



「中国科技 燃」

2026年诺安基金科技投资报告



【风险提示】

本材料由诺安基金管理有限公司提供，报告中提及的市场观点是基于目前市场情况分析得出、具有时效性。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，亦不作为任何法律文件。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。材料中提及的上市公司仅供举例说明，不代表个股推荐和具体投资建议。

本报告中图片及商标仅为示例引用，不构成任何商业推荐或品牌关联。

市场有风险，投资需谨慎。投资者不应将本报告作为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向专业人士咨询并谨慎决策。

投资者投资于本公司管理的基金时，应认真阅读《基金合同》《托管协议》《招募说明书》《风险说明书》、基金产品资料概要等文件及相关公告，如实填写或更新个人信息并核对自身的风险承受能力，选择与自己风险识别能力和风险承受能力相匹配的基金产品。投资者需要了解基金投资存在可能导致本金亏损的情形。基金管理人承诺以诚实信用、勤勉尽责的原则管理和运用基金资产，但不保证基金一定盈利，也不保证最低收益。基金管理人管理的其他基金的业绩不代表本基金业绩表现。基金的过往业绩及其净值高低并不预示其未来业绩表现。基金管理人提醒投资者基金投资的“买者自负”原则，在作出投资决策后，基金运营状况与基金净值变化引致的投资风险，由投资者自行承担。我国基金运作时间较短，不能反映股市发展的所有阶段。

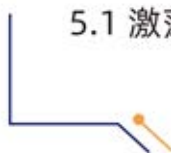
目录 CONTENTS

一、燃之势能

应用渗透：在确定性与效率边缘重构商业秩序	03
1.1 认知重构：聚焦应用价值，爆发趋势确定	03
1.2 路径思辨：从技术理想主义走向商业现实主义	04
1.2.1 范式分歧：在不确定的演进中锚定应用起点	04
1.2.2 落地共识：挺进应用深水区的工程化挑战	05
1.3 价值探索：以数据和生态为抓手的确定性应用图谱	05
1.3.1 逻辑基石：数据、生态与模型能力的阶段性协同	05
1.3.2 看好方向一：AI for Business，微笑曲线右端的营销与交付变革	06
1.3.3 看好方向二：AI for Science，微笑曲线左端的研发范式重构	08
1.3.4 看好方向三：智能助手，从指令执行向 Agent 任务协同的跨越	09
1.3.5 看好方向四：自动驾驶，物理空间智能闭环的终局一跳	09
1.4 前瞻洞察：迈向由业务定义价值的新纪元	16
端侧AI的关键转折：硬件重构与场景价值跃迁	17
2.1 AI眼镜多路线并行，手机与PC正突破性能与形态的物理瓶颈	18
2.2 新兴品类的爆发：AI硬件正剥离极客感，转向健康管理、情感计算与儿童成长的垂直领域	21
AI基础设施：从芯芯之火到全面燎原	25
3.1 算力芯片：通用GPU与专用ASIC 从竞争走向竞合，CPU受益推理拉动迎新市场	25
3.1.1 通用GPU与专用ASIC 从竞争走向竞合	25
3.1.2 AI Agent时代，CPU正从台后走向C位	28
3.2 通信：追求高集成度下的极致性能	29

3.2.1 数据中心间 (Scale-across): DCI系统及长距离光纤迎需求爆发	30
3.2.2 超节点间 (Scale-out): 1.6T高速率光/硅光模块成为主流	31
3.2.3 超节点内 (Scale-up): CPO/NPO、OCS技术从柜间互联向柜内渗透, PCB仍是短距离互联主流技术	33
3.3 存储: AI发展推动存储架构变革, 供需错配下存储迎大周期	36
3.3.1 AI推理需求推动存储架构变革: 突破内存墙, 向异构存储范式演进	36
3.3.2 存储架构变革下的底层裂变: 3D DRAM的4F ² 工艺和千层NAND	39
3.3.3 供需错配驱动, 存储步入上行大周期	41
中国半导体的决定性突破: 全链协同, 从被动替代到创新自强	43
4.1 被动替代: 封锁围堵下, 中国成功度过半导体自主自强上半场	43
4.1.1 人工智能发展强化半导体战略地位, 外部技术封锁成为国产替代核心驱动力	43
4.1.2 多维协同发力, 保障产业发展	44
4.1.3 自主可控显现成效, 国内半导体产业规模和韧性跨越式增强	45
4.2 主动创新: 实现决定性技术攻坚, 闭环中国人工智能产业生态	45
4.2.1 “十五五”规划明确全链条突破, 大基金三期注入动力	45
4.2.2 本土化工艺制程突破, 逐步构建自主可控AI芯片制造体系	46
4.2.3 “国芯+国模+国用”促国产AI芯片“可用”到“好用”	48
4.2.4 存储芯片在AI时代战略意义凸显, 国产存储技术和产能提速追赶	49
4.2.5 迈向下个十年, 计算底层技术演变提供换道超车机遇	49
二、燃动未来	
人形机器人: 具身智能的“iPhone时刻”	53
1.1 产业拐点: 从宏观叙事看周期定位	53
1.2 标杆引领: 特斯拉Optimus量产进程	54
1.3 技术攻坚: 核心博弈与关键增量	56

1.4 群雄逐鹿：中国新势力与商业化爆发	58
脑机接口：从科幻到产业化临近	61
2.1 产业链生态图谱	61
2.2 脑机接口发展空间	63
2.3 BCI的“真正生产力”来自AI	64
2.4 脑机接口与具身智能	64
2.5 核心壁垒、驱动因素、难点与风险、未来演进方向	65
太空新经济：中美太空竞速与“天基算力”新基建	67
3.1 太空经济：不可再生轨道资源与大国安全防线的终极博弈	67
3.2 终极愿景：硅谷巨头眼中人类文明的星际愿景	68
3.3 算力上天：AI浪潮下数据中心向宇宙迁移的范式革命	69
3.4 运力鸿沟：SpaceX统治力下的效率极限与中国产能突围	71
可控核聚变：人类能源的终极解决方案	73
4.1 现有能源体系无法支撑人类长期发展	73
4.1.1 化石能源面临气候和储量双重束缚	73
4.1.2 风光出力不稳定，大规模装机对电网提出较大挑战	74
4.1.3 核裂变面临铀资源和乏燃料处理双重难题	74
4.2 可控核聚变是未来能源的理想解决方案	75
4.2.1 可控核聚变技术原理	75
4.2.2 可控核聚变已成大国战略竞争的新高地	77
4.3 政策和资本双轮驱动，产业浪潮澎湃向前	78
4.3.1 聚变正引爆一个前所未有的全新生态链	78
4.3.2 政策和资本双轮驱动，中国企业奋勇争先	79
中国创新药：“燃烧的远征”	80
5.1 激荡十年：政策、生态与管线爆发（2015-2025）	80



5.1.1 政策顶层设计：从“审评积压”到“全链条支持”	80
5.1.2 “VC+IP+CXO”的生态不断强化	81
5.1.3 管线爆发：数量与质量的双重跃迁	82
5.1.4 中国资产的全球定价	82
5.2 中国药企的“比较优势”与战略机遇	83
5.2.1 核心能力：效率与成本的极致	83
5.2.2 历史机遇：填补MNC的专利悬崖	84
5.3 未来十年：全球化兑现与大国竞合	84
5.3.1 重磅炸弹的全球上市潮	84
5.3.2 2040年市场空间展望	85
5.3.3 中国商保大发展与多层次支付	85
5.3.4 中美共同引领前沿技术	85







2026年

诺安基金科技投资报告

燃

之势能

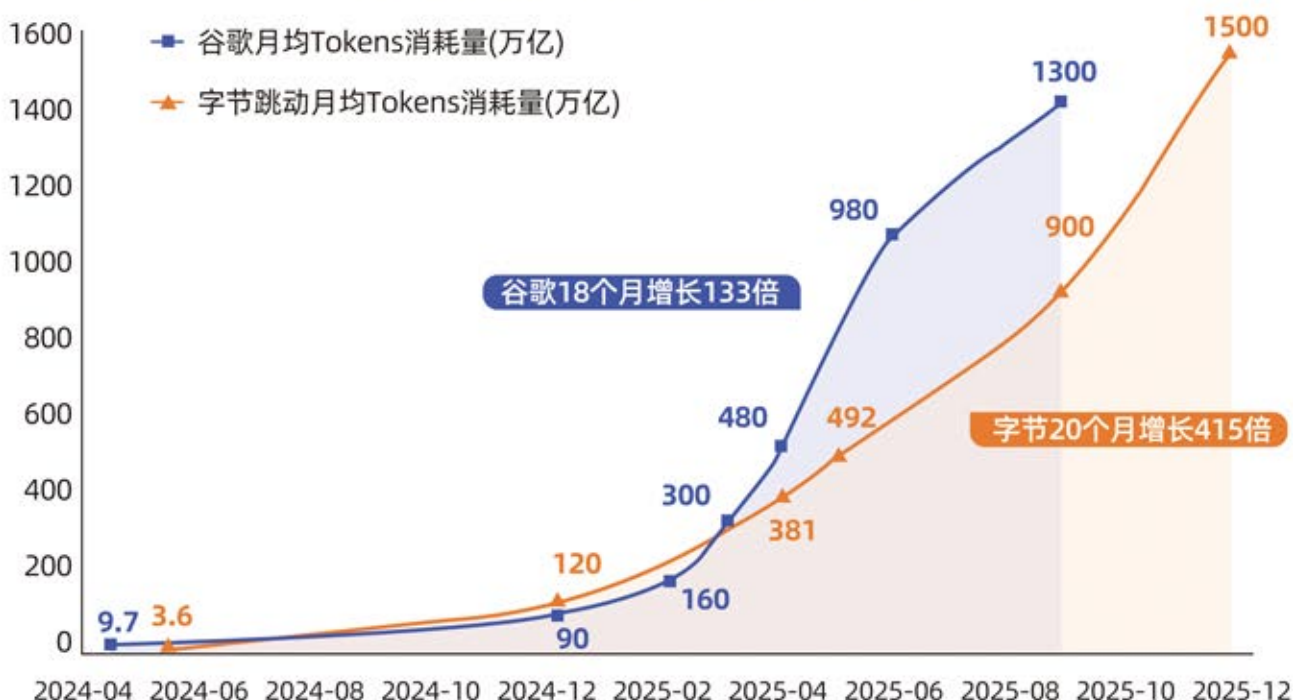
01

应用渗透 在确定性与效率边缘重构商业秩序

1.1 认知重构：聚焦应用价值，爆发趋势确定

当前AI产业正站在技术理想与商业现实的十字路口。尽管关于通用人工智能（AGI）的实现路径、底层架构的演进上限以及Scaling Law的边际效应仍存在激烈的理论博弈，同时，高质量公开数据已近枯竭、巨大的能耗负担、底层架构的计算效率上限以及模型在严谨场景下的“幻觉”与记忆缺失等问题，也均构成了智能向上突破的实质性限制，但我们认为智能的价值最终需通过ROI和利润来体现。当前，国内外模型Tokens调用量保持高速增长，反映出底层算力正加速转化为前端应用消耗。本章将剖析当前AI应用落地的路径分歧与共识，并基于数据与生态的逻辑，推演最具增长确定性的应用蓝图。

图：国内外模型Tokens调用量持续增长



资料来源：谷歌、火山引擎

1.2 路径思辨：从技术理想主义走向商业现实主义

» 1.2.1 范式分歧：在不确定的演进中锚定应用起点

目前业界对于AI未来的分歧点集中在AGI达成期限、架构演进（如Transformer是否为终点）以及Scaling Law的有效性上。但在应用层面，共识已经达成：多数商业场景并不依赖全知全能的AGI，而是追求在特定领域以低于人工成本的代价完成生产关系重构。与此同时，底层技术架构的分歧同样显著：Transformer架构很可能不是智能进化的终点，但它已无疑成为了应用的起点。Scaling Law正从单纯的“规模线性放大”转向由“架构创新与训练方式”双轮驱动的结构化扩展。在这种背景下，模型与应用的关系正在经历层级化演进，模型在短期作为单点效率工具，中期重构复杂业务流程，长期则有望进化为独立决策的主体。

表：当前AI核心分歧梳理

分歧话题	支持方	反对方
当前路径下AGI能否达成？	Ilya Sutskever (前OpenAI首席科学家) “GPT-4已接近人类水平的理解能力，只需优化数据和计算，AGI将在5-10年内实现。”	Yann LeCun (前Meta首席AI科学家) “LLM无法解决因果推理，它们只是复杂的模式匹配器。AGI需要世界模型(world models),至少需要20-30年。”
Transformer是否为生成式AI最终架构？	Andrej Karpathy (前OpenAI科学家/Tesla自动驾驶和AI负责人) “Transformer是LLM的正确架构。工具调用(tool use)和agent系统将使LLM真正实用化。”	Yann LeCun (前Meta首席AI科学家) “LLM是死胡同--它们依赖海量数据却易产生幻觉。我们需要JEPA (Joint Embedding Predictive Architecture)来学习世界模型。”
Scaling Law是否仍然有效？	Greg Brockman (OpenAI联合创始人兼总裁) “传统的Scaling Law可能在某些方面受到挑战，但推理时扩展(inference-time scaling)开启了第二增长曲线，给模型更多时间思考，它会变得更聪明”	Andrew Ng 吴恩达 (前Google Brain负责人/前百度首席科学家) “与其无休止地增加模型大小，不如转以数据为中心(Data-centric)。现在的瓶颈不是我们没有算力，而是高质量数据被用光了。”
模型是否会吞噬应用？	Andrej Karpathy (前OpenAI科学家Tesla自动驾驶和AI负责人) “模型的能力边界就是应用的边界。模型即一切，应用侧只是它的一种表现形式。”	Demis Hassabis (Google Deepmind CEO) “AI的真正考验在于它能否解决现实世界的问题，比如蛋白质折叠或核聚变控制。这是应用在引领模型的研究方向”

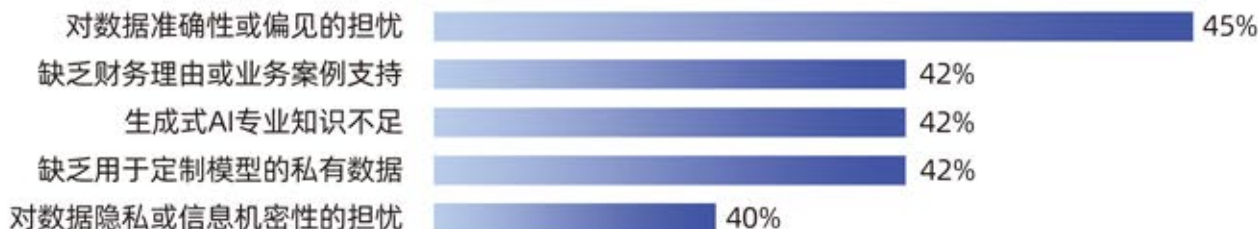
资料来源：公开资料收集

» 1.2.2落地共识：挺进应用深水区的工程化挑战

在应用落地层面，产业界已形成多项共识。首先是高质量数据的稀缺，公开域数据的利用已接近上限，未来的增长空间取决于私有域及具备逻辑因果属性的垂类数据。其次是ROI（投资回报率）的逻辑转变。虽然API价格在持续下降，但推理成本仍是企业决策的关键考量。更核心的变化在于，ROI的倍增不再仅依赖于缩减成本（分母），而是通过AI实现业务价值的增量创造（分子）。这种从“小幅优化”向“业务价值创造”的跨越，决定了AI能否真正渗透进企业的核心业务流。多智能体（Multi-Agent）系统正成为解决复杂业务逻辑的工程化公约数。与单模型输出不同，多智能体架构通过角色分工与协同决策，能够有效拆解长链业务流程，在容错率、专业深度和逻辑闭环上实现了质的突破，是AI渗透进核心生产环节的关键桥梁。

此外，产业界对AI局限性的认识更加理性。幻觉问题、长期记忆缺失以及连续反馈闭环的不足，仍是金融、医疗等严谨场景落地的障碍。目前，行业正通过产品形态的重构与工程化手段来弥补模型能力的短板。但需要注意的是，高密度的工程投入并非唯一的解决路径。随着Test-timeScaling（推理时扩展）以及边推理边学习等技术的落地，模型有望在推理阶段通过增加计算量直接提升逻辑表现，从而在底层架构层面缓解对复杂工程补丁的依赖。

图：数据/财务/幻觉/缺乏工程及产品型人才均为商业落地核心共识困难



资料来源：IBM《2025年AI五大落地障碍》

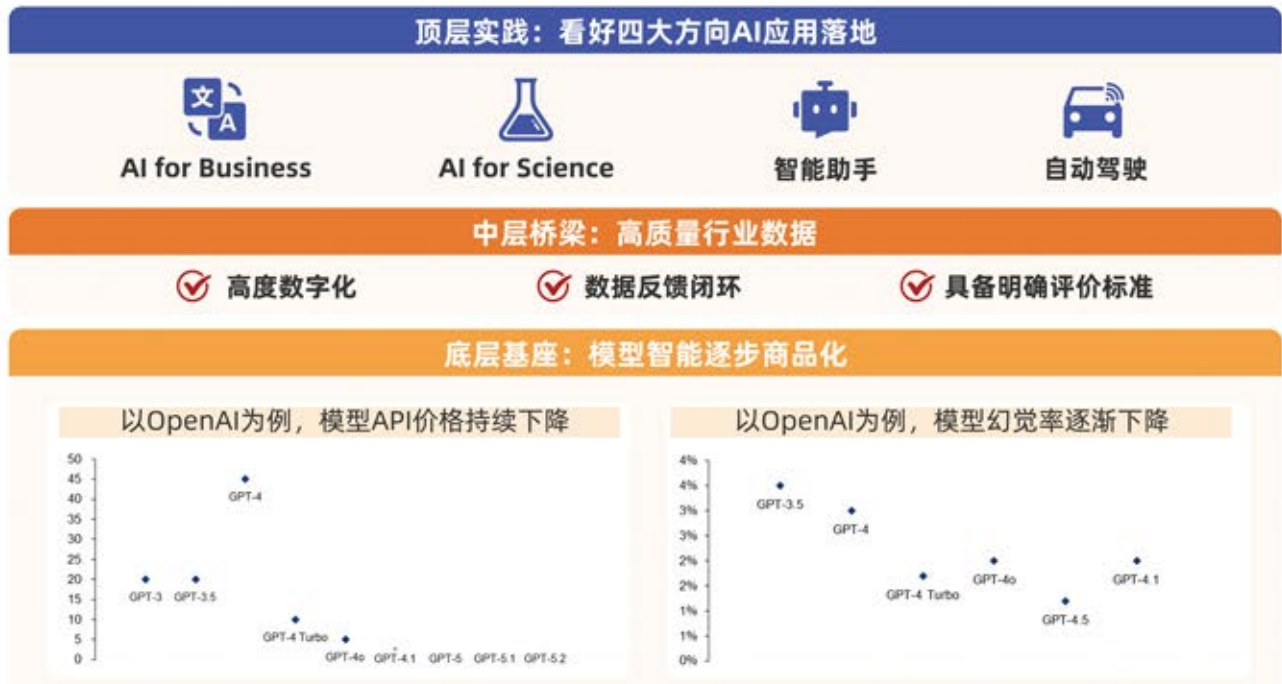
1.3 价值探索：以数据和生态为抓手的确定性应用图谱

» 1.3.1逻辑基石：数据、生态与模型能力的阶段性协同

寻找最具爆发力的AI应用，需基于数据、生态与模型能力的动态关系。展望未来，随着模型能力的演进，若自我学习能力实现突破等，对数据的依赖可能降低。但现阶段，拥有高质量行业数据及深度交互场景的方向，更易跨越从技术原型到业务流的门槛。具备这一特征的方向通常表现为：原流程本身已高度数字化、数据反馈闭环清晰且具备明确

的评价标准。基于此逻辑，我们看好AI for Business、AI for Science、智能助手与自动驾驶四个方向成为率先落地的AI应用。

图:以“数据与生态”为抓手寻找AI应用落地方向



资料来源：OpenAI、Vectara

» 1.3.2 看好方向一：AI for Business，微笑曲线右端的营销与交付变革

在商业应用侧，AI正在重构营销、广告等环节的生产力逻辑，这一方向位于微笑曲线的右端（销售与服务）。其核心价值不再是简单的“模版化生成”，而是生产逻辑的根本转变。通过降低专业劳动的准入门槛，AI能够替代原本需要多年专业教育才能胜任的复杂任务，实现交付速度与规模的数量级提升。在代码研发、法律、教育等领域，这种效率溢出已实现了从辅助工具向生产力核心的转化。

在生成式AI的首波红利中，结构化内容的产出正经历一场“工业化革命”。在代码研发、法律合同、教育课件及营销短文案等领域，内容的本质并非天马行空的灵感，而是基于特定逻辑框架的精准排布。AI在这些场景中的表现，精准击中了商业痛点：显著降低了专业劳动的准入门槛，实现了交付速度与产出规模的数量级提升。这种“模版化生成”的能力，使得原本昂贵的、依赖个体经验的精英工作，转化为可大规模复制的标准化流水线。例如，在代码领域，AI已实现了从“辅助补全”向“系统重构”的进化；在营销领域，千人千面的超个性化文案已成为标配。这种效率溢出的本

质，是AI利用了人类历史上沉淀的最具逻辑性的结构化数据，完成了从“知识检索”向“内容再造”的范式平移。

图：AI覆盖任务所需要的平均受教育年限为14.4年高于全行业任务13.2年

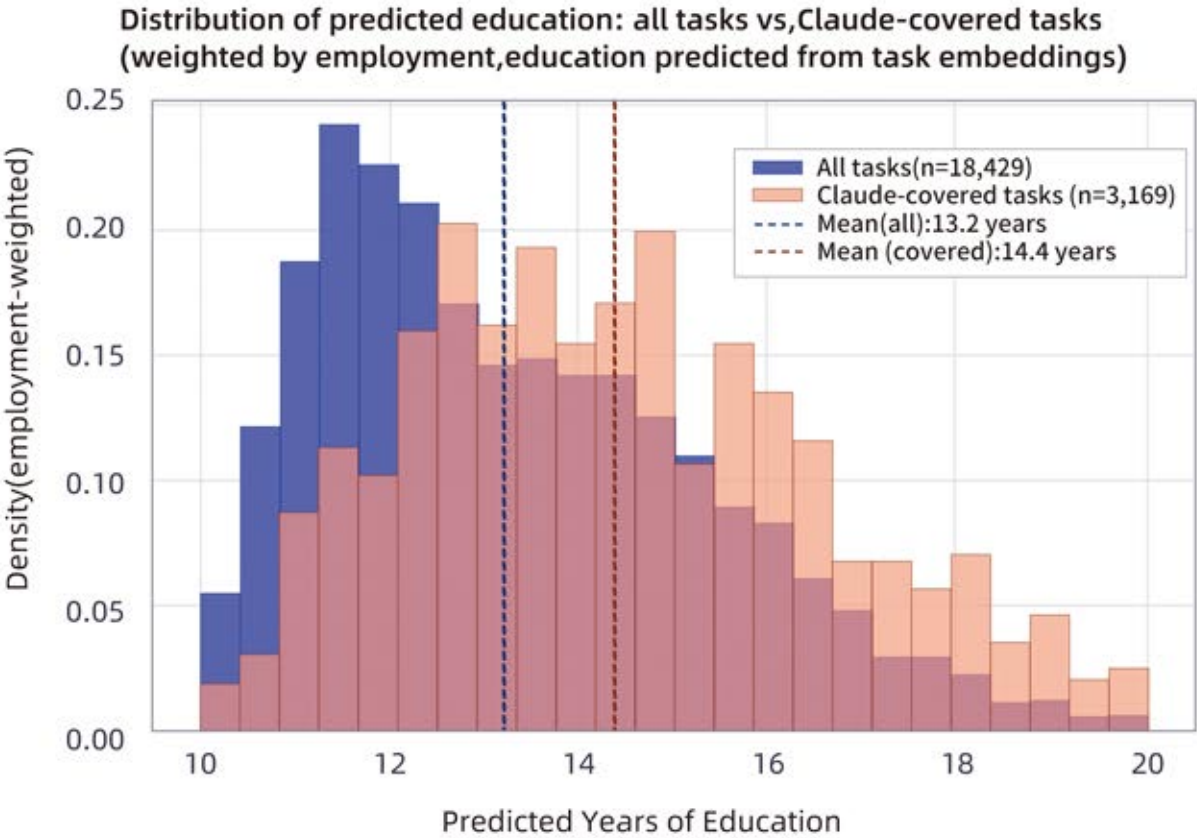


Figure 4.5: Education level of all tasks vs. Claude-covered tasks.
 This shows two histograms, The blue bars give the distribution of the predicted task-level education required for all tasks in the O*NET database, weighted by employment. The orange bars show the same, restricting to tasks that appear in Claude.ai data.

