AI 研发团队，拥有了 Ziva Dynamics、Weta Digital、Speedtree 等由 AI 赋能的创作工具，以期帮助开发者加速创作。

整机品牌：Nreal、亮亮视野、Rokid、亮风台、大朋 VR

Nreal（未上市）：推出全球首款消费级 AR 眼镜

推出全球首款消费级 AR 眼镜。根据公司官网，Nreal 成立于 2017 年，在 2019 年初推出全球首款消费级的 AR 眼镜 Nreal Light。同时，Nreal 全球首创“AR 眼镜+手机”的交互方式在国际上也获得一致好评。公司产品应用于医药、教育、文娱旅行、工业等领域。迄今为止，Nreal 拥有近 400 名员工，已与 10 多家全球知名运营商建立合作伙伴关系，并已将其产品带到日本、韩国、美国等 6 个国家。根据 IDC 的 2022 年 AR/VR 报告显示，Nreal 以 9.8 万台的成绩成为 2022 年 AR 行业出货量最大的品牌，占全球 AR 设备出货量的 37%，在消费级 AR 设备的市占率高达 57%。

公司主要有四款产品（两款眼镜+一个软件操作平台+一个开发者 SDK）。1）Nreal Light，售价 599 美元，拥有基于视觉惯性 SLAM 的自由度空间定位能力，以及优秀的精度（毫米级）、延迟（毫秒级）、稳定性。2）针对中低端市场发布的 Nreal Air，在 Light 的基础上降低部分摄像头规格，售价仅为约 330 美元。3）Nebula：专为 Nreal AR 眼镜打造的 3D 交互系统，可以安装并运行在兼容的手机上。4）NRSDK：Nreal 面向全球开发者发布了 NRSDK 1.0 Beta 版。5）XREAL Beam：2023 年 5 月 Nreal 改名为 XREAL 后推出，功能亮点包括支持无线连接，可续航四小时，可启用 3DoF 支持，引入空间显示等。

免责
更多免费行业报告 并购家 www.ipoipo.cn

亮亮视野（未上市）：AR+AI 行业解决方案提供商

亮亮视野自有产品+解决方案全栈布局，业务多样化、全球化。公司以自有品牌 AR 硬件产品为基础，深度整合 5G、IOT、大数据，拥有端到端全栈 AR+AI 产品及解决方案布局。公司产品及解决方案广泛服务于工业、能源、航空、安防、制造、交通汽车等行业领军企业，业务涉及 66 个国家和地区，开拓了 20 余个生态伙伴，服务于全球超 5000 家企业用户。

公司“听语者”AR 眼镜是全球首款 80g 以内的无线一体双目全彩光波导 AR 智能沟通眼镜，旨在将声音转写成字幕实时显示在眼前，帮助听力受损人群、听力下降的老人群体和跨国交流人群实现更高效、更平等的交流方式。它由超高透光度的阵列光波导显示镜片、高性能麦克风阵列、降噪模块、电池及无线通讯模块等组成，具有高透明度、外型时尚、无线佩戴等特点。

图表338： Nreal Light 产品展示图

资料来源：Nreal 官网，华泰研究

图表339： 亮亮视野 Leion Pro 实物图

资料来源：亮亮视野官网，华泰研究

Rokid（未上市）：AR 2B 提供行业解决方案，2C 国产产品提供商

Rokid 将 AR 眼镜与操作系统相结合提供行业解决方案。公司一方面生产 AR 眼镜等软硬件产品，另一方面构建 YodaOS 操作系统为载体的生态，通过语音识别、自然语言处理、计算机视觉、光学显示、芯片平台、硬件设计等前沿技术，将 AR 技术与 AI 相结合，为不同领域的客户提供全栈式解决方案，其 AI、AR 产品已在全球七十余个国家和地区投入使用。

Rokid 拥有多样化产品矩阵，重点发力 C 端市场。公司主要核心产品有 5G 工业防爆 AR 头带 Rokid X-Craft、单目增强现实眼镜 Rokid Glass 2 以及 C 端 AR 眼镜 Rokid Air 系列。。

2023 年 3 月 21 日，Rokid 推出的第二代的 AR 眼镜计算终端 Station Pro，是 Rokid 全新的产品种类——AR 主机，升级到骁龙 XR2+的芯片带来 30%的散热性能提升和 50%的持续性能提升，搭载 Rokid 自研的全空间化 AR 操作系统使该 AR 眼镜基于单摄像头就能实现 SLAM、3D 裸手交互、第一视角分享、视觉定位 VPS 能力，支持眼手交互、空间创作等 AR 体验，给更轻量化、普惠的 AR 眼镜带来可能。

亮风台（未上市）：掌握 AI 核心能力，国内领先的 AR+AI 行业解决方案提供商

亮风台 AI 能力突出，产品涵盖硬件终端与 AR 云系列产品。公司拥有计算机视觉、深度学习、智能交互等人工智能核心技术，自主研发 AR 终端和 AR 云，已建成业内领先的以 AR 云为数字中枢、端云结合的 AR 平台 HiAR Space。HiAR Space 基于亮风台空间计算、AR 云、自然交互和 AR 智能硬件四个方面的技术沉淀与研发能力，支持平台+垂直优化，可以将场景、空间、数据和用户连接起来。

免责

更多免费行业报告 并购家 www.ipoipo.cn

亮风台解决方案以 B+G 端为主，覆盖超十个行业。公司已有业务涉及能源、汽车、钢铁、航空航天等超十个行业，与海尔、腾讯、阿里等大型企业和公安局、医院、学校等国家机构达成良好合作关系。同时，联合 5G、IoT、大数据等新兴技术与产业合作伙伴，使 AR 更深入广泛地赋能行业、服务生活。亮风台投身文旅服务，与大理州旅游集团合作，与携程联合打造元宇宙蝴蝶泉，为用户提供端到端的元宇宙深度旅游体验。亮风台以元宇宙文旅平台“绿洲”为载体，通过 AR/VR、图像测量和三维重建等相关技术，构建元宇宙创建工具链和体验入口，形成三方生态的互补与互通。

图表340： Rokid AR 眼镜实物图

资料来源：Rokid 官网，华泰研究

图表341： 亮风台产品和解决方案部分使用场景

资料来源：亮风台官网，华泰研究

大朋 VR（未上市）：国际领先的软硬件一体化的全栈 VR 技术与产品公司

大朋 VR 是国内认可度较高的 VR 整机生产商，产品覆盖 B/C/G 三大市场。公司拥有 3 大产品体系，其中 P 系列主打行业解决方案，旗舰产品 Ultra 4K VR 一体机，采用模组化设计，搭载高通骁龙 845 和真 4K 显示屏，拥有外接脑机接口、手势识别、眼球追踪、虹膜识别等技术；E 系列主打娱乐，公司旗下多款明星产品除了支持 6000+ SteamVR 游戏，在屏幕分辨率、佩戴舒适度方面也拥有很强的产品竞争力；党建系列为智慧党建提供创新内容形式，让党建学习更深入人心。

2022 年 11 月，公司推出旗舰产品大朋 E4，是全球第一款实现 4K 显示+120Hz 分辨率的 VR 设备，并且支持 6dof 的空间定位，提供更沉浸式的使用体验；此外，大朋 E4 的重量只有 285 克，支持更长久的使用。

光模组领域：鲲游、耐德佳、至格、JBD、视涯、宏禧科技、赛富乐斯、熙泰、研鼎、蓝特光学

鲲游光电（未上市）：专注于晶圆级光学、光集成领域的高科技企业

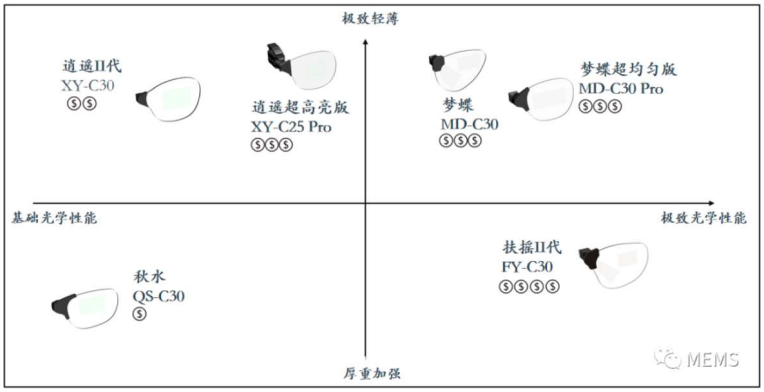
鲲游光电是专注于晶圆级光学、光集成领域的高科技企业，总部位于中国上海。公司由国际国内光电子领域顶级学者、多年光电产业经验的行业资深人员组成。依托自身深厚的技术积累、高精度的制造工艺，鲲游光电通过融合集成光学与集成电路的创新思路，致力于探索前沿光子、光电子领域的实现与创新应用，助力中国在下一代光子科技领域的弯道超车，拉开消费光子的序幕。

晶圆级光学是通过光学的波动理论和创新工艺，超越摩尔定律的下一代芯片技术，属于典型的光电子技术制高点与卡脖子环节。经过 5 年的努力，鲲游光电是目前唯一一家实现从研发设计、母版制作，到生产、检测出货全程闭环的 IDM 高科技企业。目前公司产品主要包括：1) 3D 成像光芯片，主要与华为合作；2) AR 显示衍射光波导。AR 显示光波导是元宇宙 AR 显示的核心元件；3）激光雷达 LiDAR 耦入光纤模组。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表342：鲲游光电主要产品



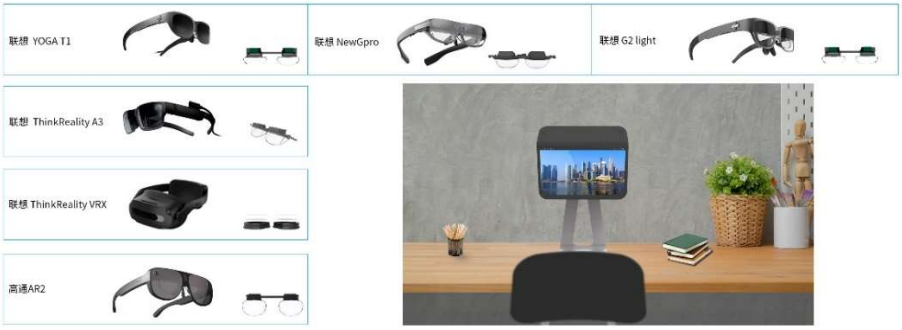
资料来源：鲲游光电官网，华泰研究

耐德佳（未上市）：以自由曲面技术底层平台技术的显示光学领军供应商

耐德佳研发团队实力强劲，在自由曲面显示光学设计、精密制造和检测方面有超十年研究经历和数十项专利及国际一流研究论文，根据公司官网，产品性能指标已达到或超过美国、日本等高科技公司相关产品，跻身国际一流水平。公司围绕显示主要产品包括自由曲面钻石系列 AR、Pancake 水晶系列 VR 近眼显示用光学模组以及桌面型视远光学护眼显示器等。

耐德佳先后于江苏镇江、江西上饶建设生产基地。先期建设的镇江基地占地近 4,000 平方米，作为近眼显示光学模组制造调校平台，拥有多个万级及千级无尘光学车间，引进了超精密单点金刚石车床、精密加工设备、高精度注塑机等生产设备，以及光学轮廓仪等超精密检测设备。

图表343：耐德佳主要产品



资料来源：耐得佳官网，华泰研究

至格科技（未上市）：专注于 AR 衍射光波导和衍射光栅领域的高科技企业

公司采用 IDM 模式，自主掌握“光栅设计、光栅母版加工、纳米压印生产”三大核心技术，拥有功能完备的光栅母版加工中心和衍射光波导生产线，致力于 AR 衍射光波导光学显示模组及衍射光栅的研发、生产和销售。

公司自主掌握光栅产业链三大核心技术。1）光栅设计，拥有具有自主知识产权的针对不同类型光栅的全套计算软件，2）光栅母版加工，掌握基于全息光刻和离子束刻蚀的完整的光栅母版加工工艺，同时拥有工艺终点在线检测、杂散光控制等先进工艺技术，3）纳米压印生产，已经打通全部生产工艺环节，图形转写率可达 95%。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表344： 至格 AR 衍射光波导产品

资料来源：至格官网，华泰研究

JBD（未上市）：Micro-LED 微显示面板领域的一流厂商

上海显耀显示科技有限公司（Jade Bird Display 以下简称 JBD）成立于 2015 年，总部位于上海浦东，在合肥设立子公司，JBD 一直在 Micro LED 微显示面板领域深耕细作，拥有自主的 IC 设计、MOCVD 材料生长、Micro LED 微显示技术加工制造、封装测试、软件硬件驱动设计等技术，其自主研发及制造世界上超微型显示面板，强力推动混合现实领域、汽车、半导体设备等领域的技术发展。基于掌握显示屏的核心技术，JBD 实现“芯屏契合”，具体产品主要包括单色/三色显示板，AM-μLED 微显示开发套件，AM-μLED 显示屏，AM-μLED 光引擎、光模组等。

图表345： JBD Micro LED 微显示器

资料来源：JBD 官网，华泰研究

视涯（未上市）：OLEDoS 微型显示器头部供应商

视涯是专注于 OLEDoS 显示器研发、生产的半导体公司，核心团队实力强劲。公司从事新一代半导体 OLEDoS 显示器研发、设计、生产和销售，致力于打造 OLEDoS 显示应用生态链，为客户提供端到端的微显示相关解决方案。一期工厂面积超过 40,000 平方米，是全球为数不多的专注于 12 英寸晶圆上生产 OLEDoS 显示器的产业基地之一。

视涯 OLEDoS 技术领先，适配范围广，已有成熟产品登陆市场。公司 OLEDoS 技术相对市场其它微显示技术拥有轻薄小巧、结构精密、亮度高、功耗低、对比度高的特点。当前公司 OLEDoS 技术可以适配 Pancake、Lens group、Birdbath、Waveguide 等主流光学方案，目前已经推出 0.32-1.03 英寸系列显示器，其中 0.49 英寸 OLEDoS 显示器被高通用于 XR2 无线 AR 智能眼镜的参考设计。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表346：视涯 0.49 英寸 OLEDoS 微显示器