资料来源：Anthropic Economic Index

图：AI for Business在代码、法律、教育及营销领域应用



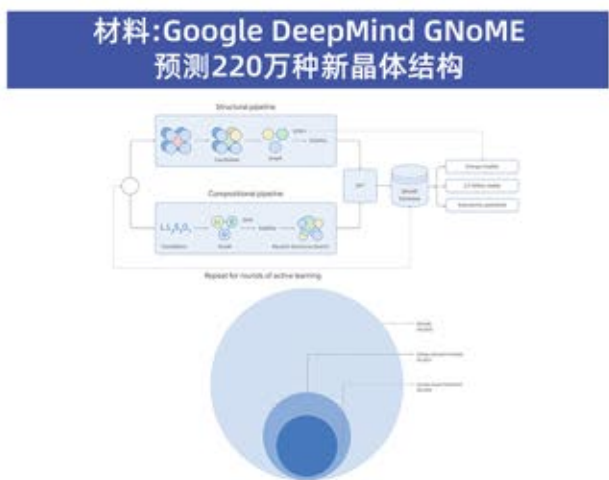


资料来源: Github、Pwc、Harvey、Khan Academy、Jasper

» 1.3.3 看好方向二: AI for Science, 微笑曲线左端的研发范式重构

AI for Science (AI4S) 正在重塑“微笑曲线”左端的研发与创新范式，其核心价值在于利用大模型处理海量科学数据与复杂知识体系的能力，突破人类个体或团队在数据检索与跨学科关联上的效率瓶颈。当前的产业应用重点在于利用其模式识别优势，将传统的“经验试错”转向“计算预测”，在生命科学、材料科学等高维参数空间中解决“大海捞针”式的筛选难题，从而显著缩短分子发现、蛋白质结构预测及合成路径分析的探索周期。然而，真正的AI4S并不止于对现状的描述或数据的辅助解析，其深层变革在于从“分析问题”向具备“自主提出问题”能力的研发主体演进。随着模型对科学规律逻辑理解的加深，研发范式将经历从“人定义课题、AI提供方案”向“AI 基于现有知识体系主动构建科学假设并闭环验证”的跨越。这种从工具性检索向主体性定义的转变，将推动基础科学研究实现指数级的突破，真正开启由AI驱动的自主科学发现时代。

图: AI4S在制药、材料领域应用



资料来源: 英矽智能、Google DeepMind

» 1.3.4 看好方向三：智能助手，从指令执行向 Agent 任务协同的跨越

智能助手的核心演进动力源于 Agent（智能体）架构、RAG（检索增强生成）及长期记忆技术的成熟。通过对用户行为数据的深度建模，助手能够实现跨场景的个性化偏好识别，并基于私有知识库提供精准的业务决策支持。在这种技术框架下，智能助手不再仅是被动响应指令的工具，而是具备了任务拆解与闭环执行能力的协同中枢。这种能力的具象化，正推动人机交互界面从单纯的“信息检索”向主动化的“智能代理服务”平移。

图：AI个人助手应用



资料来源：千问

» 1.3.5 看好方向四：自动驾驶，物理空间智能闭环的终局一跳

一、智驾演进方向：规则驱动转向数据驱动

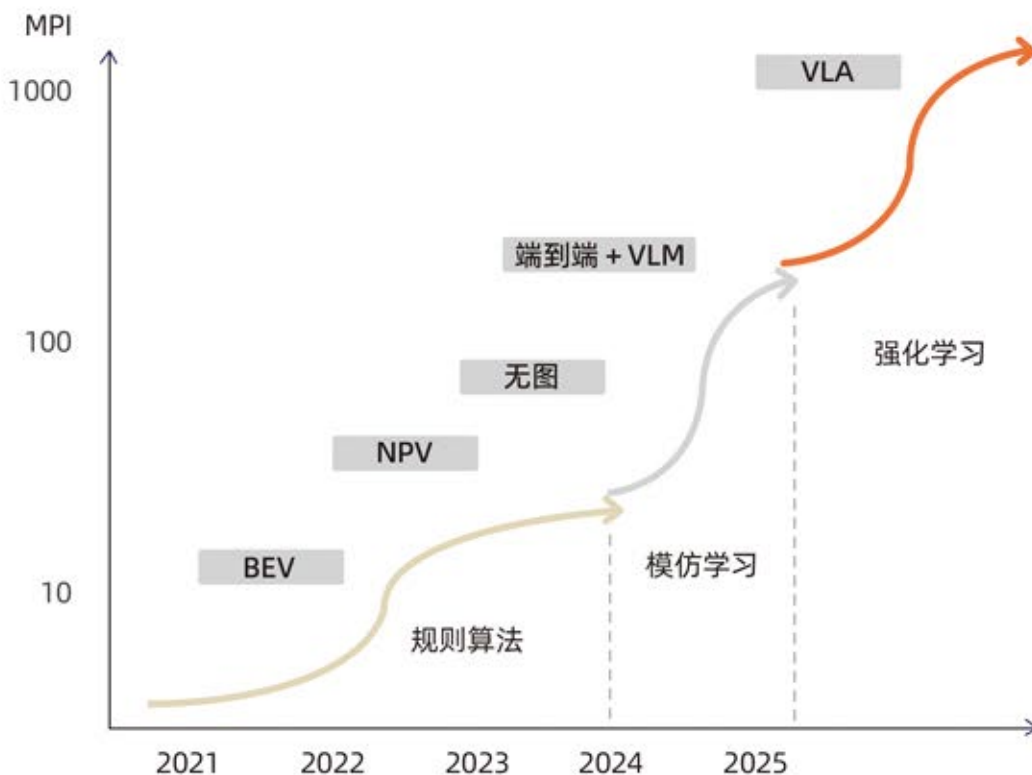
早期自动驾驶主要集中于基于规则的车辆控制，技术上限较低且可扩展性不足，2023年大模型技术开始快速被应用于自动驾驶领域，端到端神经网络架构将传统模块化的感知、决策、控制整合为统一的大模型，大幅提升了系统在复杂场景下的泛化能力和决策效率。

未来智驾技术演进将围绕算法与硬件进行创新：

1) 算法层面，端到端大模型将成为自动驾驶算法的主导方向。特斯拉智驾端到

端算法，基于深度学习，直接从摄像头等传感器数据输入预测车辆控制指令输出，简化了传统自动驾驶系统中复杂的模块化设计流程，提高了系统的响应速度和鲁棒性；Momenta的飞轮大模型创新性地采用“长期记忆+短期记忆”相结合的模式，模拟人脑“直觉推理+逻辑分析”的问题处理机制，大幅降低了大模型的训练成本，使其能够有效解决数以百万计的长尾问题。

图：智驾算法演进过程



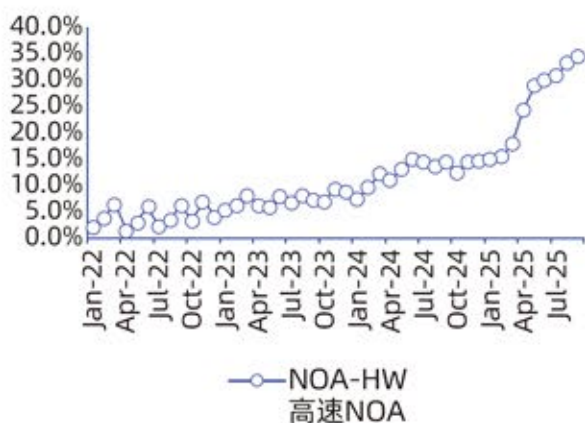
数据来源：诺安基金，智驾网

2) 硬件层面：智算芯片与线控底盘突破。目前主流车型的算力在200 TOPS以下，但未来三年为了支撑全场景城市NOA和L3级自动驾驶，单车芯片算力需求将向500~1000 TOPS甚至更高区间转移。全线控技术落地（转向、制动、悬架），未来的底盘不再由各个子系统（转向、制动、悬架）各自为战，而是通过底盘域控制器进行集中决策。

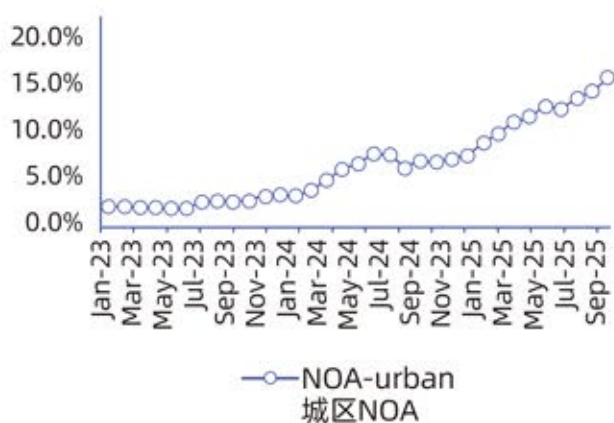
二、行业所处阶段：低阶智驾开始普及，高级智驾迎来元年

全球范围内，中国在自动驾驶行业走在前列，高速NOA渗透率超过40%，城市NOA渗透率已经超过10%，未来5年全球高阶智驾将逐步普及，各国在智驾推进过程中，关键制约因素是技术迭代与监管政策。

图：国内高速NOA渗透率



图：国内城市NOA渗透率



数据来源：高工锂电

技术方面，自动驾驶技术的发展依赖于先进的感知系统、决策算法以及可靠的执行机构，现阶段看，智驾算法已经收敛于端到端神经网络，技术已经不构成制约行业发展的因素。

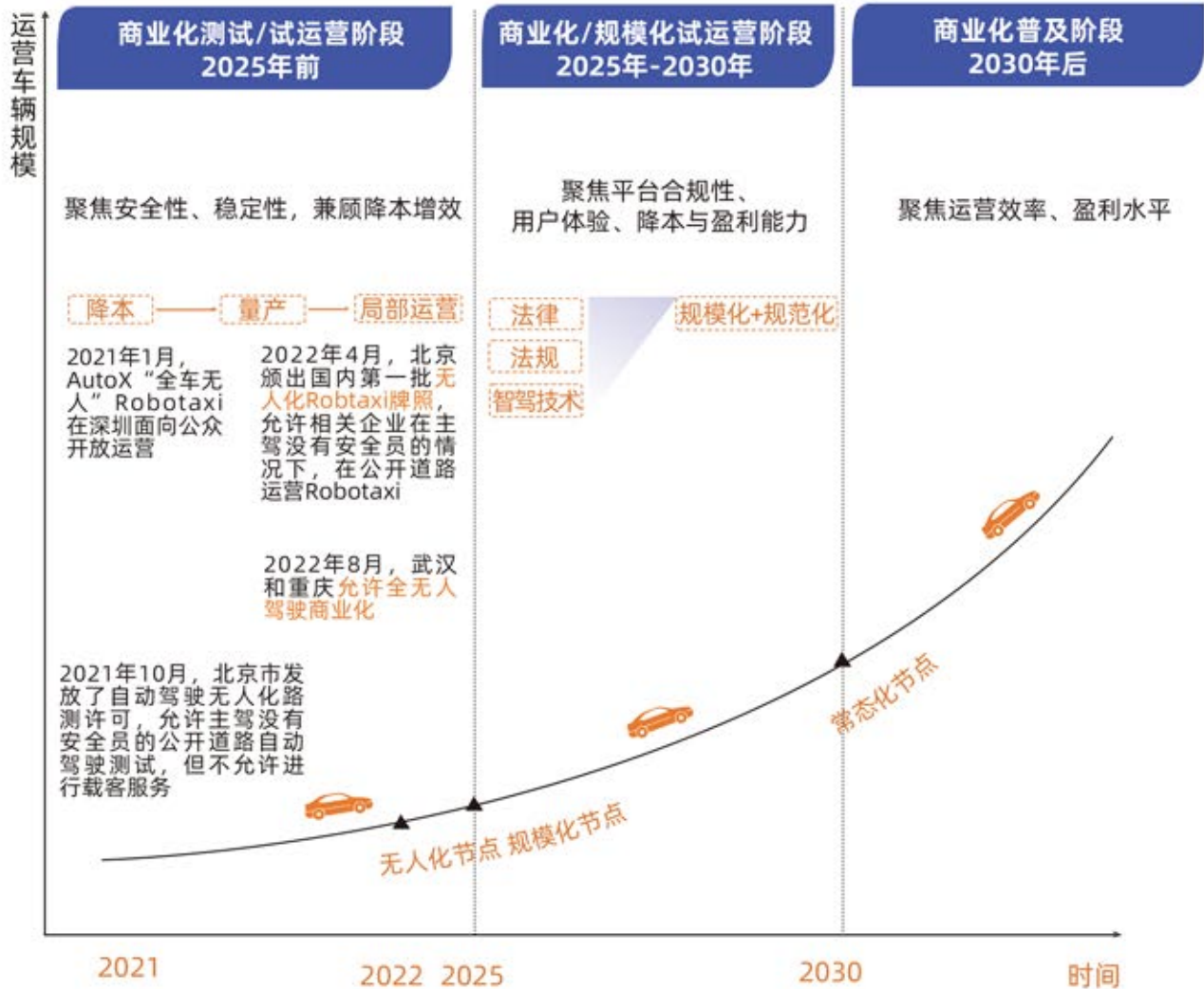
监管方面，UNECE（联合国欧洲经济委员会）和ISO（国际标准化组织）提供了统一框架，美国NHTSA（公路交通安全管理局）、欧盟委员会和中国的工信部发布了具体的指导文件和法规，确保技术的安全性和可靠性。2025年12月，工信部等八部门正式发布《汽车行业稳增长工作方案（2025-2026年）》，明确提出“有条件批准L3级车型生产准入”，标志着中国L3级自动驾驶正式进入商业化阶段。2021年美国NHTSA发布《自动驾驶系统（ADS）6.0：确保美国在自动化技术方面的领导地位》，支持自动驾驶技术的研发项目，包括传感器技术、人工智能算法和车辆控制系统，建立更多的自动驾驶测试设施和试点项目。欧盟UN R171（驾驶员控制辅助系统法规）于2024年9月22日正式生效，它专门针对L2级（SAE L2）辅助驾驶系统，填补了GSR（通用安全法规）在“人机共驾”领域的技术空白。

世界各国监管政策持续利好自动驾驶汽车的商业化，端到端算法为行车安全提供保障，激光雷达、毫米波雷达、摄像头、连接器及车规级芯片成本逐步下降，L2/L2+级自动驾驶汽车渗透率将持续提升，2026年L2级加速普及，L3级开始落地，L4级在特定场景实现商业化，不同场景的渗透率差异将更加明显。

乘用车场景，中国智驾域控制器及激光雷达等硬件市场规模超过1000亿元，2030年城市NOA渗透率将超过70%，L3渗透率约15%；全球市场乘用车智能硬件超过3000亿元，城市NOA渗透率超过40%，L3渗透率个位数。出租车场景，武汉、北京、上海、广州等一线城市将成为主要运营区域，2026年一线城市Robotaxi渗透率有望达到1-2%，全国整体渗透率仍将低于0.5%；欧美国家，Waymo、Tesla等在凤凰城、旧

金山等部分城市实现商业化，Robotaxi占当地网约车出行总量7%。

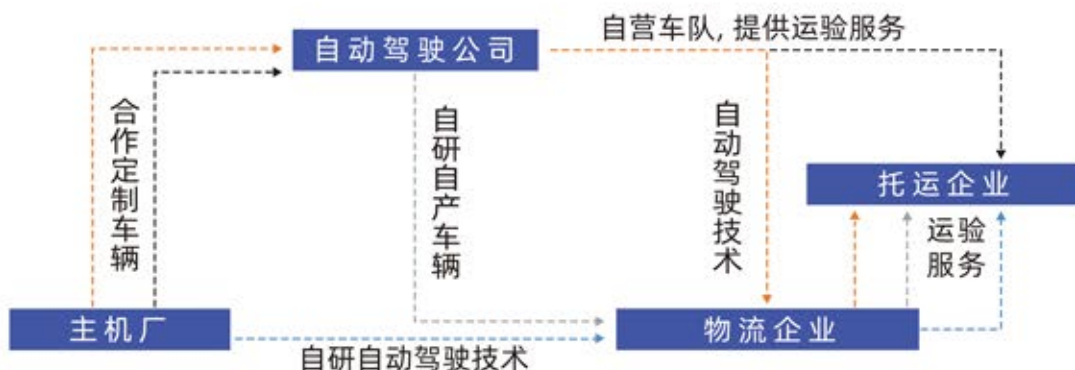
图：中国Robotaxi运营路径与发展节点



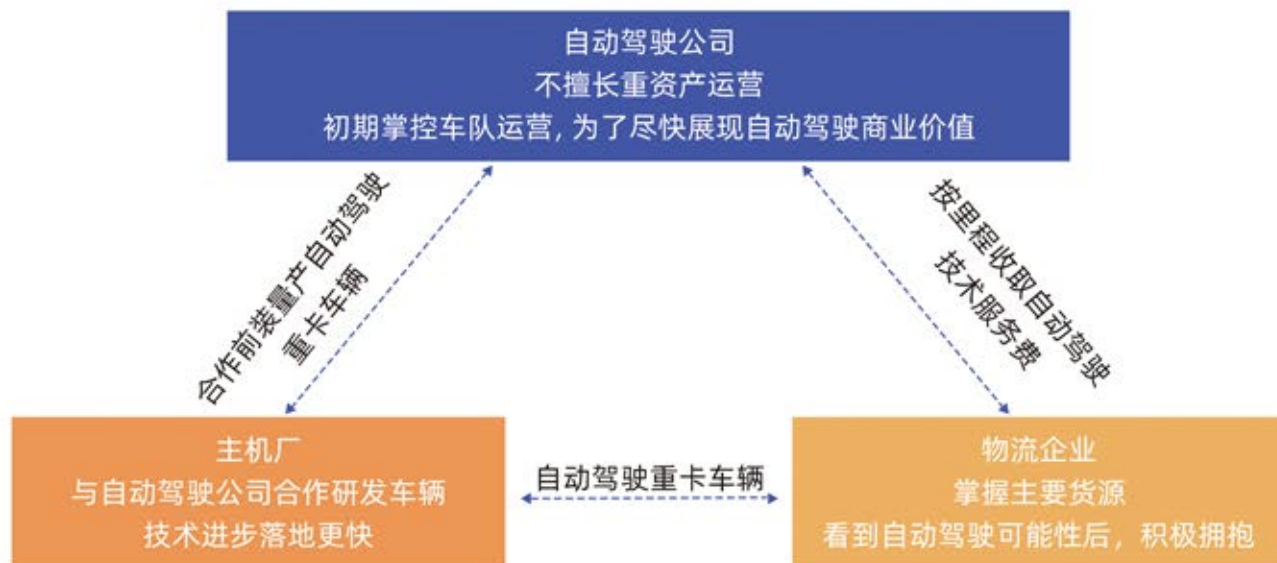
数据来源：盖世汽车，诺安基金

货车场景，中国Robovan+Robotruck潜在市场2500亿元，2030年矿山、港口等封闭场景L4级自动驾驶货车渗透率将超过40%，干线物流L4级自动驾驶渗透率将保持在10%以下，城市配送L4级自动驾驶渗透率将超过30%。公交与环卫车场景，中国Robobus+Robosweeper潜在市场600亿元，L4级自动驾驶公交车将在10个以上城市开展试点，L4级自动驾驶环卫车将在20个以上城市开展试点，未来特定场景渗透率将超过50%。

图：自动驾驶货车商业模式



- > 模式一：“铁三角”模式，主机厂与自动驾驶公司合作开发自动驾驶重卡，将车辆交于物流企业用于提供运输服务
- > 模式二：主机厂与自动驾驶公司合作开发自动驾驶重卡，自动驾驶公司自营车队，提供运输服务
- > 模式三：自动驾驶公司自研自产自动驾驶重卡，向自营车队提供运输服务，或者将车辆交于物流企业用于提供运输服务
- > 模式四：主机厂自研自动驾驶技术，生产自动驾驶重卡，将车辆交于物流企业用于提供运输服务，也可自营车队提供运输服务



铁三角模式综合各方优势，共同促进干线物流自动驾驶产业发展与落地

数据来源：盖世汽车、诺安基金

三、竞争格局：华为领头，三分天下

中国智能驾驶市场已形成三分格局——车企自研、华为方案与Momenta三分天下，这一格局反映了中国企业在智能驾驶领域的差异化发展路径：

1. 车企自研派：以小鹏、理想、比亚迪等车企为代表，通过自研或与科技公司深度合作的方式，开发智能驾驶系统。小鹏，坚持纯视觉路线（BEV+Transformer），2025年战略转型移除了激光雷达，依靠强大的算法弥补硬件差距，高速场景接管率极低，迭代速度快，是15-25万元价格带的智驾标杆；理想，端到端模型+VLM（视觉语言模型），极其擅长处理中国特有的交通场景（如识别中文标识、限时公交车道），在家庭用户最常跑的城市场景中表现稳定；比亚迪，天神之眼A和B套餐合作Momenta，天神之眼C自研，成本控制能力极强，将高阶智驾下探到10-20万元车型。

图：比亚迪天神之眼技术方案对比

	天神之眼A	天神之眼B	天神之眼C
名称	高阶智驾三激光版	高阶智驾激光版	高阶智驾三目版
硬件配置	3颗激光雷达+12颗摄像头+5颗毫米波雷达+12颗超声米波雷达	1-2颗激光雷达+12颗摄像头+5颗毫米波雷达+12颗超声波雷达	12颗摄像头+5颗毫米波雷达+12颗超声波雷达，无激光雷达
算力平台芯片	双Orin-X芯片 (508TOPS)	双Orin-X芯片 (254TOPS)	Orin-N或地平线J6M芯片(84-128TOPS)
主要功能	支持全域无图NOA(城市+高速领航)、夜间极端天气冗余感知、记忆泊车(支持100条路线)，甚至能应对暴雨、浓雾等低能见度场景。	能轻松应对90%的日常通勤，高快领航(高速+快速路)、代客泊车(支持垂直/平行车位)、AEB自动紧急制动(可识别行人、骑行者)。	车道居中保持、全速域ACC自适应巡航、基础AEB(识别车辆)、倒车影像辅助等
搭载车型	仰望U8、仰望U7全系，腾势D9旗舰版	腾势D9 EV、汉EV智驾版、唐DM-i冠军版、海豹07 EV高阶版等	海鸥、海豚、宋Pro DM-i、元UP全系等
覆盖车型价格带	30万以上高端车型，是比亚迪专为仰望品牌打造的“全域智驾标杆”，技术对标华为ADS3.0	覆盖15-30万元主流市场	下沉至7万元市场，成本仅为激光雷达版的1/5

数据来源：公开资料整理

2. 华为方案派：华为通过三种模式与车企合作——零部件供应模式、HI(Huawei Inside)模式和智选车模式。华为采用“激光雷达+视觉+毫米波”多传感器融合，配合自研昇腾芯片，在无图城区NOA覆盖上遥遥领先（超200城），且在深圳等地开启了L3级“脱手脱眼”的试点运营，华为在智能驾驶领域的长期投入和技术积累使其在技术上保持领先。

图：华为与Momenta智驾算法简介

算法公司	算法特点	合作车企
华为	<p>经历ADS1.0至ADS4.0的多版本迭代：</p> <ul style="list-style-type: none"> ●ADS1.0:采用“激光雷达+毫米波雷达+摄像头”的多传感器融合方案，搭配高精地图，核心算法为BEV+Transformer。 ●ADS2.0:引入GOD(通用障碍物检测)算法，基于OccupancyNetwork将环境分割为微小立方体。放弃外购高精地图，采用RCR(实时道路推理)算法，通过传感器实时生成道路拓扑结构。 ●ADS3.0:采用“GOD大网+PDP(预决策规划)”端到端大模型架构，实现感知与决策的深度融合。 ●ADS4.0:基于WEWA架构，云端生成海量极端场景训练数据。 	<ul style="list-style-type: none"> ●HI模式:阿维塔、岚图等。 ●鸿蒙智行:赛力斯(问界)、奇瑞(智界)、北汽(享界)、江淮(尊界)、上汽(尚界)等。
Momenta	<ul style="list-style-type: none"> ●Gen2:多任务感知且数据驱动，规划大多基于规则，包括检测、高精地图、融合&跟踪等模块。 ●Gen3:感知将时序任务整合到同一模型，逐渐从规则转型数据驱动，有FTP融合&跟踪&预测等模块。 ●Gen4:感知所有任务整合到同一模型，全部数据驱动处理，涵盖4D感知DDOD检测&融合&跟踪&预测等。 ●Gen5:感知与规划整合进一个大模型中，模仿人类长期记忆，保留DLP模型类似短期记忆，能快速学习。 	<ul style="list-style-type: none"> ●丰田、上汽大众、奥迪、奔驰、通用、本田、东风日产、上汽、智己、比亚迪、仰望、腾势、广汽、埃安、红旗等。

数据来源：智驾派、芝能科技

3. Momenta派：Momenta采用开放平台模式，专注于软件算法，提供灵活可定制的解决方案。Momenta以60.1%的市场占有率稳居中国城市NOA榜首，其“飞轮式”数据闭环策略通过量产车收集数据，反哺L4级自动驾驶研发，形成“量产数据→算法迭代→技术降本”的正向循环。