资料来源：视涯官网，华泰研究

宏禧科技（未上市）：硅基 Micro OLED 微型显示器头部供应商

专业从事硅基 Micro OLED 微型显示器件和模组研发、设计以及制造的高科技企业，产品主要应用于 AR / VR 产品领域。公司已建成全球领先的硅基 Micro OLED 生产线，拥有从 IC 设计到微型显示器封装工艺的全环节技术和管理团队，技术团队成员有着丰富的硅基 Micro OLED 微型显示器的研发及量产经验。

宏禧科技发布 0.6 英寸硅基 Micro OLED 微型显示器。宏禧科技致力于“视觉科技感触未来”企业文化，现已设计完成多种型号的硅基微型显示器产品。2023 年 01 月 19 日发布“0.6 英寸硅基 Micro OLED 微型显示器”，即将量产 0.7 寸 2k 全高清和 1.33 寸 4k 产品。

图表347：宏禧科技 0.6 英寸硅基 Micro OLED 微型显示器

资料来源：宏禧科技官网，华泰研究

赛富乐斯（未上市）：micro-LED 显示解决方案一流提供商

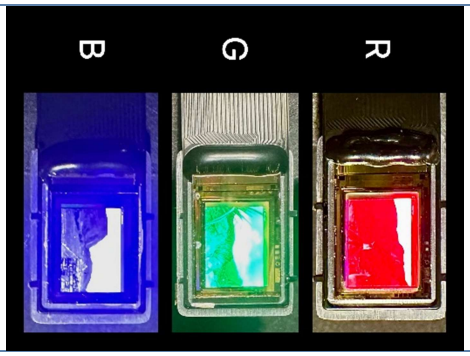
赛富乐斯是业界领先的原位量子点（QD in chip）micro-LED 显示解决方案提供商。公司脱胎于耶鲁大学，经过近 10 年的积淀，已成为 micro-LED 领域少有的拥有底层创新技术，并且成功打通商业化的企业，拥有半极性氮化镓及纳米孔量子点（NPQD）等原创技术，有望成为下一代通用级 ToC 硬件的显示技术引领者及核心模组供应商。

公司为显示领域客户设计并制造 micro-LED 芯片和微显示模组。目前产品包括应用于公共信息显示屏的 R1 系列芯片，以及应用于 AR 眼镜的 T1 系列微显示模组。目前，搭载 R1 芯片的显示屏产品已进入导产阶段。公司于 2021 年 10 月及 2022 年 8 月，分别发布 2 微米全彩 NPQD® micro-LED 及红光 NPQD® micro-LED 光引擎，首次将全彩 AR Micro-LED 的性能提升到了应用标准。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表348：赛富乐斯 T1 微显示模组



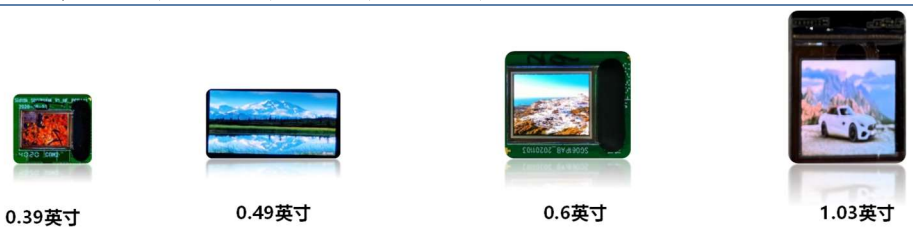
资料来源：赛富乐斯官网，华泰研究

熙泰（未上市）：硅基 OLED 领先制造商

熙泰科技专注于新型微显示技术领域，现有一条 8inch 量产线，稳定量产出货。根据 WitDisplay，总投资 60 亿元的 12inch 量产线正在建设中，引进全球顶尖设备，预计 2023 年底点亮，2024 年量产出货。

熙泰技术创始团队来自中科院长春光机所，现有专业研发人员 160 余人，硕博占比 40%以上。公司坚持自主创新，已申请专利 300 余项，其中发明专利 190 余项，已授权专利 130 余项。自有知识产权发光技术，不同于传统硅基“白光+CF”技术路线，发光效率提高 30%，DCI-P3 色域达到 110%以上。

图表349：熙泰 0.39 英寸、0.49 英寸、0.6 英寸、1.03 英寸全彩显示器



资料来源：熙泰官网，华泰研究

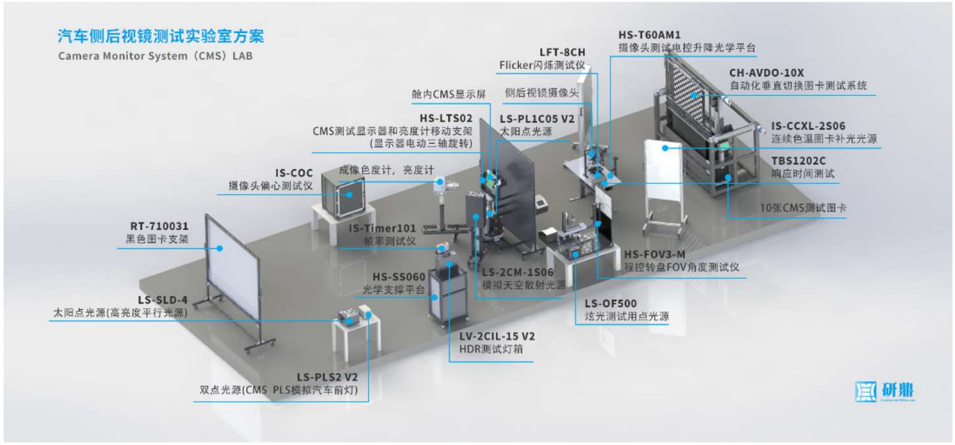
研鼎（未上市）：影像测试行业的开拓者

公司创始人及核心团队由来自海内外知名公司高级技术管理人才及资深行业技术人员组成，拥有平均 20 年以上的一线的研发和运营管理的工作经历，同时公司还拥有一支技术精湛、追逐创新的研发团队，拥有本行业国际一流研发和运营经验。

公司的主营业务为高精密仪器与量产高端装备和图像分析算法和工业软件的研发、设计和销售。通过多年的技术积累，公司在高精密光学仪器、图像质量分析与标定算法、运控软件、光学镜头、图像采集板卡、图像标板量产检测设备等相关领域具备全栈自研开发和交付能力。针对图像领域不断涌现的多摄、红外，深度相机等新兴技术的测试需求，研鼎积极的应对和响应，提供最全面的测试解决方案及相关服务，包括手机方案、安防方案和汽车方案等。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表350：研鼎图像解决方案矩阵



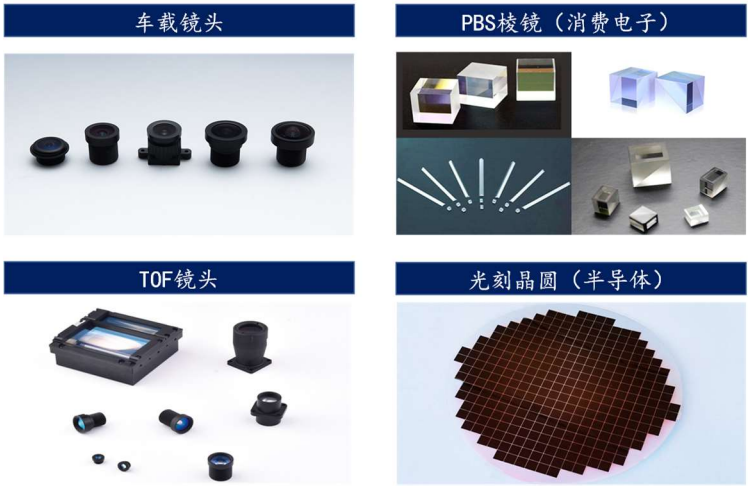
资料来源：研鼎官网，华泰研究

蓝特光学：精密玻璃光学元件领军企业

蓝特光学在精密玻璃光学元件加工方面具备突出的竞争优势和自主创新能力。在玻璃光学元件冷加工、玻璃非球面透镜热模压、高精密模具设计制造、中大尺寸超薄玻璃晶圆精密加工等领域具有多项自主研发的核心技术成果。凭借全面的研发体系、良好的技术实力，公司先后承担了 2004 年国家火炬计划“别汉屋脊棱镜”项目、2013 年国家火炬计划“汽车用后视侧视宽视野异形反光镜组”项目、2016 年国家工信部工业强基项目“年产 1125 万件精密模压非球面玻璃光学元件实施方案”等重要项目。

公司主要生产棱镜、透镜、玻璃晶圆、玻璃非球面透镜，镀膜以及镜头组装。涉足传统光学、光电显示、光伏、汽车、照明、光通讯、半导体、消费类电子、医疗、虚拟/增强现实等领域光学产品的加工和制造，在望远镜、显微镜、数码相机、手机、AR/VR、安防监控器、多媒体投影仪、瞄准仪以及影像测量设备、汽车内外饰、车载镜头等高技术产品的核心部件中均可找到蓝特光学的产品。

图表351：蓝特光学产品应用矩阵



资料来源：蓝特光学官网，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

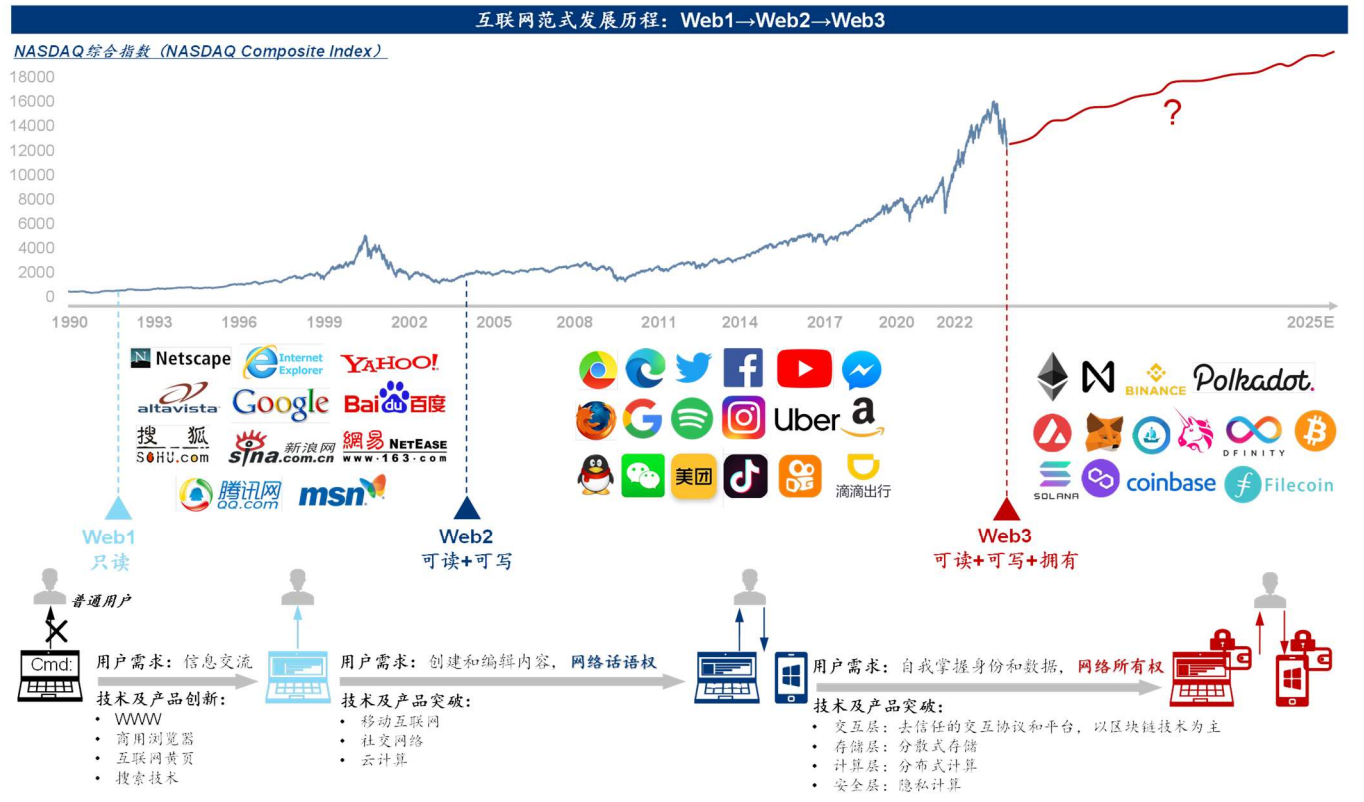
Web3：过去一年发生了什么？

Web3：以所有权为核心的价值网络

伴随着底层技术的发展突破以及用户需求的演化升级，互联网的范式也在发生着演进。为了解决原始 web 网络无法支持普通人进行信息交流和分享的难题，一个“可读”的 Web1 时代伴随着万维网的发明而到来，各类搜索引擎和门户网站将信息大量搬上网，普通用户也能轻松访问各类资源。为了解决 Web1 下用户无法主动创建内容、不享有网络发言权的难题，一个“可读+可写”的 Web2 时代伴随着智能手机的普及、社交网络的兴起和云计算的发展而到来，所有用户都可以在社交媒体分享自己的所思所想并和他人交互，互联网世界的话语权从商业巨头转移到每一位终端用户手中。而现在，在经历了 Web1 和 Web2 时代后，互联网正在进行一场 Web3 革命。

Web3 是用户拥有、用户控制的互联网。以太坊联合创始人加文·伍德（Gavin Wood）最初提出了 Web3 的概念，并将其描述为“一组保障人们在低壁垒市场中为自己行动”的兼容协议，认为其是“可执行的大宪章——个人反对暴力权威的自由的基础”；而知名区块链研究机构 Messari 的研究员江下（Eshita）及区块链技术公司 BitFury CEO 布里安·布鲁克斯（Brian Brooks），则分别将其定义为“可读+可写+拥有的互联网”、“用户可以拥有产权的互联网”。