四、未来产业展望与细分赛道挖掘

随着算法和硬件的进步，我们相信2030年后中国各类无人驾驶车辆将逐步普及，L3级别的自动驾驶将在高端车型中广泛采用，预计渗透率将达到30%以上，这些车辆能够在特定条件下实现完全自动驾驶，L4级别的自动驾驶将在特定的城市区域和封闭园区内实现商业化应用。其中，Robotaxi将在主要城市的商业区和交通枢纽实现大规模运营，渗透率有望达到10%以上，将与现有的公共交通系统和共享出行平台深度

融合，形成完整的智能出行生态系统；其次，L4级别的自动驾驶公交车将在多个城市的主要线路和特定区域实现商业化运营，根据实时交通状况和乘客需求动态调整路线和班次，与地铁站、火车站等交通枢纽无缝对接，解决“最后一公里”问题，提供从家门到目的地的全程无缝衔接服务；再者，L4级别的自动驾驶物流车将在干线物流中广泛应用，在城市最后一公里配送中逐步普及，提高配送效率和降低运营成本；最后，L4级别的自动驾驶环卫车将在城市清扫和垃圾收运中广泛应用，提高收运效率和降低运营成本，在公园、校园等封闭场景中，L4级别的自动驾驶环卫车将实现全面应用。

智能驾驶产业正站在Gartner技术成熟度曲线的“实质生产期”门槛，在此背景下，挖掘价值成长股应聚焦具备技术壁垒、绑定头部生态、且净利润增速持续高于30%的公司，我们认为产业趋势呈现四大结构性机会：一是硬件降本与国产替代加速，激光雷达、车载摄像头等传感器价格下探推动高阶智驾下沉至10-15万级车型，实现“智驾平权”；二是线控底盘作为执行层核心，受益于强制性国标实施与外资垄断打破，迎来价值量跃升与国产化率快速提升；三是国产智驾主控芯片开始大规模上车，单芯片算力不输英伟达，其自研智驾算法与芯片适配性更佳，软硬件均可升级；四是运营商将在传统整车厂、第三方算法公司及现有共享出行平台商中诞生，运营市场规模万亿，未来必将诞生市值万亿级别的运营巨头。

1.4 前瞻洞察：迈向由业务定义价值的新纪元

展望未来，AI应用将彻底摆脱对底层模型的盲目追随，进入由“场景深度”定义“商业溢价”的下半场。首先，模型能力将迅速“水电化”与“商品化”，通用模型能力将进入标准化供给阶段，智能成本的结构优化将推动大规模应用平权。这意味着，底层模型的竞争重心将从单纯的参数规模转向与具体场景的“深度耦合”能力，真正的价值将来源于：谁能将这些廉价的智能，精准地注入到那些曾经无法标准化、依赖精英人力的复杂业务流程中。未来的竞争将是一场关于“重构力”的竞赛。企业将不再询问“AI能做什么”，而是在反思“为了适配AI，我们的组织架构和业务流程需要做哪些暴力拆除”。这种由应用渗透引发的生产关系倒逼生产力变革，将催生出一批“原生AI企业”。它们从诞生之初就以智能体为组织核心，以数据闭环为生命线。这场由确定性驱动的应用革命，才刚刚拉开大幕。

端侧AI的关键转折 硬件重构与场景价值跃迁

端侧AI的发展正在跨越早期概念验证阶段，进入深度应用落地与用户体验精雕细琢的关键窗口期。对比前两年以基础端侧算力搭建、通用大模型初步接入为核心的发展路径，当前行业的核心演进方向，已转向硬件形态对AI功能的深度适配，以及垂直场景精细化解决方案的规模化落地：

AI定义硬件核心价值：AI不再是硬件的附加功能或营销噱头，而是成为定义全新硬件形态的核心要素。其底层逻辑在于端侧AI硬件的长期生命力，源于对用户非结构化数据的高频触达，以及对用户生活场景的深度嵌入——通过优化交互界面、强化场景适配，让AI能力通过分析端侧硬件获取的数据，以达成对用户需求精准耦合，最终实现从AI+硬件的“功能叠加”到AI与硬件“价值共生”的跃迁。

硬件生态边界消融：各类智能硬件的品类边界进一步模糊，跨设备协同成为行业常态，不同终端基于端侧AI能力实现数据互通、功能互补，构建无缝衔接的智能体验闭环。

算力升级逻辑迭代：端侧计算能力持续提升，但发展重心从单纯比拼 TOPS 算力数值，转向能效比优化与特定任务的算力精准匹配，例如面向实时图像处理、低延迟语音交互等场景的专项算力调校，实现性能与功耗的平衡。

图：端侧AI核心演进方向



资料来源：公开资料整理

2.1

AI眼镜多路线并行,手机与PC正突破性能与形态的物理瓶颈

极致轻量化的AI拍摄与音频眼镜正在通过“去屏幕化”率先抢占大众市场。以Meta Ray-Ban的热销和缺货为标志,市场证明了用户愿意为了无感记录和AI语音交互付费。字节跳动推出的豆包AI眼镜采用了ODM模式,以2000元以内的亲民价格,搭载高通AR芯片和摄像头,专注于AI语音助手和第一视角内容的捕捉,通过海量视觉数据反哺模型训练。百度的“小度AI眼镜Pro”同样沿用此路径,强调45g的极致轻便和旅游场景下的实时翻译与讲解能力,试图将搜索入口从手机屏幕转移到用户视野中。

Birdbath方案依然是沉浸式影音娱乐的最佳载体,正向独立计算终端进发。虽然BB方案体积相对较大,但其“便携巨幕”的特性在特定场景下不可替代。Xreal通过自研Real3D芯片将普通2D内容转为3D,强化游戏主机连接属性。而雷鸟Air2s等产品开始植入eSIM模块,试图摆脱手机束缚,成为独立的通信和娱乐终端。

AR显示模组正处于跨越量产门槛的加速冲刺期。预计2027年或将成为全彩光波导技术的产能爆发元年,当前市场正通过多路线并行策略,为这一终极形态提前预演。IDC预测,全球智能眼镜市场将在2029年突破4000万台出货量大关,其中中国市场将以55.6%的年复合增长率领跑全球,成为这一赛道的核心增长引擎。

光波导方案有望承载AR眼镜的终极未来,正在美学与显示性能之间寻找平衡点。CES 2026上展出的Even Realities G2采用了单绿双目衍射光波导技术,其做到了完全隐形和无漏光,外观与普通眼镜无异,标志着光机小型化与镜片设计已突破物理瓶颈,完美诠释了“无感佩戴”的设计哲学。展望2034年,Dimension Market Research 预测AR眼镜市场规模或将达到1112亿美元,成为继手机之后的下一个万亿级计算平台。随着光波导技术的普惠化与端侧算力的能效跃迁,未来的AI眼镜必将走向集成高效显示系统(全彩光波导)、高质量影像拍摄以及强大端侧AI计算能力的一体化方案。

图：AI眼镜主流形态发展

沉浸式娱乐独立终端

不再局限于手机配件，正通过植入高算力芯片与eSIM模块向独立计算机主机进化，以影音游戏率先打开大众市场



专注显示的AR眼镜

在美学与性能间寻找完美平衡，通过单绿/全彩衍射波导技术实现无感佩戴，将数字信息无缝叠加于现实世界，重新定义人机交互界面



配置摄像头的AI眼镜

在极致轻便（< 50g）的限制下，验证了端侧计算与多模态技术，专注于第一视角视觉捕捉与语音交互，是渗透潜力最高的形态



资料来源：Focally官网概念图（上）、Xreal官网（左）、Rokid官网（中）、极米官网（右）

AI手机与PC正在利用3nm/1.8nm制程红利带来的能效飞跃，将堪比专业工作站的推理能力下放至移动端，并催生出适应AI流的新形态。Intel发布的Panther Lake处理器基于18A工艺，内置50 TOPS NPU且能效改善显著。AMD的Ryzen AI Max+系列则通过高达128GB的统一内存架构，打破了显存墙，使得在笔记本本地运行128B参数级别的超大模型成为可能，这对于金融、法律等对数据隐私极其敏感的行业具有革命性意义。

PC硬件形态则开始主动适应AI交互需求。联想ThinkBook Plus Gen 7 Auto Twist利用AI视觉追踪算法，实现了屏幕随人转动，将笔记本变成了智能拍摄助理。华硕ZenBook Duo则利用AI助手智能分配双屏任务流，使双屏设计从单纯的显示扩展变成了真正的智能工作台。手机方面，“无AI，不旗舰”已成为行业铁律。Counterpoint数据显示，2025年全球GenAI智能手机SoC的出货量占比已达到35%，同比实现74%的爆发式增长。三星Galaxy S26 Ultra全系搭载Snapdragon 8 Elite Gen5，NPU性能较前代提升6倍，并配备16GB RAM，有力支撑Google Gemini Nano和三星Gauss2模型的本地化运行，实现如屏幕防窥视眼球追踪等底层或端侧AI功能。

图：AI手机特征及应用



Galaxy AI

Advanced intelligence

Unlock a healthier you

- Now brief**
Get a personalised briefing with suggested content and actions that change throughout the day.
- Health assist**
Get personalised insights and guidance based on your health data.

Enhance your communication

- Call assist**
Get real-time translation during voice calls.
- Writing assist**
Get tools to compose, organise, and improve text, plus chat translation and suggested replies to text messages.
- Interoreter**

The first time you use a Galaxy AI advanced intelligence feature in an app or service that requires data collection to provide results, we'll let you know before your data is collected and processed.

Continue

Unleash your productivity

- Note assist**
Auto format, summarise, spell check, and translate text. Transcribe and summarise recordings. Decorate notes with AI-generated covers and images. All with Note assist.
- Transcript assist**
Transcribe voice and call recordings into text, then summarise them for easy review.
- Browsing assist**
Get simple summaries of webpages, translate text and images, hear summarised webpages read aloud, and listen to highlights from articles and posts on select sites.

Create epic images

- Photo assist**
Reimagine images by moving or removing people and objects, creating portraits in a variety of fun styles, and sketching new stuff into their backgrounds.
- Drawing assist**
Transform simple sketches into awesome art, create images from text descriptions, and turn

The first time you use a Galaxy AI advanced intelligence feature in an app or service that requires data collection to provide results, we'll let you know before your data is collected and processed.

Continue

- Browsing assist**
Get simple summaries of webpages, translate text and images, hear summarised webpages read aloud, and listen to highlights from articles and posts on select sites.

Create epic images

- Photo assist**
Reimagine images by moving or removing people and objects, creating portraits in a variety of fun styles, and sketching new stuff into their backgrounds.
- Drawing assist**
Transform simple sketches into awesome art, create images from text descriptions, and turn photos into cartoons, paintings, and more.
- Audio eraser**
Eliminate distracting noise in videos by adjusting the volume of voices, wind, and other sounds.
- Photo ambient wallpaper**
See how AI changes your own photo based on the time and weather.

The first time you use a Galaxy AI advanced intelligence feature in an app or service that requires data collection to provide results, we'll let you know before your data is collected and processed.

Continue

资料来源：高通官网，三星官网，上述展示内容仅做案例使用

2.2

新兴品类的爆发：AI硬件正剥离极客感，转向健康管理、情感计算与儿童成长的垂直领域

健康监测硬件正在经历从“被动数据记录”到“主动干预生活”的质变，形态上彻底打破了手环/手表的局限，向指环及家居环境延伸。健康硬件不再满足于简单的计步或心率监测，而是引入了医疗级的诊断AI与更隐形的形态。IDC测算，全球智能指环出货量在过去一年实现了49%的爆发式增长，市场规模预计将在2032年突破74亿美元。Oura Ring 4以详细的体征监测和AI对话分析确立了行业标杆，而RingConn Gen 3则突破性地在戒指中实现了光电容积脉搏波血压趋势分析，这在消费级产品中尚属首创。

与此同时，AI硬件开始关注居住环境与心理健康：Sleepal AI台灯通过感知用户活动数据，动态调节光线色温以辅助阅读或专注；Lepro Ami桌面伴侣借助OLED屏幕与眼球追踪技术模拟眼神接触，有效缓解居家办公的孤独感；Nuralogix“长寿镜”则利用透皮光学成像技术，量化评估用户的生理年龄与健康风险。

图：智能指环内部密集的传感器



资料来源：IFIXIT，Oura Ring官网

情感陪伴类硬件正在开辟“去屏幕化”的交互新赛道，AI从工具变成了具备共情能力的伙伴，填补了现代人的孤独感缺口。据Global Market Insights报告，全球情感AI市场在2024年已达29亿美元，并预计将以22%的惊人年增长率狂奔。这一领域的硬件设计刻意弱化了功能性，转而强调“存在感”和“情绪价值”。Friend与ThingX等挂坠产品主打无屏设计，通过蓝牙连接充当用户的全天候情感倾听者；Motorola Qira则进化为跨端协作的“视觉记忆代理”，通过摄像头捕捉生活流并与PC及车载系统深度联动；松下NICOBO等“弱机器人”则反其道而行，通过模拟生物性的示弱与依赖，激发用户保护欲以建立深层的情感连接。

儿童教育硬件正在利用大模型实现“千人千面”的个性化陪伴，从单一的故事播放升级为深度的内容共创。全球智能教育硬件市场在2025年已达到376.2亿美元，AI彻底改变了玩具的交互逻辑。希沃Seedpace的故事机允许儿童通过语音介入剧情发展，实时生成个性化内容；零动未来“话话糖”依托通义大模型，侧重于通过多轮情感对话建立玩伴关系，而非单纯的知识灌输；听力熊则利用穿戴式视觉技术，自动识别环境并捕捉青少年的成长高光时刻，填补了家长的陪伴空窗期。

表：CES 2026 技术风向标：AI 硬件的多元化创新与场景渗透

企业/产品	类目	差异点
BreakReal	AI 调酒机器人	情绪调酒：结合面部表情识别，根据用户当前的情绪状态推荐并调制专属鸡尾酒
Chuangmo (创谟电子)	AI 桌面翻译机	商务伴侣：专为办公桌面设计的硬件，集成透明屏幕或定向收音，支持多语种实时字幕
Cresento	AI 护腿板	球类运动监测：嵌入式传感器，专门分析足球等运动中的腿部击球力度、跑动数据和冲撞情况
睿宝智能、光帆科技	AI 耳机 (配摄像头)	视觉音频双模态：耳机集成摄像头，支持第一人称视角的 AI 视觉问答
Ellie	AI 婴儿监视器	非接触监测：利用视觉或雷达技术，无需接触皮肤即可监测婴儿呼吸频率和睡眠，并识别哭声
Farseer (RocX)	AI 远景追踪相机	50倍变焦追踪：专为观鸟设计，利用 AI 锁定并平滑追踪高速移动物体 (如飞鸟)，解决长焦跟拍难题
FitButlr	AI 毛巾	接触面监测：织物中嵌入柔性传感器，监测健身房器械使用情况或用户汗液/体态数据

表：CES 2026 技术风向标：AI 硬件的多元化创新与场景渗透

企业/产品	类目	差异点
Friend	AI 挂坠	情感连接：著名的“项链”形态，无屏幕，通过蓝牙连接手机，主打倾听和发送“主动式”短信
Inupathy、PetPace	AI 狗环	宠物情绪翻译：监测犬类的心率变异性 (HRV)，分析宠物的情绪状态（焦虑、兴奋等）
IdentifyHer (Peri)	AI 可穿戴（女性健康）	围绝经期追踪：佩戴在躯干的传感器，捕捉潮热、盗汗等症状数据，辅助更年期健康管理
ININUM (NINU)	AI 香氛	智能调香：世界上第一款 AI 智能香水，通过 App 混合内置的三种基调，随心情和场合定制香味
Lepro Ami	AI 桌面伴侣	远程工作伴侣：带有 OLED 屏幕和眼球追踪，能模拟眼神接触，专为缓解居家办公孤独感设计
Limbico	AI 情绪手环	神经信号量化：结合皮质醇/多巴胺监测与手机数据，量化用户的心理压力和情绪
Looki L1	AI 可穿戴摄像头	生活记录仪：30g 轻量化设计，自动捕捉生活瞬间并生成漫画/视频日志，解决“手机拿出来太慢”的痛点
LOONA DeskMate	AI 桌面充电站	桌面助理：兼容 MagSafe 的 3 轴仿生机械运动系统，多模态感知与情感交互
Luya Tech	AI 蔬菜种植器	家庭微型农场：全封闭无土栽培，AI 视觉监控生长，自动调节环境，提高营养密度
Mooni Pro (听力熊)	AI 可穿戴摄像头	青少年陪伴：通过摄像头与语音直接理解青少年所处的真实环境，记录与沉淀成长瞬间
Nextbase (iQ)	AI 行车记录仪	座舱守护：具备车内 AI 监控功能，检测驾驶员疲劳、分心，并提供停车云端实时守护
Noul	AI 诊疗机	微流控诊断：这种设备通常用于疟疾或血细胞形态分析，利用 AI 进行快速且精准的病理诊断
Nuromova	AI 运动健康监测头带	脑电波 (EEG)：相比手环，头带能更精准地采集脑电波，分析运动专注度或辅助冥想
Razer (Project AVA)	AI 桌面伴侣	全息/沉浸：雷蛇的概念产品，结合全息投影或头戴显示，为玩家提供可视化的 AI 游戏助手
Razer (Project Motoko)	AI 游戏伴侣（头戴耳机）	视觉感知：功能逻辑与 AI 眼镜类似，对空间和重量要求更宽松，骁龙平台双目立体视觉

表：CES 2026 技术风向标：AI 硬件的多元化创新与场景渗透

企业/产品	类目	差异点
Sleepat® AI Lamp	AI 台灯	环境感知：不仅仅是照明，通过感知用户活动自动调节光线色温，甚至辅助阅读或工作专注
ThingX	AI 吊坠	便携大模型：类似 Friend 的形态，但可能更侧重于特定垂直场景（如会议记录、老人看护）的语音交互

资料来源：CES2026展会新闻

AI基础设施

从芯芯之火到全面燎原

3.1

算力芯片：通用GPU与专用ASIC 从竞争走向竞合，CPU受益推理拉动迎新市场

» 3.1.1 通用GPU与专用ASIC 从竞争走向竞合

GPU基于冯·诺依曼架构，作为通用芯片可兼容多场景，包括图像渲染等多样化应用。英伟达自2010年的Fermi架构开始设计GPU，每一代架构都承载着明确的技术突破，近几年的hopper、Blackwell、rubin架构更是在算力、存储等方面呈现了持续翻倍增长，引领了算力芯片的发展。

表：英伟达芯片演进

架构	Hopper			Blackwell		Rubin	
时间	2022	2023	2024	2025		2026	2027
Accelerator	H100	H200	B200/ GB200	GB300 (Ultra)	GB300 (B300A)	VR200	VR300 (Ultra)
GPT TDP (w)	700	700	700/1200	1400	600	1800	3600
Foundry Node	4N	4N	4NP	4NP	4NP	3NP	3NP
FP4 PFLOPS	4	4	10	15	4.6	50	100
HBM	80GB HBM3	141GB HBM3E	192GB HBM3E	288GB HBM3E	144GB HBM3E	288GB HBM4	1025GB HBM4E
HBM Stacks	5	6	8	8	4	8	16
HBM BandWidth(TB/s)	3.35	4.8	8	8	4	13	32
Packaging	CoWoS-S	CoWoS-S	CoWoS-L	CoWoS-L	CoWoS-L	CoWoS-L	CoWoS-L
SerDes speed(G)	112	112	224	224	224	224	448
Nvidia CPU	Grace			Vera			

资料来源：英伟达官网，诺安基金研究部整理

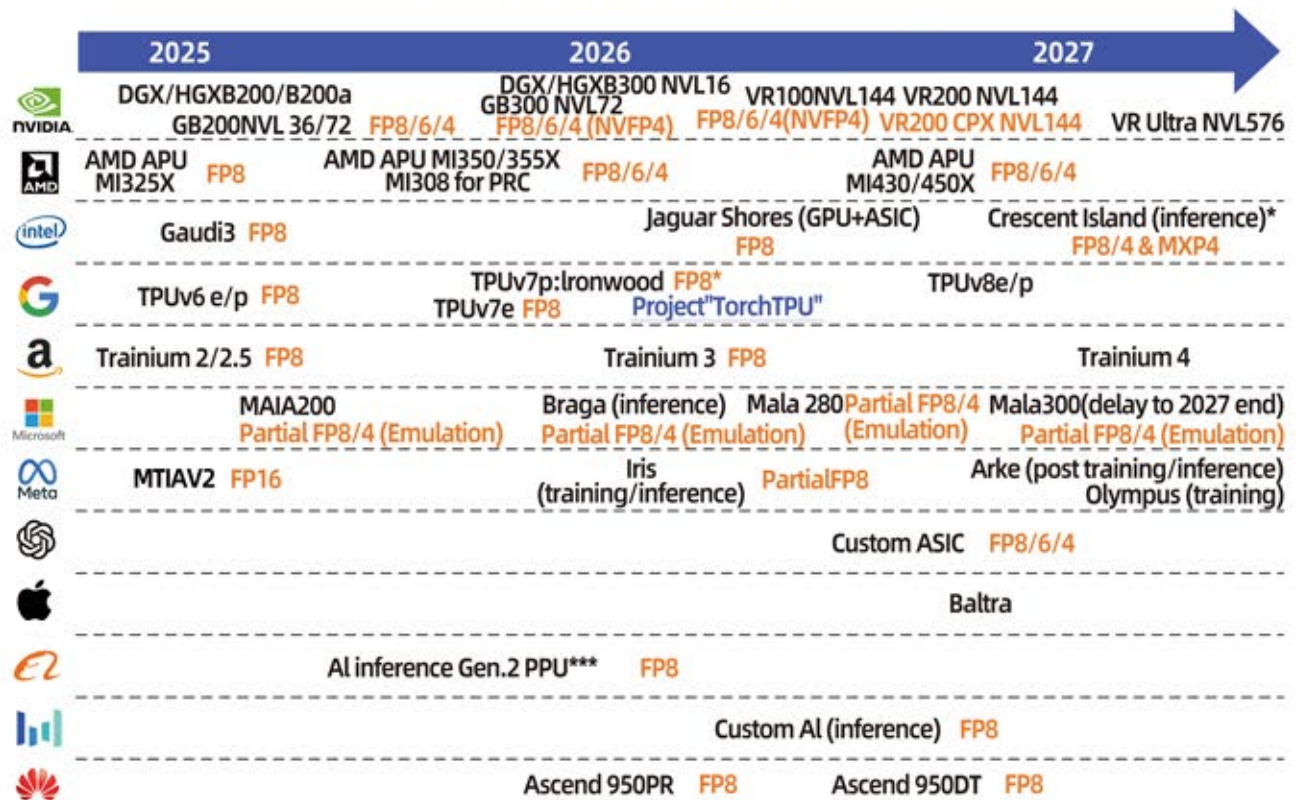
ASIC芯片采用脉动阵列架构，专为矩阵运算设计，计算结果直接传递，减少数据交互，提高效率。ASIC通过精简指令集、优化内存架构和计算单元，在特定任务上实现了数量级的能效提升。其“定制化”的优势，受到谷歌、亚马逊、META、微软等互联网大厂青睐。通过自研ASIC芯片，互联网大厂首先可以优化内部工作负载，高效地满足其内部AI推理和训练需求；其次，ASIC芯片具有更好的性价比。比如谷歌使用自研的TPU7（Ironwood）训练自家Gemini3模型，呈现出了领先业界的顶尖模型性能。

起初，ASIC芯片的出现对GPU的市场份额发起了挑战，但随着AI大规模投入的发展，不同应用场景下的芯片需求有所不同，GPU和ASIC芯片分别占据了不同细分领域：

- 英伟达强大的CUDA生态以及自研的芯片互联构成了系统级优势，其更快的“产品上市时间”和技术领先型，使其面对企业级客户保持了其市场的领先地位。

- ASIC的优势在于，对CSP这类超大规模用户，当其软件栈完全自控，能带来极致的TCO（总拥有成本）优化和能效比，ASIC适用于超大型CSP或者大模型等超大规模企业。

图：全球头部芯片厂商的芯片发展路标



*Google TPUv7 only provides native FP8 and lacks hardware support for FP4, requiring software emulation.