图表352：互联网范式发展历程总览



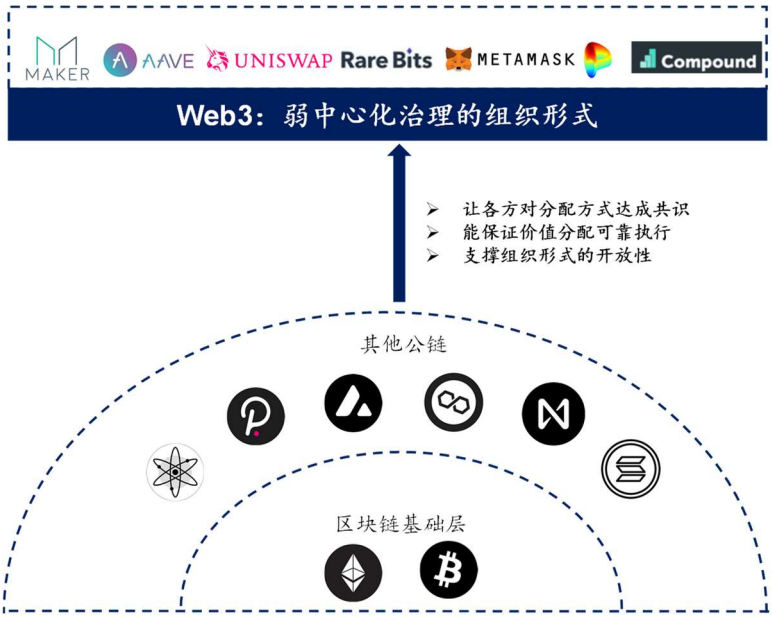
资料来源：，华泰研究

Web3 以区块链为核心，解决价值分配问题和信任问题

区块链通过解决底层价值分配问题，贯彻 Web3 去中心化的核心理念。Web3 基于一套开放的协议簇，打破 Web2 平台垄断，解决平台价值分配问题。而区块链作为底层价值分配系统，通过建立共识协议、开放源代码，保证价值分配的可靠执行。区块链作为安全性和去中心化水平都较高的网络，通过共享账本储存数据、交换价值并记录交易活动，不受任何中心化实体控制，为 Web3 提供安全的执行层，让使用者可以在其中创建、发行并交易加密资产，并且开发可编程的智能合约。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表353： 区块链解决价值分配问题，支撑 Web3 弱中心化



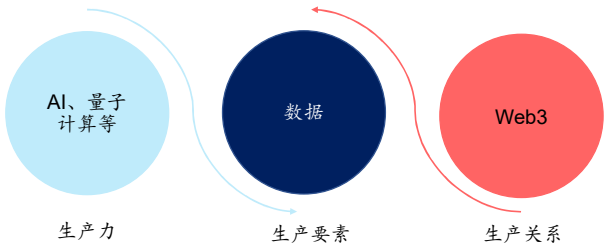
资料来源： Web3 Explorer，华泰研究

社会元素、数学和物理规则和激励结构是构成人类网络信任模型的三个要件。 Web1 时代是物理信任模型，由于开源的底层协议不受任何单一集体的控制，用户、开发者和企业都可以相信游戏规则将保持中立和公平。Web2 时代则是机构信任模型，由于协议是封闭的，商业公司接力数学和物理规则成为新的信任守门人，用户被迫信任其编写的专有和不透明的代码。而 **Web3 的信任模型则是由数学和物理规则以及激励结构共同主导的可编程信任。**开源协议以及区块链系统通过自动执行的智能合约、加密技术、共识机制、分布式技术等多种技术提供了可编程的信任；而代币发行和奖励则提供了基于激励结构的信任，用户无需再依赖信任中介。

Web3 或将成为未来数字经济基础设施，优化信任和分配机制

随着 AI 大模型的普及，我们看到办公、文娱创作、科学研究等领域效率的大幅度提升。展望未来，随着 AI、量子计算等新兴技术的发展驱动人类社会生产力提升，我们认为普通人进行专业创作的门槛将降低，而个人在网络中生产的数据将成为数字经济时代最重要的生产要素之一。如何进行个人数字资产的确权、流转以及变现和收益分配将成为数字经济时代的面临的重大问题之一。而 Web3 以区块链作为底层价值分配系统，通过共享账本储存数据、交换价值并记录交易活动，优化数字时代信任和分配机制，或将成为未来数字经济时代重要基础设施。

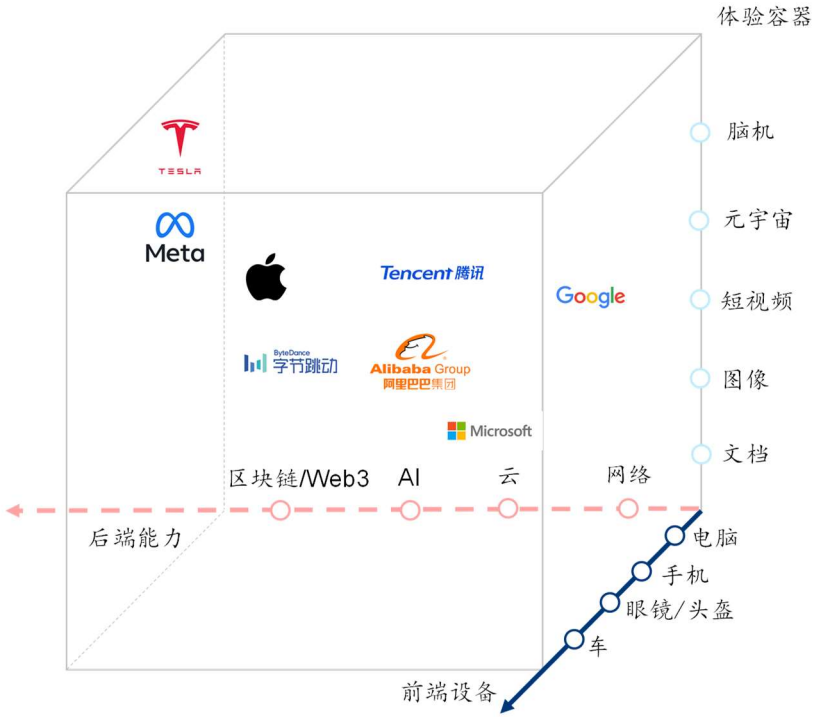
图表354： 数字经济时代生产力、生产要素和生产关系的边际变化



资料来源：华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

图表355： Web3 或将成为未来数字经济基础设施，优化信任和分配机制



资料来源：Web3 嘉年华，阿里巴巴，华泰研究

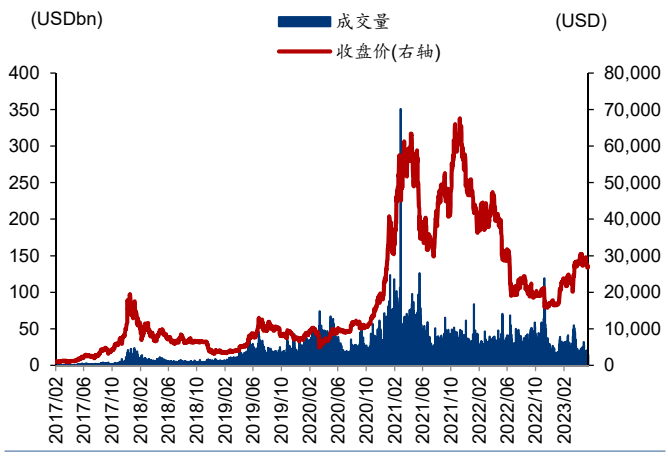
风险事件频发，加密合规时代已至

回顾 Web3 过去一年的发展，随着美联储进入加息周期，全球流动性收紧，加密资产市值从高位跌落。加密寒冬中市场风险事件频发，UST 断崖式下跌、三箭资本崩盘、加密货币交易所 FTX 破产、Silvergate 暴雷等多起风险事件的发生引发市场震荡。加速合规框架建立成为美国、欧洲、新加坡、香港等各国家/地区监管强共识。2022 年 9 月，白宫发布了加密监管框架草案，10 月，欧盟出台《加密资产市场监管法案》，形成欧盟内部统一的监管和执法体系。2023 年 1 月，香港金管局发布《加密资产和稳定币的讨论文件》，提出了对加密资产特别是支付稳定币的监管模式构思，全球加密资产监管生态开始逐步完善。

过去一年见证了 Web3 风险事件的密集发生

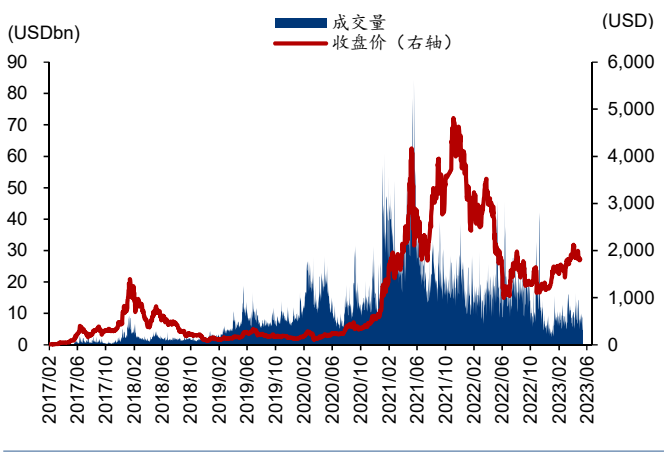
随着美联储推进加息，加密资产进入熊市。回顾过去一年，随着美联储进入加息周期，全球流动性收紧，加密资产市值从高位跌落。比特币价格从 2022 年 4 月超过 4,6000 美元的高点跌落至 2022 年 11 月的约 1,5000 美元。年初以来，加密资产价格开始小幅反弹，比特币价格已恢复到 2,8000 美元附近。

图表356： 比特币价格及日均交易量（截至 2023 年 5 月 15 日）



资料来源：CoinMarketcap，华泰研究

图表357： 以太币价格及日均交易量（截至 2023 年 5 月 15 日）

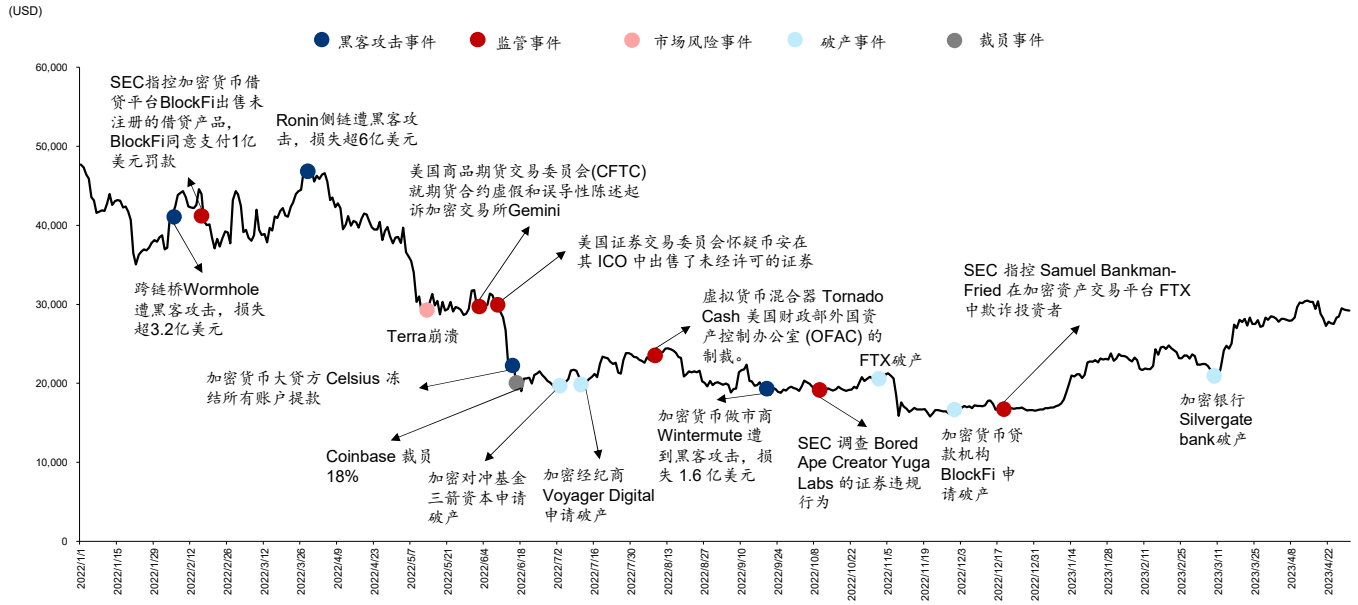


资料来源：CoinMarketcap，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

在过去一年，Web3 风险事件密集发生。过去一年的熊市中，Web3 行业发生了诸多风险事件。UST 断崖式下跌、三箭资本崩盘、加密货币交易所 FTX 破产、Silvergate 暴雷等多起风险事件的发生引发市场震荡。2022 年 3 月，Ronin 侧链遭黑客攻击，损失超 6 亿美元。2022 年 5 月，DeFi 借贷领域的领导者 Celsius 崩盘。2022 年 8 月，混币器 Tornado Cash 被列入美国海外资产管理办公室（OFAC）美国特别制定国民名单（SDN）列表，开发者 Alexey Pertsev 在荷兰被捕入狱。2022 年 11 月，头部加密货币交易所 FTX 破产，引发市场震荡。2023 年 3 月，加密银行 Silvergate Bank 破产。密集的风险事件暴露了行业的风控漏洞，也促使各地监管积极完善监管框架。

图表358： Web3 风险事件梳理



资料来源：华泰研究预测

典型风险事件复盘#1：UST-Luna 崩盘事件

UST 币是 Terra 公链上发行的算法稳定币，通过“LUNA-UST 双币套利机制”维持币值稳定。UST 发行方承诺，恒定 1 美元 LUNA = 1 个 UST，当 UST 币的价格大于 1 美元，用户可以销毁价值 1 美元的 LUNA 币，来发行 1 枚 UST 币，反之亦然。**Anchor 的大规模提款和 UST 的大量抛售打破了表面稳定，将 UST 和 Luna 拖入下行螺旋。**Anchor 是 Terra 推出的承诺约 20%年化收益率的去中心化金融协议，当 Anchor 无法维持高利率而下调收益率后，市场对 UST 信心下降。Anchor 存款的大额提取和 UST 被大量抛售向市场释放了非常消极的信号，投资者对于 UST 的信心迅速崩塌，导致 UST 和 Luna 进入下行螺旋。2022 年 5 月 8 日开始，Luna 币断崖式下跌，高峰时期 400 亿市值遭遇史诗级归零。

典型风险事件复盘#2：Celsius 崩盘复盘

挤兑潮加速流动性枯竭，Celsius 遭遇崩盘。Celsius 是 DeFi 借贷领域的领导者，根据 Dune analytics，2021 年 1 月 Celsius 活跃用户数高达 400 万。2022 年 5 月初，UST 大幅下跌大批资金撤离 Celsius，储户挤兑加剧。然而，Celsius 的流动性管理无法满足兑付需求。根据 Coin News， Celsius 持有的 73%的 ETH 锁定在 stETH 或 ETH2 中，只有 27%的 ET 具备流动性。于是在 2022 年 6 月，Celsius 宣布暂停提款， CEL 代币一小时内大幅下跌 70%。

典型风险事件复盘#3：FTX 破产事件

加密货币交易所 FTX 破产，引发市场震荡。CoinDesk 和 Binance 赵长鹏对于 FTX 业务稳定性的质疑引发了为期三天的约 60 亿美元的挤兑，FTX 陷入危机。币安同意收购 FTX，在尽职调查后又中止交易，引发市场剧烈波动。11 月 11 日，FTX 进入破产重组程序，FTX 创始人 Sam Bankman-Fried（SBF）已在美国法院被投资者起诉。

免责
更多免费行业报告 并购家 www.ipoipo.cn

监管政策逐步完善，加密合规时代已至

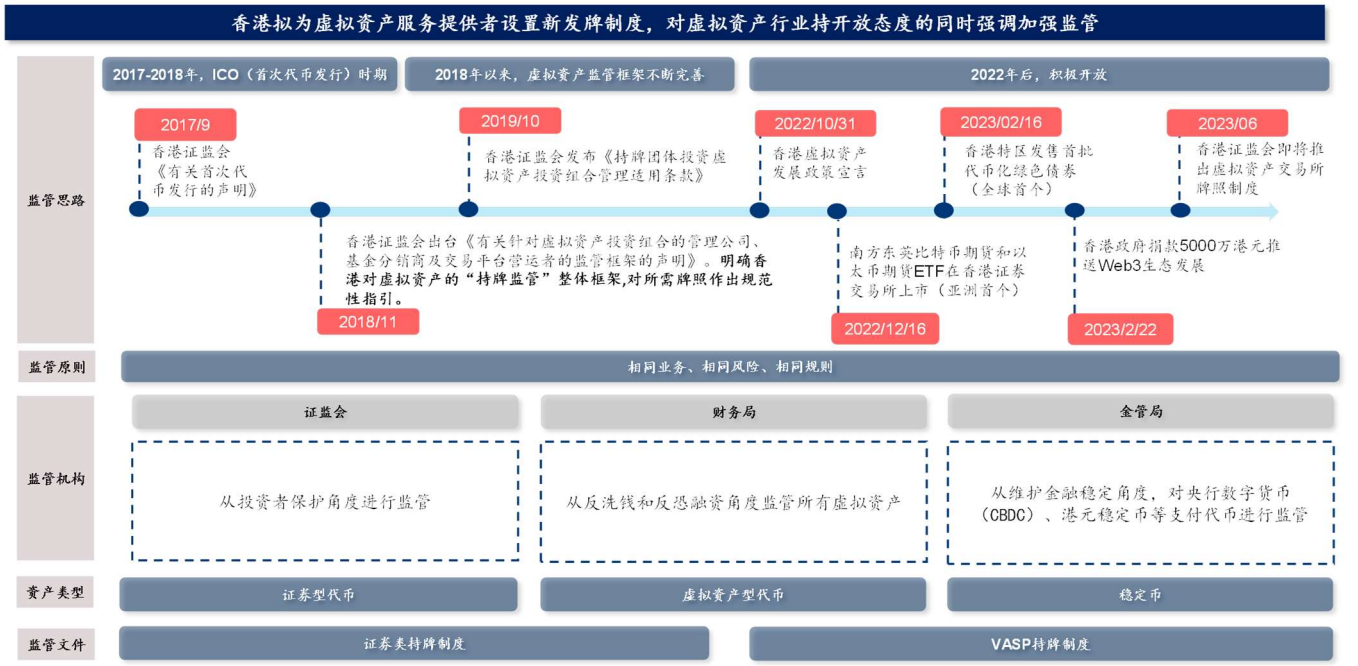
回顾 Web3 过去一年的发展，加密寒冬中市场风险事件频发，加速合规框架建立成为美国、欧洲、新加坡、香港等各国家/地区监管强共识。9 月，白宫发布了加密监管框架草案，10 月，欧盟出台《加密资产市场监管法案》，形成欧盟内部统一的监管和执法体系。2023 年 1 月，香港金管局发布《加密资产和稳定币的讨论文件》，提出了对加密资产特别是支付稳定币的监管模式构思，我们看到全球监管政策逐步完善，加密合规时代已至。

#1 香港虚拟资产监管框架：强调持牌制度，由证监会、财库局、金管局共同监管

2022 年，香港正式表示将支持 Web3 产业发展。2022 年 10 月，香港财经事务及库务局发布《有关香港虚拟资产发展的政策宣言》，明确欢迎和鼓励态度，香港证监会表示会支持元宇宙、NFT、GameFi 创新，积极探索适当的代币化资产监管框架。4 月 11 日，香港 Web3.0 协会在香港交易所举行成立仪式，表示将会推动 Web3 产业在港发展。

香港强调“适当监管”和“推动发展”并重，由证监会、财库局、金管局共同进行虚拟资产监管。香港对虚拟资产的政策经历了三个阶段：1）2017-2018 年，香港证监会出台《有关首次代币发行的声明》；2）2018 年-2022 年。以香港证监会《有关针对虚拟资产投资组合的管理公司、基金分销商及交易平台营运者的监管框架的声明》为起点，香港明确了对虚拟资产“持牌监管”的基本思路；3）2022 年以来，香港正式开始支持 Web3 产业发展。2022 年 12 月，南方东英比特币期货和以太币期货 ETF 在香港证券交易所上市；2023 年 2 月，香港特区发售首批代币化绿色债券；2023 年 6 月，香港拟正式引入虚拟资产服务供应商发牌制度，未来将继续在法例更新、监管、投资者保护等方面有序合规地推动行业发展。

图表359： 香港虚拟资产监管框架



资料来源：香港 web3 嘉年华 2023，Web3 小律，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

#2 大陆 Web3 政策：自上而下推动，重视底层基础设施投入和审慎监管

中国大陆对 Web3 产业的政策相对保守，坚持“去币存链”，重视区块链基础设施的建设，但对虚拟资产秉持审慎监管态度。以 NFT 为例，国内监管层面对 NFT 投资保持谨慎态度。NFT 的法律性质、交易方式、监督主体、监督方式等尚未明确，并且存在炒作、洗钱和金融产品化等风险，国家层面对于 NFT 投资态度谨慎，警惕“击鼓传花”式的金融骗局。2022 年 4 月 13 日，多部门联合发布《关于防范 NFT 相关金融风险的倡议》要求各会员单位坚决遏制 NFT 金融化证券化倾向，禁止通过 NFT 变相发行交易金融产品，抵制 NFT 投机炒作行为，不直接或间接投资 NFT。

图表360： 中国大陆 Web3 相关政策

时间	措施	部门	内容
2017.9	发布《关于防范代币发行融资风险的公告》	人民银行、网信办等七部门	虚拟货币是一种特定的虚拟商品，不由货币当局发行，不具有法偿性与强制性等货币属性，不是真正的货币，不应且不能作为货币在市场上流通使用。
2021.5	发布《关于防范虚拟货币交易炒作风险的公告》	中国互联网金融协会、中国银行业协会、中国支付清算协会	虚拟货币相关业务活动属于非法金融活动
2021.7	《关于支持浦东新区高水平改革开放打造社会主义现代化建设引领区的意见》	中共中央、国务院	构建贸易金融区块链标准体系，开展法定数字货币试点
2021.9	发布《国家发展改革委等部门关于整治虚拟货币“挖矿”活动的通知》	国家发展改革委等 11 部门	禁止以任何名义发展虚拟货币‘挖矿’项目；加快有序退出存量项目
2021.10	《国家标准化发展纲要》	中共中央、国务院	在两化融合、新一代信息技术、大数据、区块链、卫生健康、新能源、新材料等应用前景广阔的技术领域，同步部署技术研发、标准研制与产业推广，加快新技术产业化步伐。
2022.4	《关于防范 NFT 相关金融风险的倡议》	中国互联网金融协会、中国银行业协会、中国证券业协会	防范金融风险，遏制 NFT 金融化证券化倾向；不直接或间接投资 NFT，不为投资 NFT 提供融资支持
2022.6	《数字藏品行业自律发展倡议》	中国文化产业协会等	提出数藏平台应依法具备相应资质、确保区块链技术安全可控、坚持实名制、加强知识产权保护能力建设、坚决抵制防范金融化和恶意投机炒作、倡导理性消费等。
2022.7	《关于推进对外文化贸易高质量发展的意见》	商务部等 27 部门联合发布	发挥国内大市场和丰富文化资源优势，加强数字文化内容建设，促进优秀文化资源、 文娱模式数字化开发。推动文化和科技深度融合，促进大数据、云计算、人工智能、 区块链等新技术应用，赋能文化产业和贸易全链条，带动传统行业数字化转型，提升企业数字化运营能力。
2022.10	“剑网 2022”专项行动	国家版权局等四部门	联合相关部门对重大案件进行挂牌督办，通过推动规范发展与打击惩治并举，加强网络版权全链条保护，重点聚焦文献数据库、网络视频、网络文学、NFT 数字藏品、“剧本杀”等领域业态版权监管。

资料来源：各部门网站，华泰研究

#3 美国、欧盟相继推出加密资产监管框架，全球加密资产监管生态逐渐完善

2022 年 9 月，白宫发布了加密监管框架草案，草案列出的主要原则包括采取消费者保护措施、维持金融稳定、防止非法使用加密货币、保持美国在全球金融领域的领导地位以及负责任的技术创新、探索美国中央银行数字货币 (CBDC)等。10 月，欧盟出台《加密资产市场监管法案》（简称 MiCA），形成欧盟内部统一的监管和执法体系。这一监管框架将保护投资者并维护金融稳定，同时允许创新和培养加密资产行业的吸引力。

更多免费行业报告 并购家 www.ipoipo.cn

图表361： 全球虚拟资产监管政策

时间	国家/地区/机构	政策/法案/项目	详情	态度
2023.3	美国	《加密挖矿保护法》	美密苏里州众议院委员会通过《加密挖矿保护法》，该法案禁止州和政治部门停止加密货币挖矿节点的工作，确保加密矿业公司与其他企业一样受到相同法律的约束，主要目的是防止能源公司对矿业公司适用歧视性税收规则	😊
2023.2	中国香港	香港证监会就有关监管虚拟资产交易平台的建议展开咨询	此次咨询中，证监会特别就以下事项征询市场意见：是否准许持牌平台营运者向零售投资者提供服务；如准许的话，除了所建议一系列妥善的投资者保障措施外，还应实行哪些措施	😊
2023.1	中国香港	-	香港证监会行政总裁：将为散户投资者制定虚拟资产交易标准，仅限于高流动性资产	😊
2022.12	美国	美国加州监管机构对 17 家涉嫌欺诈的加密网站发出警告	美国加州金融保护与创新部（DFPI）针对其怀疑存在欺诈行为的 17 家加密货币经纪公司和网站发出警告，其中包括 Tahoe Digital Exchange、TeleTrade Options、Tony Alin Trading Firm 等，敦促消费者在回应任何提供投资或金融服务的邀约之前要格外谨慎，检查投资或金融服务提供商是否在加州获得许可证	😐
2022.11	新加坡	-	新加坡金融监管局（MAS）宣布上线首个数字资产和去中心化金融（DeFi）行业试点，已完成第一笔实时交易	😊
2022.11	中国香港	《有关虚拟资产期货交易所交易基金的通知》	为加密货币期货 ETF 发行商制定了新规则，具体内容包括：ETF 的管理公司需具有良好的监管合规记录、展示至少三年的业绩记录；发行商需证明虚拟资产 ETF 有足够的流动性；衍生品净风险敞口不能超过 ETF 总资产净值的 100% 等	😊
2022.10	欧盟	《加密资产市场监管法案》（简称 MiCA）	欧盟内部将形成统一的监管和执法体系。MiCA 的执法将由每个成员国指定的监管机构管理，欧盟层面的将由欧洲银行管理局（EBA）和欧洲证券及市场管理局（ESMA）进行监管	😐
2022.09	美国	加密货币监管框架草案	美国白宫发布加密货币监管框架，包括金融服务行业应该如何发展以使跨国交易更容易，以及如何打击数字资产中的欺诈行为等	😐

资料来源：The New York State Senate，新加坡金管局，香港证券及期货事务监察委员会，The Block，欧盟，瑞士金融市场监管局，国际清算银行，The White House，华泰研究