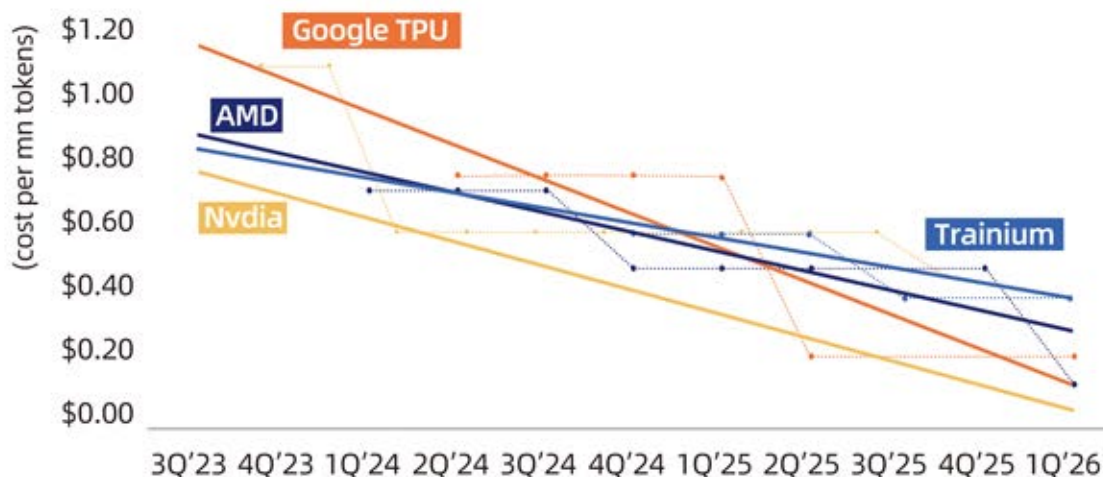
资料来源：IDC

从单芯片成本分析：根据研究机构伯恩斯坦今年1月数据分析，谷歌的TPU (ASIC) 在原始计算成本性能上正迅速缩小与英伟达 (Nvidia) GPU的差距。从TPU v6到v7，每token成本降低了约70%，其绝对成本已比英伟达的GB200 NVL72相当甚至更优。

从IDC预测数据整体市场规模看，全球市场GPU芯片和ASIC芯片 (下图Other Accelerated) 规模齐头并进。

图：每百万tokens的推理成本（美元）

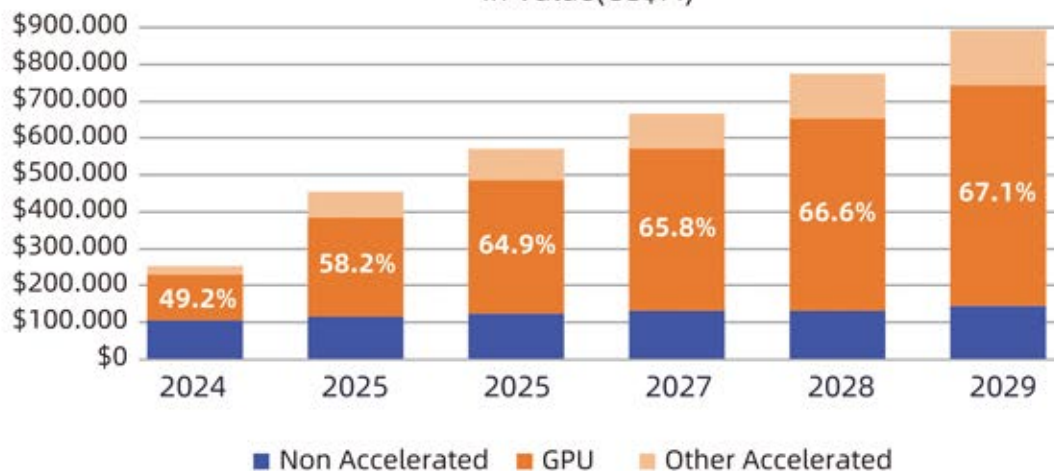
Exhibit 1: Inference cost per million tokens (\$)



数据来源：伯恩斯坦

图：全球AI芯片市场发展

In Value(US\$M)



数据来源：IDC

» 3.1.2 AI Agent时代，CPU正从台后走向C位

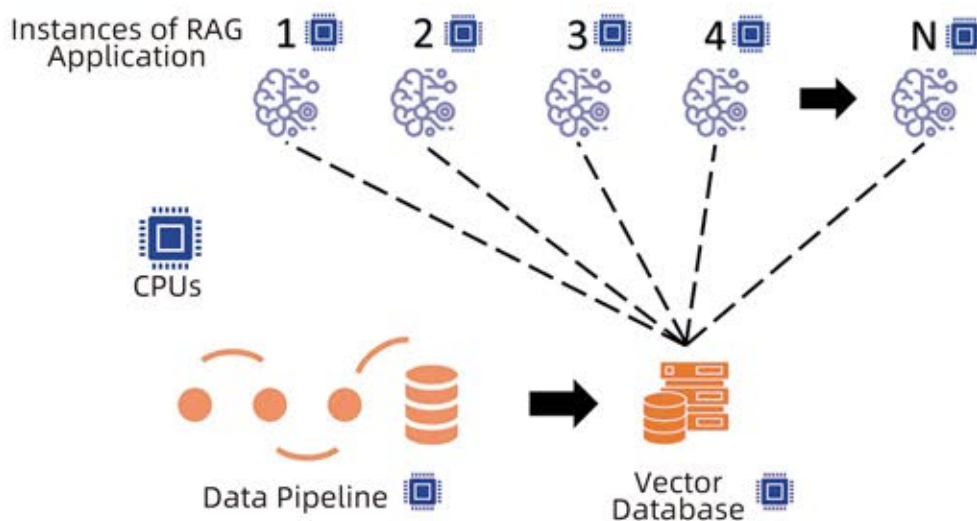
CPU正成为让AI Agent跑起来的核心芯片。随着大模型应用从“生成答案”走向“执行任务”，预计2026年AI产业全面进入AI Agent（智能体）时代。AI Agent不只是生成token，而是围绕目标持续完成任务拆解、检索、记忆、调用工具、连接企业系统、执行动作与多agent协同。Google Cloud将agentic AI定义为围绕“编排与执行”的系统；AWS把agent描述为能够连接公司系统、API和数据源，自动完成多步骤任务的智能体。因此，AI基础设施的压力不再只落在GPU推理上，CPU将承担更多的资源调配职责。

从模型发展角度看：

- 在用于模型改进的RL（强化学习）训练闭环里，“RL环境”需要执行模型生成的动作，并计算相应的奖励。为此，在编程与数学等领域，需要大量CPU并行完成代码编译、验证、解释与工具调用。CPU也深度参与复杂的物理仿真，以及以高精度验证生成的合成数据。

- 在推理侧，检索增强生成（Retrieval Augmented Generation, RAG）模型以及智能体工具调用，需要CPU处理海量网络请求、数据库查询和多源数据整合。由于可以同时向多个来源发起API调用，每个agent本质上都能比人类更密集地使用互联网——不再是简单的Google搜索。AWS与Azure正在对其自研的Graviton与Cobalt CPU系列进行大规模扩建，同时也在采购x86通用服务器，以应对互联网流量在这一“阶跃式增长”下的CPU需求。

图：CPU支持RAG模型应用的架构



数据来源：腾讯云

从硬件架构看：每单位GPU需要更强CPU配套。在经典的训练服务器中，1颗CPU往往配置8颗GPU。之后在英伟达的GB200/300 NVL72架构中，配比已经提升到了36颗Grace CPU对应72颗Blackwell GPU（即 1:2 配比）。随着智能体推理和强化学习模拟的快速增长，CPU与GPU的配比有望将进一步提升。未来GPU和CPU将组成一套联合架构，GPU负责生成智能，CPU负责组织智能。

图：微软Fairwater数据中心发展（CPU配套逐步提升）



数据来源：Semianalysis

3.2 通信：追求高集成度下的极致性能

在追求AI芯片性能提升和成本下降的同时，网络、存储的发展相辅相成。网络方面，AI算力基础设施的互联扩展从传统三层组网架构向立体三维度Scale-Up（纵向扩展）、Scale-Out（横向扩展）、Scale-Across（跨区域扩展）演进。

图：META在建的全球规模领先的数据中心 Prometheus和Hyperion

图：Scale-in/Scale-up/Scale-out组网



资料来源：2025 OCP Meta 《leading the future of AI》

资料来源：2025 OCP Meta 《leading the future of AI》

» 3.2.1 数据中心间（Scale-across）—— DCI系统及长距离光纤迎需求爆发

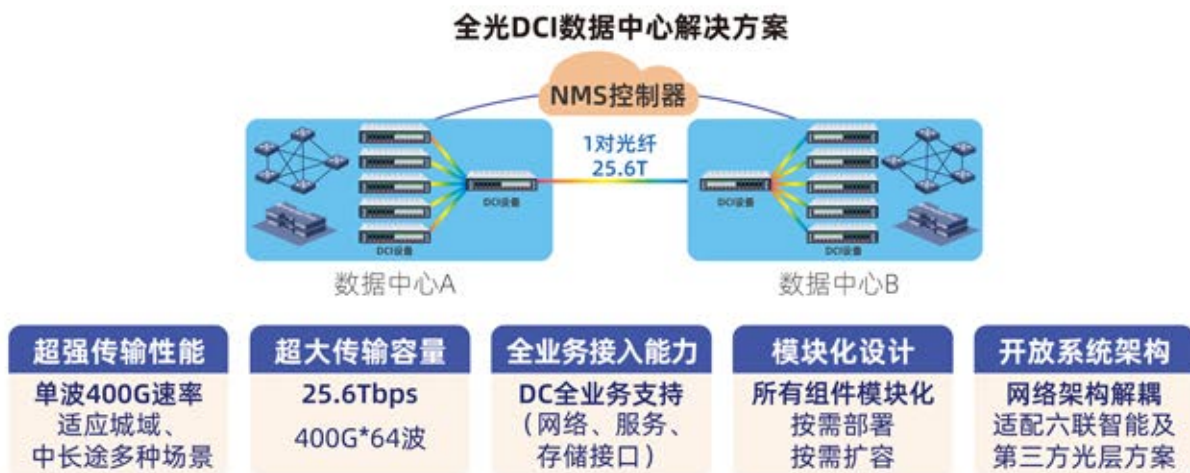
数据中心间互联（Data Center Interconnect, DCI）是一种实现多个数据中心之间互联互通的网络解决方案。其核心在于高速光网络技术的持续升级。目前商用系统的单波最高速率已达到1.6T，采用3nm工艺和100GHz光电子器件。在单纤容量方面，C+L双波段系统已成功实现商用，使光纤可用频谱达到了12 THz，可支撑近百T容量。

图：DCI市场规模预测



资料来源：Trendforce

图：DCI数据中心互联解决方案



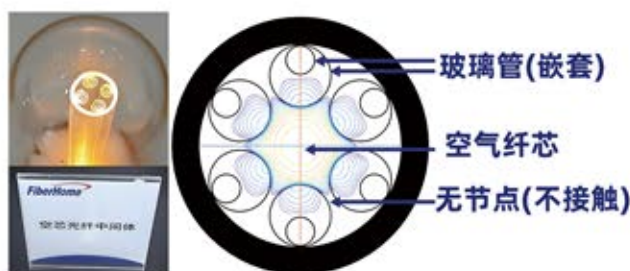
资料来源：六联智能

空芯光纤是数据中心长距离互联未来重要技术方案。空芯光纤（HCF, Hollow-core fiber），是以空气为传输介质，替代传统以“玻芯”（石英玻璃，主要成分是二氧化硅）作为传输媒介的光纤。空芯光纤的导光原理是以特殊微结构形成的光子禁带或反谐振效应

来抑制纤芯中光的横向泄漏, 实现低折射率纤芯中的轴向光传输, 因此光传输受光纤材料吸收影响少, 非线性效应和延迟较小, 损伤阈值高, 热稳定性好。

微软率先布局空芯光纤技术商业化, 其或将长期引领空芯行业技术、市场及商业化进程。Lumen Technologies早在2024年8月宣布, 人工智能的需求推动其获得50亿美元订单。2025年7月中国移动在广东开通了我国首条反谐振空芯光纤商用线路。

图：空芯光纤对比实芯光纤减少约30%延迟



资料来源：烽火通信

图：我国“首个”反谐振空芯光纤商用线路，损耗达到0.085dB/km



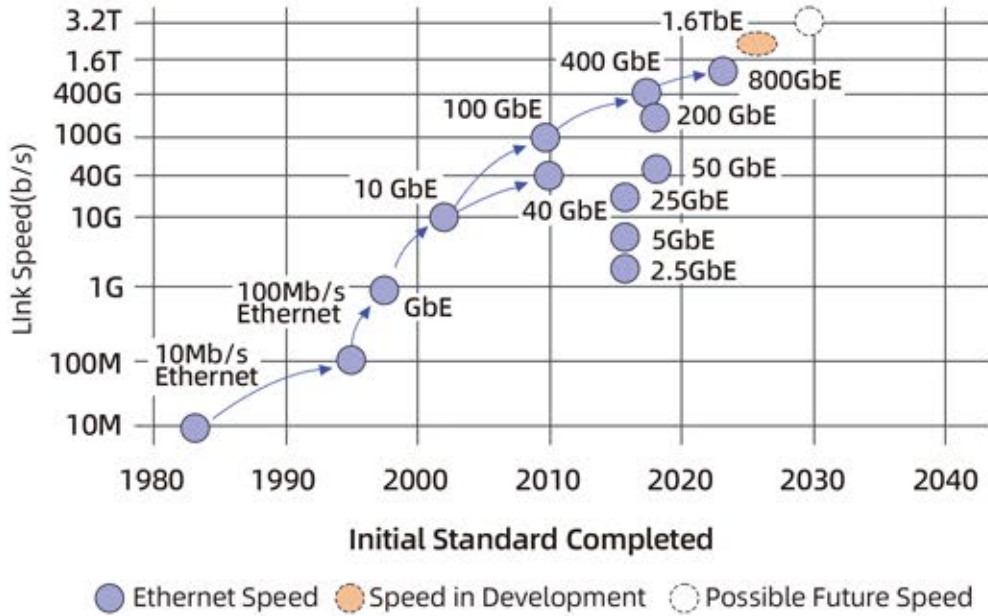
资料来源：中国移动、长飞光纤

» 3.2.2 超节点间 (Scale-out) —— 1.6T高速率光/硅光模块成为主流

AI基础设施发展持续加大了对数据通信光学需求, 1.6T是400G和800G之后的下一代光通信主流速率节点。传统分立器件光模块的上游光芯片目前处于供不应求状态, 而硅光技术凭借低功耗、高集成度、低成本的核心优势, 有望解决传统光模块上游缺芯的问题, 也助光模块不断突破性能瓶颈。

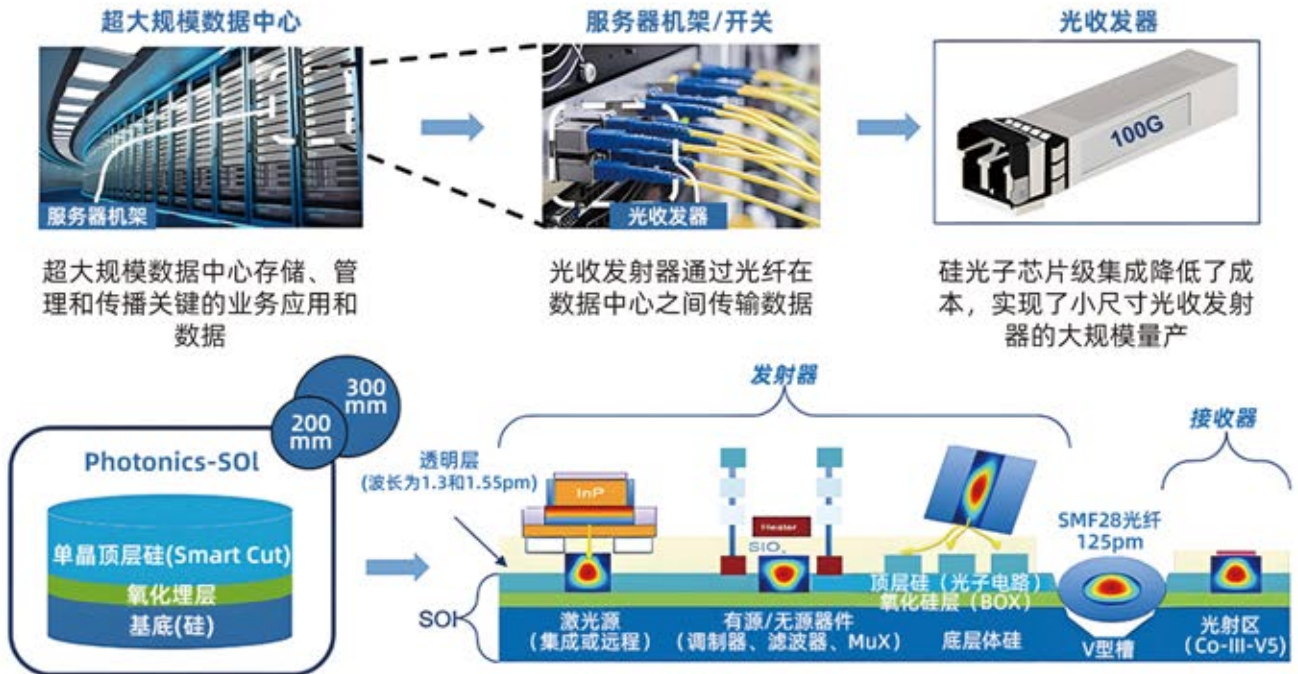
硅光技术优势是高度集成化, 以半导体制造工艺将硅光材料和器件集成在同一硅基衬底上, 形成由调制器、探测器、无源波导器件等组成的集成光路, 优化了光模块结构件/PCB用料, 同时节省了温控器件TEC, 并且用CW光源替代传统光模块的EML等, 使得光模块总体成本节省约20%, 功耗优化近40%。

图：光模块速率升级路径



资料来源：Dell'Oro

图：硅光模块应用及产品工艺概览



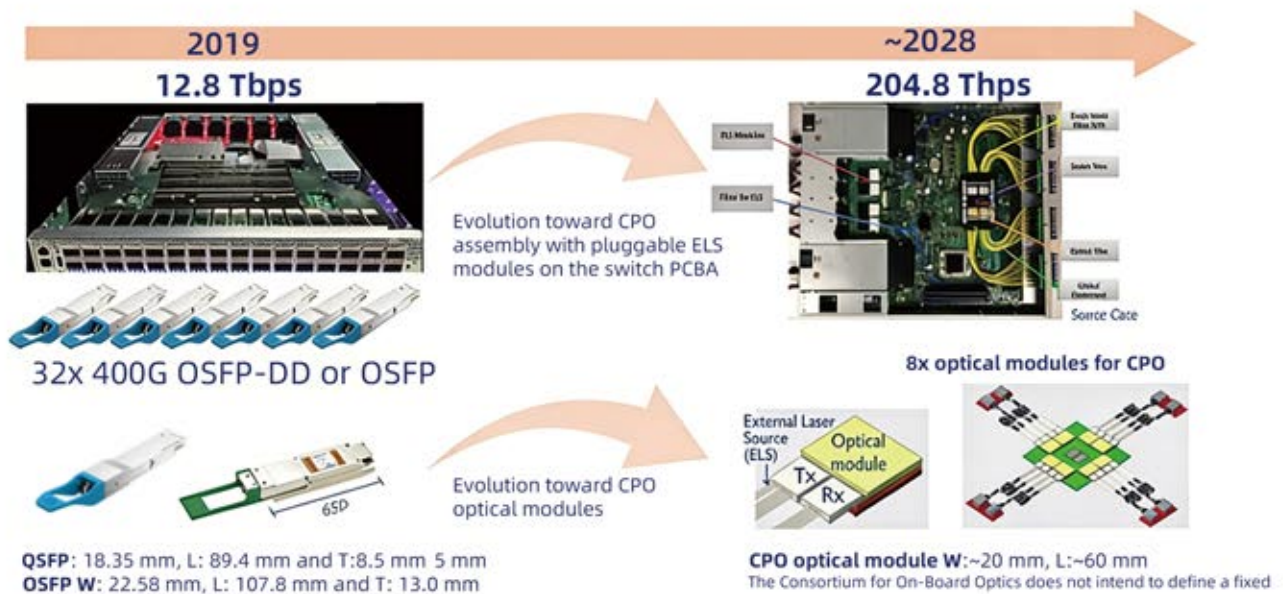
资料来源：Soitec官网

3.2.3 超节点内 (Scale-up) ——CPO/NPO、OCS技术从柜间互联向柜内渗透, PCB仍是短距离互联主流技术

光电共封装CPO(Co-packaged Optics)是一种在数据中心互连领域应用的光电集成技术, 目前主要用在交换机接口中。通过将交换芯片和光器件共同装配在同一个插槽上, 使得光信号和电信号在同一芯片上进行转换和处理, 该技术缩短了芯片和模块之间的走线距离, 最大程度地减少高速电通道损耗和阻抗不连续性, 也在传输速率提高的同时大大缩减功耗, 行业头部企业博通 (Broadcom) 和扇港 (SENKO) 均分析过CPO技术可助力交换机整体系统减少功耗25%-30%。

在本轮AIGC革命的驱动下, CPO技术产业化进程继续加速。博通于2025年6月发布了基于Thomahawk6 (102.4T交换容量) 版本的CPO芯片, 英伟达于2025年3月GTC大会发布其Spectrum系列 (112.4T交换容量) 的CPO交换机。国内交换机头部厂商新华三、锐捷网络近两年也陆续发布了商用CPO交换机。而随着SerDes速率224G乃至更高速率的快速升级, 以及硅光材料和3D封装技术的逐步成熟, CPO技术的快速迭代, 也为光互联Optical IO技术发展奠定了良好基础。

图：CPO交换机对比传统交换机



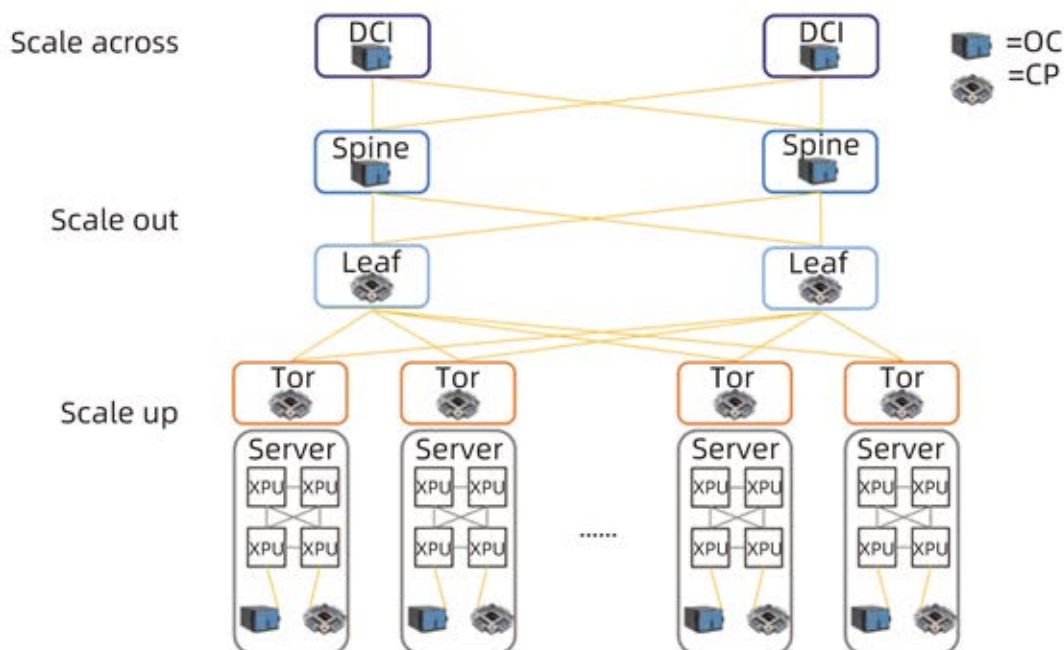
资料来源: Yole

OCS光交换是一种无需光电/电光 (O/E/O) 转换, 直接实现光信号在光纤端口间切换的技术。OCS原理是直接对光信号进行物理路径的重构, 从而在输入/输出端口之间建立专用光路。OCS核心优势:

- 1) 低时延、低功耗;
- 2) 作为物理层传输设备, 对底层协议不敏感;
- 3) 速率无关, 无需频繁更换升级。

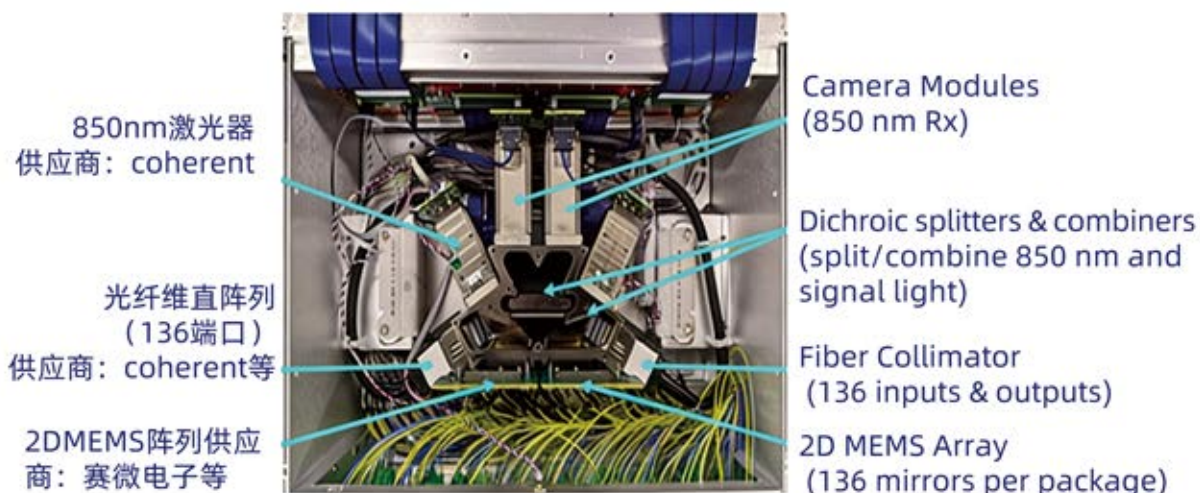
谷歌早在TPU4时代开始部署OCS交换机, 目前已达到年产近万台规模。在2025年4月的OCP EMEA SUMMIT上, 英伟达首席科学家Benjamin Lee分享了AI数据中心CPO与OCS搭配使用的方案。引入OCS可将数据中心网络功耗降低30%-40%。