基础设施层：寒冬孕育机遇，Web3 基础设施创新持续

在过去一年的熊市中，Web3 基础设施领域的建设仍持续。我们看到：1）公链：以太坊相继完成了合并、上海/卡佩拉升级，由 PoW 顺利过渡到 PoS；2）Layer2：业内继续聚焦以太坊扩容方案，Layer2 Rollup 赛道的发展成为热点话题。根据 l2Beat 数据，年初至今，Layer2 总锁仓金额迅速增长约 1.5 倍至 USD10.5bn；3）钱包和 DID：钱包是用户进入 Web3 领域的第一站，其性能和体验直接影响 Web3 能否步入大规模应用，而安全和易用性之间的冲突是钱包解决方案的主要瓶颈之一，行业关注 MPC 钱包、设备账户等方案；4）央行数字货币：全球已有 131 个国家/地区的 CBDC 处于研究/概念验证/试点/运行阶段，全球 CBDC 推广进展迅速。

Layer1：以太坊已完成合并及上海-卡佩拉升级

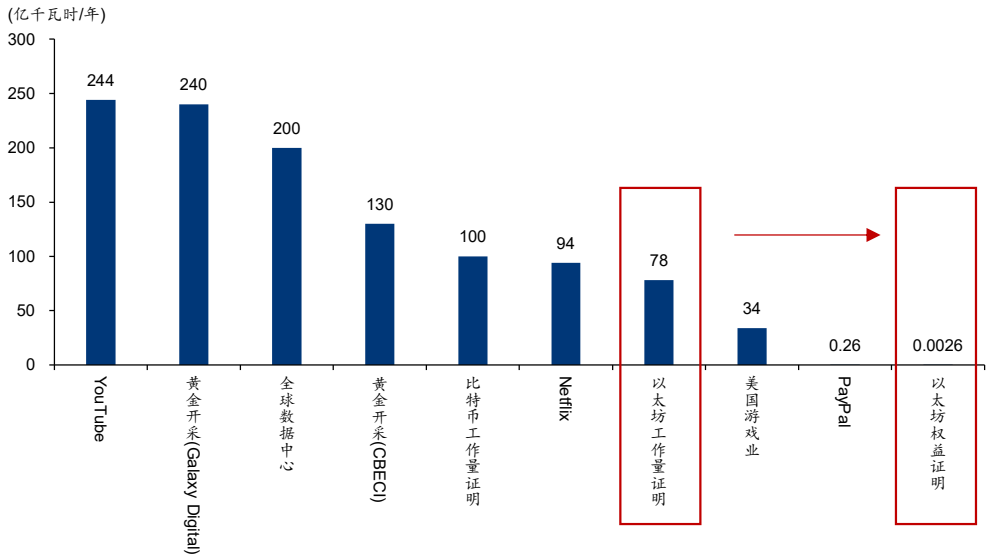
为了提升可扩展性，以太坊通过链上与链下两条路径积极进行扩容。链上扩容通常被称为 Layer1，通过直接修改区块链规则来加快数据处理速度，扩充区块链容量。比如以太坊 2.0 通过将共识机制由 PoW 转向 PoS、分片链等方式来解决可扩展性问题。而链下扩容则不改变区块链基础规则，转而通过架设外部通道，实现信息的传播与扩展，从而达到扩容效果。其中，Layer2 解决方案的安全性直接建立在主网，主要包括状态通道、Plasma、Rollups、validiums 等。其他解决方案包括各种形式的新链，其安全性独立于主网，如侧链等。

2022 年 9 月，以太坊合并统一了以太坊的共识和执行层，实现了从使用工作量证明 (PoW) 共识机制到权益证明 (PoS) 共识机制的升级。合并后，年能源消耗量由工作量证明机制的 78 亿千万时/年降低为 0.0026 亿千万时/年，实现能源消耗量的显著降低。同时，在 PoS 机制下，验证节点不需要进行计算难题的计算，而是根据自己所持有的 ETH 数量来获得权益奖励，大大降低了 ETH 的日均增发量。而以太坊的上海/卡佩拉升级允许质押者提取他们所投的 ETH，减少与质押 ETH 相关的流动性风险，促进流动性质押供应商的增长。

免责

更多免费行业报告 并购家 www.ipoipo.cn

图表362：各个行业估算年能源消耗



资料来源：Etherscan，华泰研究

Layer2：以 Rollup 为中心的扩容路线图已确立，ZK 和 OP 技术路径成为主流

目前以太坊已确立起以 Rollup 为中心的扩容路线图，主要包括 ZK Rollup 和 Optimistic 两种类型。两者在技术路径上存在差异，主要体现在以下方面。1）交易速度：ZK Rollup 理论上可以实现每秒数千笔的交易量，高于 Optimism 的每秒数百笔。2）交易费用：ZK Rollup 需要高端昂贵的硬件和证明验证，交易费用更高。 3）智能合约可执行性：ZK Rollup 将智能合约逻辑转换为零知识证明，提高了智能合约的可执行性；Optimistic 则需要更多的存储和计算资源来支持智能合约的执行。4）安全性：ZK Rollup 建立在零知识证明和密码学算法的基础上，可以保证交易隐私和链上安全；Optimistic 则存在一定的安全风险。 5）去中心化程度：ZK Rollup 需要更多的节点参与来确保去中心化程度。

图表363：Rollup 技术路径梳理

Roll-up技术路径梳理	交易速度	交易费用	智能合约可执行性	安全性	去中心化程度
ZK rollup	通过将多个交易打包到一个区块中，可以大幅提高交易的吞吐量，理论上可以实现每秒数千笔的交易量	需要高端昂贵的硬件和证明验证，交易费用更高	将智能合约逻辑转换为零知识证明，并在链上存储证明的哈希值，从而提高了智能合约的可扩展性和可执行性	建立在零知识证明和密码学算法的基础上，可以在保证交易隐私和链上安全的同时，实现高吞吐量和低成本的交易	在验证和确认过程中依赖于验证节点和存证人，因此需要更多的节点参与来确保去中心化程度
Optimism	采用区块时间缩短和交易批量优化等技术，可以实现每秒数百笔的交易量	不需要交易证明和数据发布，交易费用更低	将交易验证和执行的过程推迟到链上的状态确认阶段，因此需要更多的存储和计算资源来支持智能合约的执行	依赖于多方共识和交易确认机制，需要更多的节点参与验证和确认，因此存在一定的安全风险	使用了更加去中心化的验证和确认机制，可以实现更高的去中心化程度

资料来源：MarsBit，Forsight News，Zonff Partners，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

钱包：进入 Web3 的第一站，关注安全和易用性的平衡

钱包在 Web3 世界是资产、身份、账号的统一体，也是用户进入 Web3 领域的第一站，其性能和体验直接影响 Web3 能否步入大规模应用。随着过去一年中心化交易所 FTX 破产等风险事件的密集发生，用户对于资产自托管的关注逐步提升，对于钱包应用的需求攀升钱包技术路径主要包括多方计算、账户抽象和设备账户等。多方计算（Multi-Party Computation）通过多个参与方共同计算，实现保护隐私数据的目的；账户抽象（Account Abstraction）将智能合约功能从特定账户中抽象出来，使智能合约可以被任何账户调用；设备账户（Device account）将设备作为账户，设备可以通过私钥签名的方式执行交易，并接收来自其他账户的交易。**我们看到，安全和易用性之间的平衡是钱包方案的发展关键。**

图表364：钱包技术路径：多方计算、账户抽象和设备账户

	多方计算（Multi-Party Computation）	账户抽象（Account Abstraction）	设备账户（Device account）
原理	通过多个参与方共同计算，实现保护隐私数据的目的	将智能合约功能从特定账户中抽象出来，使智能合约可以被任何账户调用	将设备作为账户，设备可以通过私钥签名的方式执行交易，并接收来自其他账户的交易
优点	在不泄露敏感信息的前提下完成计算任务，同时也能够保护数据隐私和防止数据泄露	简化智能合约开发和提高智能合约的安全性，同时也可以提高以太坊的可扩展性和交互性	提高了设备的自主决策能力和交易灵活性，同时也能够为设备提供更好的安全性和信任性
缺点	多方计算的运算速度较慢，需要更多的计算资源和通信开销，同时也需要更复杂的协议和算法来支持这一特性	可能会导致开发者误用智能合约功能，同时也需要更复杂的编程语言和智能合约编写工具来支持这一特性	设备账户管理和维护成本较高，并且需要解决设备私钥安全性和使用权限控制等问题
项目	enigma、KEEP	ethereum、HYPERLEDGER FABRIC	vechain、fetch.ai

资料来源：SevenUp DAO, Ethereum 官网，华泰研究

央行数字货币：2022 年各国央行数字货币进展迅速

2022 年，各国央行数字货币纷纷开启试点。截至 2022 年 7 月，共有近 100 个 CBDC（国家数字货币）处于研究或开发阶段；而截至 2023 年 5 月 15 日，全球已有 131 个 CBDC 处于研究/概念验证/试点/运行阶段，其中巴哈马 Sand Dollar 和牙买加 JAM-DEX 已全面运行。反映出全球 CBDC 推广进展迅速，许多国家对数字货币持开放态度，在技术开发、法规体系和监管方面积极探索。目前，中国央行数字人民币自 2020 年 8 月推出后，目前已进入试点阶段。根据 Juniper Research 预测，到 2030 年，通过 CBDC（中央银行数字货币）支付的价值将达到每年 2130 亿美元，高于 2023 年的 1 亿美元。我们认为，央行数字货币是区块链和代币化的重要用例，构建了未来全球数字经济发展的基石。

应用层：市场回归冷静，关注 Web3 与真实世界连接

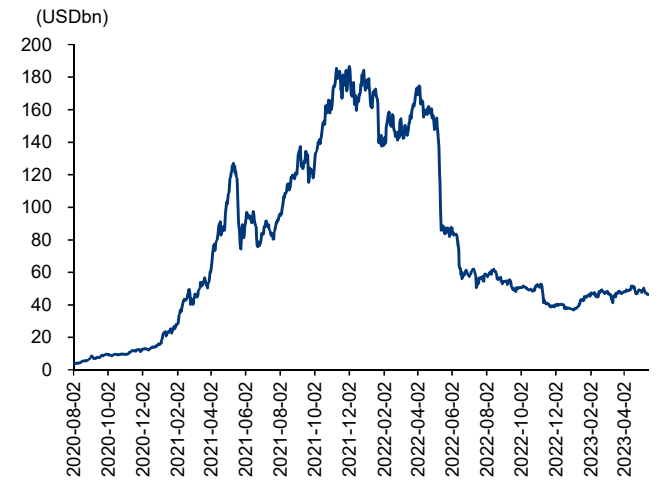
Web3 应用层，继过去一年 NFT 冲高后跌落，市场逐渐从泡沫破裂期回归冷静建设期。目前阶段，各赛道应用的大规模普及依然存在很长一段距离，我们观察到以下趋势：1）DeFi：以太坊合并后，质押取代挖矿，催化 LSD（流动性质押衍生品）市场增长。现实世界资产（RWA）代币化赛道兴起，架设虚拟资产和现实资产的桥梁；2）NFT：泡沫破裂后，市场关注 NFT 与音乐、社交、游戏等具体场景结合；3）GameFi：P2E（Play to earn）模式热度消退，游戏性受到关注；4）SocialFi：旨在实现社交权益保障/价值分配，但各生态间存在数据孤岛问题，不同协议间的用户数据自由迁移仍待完善。

免责
更多免费行业报告 并购家 www.ipoipo.cn

DeFi：锁仓金额大幅下降，衍生品逆势发展

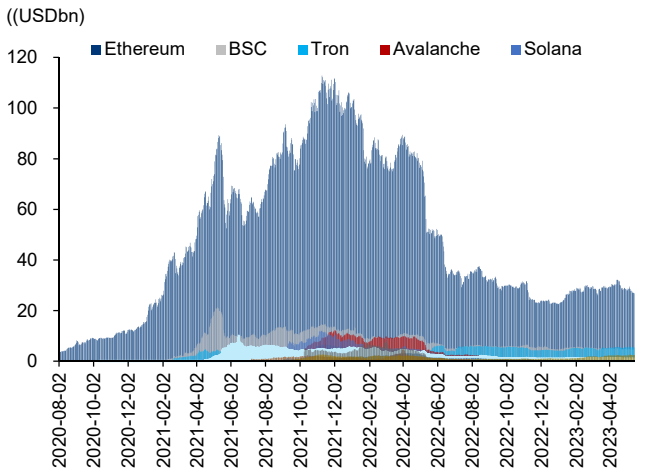
过去一年，随着 Web3 熊市来临，DeFi 遭遇重创，锁仓金额（TVL）大幅下降。部分知名机构和项目，如加密货币贷款机构 BlockFi 等由于风险事件而申请破产。同时，流动性质押衍生品（LSD）、现实世界资产（RWA）代币化等新赛道开始兴起，侧面印证 DeFi 在逐步降低普通人进入门槛，以及完善与真实世界的连接，我们认为这两个因素是通向大规模应用的必要条件。

图表365：DeFi 项目总锁仓规模（截至 2023 年 5 月 15 日）



资料来源：DeFi Llama，华泰研究

图表366：DeFi 项目锁仓规模（按公链，截至 2023 年 5 月 15 日）



资料来源：DeFi Llama，华泰研究

新兴应用#1：流动性质押衍生品（LSD）-以太坊转 PoS 机制驱动下，LSD 赛道蓬勃发展

以太坊从 PoW 机制转向 PoS 机制后，至少需要 32 个 ETH 质押才能成为节点验证者，普通用户难以通过以太坊质押受益。此外，大量被质押的 ETH 也降低了市场流动性。而流动性质押衍生品（LSD）能够降低用户质押门槛，让普通用户参与质押并获得奖励。流动性质押衍生品的商业模式中，用户将 ETH 质押给 LSD 协议，协议将收集的 ETH 交给节点运营商，由运营商来执行信标链的签名验证来获取奖励，LSD 协议则从中抽成一部分质押收益。以这样的方式，即使拥有 ETH 数量不够，无法成为节点验证者的用户，也能够参与质押，获取受益。对于生态而言，LSD 也能够促进一个生态的网络质押，提高网络安全性，同时为整个生态提供流动性。