图：OCS+CPO组网实例



资料来源：英伟达

图：Google的OCS交换机实物



资料来源: Google

在机柜内，芯片互联向正交化、无线缆化演进，信号链条更短、对材料损耗更敏感。与传统背板中信号需经长距离、多过孔、复杂转折的布线方式不同，正交架构实现了：

- (1) 信号路径最短化，芯片间通信经最短垂直路径直达对端，减少传输损耗与串扰；
- (2) 布线层简化，消除大量蛇形走线与过孔转换，背板层数得以优化；
- (3) 连接器革新，采用高密度、低插损正交连接器，支持更高速率的差分信号。

正交架构下PCB背板向高速率升级，高端PCB呈现紧俏态。其上游材料到制造工艺的全面升级，例如英伟达在2026年量产的Rubin系列芯片将采用M9/M10级覆铜板(支持224G级更高传输速率)。

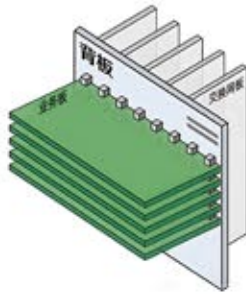
M9覆铜板的三大核心组成部分树脂、玻璃纤维布和铜箔将进行全面的技术革新。

(1) 树脂方面，M9/M10级材料转向介电性能更优异碳氢树脂(PCH)或性能更优的萜烯树脂(EX)。聚四氟乙烯(PTFE)也是候选材料之一。

(2) 玻璃纤维布的介电性能会直接影响Dk/Df值。M9/M10级材料采用第三代低介电(Low DK)玻璃纤维布，即glass或石英布。

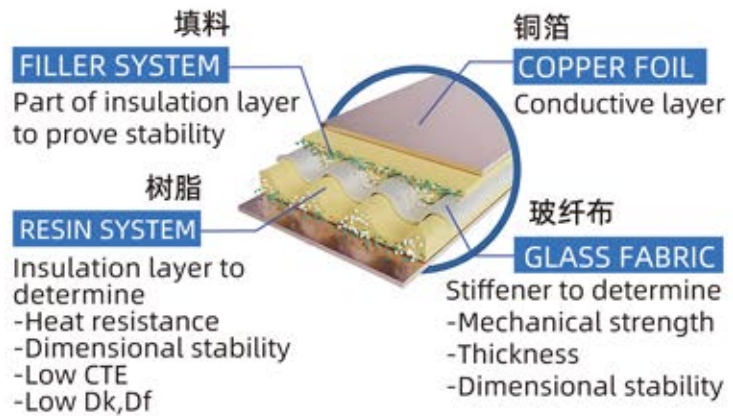
(3) 铜箔方面，M9/M10级材料须采用HVLP4(Very Low Profile4)级别的超低轮廓铜箔。这种铜箔表面极其光滑，能最大程度地减少高频信号的插入损耗。

图：正交背板架构示意图



资料来源：华丰科技

图：M9级覆铜板材料组成



资料来源：电子工程网

3.3

存储：AI发展推动存储架构变革，供需错配下存储迎大周期

» 3.3.1 AI推理需求推动存储架构变革：突破内存墙，向异构存储范式演进

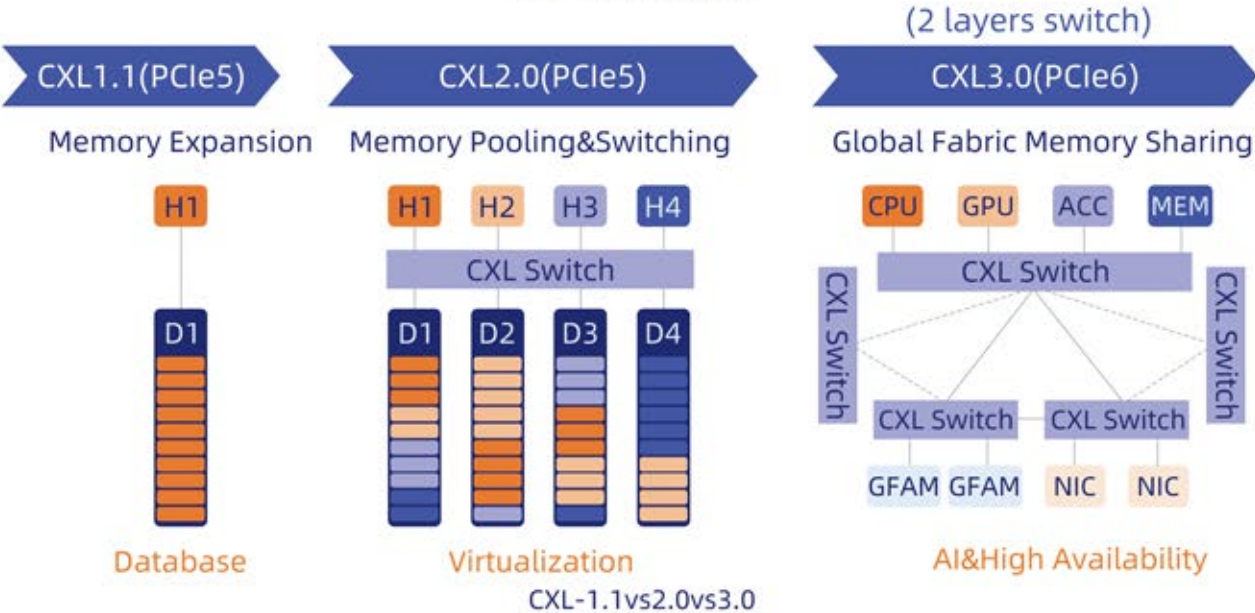
随着人工智能产业在2025年至2026年间实现从大规模模型训练向大规模推理应用的战略重心转移，推理——尤其是面向智能体（Agentic AI）的多轮对话、长文本理解和复杂推理任务——对“上下文（Context）”的保留与快速调取提出了严苛的要求，算力的核心矛盾转移成指数级数据增长所堆积出现的“内存墙”。解决如上“内存墙”问题催生了存储架构的两种变革：

·英伟达利用“NAND缓存池”承载大规模KV Cache负责，把“context memory”单独做成一类AI-native存储基础设施。2026年1月，NVIDIA发布了推理上下文内存存储平台ICMSP（Inference Context Memory Storage Platform）。传统的G1-G3层级（本地HBM、本地内存、传统存储）已无法满足大规模推理的弹性需求，ICMS引入了“G3.5”层级，即专门为KV Cache设计Pod级共享存储层，当GPU本地HBM由于上下文过长而溢出时，NIXL（推理传输库）会自动将“温”KV数据异步转移到ICMS池中；当需要再次读取时，系统能以亚毫秒级的延迟重新加载这些数据。这种机制为智能体AI的“长期记忆”提供了硬件底座。

·CSP云厂用“CXL内存池化”把DRAM变成可以共享、池化和按需分配的系统资源，实现虚拟内存池灵活互联与调度。CXL 2.0及3.0协议通过支持内存池化和交换机架构，允许成百上千台服务器共享一个统一的动态内存池。对于AI推理任务而言，

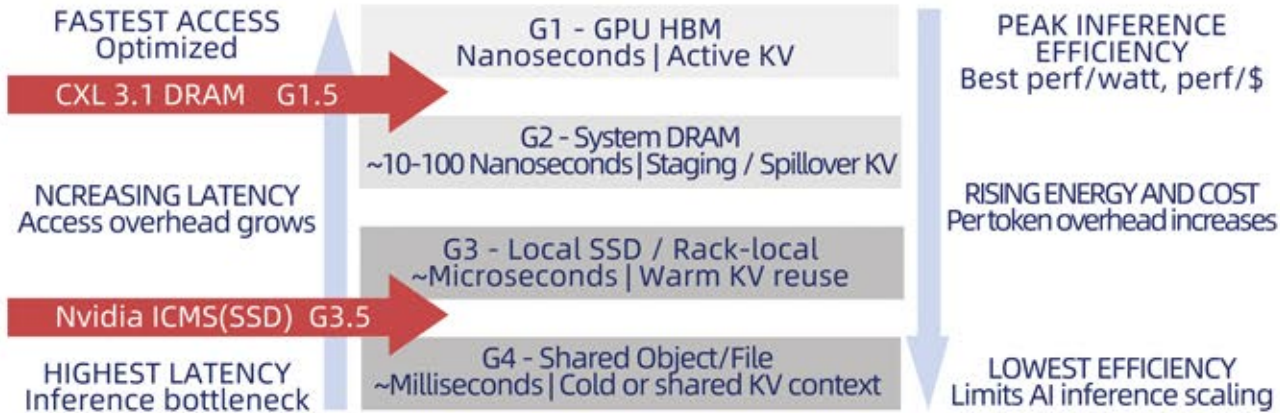
CXL不仅提供了容量扩展的路径，还引入了“近数据处理（Near-Data Processing, NDP）”的可能性。通过在CXL控制器中集成计算智能，可以在内存池侧直接完成KV Cache的压缩与量化。“CXL内存池化”在系统层为AI新增了一层“可调度、可共享、可扩容”的缓冲带。这层缓冲带，将成为未来AI数据中心里非常重要的中间层。

图：CXL内存池化



数据来源：CXL联盟

图：KV 缓存内存层次结构，从 GPU 内存 (G1) 到共享存储 (G4)

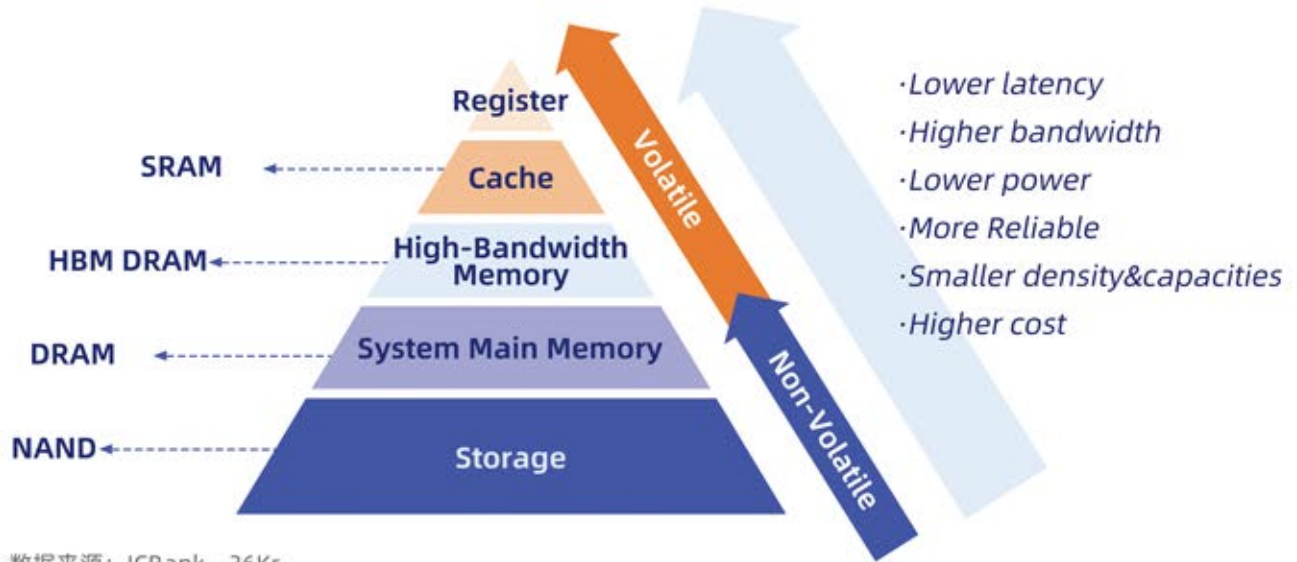


数据来源：英伟达

为了解决未来AI推理任务更极致的低延迟性能和实时响应，以及TB级甚至PB级的存储容量支撑，未来的GPU正在从单纯追求训练性能转向训练与推理能力的平衡，面对不同层级的存储器下不同性能-成本的表现，通过协同各层级的存储—如SRAM、DRAM

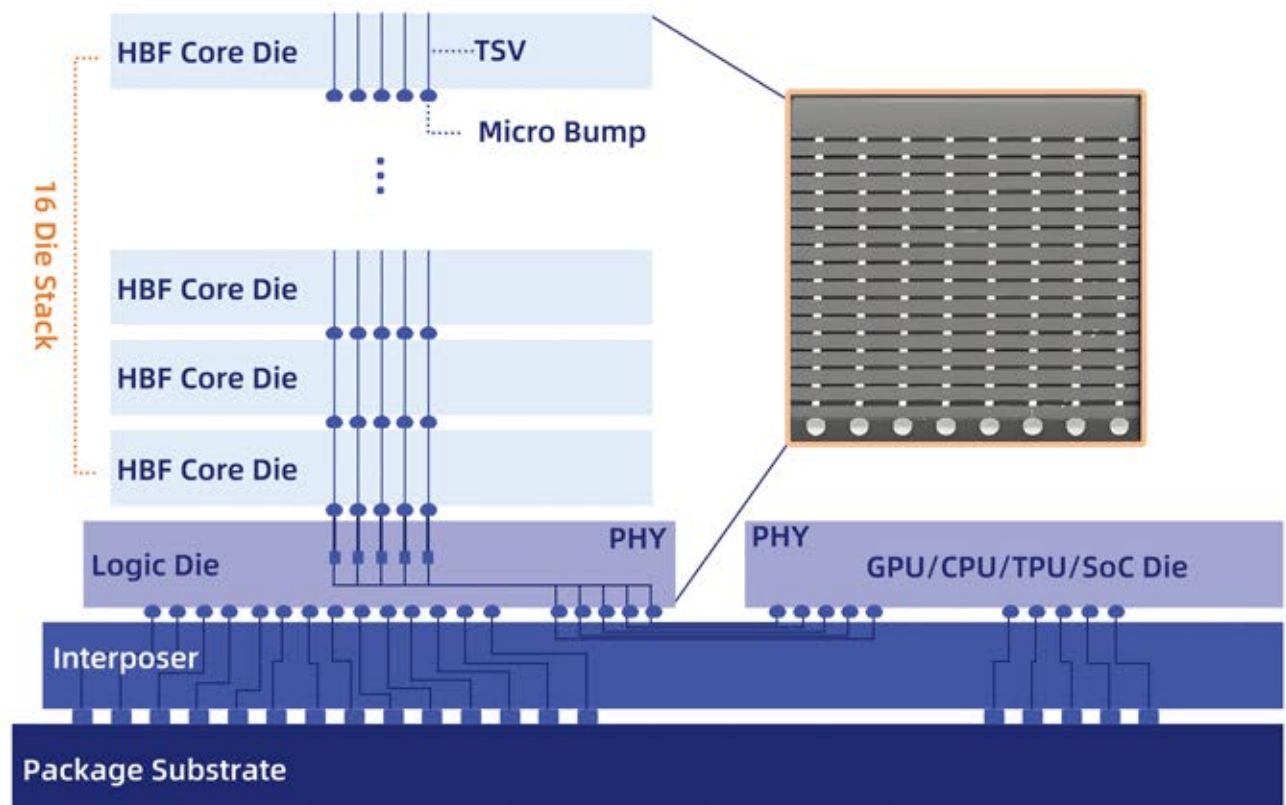
(HBM)、NAND(HBF)等—进行分工协作和融合, 可以实现GPU算力利用效率的最大化。

图: 分层存储架构



数据来源: ICBank、36Kr

图: HBF stack



数据来源: SanDisk

其中：

- SRAM层专注于处理“热”Token生成和最频繁访问的KV Cache状态。
- DRAM（HBM）层作为主带宽层，负责存储模型的主体权重和“温”上下文。
- NAND（HBF）层作为系统级的扩展内存，定位介于 HBM 和 SSD 之间的新 memory layer，兼顾 AI inference 所需的容量扩展和能效，结合 3D NAND 闪存与先进封装和互连技术，以提供接近高端内存的。

存储层级	存储类型	物理位置	核心功能	延迟特性
芯片内（LPU）	3D SRAM	GPU 核心正上方	瞬时执行层	<1ns
封装内	HBM	环绕 GPU 核心	核心工作层	~100ns
系统内	HBF	系统级存储	深度背景层	1μs-10μs

» 3.3.2 存储架构变革下的底层裂变：3D DRAM的4F²工艺和千层NAND

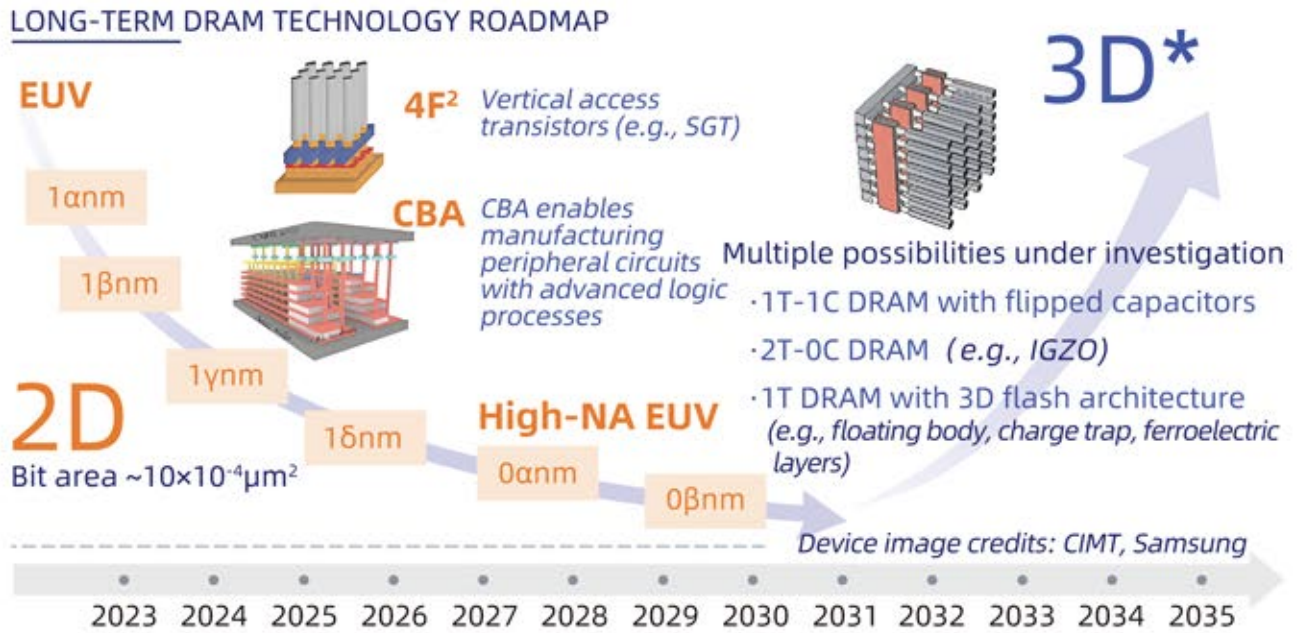
存储架构的创新将会落实到晶体管和比特位的效率提升上。为了在有限的物理空间内实现前所未有的高密度存储，存储介质正经历一场从二维平面向三维立体的升级。

DRAM方面：当前的DRAM技术已接近10nm工艺节点的物理极限，在制程微缩边际放缓的背景下，DRAM芯片正由制程演进转向以4F²与CBA为代表的架构创新。4F²技术通过将晶体管和电容器进行3D垂直堆叠，实现了平面内Cell（电容器+晶体管）尺寸的极限缩小。CBA技术实现了对DRAM性能的系统级优化，通过将存储阵列晶圆与逻辑控制单元晶圆分离制造并最终键合，使得逻辑和存储单元可以各自采用最优化的工艺流程。3D DRAM则更进一步，让位线或字线之一竖直排列，晶体管和电容器横向堆叠，层层叠加，该技术具备高带宽、低功耗、微型化和低成本的显著优势。

NAND方面：3D NAND堆叠层数努力向千层迈进，其扩展主要通过堆叠一层又一层的氧化物和氮化物层并在其中蚀刻孔（存储通道）来实现。美光第九代（G9）产品实现了276层堆叠，存储密度达35 Gbit/mm²。未来，在向千层3D NAND发展的路径上，每个晶圆都可以通过四种缩放实现：

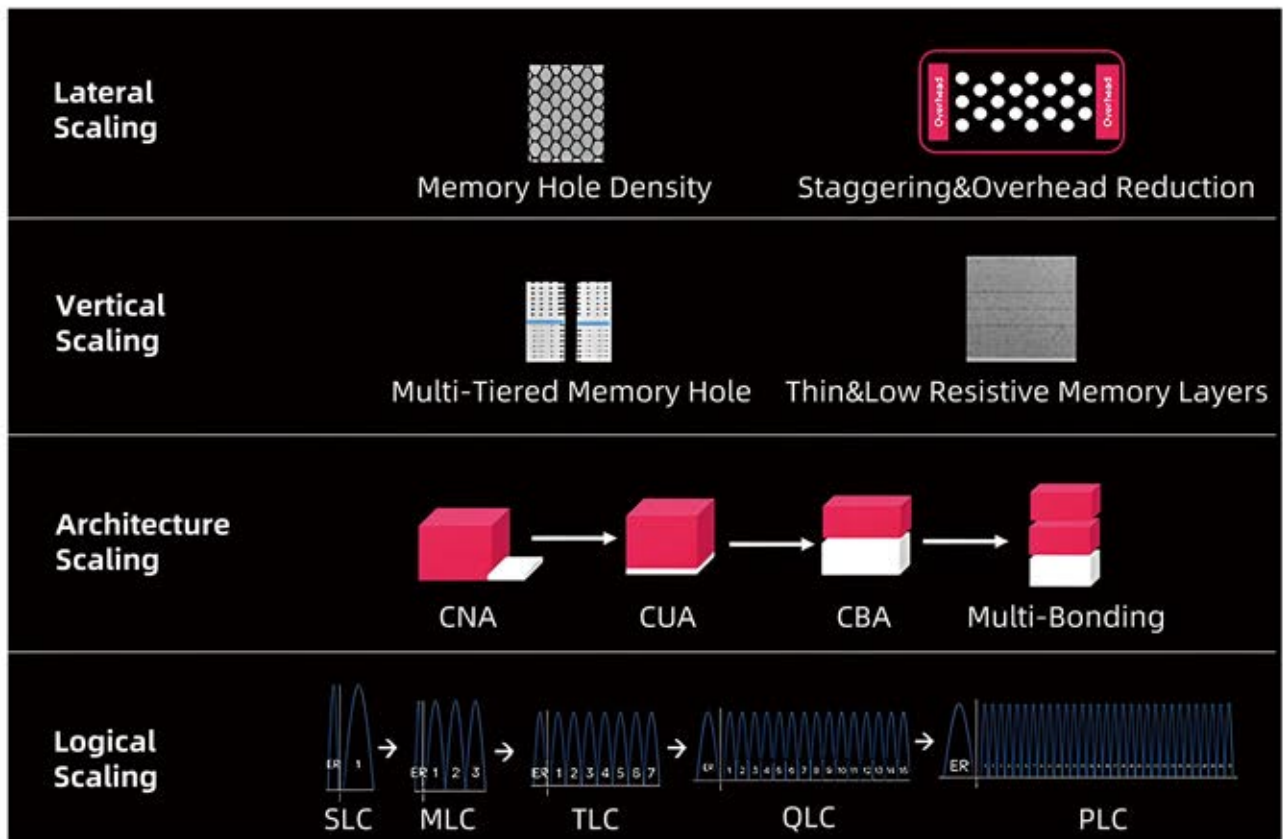
- 逻辑缩放（Logical scaling）：每个存储单元存多少 bit。需要在每个单元里存储 2^n 个电压电平。
- 垂直缩放（Vertical scaling）：NAND 单元在垂直方向堆叠的层数。
- 横向缩放（Lateral scaling）：在二维平面上，单元尺寸更小/可放入更多单元。
- 架构缩放（Architecture scaling）：各种提高密度、减少单元与外围开销的技术。

图：3D DRAM技术发展路径



数据来源：Yole

图：3D NAND技术发展路径



数据来源：西部数据

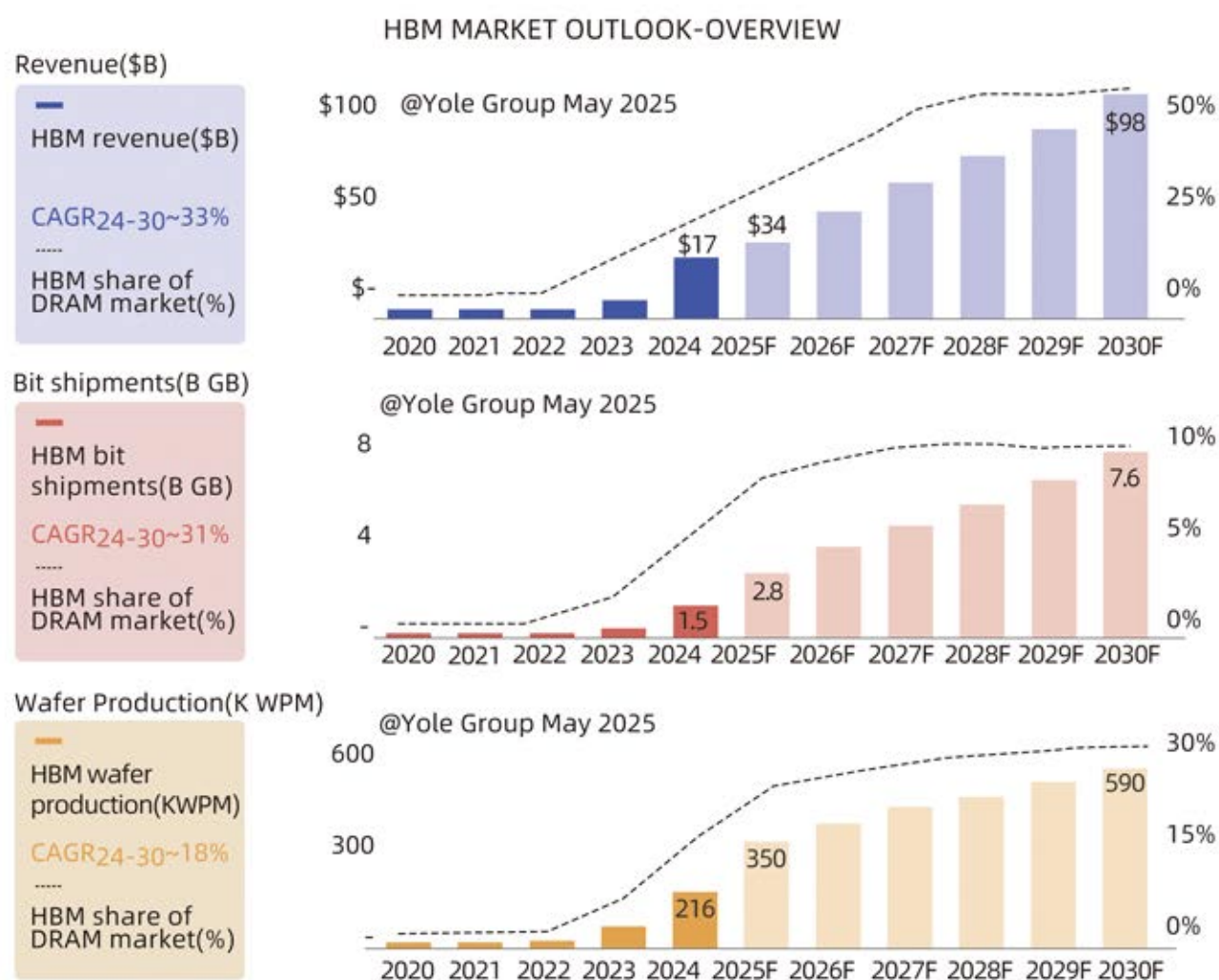
» 3.3.3 供需错配驱动，存储步入上行大周期

面对AI端显著的需求激增，目前存储呈现供不应求状态。根据Omdia数据，全球三大DRAM原厂——三星电子、SK海力士与美光在2026年的晶圆总产出预计将达到1800万片，同比增长约5%，远满足不了AI拉动的爆发式需求。目前DRAM的产能仅是订单的60%，服务器专用DRAM的满足率低于50%。

半导体制造存在固有周期。新建DRAM晶圆厂需2.5-4年，现有产线转产HBM或DDR5也需12-18个月。光刻机交期长达12-18个月，洁净室资源稀缺进一步限制扩产速度。部分厂商已在2025年大幅上调资本开支，新增有效产能要到2026年Q4才能显著释放。

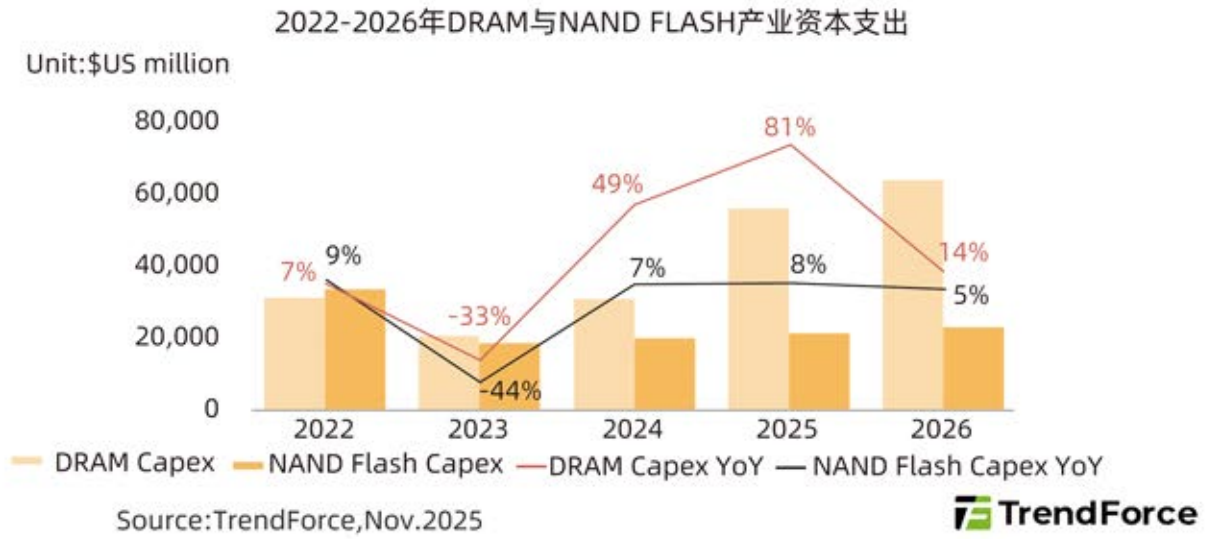
不仅存储供给总量不足，还涉及供给端结构性困境。三大原厂将先进制程与新产能分配给HBM和服务器DRAM，导致PC和智能手机等消费电子领域的供应被大幅挤压。

图：全球HBM市场预测



数据来源：Yole

图：全球存储厂商DRAM和NAND flash资本支出预测



数据来源：Trendforce

中国半导体的决定性突破

全链协同, 从被动替代到创新自强

4.1

被动替代: 封锁围堵下, 中国成功度过半导体自主自强上半场

» 4.1.1 人工智能发展强化半导体战略地位, 外部技术封锁成为国产替代核心驱动力

AI技术的爆发式增长让半导体成为全球科技竞争的核心抓手, 中美两国围绕这一领域的战略博弈全面升级, 从高端芯片到全产业链、从政策管控到生态构建, 较量贯穿始终, 深刻重塑了全球半导体产业格局。半导体作为AI算力的核心硬件基石, 其技术主导权直接决定了AI产业的发展主动权, 成为两国科技竞争的必争之地。高端AI芯片领域的对抗尤为激烈。2020年起, 为维持霸权, 美国持续加码出口管制, 多次修订《出口管理条例》, 将先进AI芯片及制造设备纳入禁售清单, 甚至通过“外国直接产品规则”实施长臂管辖, 将管制范围拓展至下级供应链, 试图切断中国高端算力供给和半导体产业发展路径。

图: 2020年以来美国对中国实行科技高压制裁

2020	2021	2022
4月 要求全球使用美国设备生产芯片的公司, 如果向华为供应产品, 必须先获得美国的许可	4月 美国总统拜登召集英特尔、台积电、三星等10家企业, 提出在芯片产业投入500亿美元重振美国芯片制造	2月 《2022美国竞争法案》在众议院通过
5月 美国商务部限制华为使用美国技术设计和生产的产品	6月 美国参议院通过《2021年美国创新与竞争法案》, 美欧成立美国-欧盟贸易和技术委员会	3月 美国政府着手联合韩国、日本和中国台湾组建Chip4芯片四方联盟
8月 收紧对华为获取美国技术的限制, 将华为列入“实体清单”	12月 美国通过了2022财年国防授权法案》(NDAA), 其中包含限制与中国军事和监视相关实体交易的条款	8月 美国总统拜登签署《2022芯片与科学法案》
9月 美国商务部对其华为及其子公司芯片升级禁令生效		10月 美国商务部发布针对中国先进计算和半导体制造出口管制新规
12月 美国众议院中美科技关系专家组发布《如何应对中国的挑战的限制》		12月 美国商务部宣布将长江存储等36家中国高科技企业列入“实体清单”

图：2020年以来美国对中国实行科技高压制裁

2023		2024		2025	
1月	美国政府向荷兰发出了强制指令，限制该国对中国的光刻机出口	3月	BIS公布出口管制政策修订预览版文件升级对华芯片出口限制	3月	BIS“实体清单”新增54家中国企业，含芯片设计、IP、EDA、量子、高超音速等
3月	将28家中国大陆企业和研究机构列入所谓“实体清单”	10月	宣布了禁止和限制美国对华投资“半导体和微电子技术、量子信息技术、人工智能技术”三个领域的《最终规则》	5月	推定全球使用华为昇腾等中国先进计算芯片违反GP10禁令；要求美国企业严防供应链转移；禁止美国EDA厂商为中国提供服务，管制授权密钥
6月	将43家公司添加到出口管制名单，其中31家实体的总部在中国	11月	美国已要求台积电从11月11日开始，停止向中国大陆客户运送常用于人工智能的先进芯片	10月	宣布“50%规则”扩大“实体清单”“连坐式”管制规则
8月	美国总统拜登授权财政部长监管美国在半导体、微电子和某些人工智能等领域对中国企业投资	12月	BIS将140家中国半导体相关公司放入“实体清单”分别是136家中国实体和4家海外关联企业	12月	对华半导体301调查结果发布，暂缓成熟制程芯片额外关税18个月保留后续加关税；并行推进232条款调查
10月	美国商务部工业和安全局(BIS)发布了针对芯片的出口禁令新规				

资料来源：公开资料整理

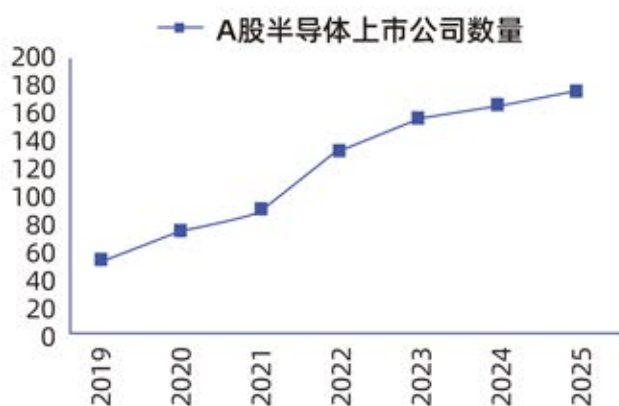
» 4.1.2 多维协同发力，保障产业发展

面对外部技术封锁与产业升级需求，中国从资金、政策、人才多维度精准施策，搭建起全链条、多层次的半导体产业支持体系。政策上构建“国家战略+地方协同”立体网络，形成全国统筹、区域联动格局。人才方面打通“产教融合+人才引进”通道，教育部推进集成电路一级学科建设，支持校企共建微电子学院；国家以专项计划引顶尖团队，地方给予人才表彰激励，建设国际培训基地，缓解行业人才缺口，为产业创新提供智力支撑。资金赋能层面，形成“国家基金+社会资本+资本市场”的多元支撑格局。国家集成电路产业投资基金持续加码，前两期基金已累计带动千亿级社会资本涌入，聚焦制造、设计、EDA、等核心环节，助力中芯国际、长江存储等企业扩产与技术攻坚。

科创板自2019年设立以来，显著推动了中国半导体产业的发展。科创板允许未盈利科技企业上市，有效缓解了半导体行业“高投入、长周期”的融资难题。截至2025年底，已有超130家半导体企业登陆科创板，涵盖设计、制造、设备、材料等全产业链，累计募资逾3500亿元，占科创板总融资额近30%。至2025年底，根据申万行业划分，A股上市半导体企业172家，总市值7.1万亿元人民币。自2020年初至2025年底，

半导体（申万）年化收益率15.6%，显著超过沪深300的2.13%，亦为投资者提供较好长期超额收益。

图：A股半导体上市公司数量（以申银万国行业划分）



资料来源：Wind

图：中国半导体自足率预测



资料来源：Morgan Stanley预测

4.1.3 自主可控显现成效，国内半导体产业规模和韧性跨越式增强

产业生态层面，面对美国持续的技术封锁，中国依托庞大内需市场和政策支持，系统性推动产业链自主可控渐显成效。国内头部晶圆代工企业实现28nm及以上成熟制程的稳定量产，产能快速扩张，基本满足国内需求；先进工艺也已进入规模化生产阶段。在半导体设备方面，刻蚀、清洗、薄膜沉积等关键设备国产替代顺利，本土企业技术突破和规模提升显著。材料环节如硅片、光刻胶、电子特气等也取得“0到1”或“1到N”的进展，逐步替代进口。受制于先进制程受限，中国同步布局2.5D/3D封装、Chiplet（芯粒）、晶圆级封装（WLP）、扇外型封装（Fan-Out）等先进封装技术，头部封测企业已具备大规模量产能力。此外，EDA工具、IP核、设备零部件等“深水区”领域开始布局。

4.2

主动创新：实现决定性技术攻坚，闭环中国人工智能产业生态

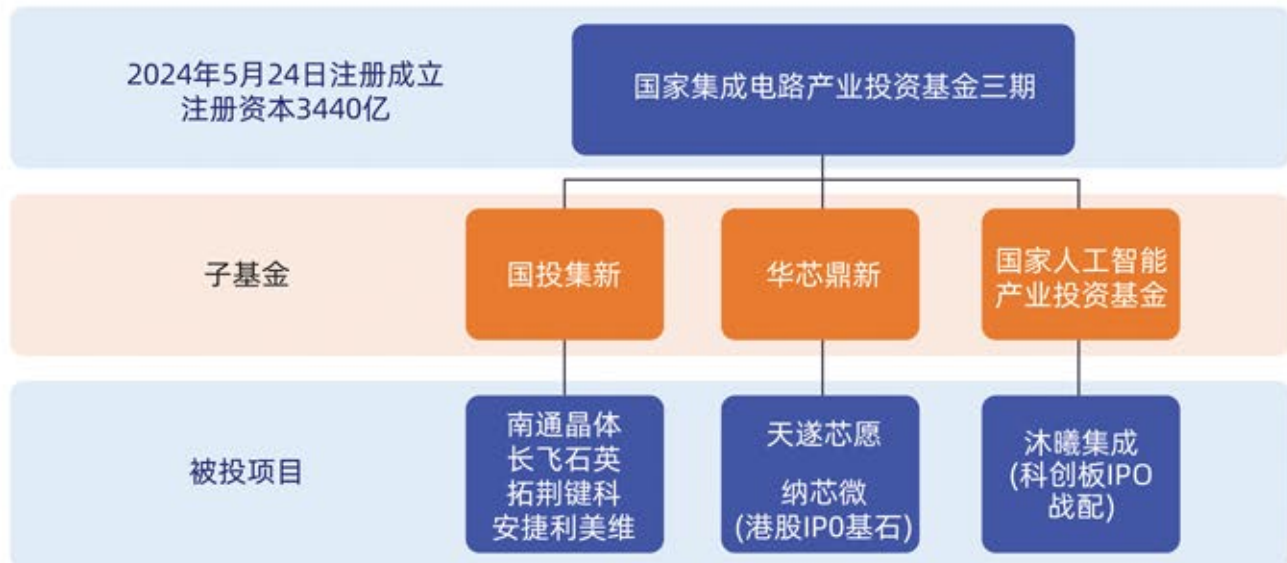
4.2.1 “十五五”规划明确全链条突破，大基金三期注入动力

半导体作为人工智能产业的核心算力根基，构成闭环的底层支撑，“十五五”规划推动实现半导体全链条自主创新，高质量中国人工智能生态因此将得以闭环。《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》将集成电路与工业母机、

高端仪器、基础软件等并列为“卡脖子”领域，以全链条突破、新型举国体制、超常规措施为核心路径，目标是实现决定性技术攻坚与产业生态自主可控，支撑科技自立自强与新质生产力发展。

大基金三期启动投资，将为未来五年中国半导体产业注入核心动力，加速自主可控进程与全球竞争力提升。截至2025年底，国家集成电路产业投资基金三期（简称“大基金三期”）已通过“母基金+子基金”架构，主导设立了至少3支核心子基金，并参与多支地方及专项基金。子基金中国家人工智能产业投资基金的设立，更清晰体现专业化子基金精准布局，重点投向半导体设备、关键材料、AI芯片与先进制造封装等方向，标志着中国半导体投资进入“强能力、攻卡点、建生态”的新阶段。

图：国家集成电路产业投资基金三期投资情况



资料来源：公开信息整理

» 4.2.2 本土化工艺制程突破，逐步构建自主可控AI芯片制造体系

AI产业的爆发式增长，正遭遇晶圆代工与先进制程的核心瓶颈，台积电的垄断地位被用于遏制中国人工智能。先进制程是AI芯片效率的核心支撑，7nm及以下节点能大幅降低芯片功耗、提升算力密度，而当前全球仅台积电具备规模化量产能力，成为瓶颈中的最大受益者。台积电凭借技术与产能壁垒，垄断全球95%以上的AI加速器芯片代工订单，包揽英伟达、谷歌等巨头的核心产能需求，市场份额不断提升。截至2025年第三季度，台积电营收占全球晶圆代工产业70.2%，市场份额相比本轮人工智能爆发前夕（2022年第三季度）提升14个百分点。同时，美国通过对台积电实行“长臂管辖”，制止中国芯片厂商使用台积电先进制程，达到遏制中国发展人工智能的目的。

图：台积电在AI时代市场份额大幅提升



资料来源: Trendforce

尽管由于技术封锁，中国在先进制程上与台积电仍有代差，但在前期我们已经完成“设计-制造-设备-材料”布局，依托本土化创新工艺制程，正逐步构建自主可控的AI算力芯片制造体系。根据SEMI预测，至2027年中国大陆仍将稳居全球半导体设备支出第一，主因本土晶圆厂仍在成熟制程及部分先进节点持续推进投资。随着制程节点步入美国限制范围，国产设备、材料份额预计将大幅提升。

图：近期国内晶圆厂动向

SMIC 中芯国际	华虹公司	晶合集成	华润微	behc 北电集成
<ul style="list-style-type: none"> ·2025年12月，公司公告拟作价406亿元全资控股中芯北方 ·2025年12月大基金三期领投、联合向中芯南方注资77.8亿美元 	<ul style="list-style-type: none"> ·2025年12月，华虹半导体公告定增募资76亿元收购华力微 ·2026年1月，无锡三期项目（Fab 9B）正式启动招标，规划12英寸月产能5.5万片 	<ul style="list-style-type: none"> ·2026年1月，四期项目正式开工，总投资355亿元，规划12英寸月产能5.5万片，锁定40nm/28nm节点 	<ul style="list-style-type: none"> ·截至2025年1月，深圳12英寸线已跨过建设期，正式进入运营阶段，聚焦40nm以上模拟及功率器件 	<ul style="list-style-type: none"> ·2024年11月，燕东微与京东方联合投资330亿元，预计2026年底实现量产，规划月产能5万片

资料来源: 各公司公告, 公开信息整理

» 4.2.3 “国芯+国模+国用”促国产AI芯片“可用”到“好用”

“国芯-国模-国用”战略的加速落地，为国产AI芯片搭建了从技术研发到场景验证的完整链路，推动其实现从“可用”到“好用”的关键性跨越。这一转变并非单纯的性能提升，而是芯片、模型、应用三方协同赋能，构建自主可控生态的系统性升级，彰显我国AI产业自主化的核心路径。

“可用”阶段聚焦技术破冰与替代补位，解决“有无”问题。此前国产AI芯片多能满足基础算力需求，但在性能、能效比及生态适配性上存在短板。“好用”的核心在于“国芯”与“国模”深度耦合、“国用”场景反向迭代。DeepSeek于2025年率先推动FP8低精度格式适配国产芯片，发布的DeepSeek-V3及V3.1模型不仅采用UE8M0 FP8参数精度，一定程度上也带动国内诸多芯片厂商推出原生支持FP8的产品。2026年1月14日，智谱发布全球首个完全基于国产算力底座训练的多模态SOTA模型GLM-Image，全程依托昇腾Atlas 800T A2设备及昇思MindSpore框架完成训练，通过动态图多级流水下发、多流并行等软硬协同优化，开源不到24小时即登顶全球知名AI开源社区Hugging Face Trending榜第一。

国内AI推理算力需求同样大幅扩张，看好国内ASIC发展：1) 通过深度定制化ASIC芯片很大程度弥补我们难以获取最先进制程的劣势；2) 国内互联网、汽车、智能终端等系统厂商也具备较强的研发能力和资本实力，有能力为ASIC研发配套的全栈软件生态，开发了一系列的编译器、底层中间件等，因此同样积极致力于自研ASIC芯片；3) 国内有较为成熟的ASIC定制业务产业链。

图：国产AI芯片在与国内模型、应用适配中加速迭代成熟

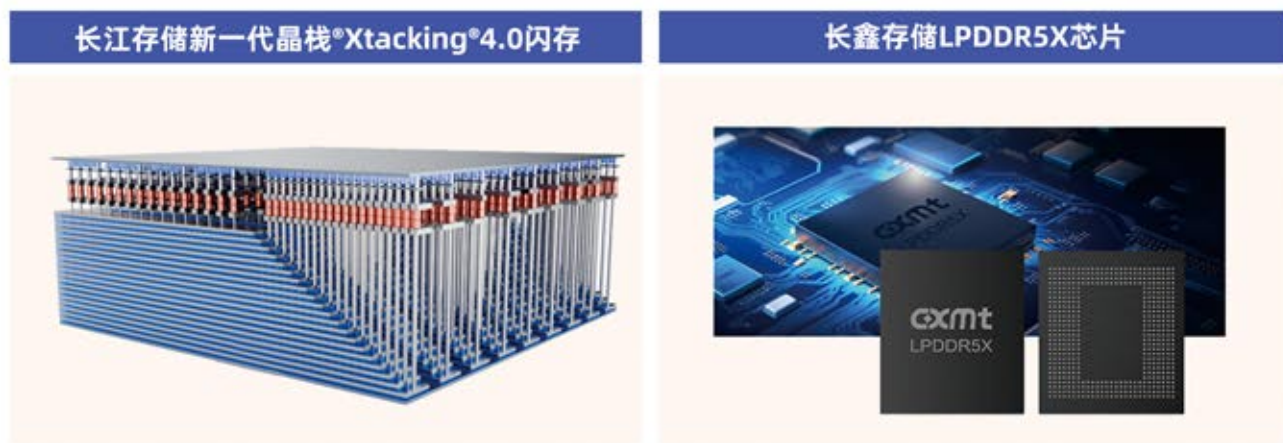


资料来源：新闻联播，DeepSeek，华为，公开信息整理

» 4.2.4 存储芯片在AI时代战略意义凸显, 国产存储技术和产能提速追赶

长江存储和长鑫存储技术逐步追上, 部分产品开始领先全球, 产能提速扩充。2025年8月, 长江存储晶栈Xtacking4.0创新架构在2025年未来内存与存储峰会(the Future of Memory and Storage, FMS)中荣获3D NAND“最具创新存储技术奖”, 拥有世界领先的存储密度、更快的I/O速度与更可靠的品质等优势。2025年10月, 长鑫存储在IEEE 2025 ASICON会议上公布其8533Mbps和9600Mbps速率的LPDDR5X产品已于5月量产。10667Mbps速率的产品已经启动客户送样。11月, 长鑫存储以“双芯共振, 5力全开”为主题, 首次全面展示DDR5和LPDDR5X两大产品线最新产品。与之相对应, 两家公司因产能瓶颈, 市占率仍低, 不能满足国内自主化需求的同时, 也难参与到全球定价当中, 先进存储产能扩充正在加速进行。

图: 长江存储、长鑫存储密集披露先进存储芯片



资料来源: 长江存储, 长鑫存储

» 4.2.5 迈向下个十年, 计算底层技术演变提供换道超车机遇

从2025到2035年, 十年技术迭代将驱动计算领域完成历史性蜕变, 也为中国半导体的追赶之路铺就了换道超车机会。华为在《智能世界2035》中提到, 未来十年间, 计算领域将迎来历史性变革——技术演进路径将从摩尔定律的延伸曲线, 逐步脱离传统冯·诺依曼架构的框架束缚, 最终催生新型计算范式的全面兴起。这一变革并非单点突破, 而是将在材料器件、工程工艺、计算架构、计算范式四大核心层面, 协同实现颠覆性技术创新, 重塑“后摩尔时代”的算力生态, 这正是中国半导体的超车契机。