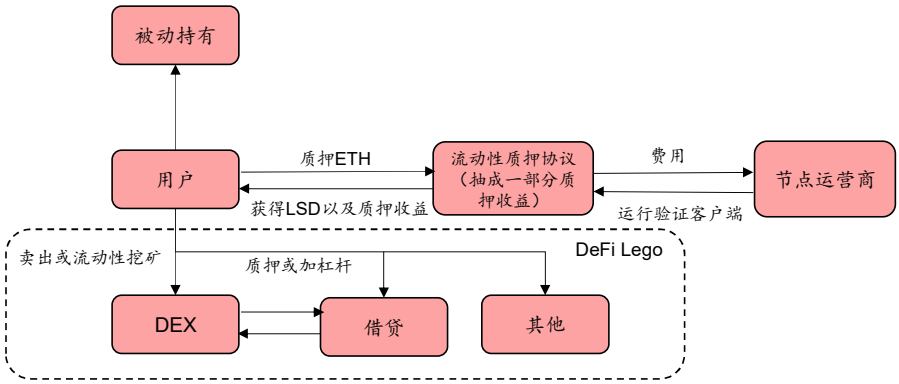
图表367：以太坊质押方式比较

质押方式	单独质押	质押即服务	联合质押（LSD）	中心化交易所
掌握密钥	验证密钥&提现密钥	提现密钥	无	无
是否需要部署硬件	是	否	否	否
需要 ETH 数量	≥32	≥32	任意	任意
资产托管	否	否	是	是
信任风险	无	低	高	很高
收益率	1	2	3	不确定
参与的难易程度	4	3	2	1

资料来源：DODO Research，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表368：流动性质押衍生品（LSD）商业模式

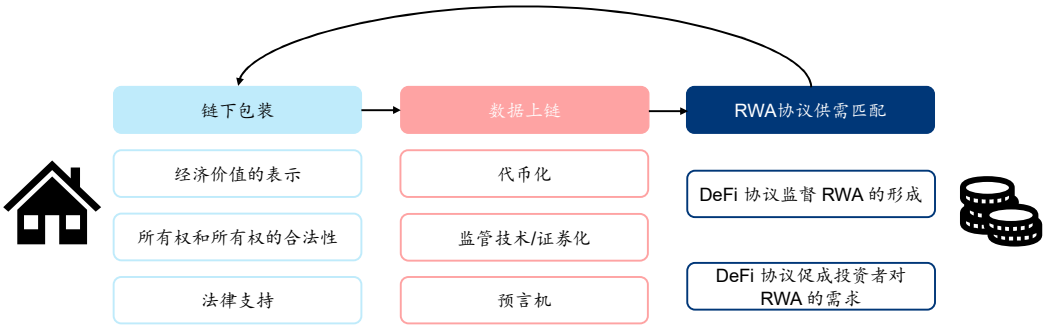


资料来源：Foresight news，华泰研究

新型应用#2：现实世界资产（RWA）代币化-架设虚拟资产和现实资产的桥梁

RWA 代币化是将有形资产转换为数字代币或 NFT 的过程，使它们能够在区块链上进行交易，在现实经济和虚拟经济之间搭建了桥梁。RWA 的流程分为三个步骤：（1）链下包装：明确资产的价值、资产所有权、资产权益的法律保障等；（2）数据上链：有关资产的经济价值和所有权及权益的信息在数据化之后被带到链上，存储在区块链的分布式账本中；（3）RWA 协议需求和供应：专注于 RWA 的 DeFi 协议推动了现实世界资产代币化的整个流程。在供应端，DeFi 协议监督 RWA 的形成。在需求端，DeFi 协议促成投资者对 RWA 的需求。

图表369：现实世界资产（RWA）代币化过程



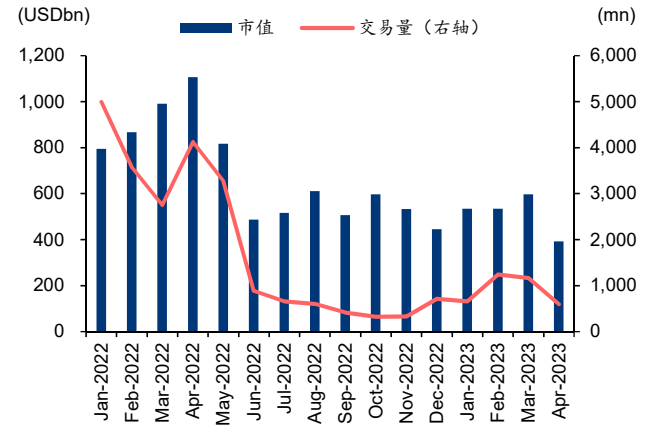
资料来源：Binance Research，华泰研究

NFT：行情消退，关注与具体场景的结合

在经历了 22 年 4 月的高峰后，NFT 整体市值呈现下降趋势，到目前为止一直呈现持续低迷的状态。根据 Footprint Analytics，NFT 交易量在经历去年年初的高峰后便陷入持续下滑，今年第一季度有小幅回温，但总体仍保持低迷。同时，NFT 买家卖家数量下降，表明 NFT 投资逐渐回归理性。

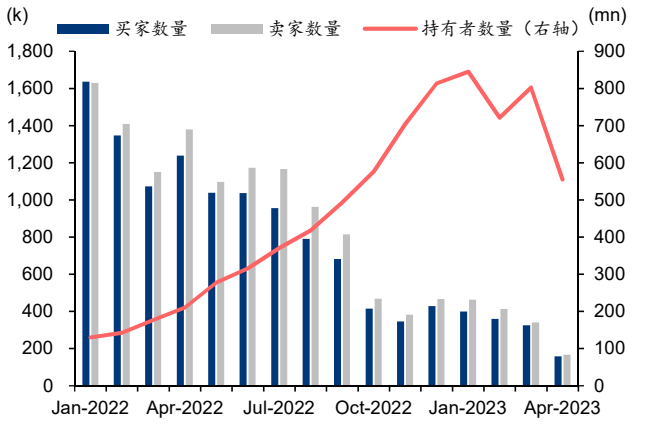
免责
更多免费行业报告 并购家 www.ipoipo.cn

图表370：NFT 交易量及市值



资料来源：Footprint Analytics，华泰研究

图表371：NFT 买家/卖家数量及持有者数量



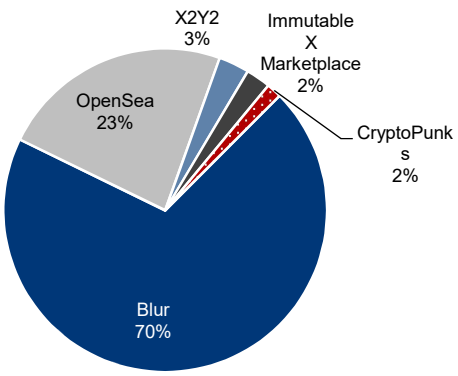
资料来源：Footprint Analytics，华泰研究

#1NFT 交易：Blur 崛起，挑战 OpenSea 龙头地位

过去一年，随着流动性大幅下降，NFT 交易平台竞争激烈程度上升，竞争格局迎来洗牌。根据 DappRadar 数据，截至 2023 年 5 月 8 日，NFT 交易平台 Blur 近 30 日的交易量达到 6.9 亿美元，成交规模超过 OpenSea 的 2.4 亿美元，在市场排名中居第一位。从周频交易量数据来看，22 年 12 月之前，OpenSea 市场份额基本保持领先地位，自去年 12 月之后，Blur 逐渐挤占 OpenSea 的市场份额，开始占据主导。

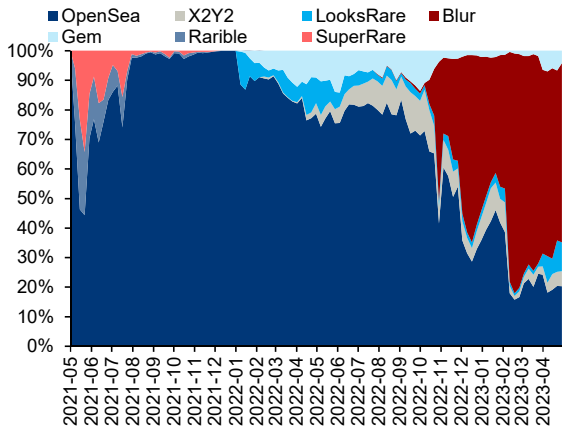
不同于 OpenSea，Blur 更为交易导向，以吸引 NFT 交易员为卖点。Blur 提供非强制版税、零交易费的交易平台，来吸引 NFT 交易员，提升平台流动性，撼动 OpenSea 龙头地位。其优势包括：1）速度快，能够快速更新 Listing 信息；2）符合交易员习惯，提供智能扫货、智能批量挂单等功能；3）交易成本低，具备可选版税+零交易费的优势。我们认为，以交易为核心的 Blur 能够在激烈的市场竞争中迅速打开局面，侧面印证了熊市中流动性的不足以及市场对于流动性的需求。但长期来看，流动性只是市场健康发展的充分不必要条件，缺乏优质 NFT 项目的持续供给仍然是行业痛点。

图表372：近 30 天 NFT 交易平台交易量占比



注：数据截至 2023 年 5 月 8 日
资料来源：DappRadar，华泰研究

图表373：主流 NFT 交易平台市场份额分布（按周频交易量）



注：数据截至 2023 年 5 月 1 日
资料来源：Dune analytics，华泰研究

#2NFT 应用：与音乐、游戏等具体场景的结合逐步深入

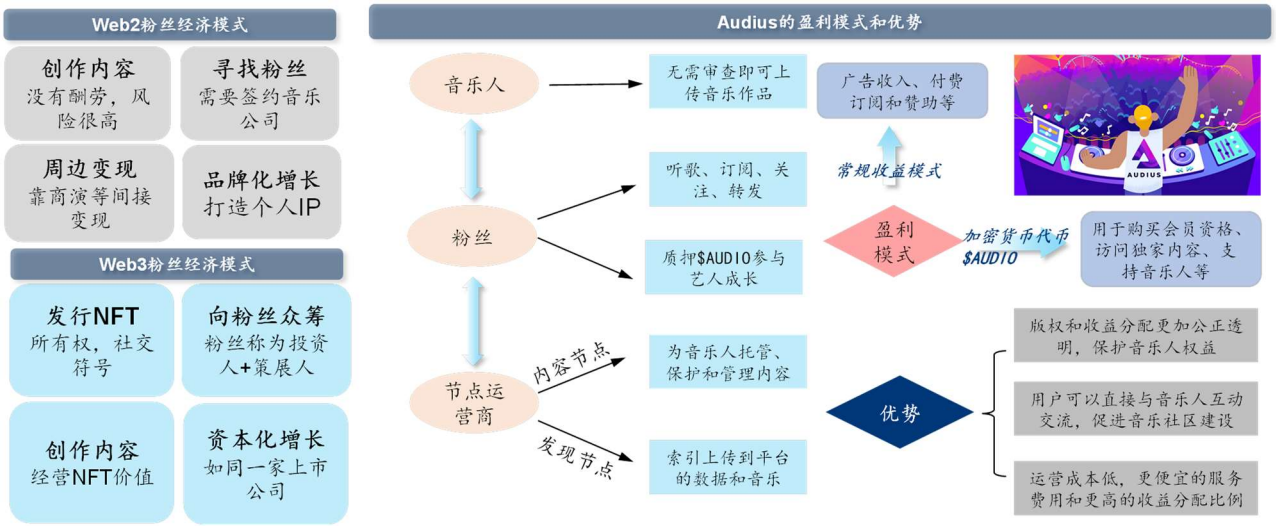
过去一年，市场关注 NFT 与音乐、游戏等具体场景的结合。Web3 独立音乐人 Daniel Allan 发行音乐 NFT 来进行创作变现，于 2021 年 4 月制作了第一款音乐 NFT。2023 年 3 月，Daniel 已在种子轮筹集 100 万美元资金，用于建立自己的音乐 NFT 项目 Daniel Allan Entertainment。在过去的一年半里，其 NFT 的一级销售总收入 228ETH，二级销售总收入 105ETH。

更多免费行业报告 并购家 www.ipoipo.cn

NFT 通过实现资产所有权认证和去中心化交易，帮助创作者建立 Web3 粉丝经济。Web2 世界中的粉丝经济有较大的能量，但模式陈旧，严重内卷；而 Web3 将建立一种新的粉丝经济模式。在新的模式下，创作者能够：1）发行 NFT：创作者对创作内容拥有所有权，NFT 同时也是社交符号的象征；2）向粉丝直接众筹：粉丝称为投资人和策展人，跳过第三方平台/公司，直接参与收益的分配；3）创作变现：创作者通过持续创作来经营 NFT 价值，从而实现创作变现。

Audius 是 NFT 与音乐结合的一个典型应用。作为一个去中心化的音乐流媒体平台，Audius 基于区块链技术构建，使用去中心化的存储方式和智能合约来实现音乐版权和收益分配等功能，降低了中间商赚取的服务费用，增加了艺术家的利润，为音乐人和用户提供更公正、透明和高效的服务。Audius 的商业模式主要基于音乐流媒体平台的常规收益模式，例如广告收入、付费订阅和赞助等。此外，Audius 还推出了自己的加密货币代币$AUDIO，该代币可用于购买会员资格、访问独家内容、支持音乐人等。相对于传统音乐流媒体平台，Audius 在版权保护、收益分配、运营成本等方面存在优势。

图表374： NFT 与具体场景结合的典型应用

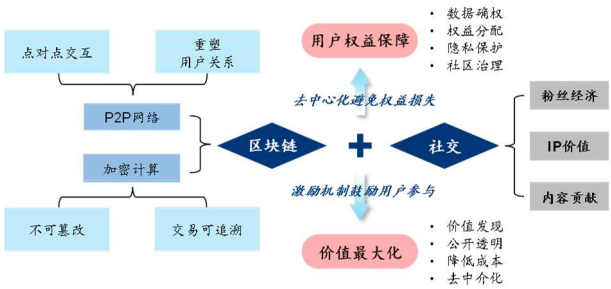


资料来源：Audius 官方网站，CoinCentral，Web3 嘉年华，华泰研究

去中心化社交 SocialFi：旨在实现社交权益保障/价值分配

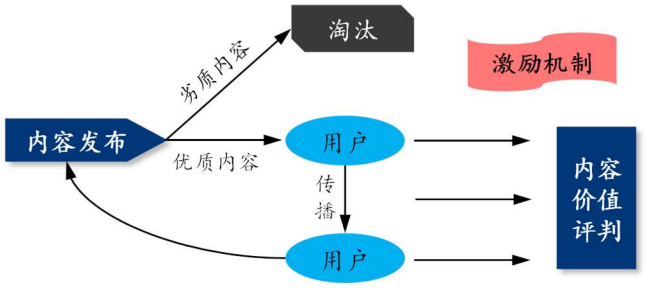
使用区块链技术构建的 Web3 去中心化社交应用能够实现社交领域的价值最大化和用户权益保障。SocialFi 的优势主要表现在：1）去中心化避免权益损失，充分保障用户权益，实现数据确权、权益分配、隐私保护和社区治理；2）激励机制鼓励用户参与，促成价值最大化，实现价值发现、公开透明、降低成本和去中介化。SocialFi 平台的内容发布后，接收到优质内容的用户在激励机制的作用下会进行用户之间的传播，而劣质内容则会逐渐被淘汰，由此实现高效的内容价值评判。现阶段，SocialFi 仍处于早期阶段，各生态间存在数据孤岛问题，不同协议间的用户数据自由迁移仍待完善。

图表375： Social-Fi 平台的价值发现和用户权益保障逻辑



资料来源：WebX 实验室，华泰研究

图表376： Social-Fi 平台的内容发现路径



资料来源：WebX 实验室，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

GameFi：P2E 热度消退，游戏性受到关注

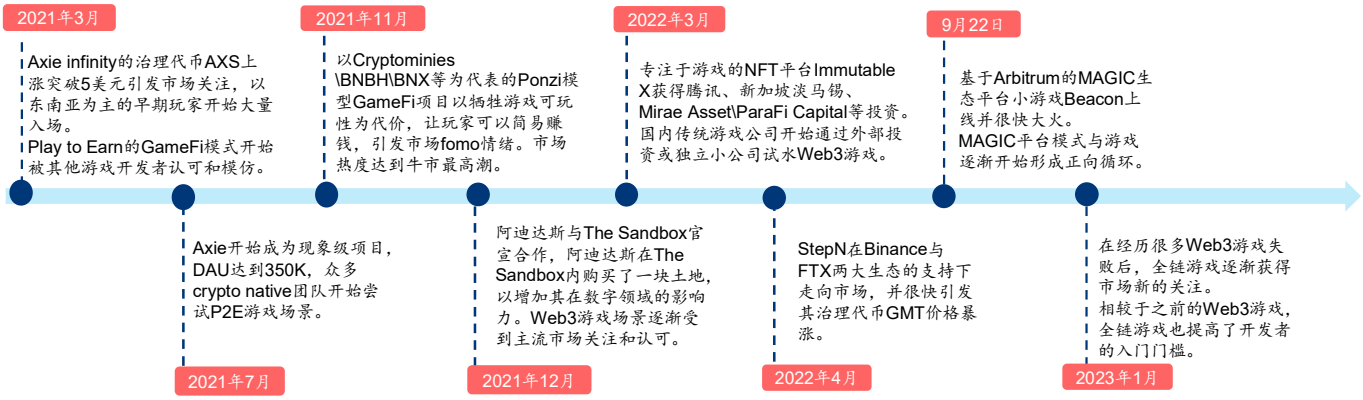
GameFi（Game Finance）是运行在区块链上的去中心化游戏，可以理解为 Game、DeFi 和 NFT 的组合产物。GameFi 将 DeFi 的规则游戏化，通过区块链系统把游戏中的道具及衍生品 NFT 化，使玩家可以在去中心化的游戏中完全拥有这些资产的所有权，并可以合法地、受保护地自由交易，玩家既是参与者，也是拥有者。现阶段主流 GameFi 项目包括以赚取经济收益为出发点推动玩家参与的 P2E 游戏（Play-to-Earn，边玩边赚）；将传统游戏、传统模型通过资产上链实现区块链改造的链改游戏；直接在区块链上进行开发的高可玩性 3A 大作；以及包含创造型 NFT，如提供土地和地图编辑器等支持用户创建内容的游戏或 IP。

图表377： Game-Fi 的四种类型及其举例

资料来源：各公司官网，华泰研究

随着 P2E 热度的消退，游戏性逐渐受到关注。21 年 3 月，Axie infinity 的治理代币 AXS 上涨突破 5 美元引发市场关注。21 年 7 月，Axie 开始成为现象级项目。21 年 11 月，以 Cryptominies 等为代表的 Ponzi 模型 GameFi 项目市场热度上升。21 年 12 月，阿迪达斯与 The Sandbox 官宣合作，Web3 游戏场景逐渐受到主流市场关注和认可。22 年 3 月，专注于游戏的 NFT 平台 Immutable X 获得腾讯、新加坡淡马锡等投资。22 年 4 月，StepN 在 Binance 与 FTX 两大生态的支持下走向市场，并引发其他治理代币 GMT 价格暴涨。23 年 1 月，在经历很多 Web3 游戏失败后，全链游戏逐渐获得市场新的关注。相较于之前的 Web3 游戏，全链游戏也提高了开发者的入门门槛。

图表378： Web3 游戏发展时间线



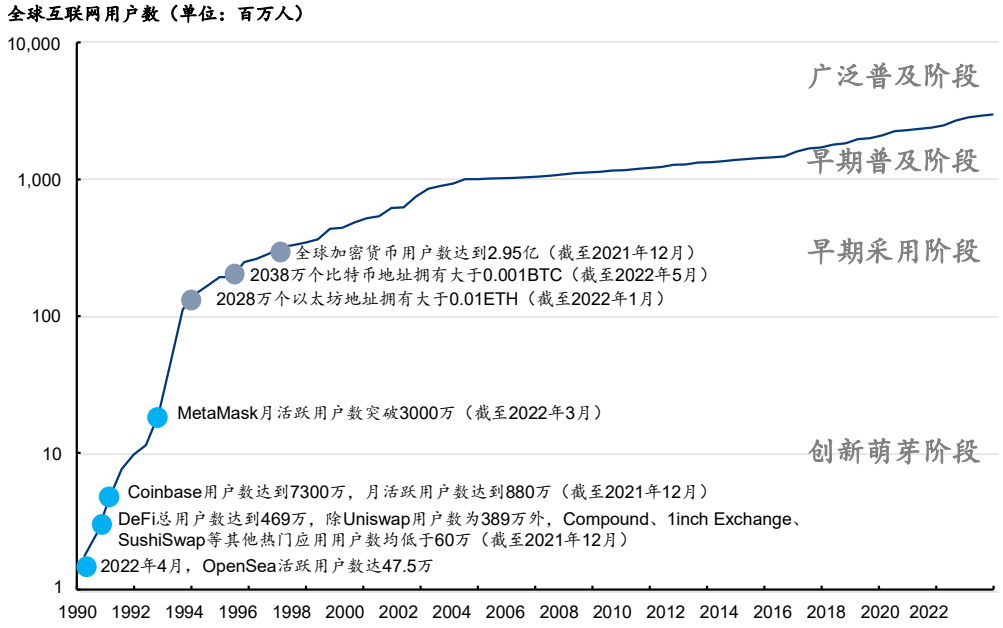
资料来源：Zonff Partners，华泰研究

更多免费行业报告 并购家 www.ipoipo.cn

未来展望：Web3 离我们还有多远？

从用户数的指标衡量，Web3 可对标 1990 年代 Web1 的创新萌芽阶段。在当前的技术水平和基础设施条件下，Web3 还无法承载对用户和开发者友好的应用程序和服务，整体生态较为贫瘠，大部分现有服务均集中在融资、加密货币和 NFT 交易上，用户进场主要为投机交易或通过 X-to-Earn 模式赚取经济收益，除部分头部项目外，用户数量较少且粘性较差。如果以用户数作为衡量互联网范式发展和应用程度的标尺，那么 Web3 目前的推进状态，应类似于 1990 年代 Web1 正在经历的创新萌芽阶段，可以说还处于一个非常早期的阶段。

图表379：Web3 目前所处阶段以及比特币和加密交易同互联网发展历程的比较



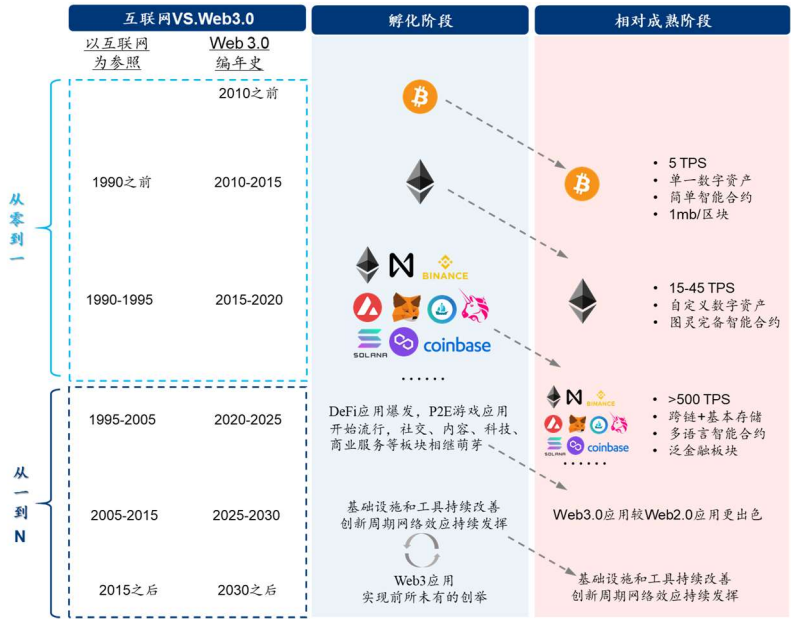
资料来源：Dune Analytics，glassnode，BitInfoCharts，Backlinko，Decrypt，World Bank，World Internet Stats，Folius Venture（壹葉創投），华泰研究

当前，我们正站在从 1 到 N 的前夜。2020 年 DeFi 应用的飞速发展实现了 Web3 从零到一的突破性发展，特别是 Opeasea、Uniswap 等应用的崛起，第一次为 Web3 导入了大批量活跃用户；同时，基础设施、数据和工具质量的大幅提高，吸引大量优秀创业者和优质资本涌入参与 Web3 建设，帮助行业真正开启尝试投机之外的实际商业应用。自金融板块后，泛文娱、NFT 消费品、科技以及商业服务类板块正在萌芽，尽管受技术研发、人才储备、资金投入、政策支持、标准建设等多方面因素的制约，当前 Web3 的发展速度仍不算理想，但我们也确实已经来到了从 1 到 N 的前夜。

展望未来，随着基础设施和工具的持续改善，Web3 将借助区块链底层协议创新周期的网络效应实现从 1 到 N 的增长。未来随着 Web 3 技术栈堆进一步成熟，第一批可行的 Web3 应用有望在 2025 年实现落地。此后，当这些应用慢慢开始从消费者和企业身上获得经济价值，从而产生真实、可持续的现金流，将逐步推动 Web3 行业超越现有投机主导的生态，走上真正良性、健康的发展路径。而随着区块链底层协议模块化可组合性的发挥，各类应用通过相互调用和嵌套实现快速创新，在网络效应的驱动下，行业将经历指数级增长，实现从 1 到 N 的跨越式发展，届时，不可预见的全新创举或将到来，海量用户将尽情徜徉在 Web3 世界，享受前所未有的服务体验。

免责
更多免费行业报告 并购家 www.ipoipo.cn

图表380： Web3 发展展望



资料来源：华泰研究

Web3 产业链相关公司：趣链科技、矩阵元

趣链科技：联盟区块链底层平台独角兽

趣链科技（未上市）是国际领先的区块链技术及解决方案提供商，致力于为政府及企业客户加速实现数字化转型，做好数字经济“新基建”。公司成立于 2016 年，总部位于杭州，于 2021 年完成数亿元的 C 轮融资，成为区块链领域首家独角兽企业。公司核心团队均来自国内外一流高校，由中国工程院陈纯院士担任首席科学家。趣链科技拥有全球首个区块链 3.0 全栈全生态能力体系，涵盖国产自主可控的底层技术能力、开箱即用的运维及研发能力、开放互联的跨链能力及高效的数据共享和隐私计算能力等，助力客户实现高效率、低成本地布局应用区块链。其中趣链区块链底层平台自 2017 年起持续参与工信部信通院每年组织的“可信区块链技术平台”测评，均 100%通过所有测评项。

2022 年 6 月，趣链科技联手易马达科技推出全球首个区块链换电 SaaS 平台。2022 年 11 月，子公司产链数字科技建设的海曙区数字房票管理系统开出全国首张区块链房票。趣链科技作为项目牵头单位承担 2022 年国家重点研发计划，并参与科技部、工信部等发起的国家重点研发项目 10 余项，目前已申请区块链相关专利近 900 件，获得软著权 130 余项，掌握关键核心技术发展主导权。

矩阵元：数字时代的私有计算和区块链技术提供商

矩阵元（未上市）成立于 2014 年，创始人及 CEO 为孙立林，目前有 110 多名企业员工，1.5 亿融资金额和 81 项发明专利。矩阵元定位为数字时代的隐私计算和区块链技术提供商，专注于现代密码学、前沿密码学、计算复杂性理论、分布式计算等科研与工程实践，协同大数据、人工智能、物联网、区块链等技术，赋能政务、金融、司法、医疗等产业应用。希望打造新一代数据要素市场和数字化基础设施，以此解决数据安全和隐私保护问题，同时利用大数据、物联网等技术，将区块链赋能在金融、医疗、政务、司法等方面。矩阵元主要产品包括隐私计算产品 PCS Box 和区块链产品 JUICE。PCSBox 是高性能聚合隐私计算商用硬件产品，基于隐私计算、数据安全、人工智能等技术为客户提供软硬一体标准化解决方案，其操作简洁、开箱即用，支持本地化部署、多系统适配及多配置选择等功能，实现集存储、加密、计算、认证「四位一体」的全数据链隐私保护，充分做到数据可用不可见，助力企业、机构、政府等各类客户在不同场景中解决数据隐私、数据安全问题。JUICE 是由矩阵元推出，符合国密标准，可服务于金融保险、医疗健康、政务服务、社会治理、物流航空等行业的实用型联盟链平台。