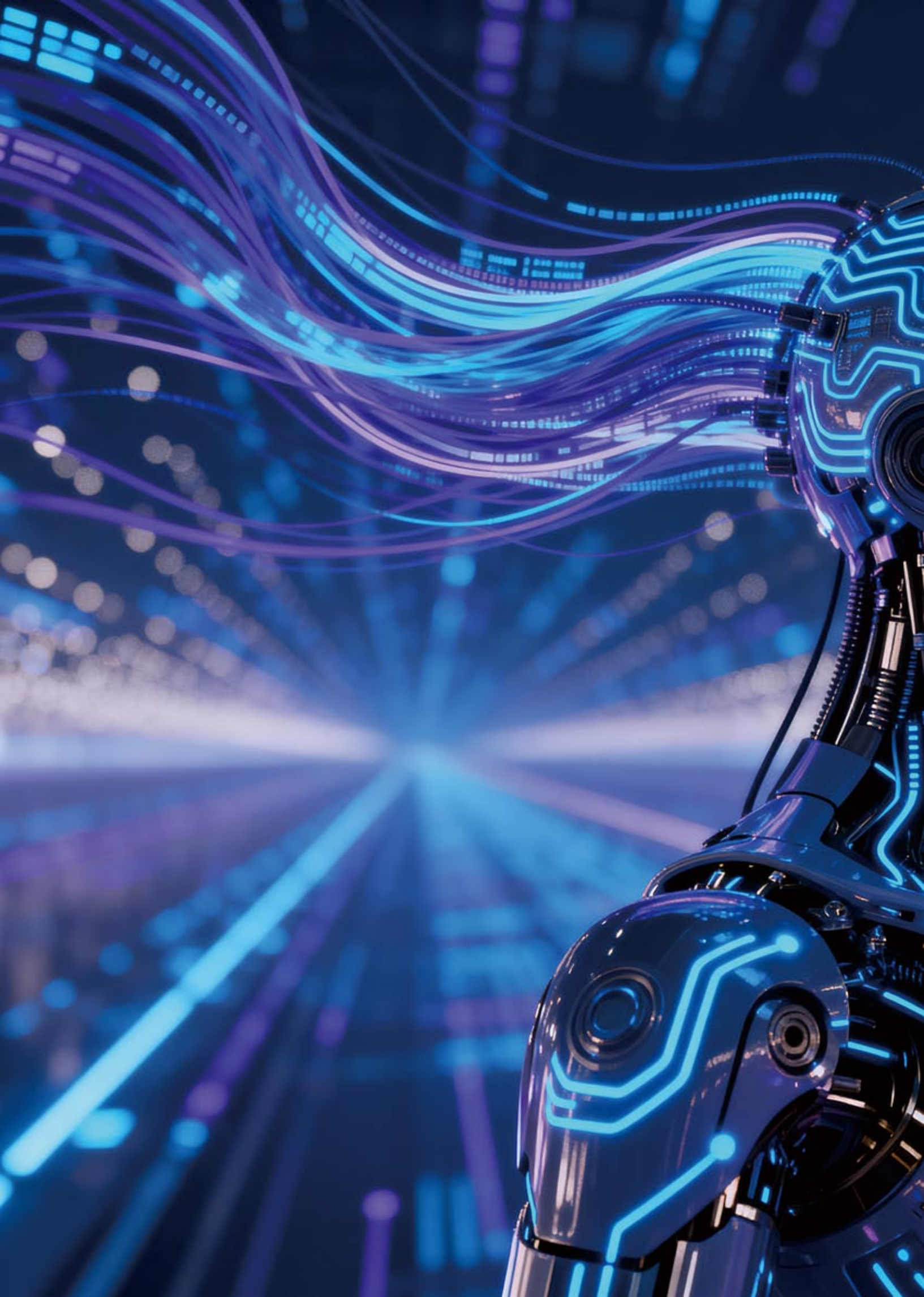
根据华为预测, 半导体材料器件将从“硅基主导”到“多元融合”, 即新型材料

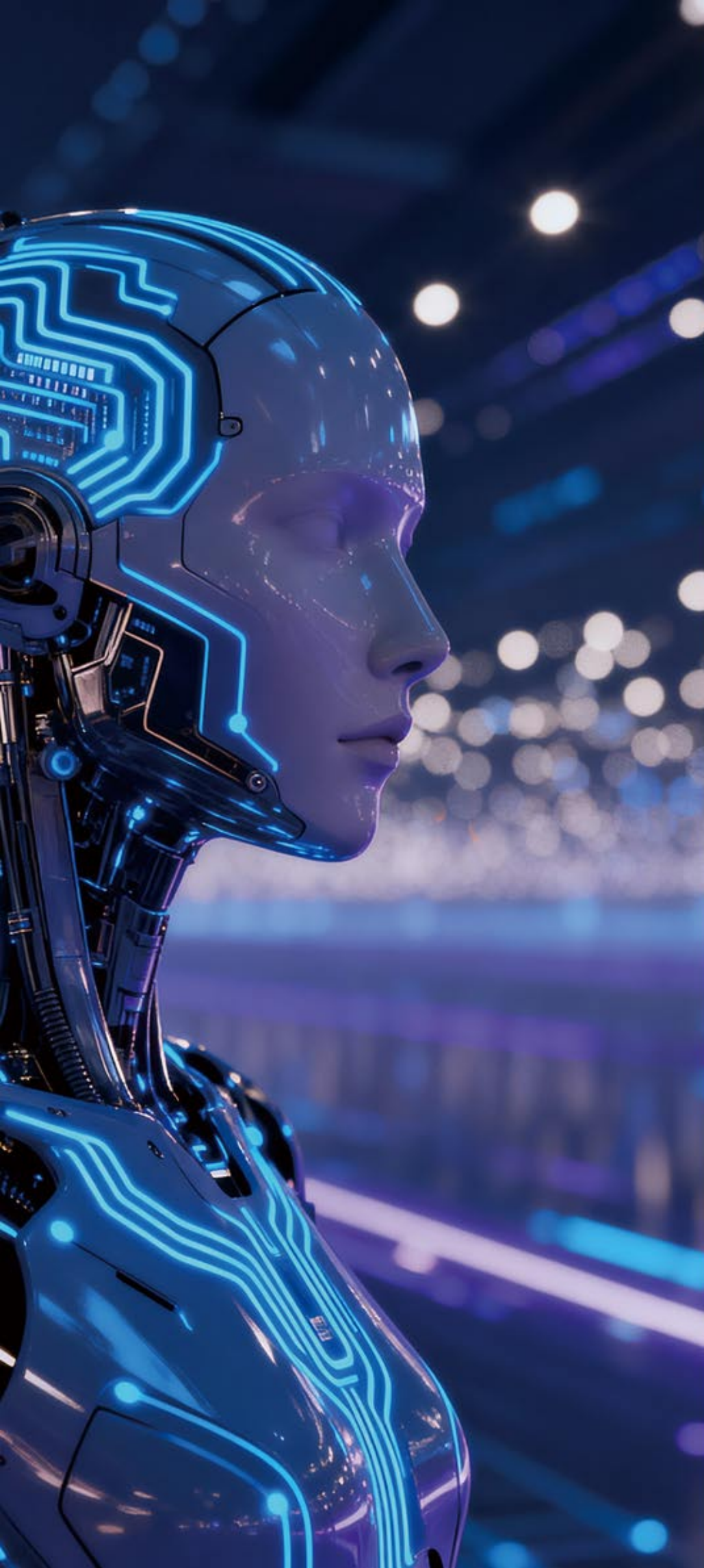
赋能功率与存储器件，低维材料突破硅基物理极限，解决尺寸瓶颈问题；工程工艺从“制程依赖”到“多维创新”：亚纳米时代，制程借CFET、M3D技术突破物理极限；算力密度提升转向工程多维革新，核心抓3D异质集成堆叠、晶圆级计算、堆叠配套技术三大方向；计算架构从“存算分离”到“存算协同”：冯·诺依曼架构的存算分离瓶颈将被突破，技术路径分三步推进：异构算力对等化提效、近存计算缩短数据搬运距离、存内计算消除“内存墙”，大幅提升计算速度与能效。

图：研究新材料、新器件、工程工艺、新架构、新范式



资料来源: 华为《智能世界2035》





2026年

诺安基金科技投资报告

燃

动
未
来

02

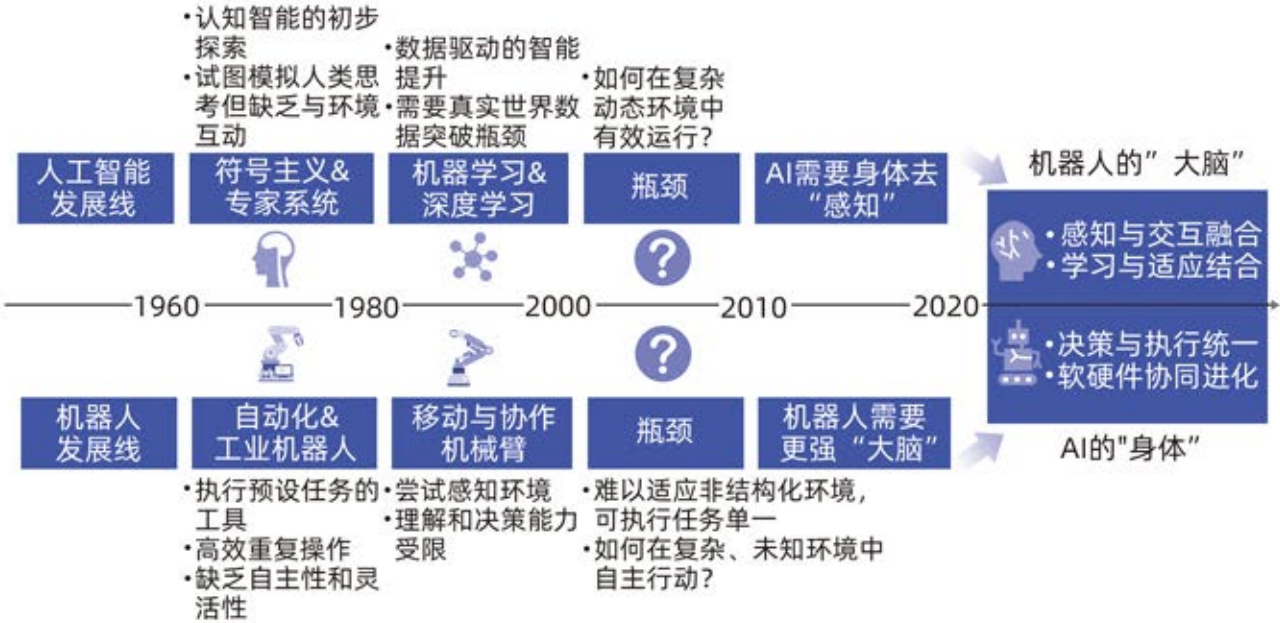
人形机器人 具身智能的“iPhone时刻”

当前行业正处于由“概念导入期”向“产业化突破期”过渡的关键节点，已经跨越了“死亡之谷”，对应Gartner技术成熟度曲线的爬坡段。当前技术方向逐步收敛、产品形态趋于稳定，市场认知从“猎奇式关注”转向“实用性评估”，行业发展不再依赖噱头，而是进入技术落地和市场验证的阶段。

1.1 产业拐点——从宏观叙事看周期定位

软硬双重突破，奇点临近。AI赋予“大脑”，供应链重塑“骨骼”，商业闭环源于软硬件质变：软件端，VLA大模型赋予机器人理解与拆解任务的泛化智能；硬件端，复用电动车产业链（电池、热管理）并依托中国制造规模优势，正推动量产成本大幅下探。

图：具身智能是机器人与人工智能发展的交汇，也是人形机器人通用化的关键

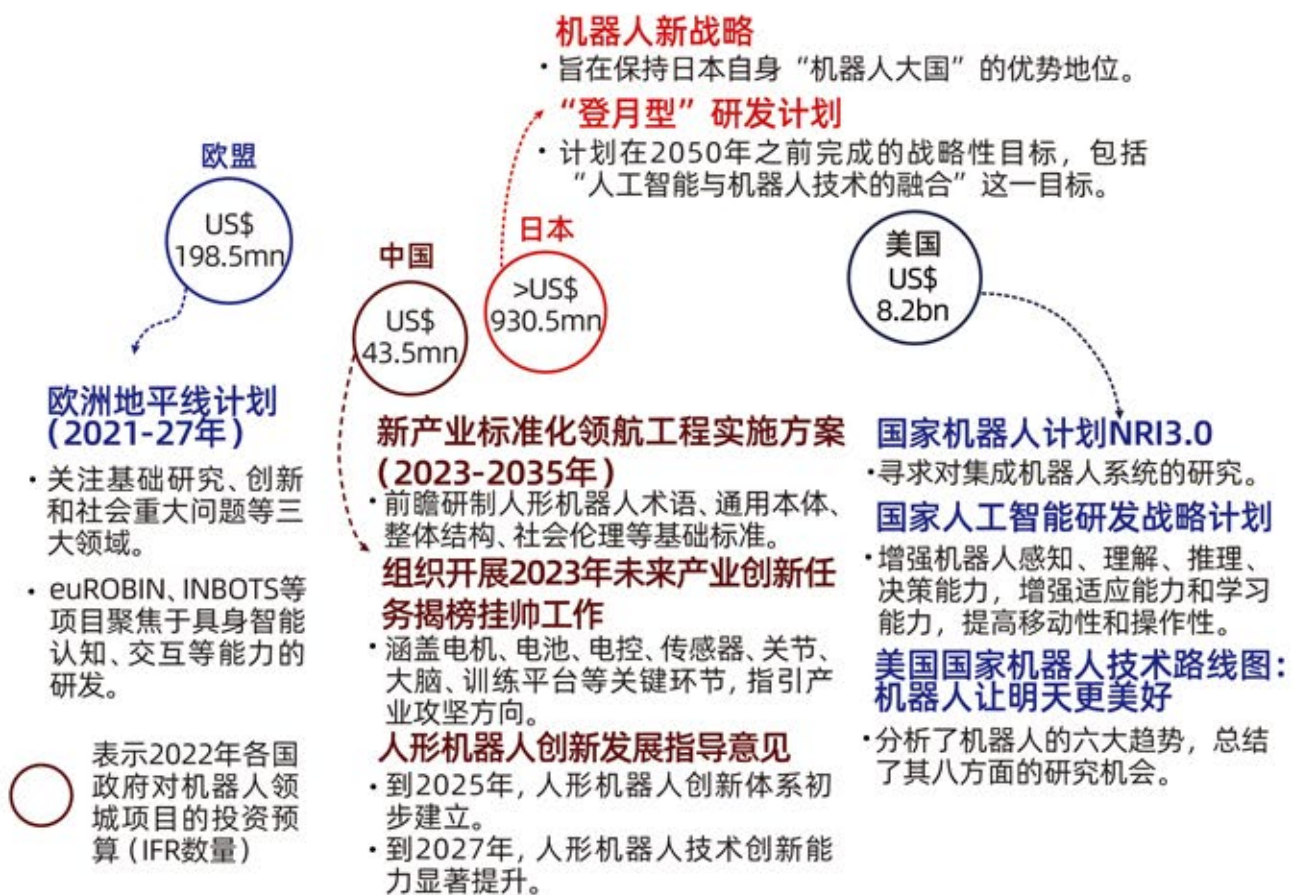


资料来源：甲子光年《具身智能行业发展研究报告》，2025.1，觅途咨询官网，中金公司研究部

全球人口结构倒逼，“机器人红利”接棒。机器人是填补全球劳动力缺口的必要工具。随着青壮年劳动力持续减少，预计2030年全球劳动力缺口将导致约8.5万亿美元产出损失。“机器人红利”将接棒“人口红利”，成为应对老龄化与弥补生产力的关键解法。

新大国重器：一场不容有失的“国力之战”。国家战略将人形机器人定义为继计算机、手机、新能源车后的“颠覆性产品”。随着“具身智能”确立为新质生产力代表，顶层设计确立了2027年构建国际竞争力生态的目标，为产业连续性提供强力支撑。

图：中国机器人产业政策梳理（截至2025年12月，不完全统计）



资料来源：工信部，发改委，国家标准委，财政部，民政部，国家卫健委，科技部，质检总局，中国机械工业联合会，中金公司研究部

1.2 标杆引领——特斯拉Optimus量产进程

马斯克构建文明闭环的最后拼图。Optimus是马斯克未来文明构想的基础设施，与能源、脑机接口、太空探索构成逻辑闭环。其终极目标是作为火星殖民先遣队建设生存设施，并消除人类贫困与重复劳动。

量产时间表与产能规划。Optimus V3定于2026年Q1发布，产品定义进入最后打磨。马斯克规划2026年底建成100万台V3产线，远期目标售价控制在2万美元以内（低于美国人均年薪），具备极强经济替代效应。

图：特斯拉人形机器人发展历史复盘

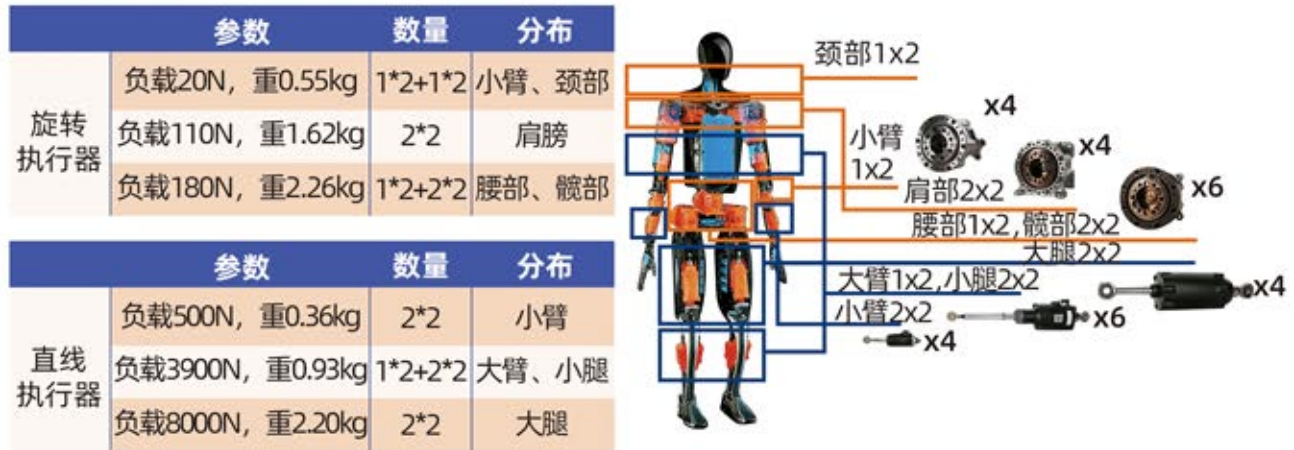


资料来源：特斯拉AIDAY, TeslaAI官方社交媒体平台X（前称Twitter）账号，特斯拉官网，中金公司研究部

供应链筛选：从宽泛到严苛。T链筛选标准转向“量产一致性”与“极致成本”。供应商名单进入“窄门筛选”期，马太效应加剧，资金加速向进入核心圈的Tier 1企业集中，核心供应商地位逐步固化。

中国制造的“入场券”。中国供应链凭成本与响应优势，有望承接T链核心份额。丝杠（行星滚柱/微型滚珠）与执行器总成（关节/头部）是国产化率较高且确定性高的赛道，未来或有望在放量期实现业绩爆发。

图：特斯拉OptimusV2关节总成示意图



资料来源：特斯拉AIDay，中金公司研究部

1.3 技术攻坚——核心博弈与关键增量

标准化前夕的技术分歧。当前，人形机器人产业正处于技术标准化的关键转折期。尽管工信部已成立标准化技术委员会，但截至目前，机器人“身体”尚未完全定型。在产业标准化定型前夕，各厂商技术路线的分歧（如灵巧手、电子皮肤、机器视觉方案）正是创造超额收益的博弈点。

灵巧手是差异化竞争的核心抓手，其工程量占整机近半，是机器人从单纯“行走”迈向实质“工作”的关键跨越。作为高价值量、高壁垒的核心子系统，其成熟度直接决定了产品的商业化上限。**触觉传感器是灵巧手的“灵魂”。**面对物理交互的复杂性，纯视觉方案局限明显，必须引入电子皮肤与MEMS力控技术。这些技术能精准感知多维力与温度，使机器触觉无限趋近人类，是解决机器人“能干”问题的核心解法。轻量化是解决续航瓶颈的必由之路。在电池技术尚未飞跃的当下，以PEEK为代表的高强度、自润滑新材料，正逐步替代传统金属与塑料，成为构建机器人高效“骨骼”与“皮肤”的关键选择，为提升整机运动性能提供支撑。

图：灵巧手主要应用场景



资料来源：因时机器人官网，中金公司研究部

表：电子皮肤技术路径优缺点对比

技术	调节参数	优点	缺点
压阻式(FSR)	电阻值	动态范围大，频率响应高，空间分辨率高，耐用，原材料易得，测量电路简单	涂层剂纯度低，填料粒径误差大；国内油墨还没有实现量产
电容式	电容值	灵敏度高，空间分辨率高，动态范围大	存在寄生电容，对噪声敏感，测量电路复杂；可弯曲性不足
压电式	电荷	灵敏度高，动态范围大，经久耐用，材料机械性能好，同时感应力和温度变化	仅适用性高动态检测：空间分辨率差；灵敏度过高，易受非目标物体和高频信号影响；测量电路复杂；材料特性衰减
光电式	光强度	空间分辨率高，无电气干扰，响应速度快，成本低	整体结构缺乏柔性，对弹性体依赖性强
电磁式	磁场强度	灵敏度高，体积小，可感知切向力	结构设计复杂，分辨率低
超声式	超声波	空间分辨率高，不受电磁干扰	布线困难，存在滞后和非线性，易受外部超声干扰

资料来源：《柔性触觉传感器在机器人上的应用综述》（传感器与微系统，2015），中金公司研究部

1.4 群雄逐鹿——中国新势力与商业化爆发

中国市场的“百花齐放”。不同于特斯拉一枝独秀，中国呈现多元竞争格局。技术派初创（智元、宇树）、传统转型力量与跨界巨头（小米、华为）并存，配合国家创新平台，形成独特的产业生态。

图：人形机器人竞争主体一览（截至2025年12月，不完全统计）



中国企业的差异化策略。中国厂商策略务实：一是极致性价比，利用国内供应链控本；二是场景驱动，优先切入汽车、3C、商超积累数据；三是坚持供应链自主可控，防范“卡脖子”风险。

表：国内外人形机器人融资事件梳理（截至2025年12月，不完全统计）

厂商	时间	轮次/进程	资金额度/募资规模	投资机构/上市相关	投后估值
银河通用机器人	2024年6月	天使轮	7亿元人民币	美团战投、北汽产投、启明创投、蓝驰创投等	未披露
	2024年11月	战略轮	5亿元人民币	上汽恒旭投资、HKIC、上海人工智能产业基金、深创投、IDG等	未披露
	2025年6月	战略轮	11亿元人民币	宁德时代领投，老股东跟投	未披露

表：国内外人形机器人融资事件梳理（截至2025年12月，不完全统计）

厂商	时间	轮次/进程	资金额度/募资规模	投资机构/上市相关	投后估值
银河通用机器人	2025年12月	战略轮	超3亿美元(约21.5亿元人民币)	中国移动链长基金领投, 中金资本、中科院基金、苏创投、央视融媒体基金、天奇股份及新加坡/中东国际资本等	未披露
宇树科技	2024年2月	B2轮	近10亿元人民币	美团、金石投资、源码资本, 老股东深创投、中网投等跟投	未披露
	2024年9月	C轮	数亿元人民币	北京机器人产业基金、成都龙珠基金、中信证券投资等	未披露
	2025年Q4	Pre-IPO辅导完成	拟募资	中信证券辅导, 具备境内IPO申报资格	30亿美元(约210亿元人民币)
智元机器人	2024年全年	多轮战略投资	超20亿美元	腾讯、京东、比亚迪、上汽、北汽、TCL、正大机器人等	未披露
	2025年9月	反向收购	未披露具体金额	完成对上市公司上纬新材的收购, 获63.6232%股份	未披露
星动纪元	2024年10月	Pre-A轮	/	清流资本、元璟资本、阿里巴巴联合领投, 策源资本等跟投	120亿元人民币
优必选	2023年12月29日	港交所上市	近3亿元人民币	公开发售+国际配售	未披露
	2024年8月起	多次配售+信贷	IPO募资约10亿港元	Infini Capital、多家银行	150亿元人民币
	2025年12月	反向收购	获InfiniCapital	拟通过“协议转让+要约收购”方式受让锋龙股份43%股权实现控股	未披露
越疆科技	2024年12月23日	港交所上市	10亿美元战略融资授信、多轮银行信贷	公开发售+国际配售	未披露
鹿明机器人	2025年12月	Pre-A1、PreA2轮	/	Pre-A1:鼎晖投资领投;Pre-A2:申能诚毅投资	未披露
星尘智能	2025年11月	A++轮	IPO募资	国科投资、蚂蚁集团联合领投, Bloom Advance Capital等跟投	16.65亿元人民币
乐聚智能	2025年10月	Pre-IPO辅导备案	约5亿港元	深圳证监局辅导备案, 正式开启IPO	未披露
七腾机器人	2025年12月	反向收购	数亿元人民币	拟以“协议转让+部分要约”方式受让胜通能源44.99%股权实现控股	未披露
追觅科技	2025年12月	反向收购	数亿元人民币	创始人拟通过“协议转让+要约收购”的方式受让嘉美包装54.9%股权实现控股	未披露
Robo-Science	2025年3-7月	天使轮	未披露募资规模	京东、招商局创投、商汤国香资本、零一创投	未披露
FigureAI(美国)	2024年2月	B轮	/	微软、OpenAI Startup Fund、英伟达等	/
	2025年9月	C轮	/	Parkway Venture Capital领投, 英伟达、英特尔资本、高通创投、Salesforce等	/
1X Technologies(挪威)	2024年初	B轮	近2亿元人民币	微软、OpenAI Startup Fund等	未披露