免责
更多免费行业报告 并购家 www.ipoipo.cn

风险提示

宏观经济波动。在美联储持续加息、地缘政治等因素影响下，全球宏观仍然处于下行周期，会对电子终端等可选消费种类需求产生负面影响。在海外通胀高企的情况下，宏观经济下行或导致消费二次探底，科技产业变革及新技术的落地节奏或将受到影响。

新技术渗透进度不及预期。我们基于 AI、区块链等各领域技术持续渗透的假设进行分析。如果新技术渗透进度不及预期，各细分市场规模可能低于估算水平，导致相同竞争格局下，公司业绩低于预期水平。

中美贸易摩擦加剧。随着美国 BIS 宣布对中国半导体限制升级，全球科技产业链供给与需求端均产生了较大不确定性。如果中美贸易摩擦激化，短期内国内科技行业公司可能会受到影响，部分公司或面临终端需求萎缩或合作受阻导致业绩增长不及预期的风险。

本研报中涉及到未上市公司或未覆盖个股内容，均系对其客观公开信息的整理，并不代表本研究团队对该公司、该股票的推荐或覆盖。

更多免费行业报告 并购家 www.ipoipo.cn

提及公司列表

图表381：提及公司列表

代码	公司	代码	公司
MSFT US	Microsoft	600487 CH	亨通光电
GOOG US	谷歌	600522 CH	中天科技
NVDA US	英伟达	002415 CH	海康威视
META US	Meta	002236 CH	大华股份
AMZN US	亚马逊	HPQ US	惠普
9888 HK/BIDU US	百度	DELL US	戴尔
9988 HK/BABA US	阿里巴巴	2382 TW	广达
700 HK	腾讯	6669 TW	纬颖
20 HK	商汤	2356 TW	英业达
002230 CH	讯飞	8210 TW	勤诚兴业
9888 HK/BIDU US	百度	4062 JP	Ibiden
9988 HK/BABA US	阿里巴巴	7120 JP	Shinko
700 HK	腾讯	ROG US	Rogers
20 HK	商汤-W	NVDA US	英伟达
300364 CH	中文在线	AMD US	AMD
601858 CH	中国科传	INTC US	英特尔
000681 CH	视觉中国	5274 TW	ASPEED
002415 CH	海康威视	MPWR US	MPS
002236 CH	大华股份	2308 TW	台达电
600050 CH/762 HK	中国联通	STM US	意法半导体
601728 CH/728 HK	中国电信	COHR US	Coherent
688325 CH	经纬恒润	LITE US	Lumentum
002920 CH	德赛西威	COHR US	Coherent
9866 HK/NIO US	蔚来	AVGO US	博通
9868 HK/XPEV US	小鹏汽车	601869 CH	长飞光缆
2015 HK/LI US	理想汽车	未上市	Bloomberg
300033 CH	同花顺	未上市	OpenAI
300059 CH	东方财富	META US	Meta
GOOG US	谷歌	GOOGL US	谷歌
MSFT US	微软	9888 HK	百度集团-SW
META US	Meta	601360 CH	三六零
APX AU	Appen	002230 CH	科大讯飞
TIXT CN	Telus international	300418 CH	昆仑万维
未上市	Open AI	0020 HK	商汤科技
未上市	华为	688375 CH	国博电子
未上市	智谱华章	001270 CH	铖昌科技
未上市	MiniMax	002465 CH	海格通信
未上市	第四范式	300342 CH	天银机电
未上市	聆心智能	688418 CH	震有科技
未上市	澜舟科技	688387 CH	信科移动
未上市	云知声	601698 CH	中国卫通
未上市	扩博智能	688066 CH	航天宏图
未上市	标贝科技	688568 CH	中科星图
未上市	慧听科技	603236 CH	移远通信
未上市	Stability AI	RKLB NASDAQ	Rocket Lab
未上市	OpenAI	AMZN NASDAQ	Amazon
未上市	它思科技	QCOM NASDAQ	Qualcomm
688206 CH	概伦电子	AAPL NASDAQ	Apple
002415 CH	海康威视	ASTS NASDAQ	AST SpaceMobile
002410 CH	广联达	TMUS NASDAQ	T-mobile
301269 CH	华大九天	未上市	蓝箭航天
600845 CH	宝信软件	未上市	天兵科技
688777 CH	中控技术	未上市	深蓝航天
688111 CH	金山办公	未上市	中国星网
600588 CH	用友网络	未上市	银河航天
688396 CH	致远互联	未上市	星辰空间
9888 HK	百度集团	未上市	航天科工火箭技术有限公司
2121 HK	创新奇智	未上市	中科天塔
MSFT US	微软	未上市	星启空间
NVDA US	英伟达	未上市	极光星通
CDNS US	Cadence	未上市	星际荣耀
BMW GR	BMW（宝马）	未上市	星河动力
未上市	第四范式	未上市	中科宇航
未上市	旷视科技	未上市	凌空天行
未上市	Lost Lore	未上市	微纳星空
未上市	Midjourney	未上市	长光卫星

免责

更多免费行业报告 并购家 www.ipoipo.cn

代码	公司	代码	公司
未上市	GitHub	未上市	时空道宇
未上市	OpenAI	未上市	工大卫星
700 HK	腾讯	未上市	航天驭星
9999 HK	网易-S	未上市	华为
MSFT US	微软	未上市	SpaceX
AMZN US	亚马逊	未上市	Relativity Space
9988 HK	阿里巴巴-SW	未上市	Blue Origin
9888 HK	百度集团-SW	未上市	OneWeb
300624 CH	万兴科技	未上市	Lynk
未上市	Adspert	未上市	Bloomberg
300592 CH	华凯易佰	未上市	OpenAI
未上市	华为	META US	Meta
JD US	京东	MSFT US	微软
002315 CH	焦点科技	GOOGL US	谷歌
GOOGL US	谷歌	9888 HK	百度集团-SW
300133 CH	华策影视	601360 CH	三六零
AAPL US	Apple	002230 CH	科大讯飞
HUBS US	HubSpot	300418 CH	昆仑万维
未上市	Jasper AI	0020 HK	商汤科技
未上市	DeepMind	未上市	BitFury
未上市	海马轻帆	未上市	Yahoo
NFLX US	Netflix	未上市	altavista
未上市	中影基地	GOOGL US	Google
002230 CH	科大讯飞	9888 HK	Baidu
未上市	AIVA	SOHU US	Sohu
未上市	Runway	未上市	Sina
ADBE US	Adobe	9999 HK	NETEASE
未上市	聚力维度	0700 HK	Tencent
未上市	DeepMusic	MSFT US	Microsoft
未上市	Snowfox Books	未上市	Mozilla
300364 CH	中文在线	SPOT US	Spotify
688111 CH	金山	UBER US	Uber
未上市	飞书	AMZN US	Amazon
SONY US	索尼	3690 HK	美团
PDD US	拼多多	未上市	字节跳动
EBAY US	ebay	1024 HK	快手
688981 CH	中芯国际	未上市	Ethereum
688256 CH	寒武纪	未上市	Binance
688041 CH	海光信息	未上市	Polkadot
688047 CH	龙芯中科	未上市	OpenSea
603893 CH	瑞芯微	未上市	Avalanche
688200 CH	华峰测控	未上市	Uniswap
300480 CH	光力科技	COIN US	Coinbase
600584 CH	长电科技	未上市	AAVE
002156 CH	通富微电	未上市	Rare Bits
002463 CH	沪电股份	未上市	Compound
600183 CH	生益科技	TSLA US	Tesla
300476 CH	胜宏科技	META US	Meta
002916 CH	深南电路	AAPL US	Apple
002436 CH	兴森科技	9988 HK	Alibaba
300308 CH	中际旭创	未上市	Optimism
300394 CH	天孚通信	未上市	STARKWARE
300502 CH	新易盛	未上市	LOOPRING
000988 CH	华工科技	未上市	MATTER Labs
688498 CH	源杰科技	未上市	Vechain
000938 CH	紫光股份	未上市	Fetch.ai
000063 CH	中兴通讯	未上市	X2Y2
301165 CH	锐捷网络	未上市	Immutable X
601138 CH	工业富联	未上市	LooksRare
0992 HK	联想	未上市	Rarible
688515 CH	裕太微	未上市	FTX
002837 CH	英维克	未上市	CoinDesk
		未上市	Celsius Network

资料来源：Bloomberg，华泰研究

免责
更多免费行业报告 并购家 www.ipoipo.cn

免责声明

分析师声明

本人，黄乐平、谢春生、朱珺、余熠、陈旭东、丁宁、张皓怡，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构(以下统称为“华泰”)对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人(无论整份或部分)等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作,在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题,请与华泰金融控股（香港）有限公司联系。

更多免费行业报告 并购家 www.ipoipo.cn

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 中控技术（688777 CH）、华峰测控（688200 CH）、中际旭创（300308 CH）、英维克（002837 CH）、亨通光电（600487 CH）、海格通信（002465 CH）、航天宏图（688066 CH）、中科星图（688568 CH）、菲利华（300395 CH）、正帆科技（688596 CH）、TCL 中环（002129 CH）、晶瑞电材（300655 CH）、路维光电（688401 CH）、鼎龙股份（300054 CH）、芯源微（688037 CH）、中微公司（688012 CH）、拓荆科技（688072 CH）、华海清科（688120 CH）：华泰金融控股（香港）有限公司、其子公司和/或其关联公司实益持有标的公司的市场资本值的1%或以上。
- 沪硅产业（688126 CH）、盛美上海（688082 CH）、中微公司（688012 CH）：华泰金融控股（香港）有限公司、其子公司和/或其关联公司在本报告发布日担任标的公司证券做市商或者证券流动性提供者。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure

其他信息请参见下方 **“美国-重要监管披露”**。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师黄乐平、谢春生、朱珺、余熠、陈旭东、丁宁、张皓怡本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 中控技术（688777 CH）、科大讯飞（002230 CH）、神工股份（688233 CH）、江化微（603078 CH）、机器人（300024 CH）、禾赛（HSAI US）：华泰证券股份有限公司、其子公司和/或其联营公司在本报告发布日之前的 12 个月内担任了标的证券公开发行或 144A 条款发行的经办人或联席经办人。
- 中控技术（688777 CH）、华特气体（688268 CH）、江化微（603078 CH）、禾赛（HSAI US）、电连技术（300679 CH）：华泰证券股份有限公司、其子公司和/或其联营公司在本报告发布日之前 12 个月内曾向标的公司提供投资银行服务并收取报酬。
- 中控技术（688777 CH）、禾赛（HSAI US）：华泰证券股份有限公司、其子公司和/或其联营公司预计在本报告发布日之后 3 个月内将向标的公司收取或寻求投资银行服务报酬。
- 中控技术（688777 CH）、华峰测控（688200 CH）、中际旭创（300308 CH）、英维克（002837 CH）、亨通光电（600487 CH）、海格通信（002465 CH）、航天宏图（688066 CH）、中科星图（688568 CH）、菲利华（300395 CH）、正帆科技（688596 CH）、TCL 中环（002129 CH）、晶瑞电材（300655 CH）、路维光电（688401 CH）、鼎龙股份（300054 CH）、芯源微（688037 CH）、中微公司（688012 CH）、拓荆科技（688072 CH）、华海清科（688120 CH）：华泰证券股份有限公司、其子公司和/或其联营公司实益持有标的公司某一类普通股证券的比例达1%或以上。
- 沪硅产业（688126 CH）、盛美上海（688082 CH）、中微公司（688012 CH）：华泰证券股份有限公司、其子公司和/或其联营公司在本报告发布日担任标的公司证券做市商或者证券流动性提供者。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。
- 本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽，亦不试图促进购买或销售该等证券。如任何投资者为美国公民、取得美国永久居留权的外国人、根据美国法律所设立的实体（包括外国实体在美国的分支机构）、任何位于美国的个人，该等投资者应当充分考虑自身特定状况，不以任何形式直接或间接地投资本报告涉及的投资者所在国相关适用的法律法规所限制的企业的公开交易的证券、其衍生证券及用于为该等证券提供投资机会的证券的任何交易。该等投资者对依据或者使用本报告内容所造成的一切后果，华泰证券股份有限公司、华泰金融控股（香港）有限公司、华泰证券（美国）有限公司及作者均不承担任何法律责任。

免责

更多免费行业报告 并购家 www.ipoipo.cn

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期

（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15%以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15%以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国：华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J

香港：华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：AOK809

美国：华泰证券（美国）有限公司为美国金融业监管局（FINRA）成员，具有在美国开展经纪交易商业务的资格，经营业务许可编号为：CRD#:298809/SEC#:8-70231

华泰证券股份有限公司

南京

南京市建邺区江东中路 228 号华泰证券广场 1 号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同 28 号太平洋保险大厦 A 座 18 层/邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区益田路 5999 号基金大厦 10 楼/邮政编码：518017

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路 18 号保利广场 E 栋 23 楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com

华泰金融控股（香港）有限公司

香港中环皇后大道中 99 号中环中心 58 楼 5808-12 室

电话：+852-3658-6000/传真：+852-2169-0770

电子邮件：research@htsc.com

http://www.htsc.com.hk

华泰证券（美国）有限公司

美国纽约公园大道 280 号 21 楼东（纽约 10017）

电话：+212-763-8160/传真：+917-725-9702

电子邮件: Huatai@htsc-us.com

http://www.htsc-us.com

©版权所有2023年华泰证券股份有限公司

更多免费行业报告 并购家 www.ipoipo.cn