表：国内外人形机器人融资事件梳理（截至2025年12月，不完全统计）

厂商	时间	轮次/进程	资金额度/募资规模	投资机构/上市相关	投后估值
1X Technologies (挪威)	2025年5月	C轮	2.5亿美元	微软、OpenAI Startup Fund等	未披露
Agility Robotics (美国)	2024年10月	C轮	1.5亿美元	亚马逊、福特汽车等	未披露
Apptроник (美国)	2024年	多轮融资	累计超3亿美元	谷歌、亚马逊等	未披露
	2025年7月	D轮	3.2亿美元	谷歌、亚马逊、福特汽车等	未披露
Physical Intelligence (美国)	2025年11月	大额融资	6亿美元	未披露详细机构	未披露

注：粉色行为机器人企业方向收购上市公司股权的事件

资料来源：36氪，证券时报网，中证网，中国证券报，中国财富网，港交所披露易，银河通用官方公众号，宇树科技官网，智元机器人官网，FigureAI官网，1XTechnologies官网，AgilityRobotics官网，Apptроник官网，TechCrunch，路透社，相关公司公告，中金公司研究部

商业化订单与交付破局。2025年上半年全行业订单近3万台，头部企业订单破5亿，且从框架协议转向实质履约。千台级量产出货能力的形成，标志着工厂制造与商业服务场景正式开启商业化落地。

资本化浪潮加速。2025年一级市场融资达235.98亿元。随着越来越多企业的上市与资本运作，上市潮不仅为企业输血，更为投资人打开退出通路，重塑板块估值体系。

未来展望

2026年行业步入爆发前夜，人形机器人投资应聚焦“量产确定性”与“技术高壁垒”。特斯拉核心供应链龙头，新技术关键增量与国产独角兽上市将带来行业的估值重塑。

脑机接口

从科幻到产业化临近

脑机接口（BCI）是把脑信号转译为可执行的数字/物理指令（或反向把刺激写回神经系统）的“神经信息通道”，短期以医疗康复为主，长期可能成为人类与AI/机器人交互的底层接口。BCI的“产业本质”是：高端医疗器械 + 手术/耗材 + 长期软件订阅/随访服务 + AI模型+场景的组合商业模式。BCI不像纯软件，其规模化取决于手术吞吐量、长期安全性、监管路径、医保/商保支付、随访与软件服务定价；医生/医院容量可能是一种瓶颈，即使在手术机器人的加持下。BCI已进入“政策-临床-产业化共振”的早期窗口，但未来3-5年更像“高强度研发+临床证据竞赛”，而非立刻大规模放量。

2.1 产业链生态图谱

四层结构：



神经信号“入口”硬件
——核心器件壁垒



植入与术式体系
——规模化的“制造业化手术”



解码算法与“神经AI”
——数据飞轮壁垒



应用与分发渠道
——生态决定长期收入

图：产业链生态图谱



资料来源：公开资料整理

·未来技术将向“高通量、无线化、柔性化”演进。例如，Neuralink计划在2026年将电极通道数扩展至3000个，2027年达10000个，以实现全脑高带宽通信。

·BCI最先落地的场景，几乎一定是医疗健康。因为它解决的是刚需：沟通、运动、康

复、神经调控。其次是一些B端场景，比如工业监测与人机协作，但要非常谨慎对待伦理与合规。最后是消费级：XR、游戏等，但兑现更慢，因为它受生态入口与体验一致性制约。

·真正的“生态”构建，至少具备五个元素：1) 硬件与手术体系；2) 算法与数据平台；3) OS级输入与应用层；4) 医疗服务网络；5) 支付与合规。

2.2 脑机接口发展空间

1) 增长路径：

- 第一阶段（2025-2030）：严肃医疗为主。侵入式BCI在瘫痪、盲人复明等重症领域实现小规模临床应用；非侵入式在康复、睡眠领域提高渗透率。

- 第二阶段（2030-2040）：消费级产品普及加速。随着成本下降和手术微创化，BCI进入游戏、办公等大众市场。

- 第三阶段（2040+）：人机共生。双向高带宽通信实现，人类智力增强，与AI深度融合。

2) 严肃医疗预期规模

- 2030-2045：植入硬件累计收入约120亿美元，累计渗透约2.8%，约27万人植入，潜在候选约970万人。预估植入体硬件 ASP 可能在2.5万-6万美元。

- BCI的ROI不只来自“硬件毛利”，其价值捕获至少来自三条现金流：1、一次性的植入器械与手术服务费；2、重复性的软件/算法升级、个体化训练与随访、云端服务；3、长周期的更换与迭代（10-15年更换周期是合理假设区间）。

3) 医疗端细分病症价值评估（累积销售额/市场价值，美元）

适应症	早期患者基数	2035累积市场价值	2045累积市场价值
MND/ALS	12,500	\$1.14 亿	\$3.83 亿
中风	10,000	\$0.61 亿	\$1.84 亿
脊髓损伤	45,000	\$1.36 亿	\$6.89 亿
多发性硬化	7,500	\$0.45 亿	\$1.61 亿
脑瘫	5,000	\$0.24 亿	\$0.77 亿
截肢(幻肢痛)	75,000	\$4.55 亿	\$11.48 亿
重度抑郁症	2,500,000	\$3.16 亿	\$25.50 亿

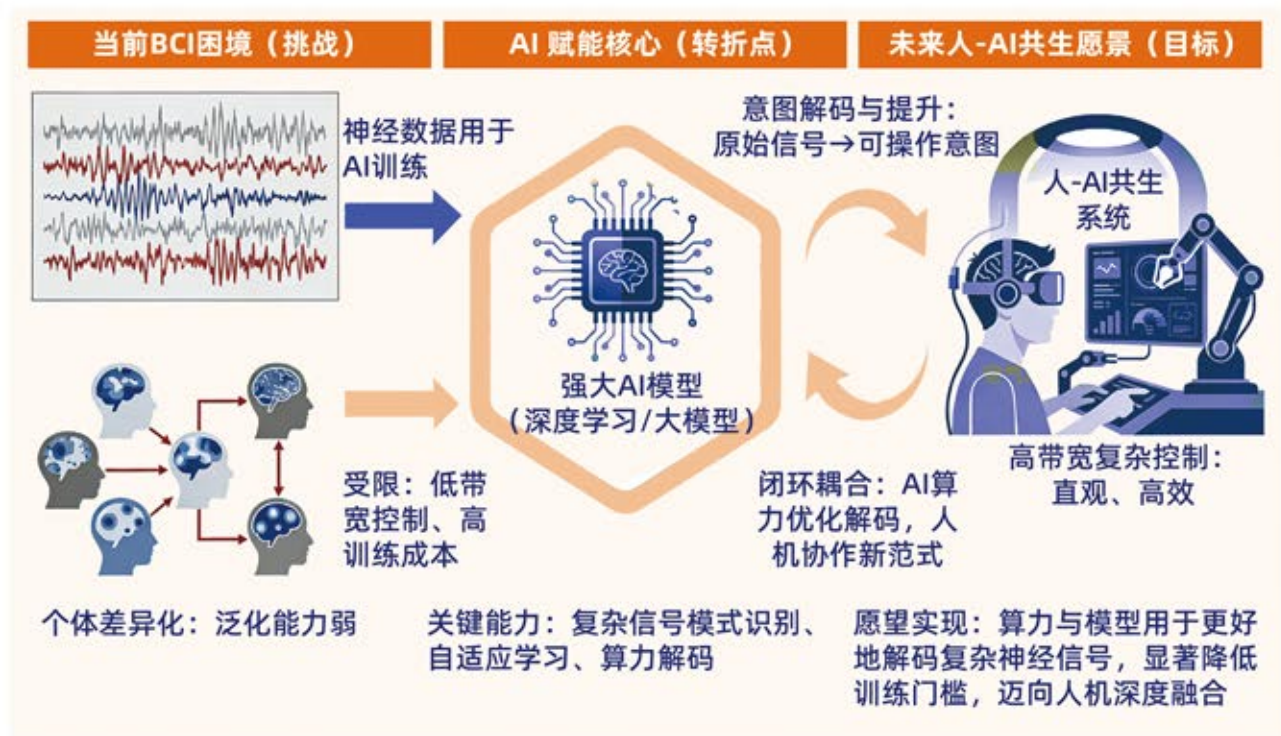
数据来源：公开资料整理

2.3 BCI的“真正生产力”来自AI

1) AI是BCI的“放大器”，BCI最大难点之一是脑信号高度噪声化、非平稳、个体差异大；没有足够强的AI，BCI往往只能做低带宽控制；而AI能把“原始信号”提升为“可操作意图”，并显著降低训练成本。未来假设神经数据可用于训练AI模型，AI算力与模型用于更好地解码复杂神经信号，并指向“人-AI共生”愿景。

2) “神经基础模型”可能成为下一代护城河，类比语音/视觉，当行业从“小数据+个体化训练”走向“大数据+基础模型”，会出现平台型赢家。关键前提是：大规模、高质量、可标注的神经数据集；统一的数据标准与监管认可。

图：BCI的未来发展



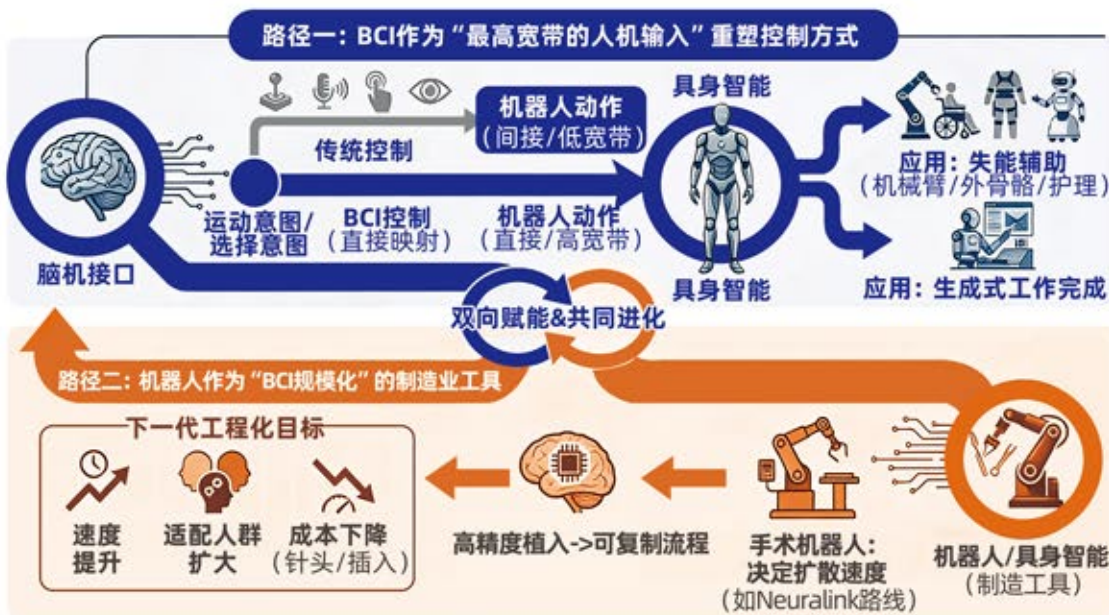
资料来源：公开资料整理

2.4 脑机接口与具身智能

BCI与机器人有两条主线关系：1) BCI作为“最高带宽的人机输入”，将重塑人对机器人的控制方式；2) 机器人反过来是BCI规模化的“制造业工具”，手术机器人决定扩散速度，Neuralink路线最典型：用机器人把高精度植入变成“可复制流程”。其下

一代机器人在速度、适配人群、成本等方面的工程化目标（针头成本显著下降、插入速度提升）。

图：脑机接口与具身智能的未来发展

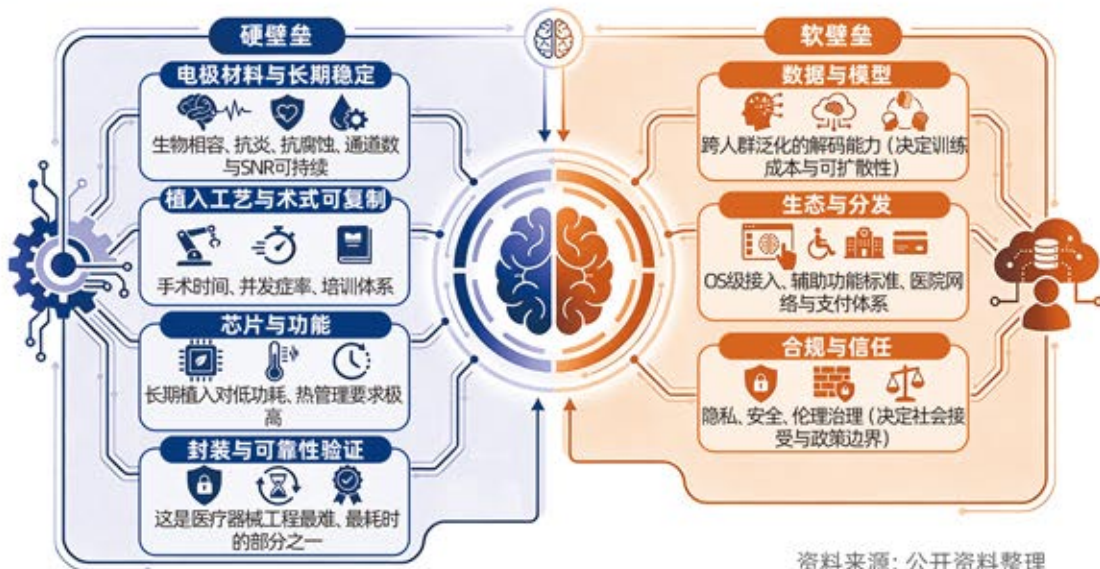


资料来源：公开资料整理

2.5 核心壁垒、驱动因素、难点与风险、未来演进方向

1) 核心壁垒

把壁垒分成“硬壁垒”和“软壁垒”：



2) 产业发展的驱动因素（未来5-10年“为什么会加速”）

- 需求端：重度失能人群对沟通与自主性的强刚需（替代方案有限）
- 供给端：AI解码能力提升+端侧算力/低功耗芯片进步+手术机器人推进
- 制度端（中国）：国家层面产业意见明确时间表与目标（2027/2030）；医保先立项、地方先试价，打通未来收费路径；药监体系推动数据与算法标准、并将“植入式脑机接口医疗器械”纳入高端器械优先审批目录。

3) 难点与风险

- 临床风险：侵入式手术并发症；电极衰减/位移导致功能下降甚至二次手术。
- 医生/医院容量：神经外科培训与吞吐量决定扩散上限。
- 支付风险：硬件昂贵+软件持续成本，软件如何定价/报销不清晰。
- 网络安全与“神经隐私”：潜在的神经数据被读取/篡改风险，并延伸到社会层面的“神经精英”担忧。
- 监管不确定性：虽然中美都制定了指导文件与加速计划，但植入式BCI仍需要长期证据与严格试验设计。

太空新经济

中美太空竞速与“天基算力”新基建

太空经济由单纯的科研探索或商业冒险，正演变为一场关乎未来50年国家战略安全与经济发展主动权的“太空圈地运动”。这一转变并非孤立的技术突破，而是地缘政治逻辑与全球顶级科技领袖商业愿景高度共振的结果。

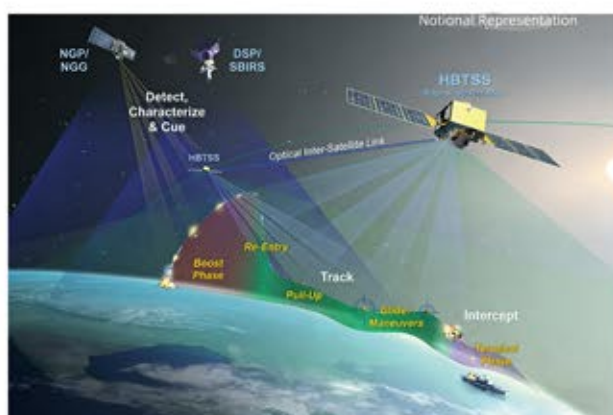
3.1

太空经济——不可再生轨道资源与大国安全防线的终极博弈

太空经济的底层竞争逻辑，本质上是对稀缺战略资源的抢占。依据国际电信联盟（ITU）“先登先占”原则，低轨（LEO）空间存在明确的容量天花板。先发国家通过密集发射构建星座，不仅确立技术优势，更在物理层面永久圈占了战略资源，为后发者构筑了难以逾越的准入壁垒。

安全叙事维度随之升级。低轨卫星群正重塑全球防御格局，如美军“天穹”系统利用分布式星座实现导弹助推段拦截。商业航天设施因此具备了实战化的战略防御价值，成为国家安全防线的前沿。

图：美国“金穹”天基导弹防御系统示意图



资料来源：兴业证券

面对“轨道挤压”，中美模式迥异：美国采取“私营领军”，依托SpaceX的可回收技术以“一周一射”的节奏快速填充轨道，形成既成事实的霸权；中国则采取“国家主导+体系突围”，以GW星座等“国家队”协同组网，辅以民营火箭作为敏捷运力，力争在窗口期内筑牢太空主权根基。

3.2 终极愿景——硅谷巨头眼中人类文明的星际愿景

太空经济由两位硅谷领袖的愿景驱动：马斯克致力于“生命多行星化”，将Starlink作为“印钞机”为火星移民积累资金，实现“以战养战”。贝索斯则主张“保护地球”，通过蓝色起源构建天地运输骨干，旨在将重工业移至太空“轨道工业园”，利用无限能源开启“零重力革命”。

殊途同归，两者都致力于将太空转化为工业基地。硅谷共识认为，太空是继互联网后的最大增量市场，已具备万亿级潜力，是可自负盈亏的商业疆域。

图：星链技术逐步突破



资料来源：《Starlink 2025 Progress Report》

图：星链已实现商业闭环（数据截至2025年末）



资料来源：《Starlink 2025 Progress Report》

3.3 算力上天——AI浪潮下数据中心向宇宙迁移的范式革命

随着AI爆发，航天技术重心正从“传输”向“计算”转移。传统“天感地算”受限于传输带宽，难以满足实时性需求；“地数天算”成为行业共识，即直接将数据中心部署于轨道。

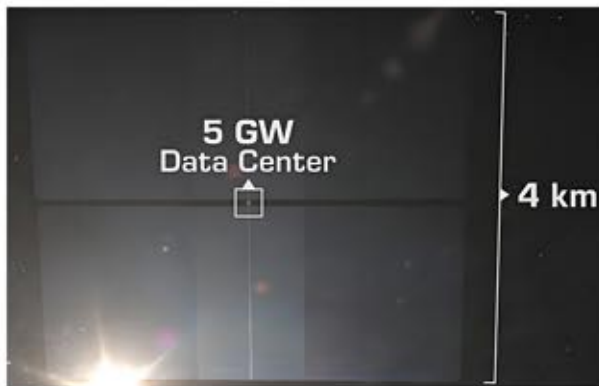
国别	主体	进展
美国	SpaceX	目标在4-5年将通过星舰完成每年100GW的数据中心部署
	Starcloud	2025年11月2日将其首颗卫星Starcloud-1送入太空，并成功在轨训练大型语言模型NanoGPT
	Google	拟在2027年初发射两颗搭载Trillium TPU的卫星，将AI算力直接部署到太空



资料来源: 公开资料整理

科技巨头正推动太空成为第三大数字基础设施。SpaceX计划通过星舰部署每年100GW的太空数据中心；Google与Nvidia相继推进TPU/GPU上天计划。商业航天已越过“工业红线”，成为数据中心的延伸。

图：Starcloud计划建设的5GW太空数据中心



图：Starcloud规划的集装箱式太空数据中心



资料来源: 英伟达官网

构建天基算力需突破三大物理瓶颈，催生了结构性机会：一是通信，星间激光链路成为标配，实现Gbit/s级高速抗干扰传输；二是能源，算力即电力，高性能芯片需依赖轻量化、大面积的柔性太阳翼供能；三是散热，真空环境下风冷失效，必须依赖先进热控涂层及利用柔性翼作为“外挂散热片”来解决高热流密度芯片的散热难题。

3.4

运力鸿沟——SpaceX统治力下的效率极限与中国产能突围

运力成本下降是产业繁荣基石，也是中美差距核心。SpaceX凭借猎鹰9号展现绝对统治力，截至2025年末，单枚火箭已刷新32次复用纪录，复用30次后发射成本降至约926美元/kg，彻底颠覆传统商业模式。

“星舰”的工程化更具颠覆性：2025年“筷子”机械臂成功回收助推器，标志着快速复用技术突破，其150吨级运力有望将成本再降数量级。SpaceX通过复用将硬件成本摊薄至数十次发射，仅需承担燃料等边际成本。这种高利润与低报价并存的经济模型，为其构建了极宽的护城河。

图：SpaceX “筷子” 回收火箭



资料来源：SpaceX官网

中国航天正处于产能爬坡的关键攻坚期。面对GW、千帆等超20万颗卫星的组网规划，年发射需求跃升至数千颗，而当前年产能仅约200颗，供需缺口超5倍。同时，国内主流报价（5万-10万元/kg）与SpaceX存在数量级差距，高成本严重制约落地速度。为突破瓶颈，行业正加速追赶：技术上，2025年朱雀三号、长征十二号甲等通过密集测试获取了关键回收数据；政策上，科创板标准扩围为企业注入资本动力。在双重驱动下，中国太空经济正加速从“技术验证”向“产能突围”跨越，力争填补运力鸿沟。

图：朱雀三号遥一运载火箭发射



资料来源：央视新闻

表：我国大型星座计划与星链（Starlink）的比较（截至2025年末）

星座计划	运营主体	入轨卫星数量(颗)	计划卫星数量（颗）	频段
GW星座	中国星网	136	12992	Ka、V
千帆星座	上海垣信	108	15000	Ku、Q、V
鸿鹄-3星座	蓝箭鸿擎	-	10000	-
吉利未来出行星座	时空道宇	64	6012	-
CTC-1/2	无线电创新院	-	193428	-
星链（Starlink）	SpaceX	10095	42000	Ku、Ka

资料来源：兴业证券

未来展望

2026年太空经济将经历“卖铲子-整车厂-应用生态”三段演进。短期上游确定性最高，中期聚焦规模化整车厂，长期看应用爆发。当前核心机遇在于布局填补运力鸿沟及解决算力瓶颈的方向。

可控核聚变

人类能源的终极解决方案

在人类文明的发展史上，能源始终是驱动社会进步和塑造世界格局的核心动力。从木柴到煤油气再到电力，每一次能源革命都深刻改变了人类的生产与生活方式。然而，传统化石能源的有限储量、环境污染与地缘政治风险，以及现有核裂变的安全与废料难题，共同构成了制约可持续发展的能源天花板。面对这一挑战，全球将目光投向了宇宙中最常见的能量源泉——核聚变。模拟太阳的可控核聚变，因其近乎无限的燃料储备、本质安全性和清洁零碳的突出优势，被誉为“人造太阳”和“终极能源”，承载着人类实现能源自由、迈向可持续发展的终极梦想。

4.1 现有能源体系无法支撑人类长期发展

4.1.1 化石能源面临气候和储量双重束缚

当前全球能源体系中，化石能源依旧占据绝对的主导地位，煤油气合计占比84%，风光等可再生能源占比12%，核能占比仅4%。中国能源结构中煤炭占据59%，油气30%，风光等可再生能源占比10%，核能占比仅1%。

表：全球能源体系份额统计

份额	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2025E
原油	34%	34%	34%	33%	33%	34%	34%	34%	34%	34%	32%	32%	32%	33%	33%
天然气	22%	22%	23%	23%	23%	23%	23%	24%	24%	24%	25%	25%	24%	24%	24%
煤炭	30%	31%	31%	30%	30%	29%	29%	28%	28%	28%	28%	28%	28%	28%	27%
可再生能源	9%	9%	9%	9%	9%	10%	10%	10%	10%	10%	11%	11%	11%	11%	12%
核能	4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	5%	4%	4%	4%	4%

数据来源：Bp能源统计年鉴

根据巴黎气候协定，2050年全球相对于工业革命前的气温提升要控制在2摄氏度以内，否则会引发全球性的生态灾难。碳中和是全球共识，碳税会越来越贵，制度性地限制化石能源使用，并加速风光和核聚变等新型能源的开发利用。

此外按照目前勘探储量，全球油气资源基本在百年左右枯竭，煤炭预计在两百年。且我国油气资源严重依赖进口，原油进口占比72%，天然气进口占比40%，对国家能源战略安全造成比较大的掣肘，加速开发新型能源显得尤为紧迫。

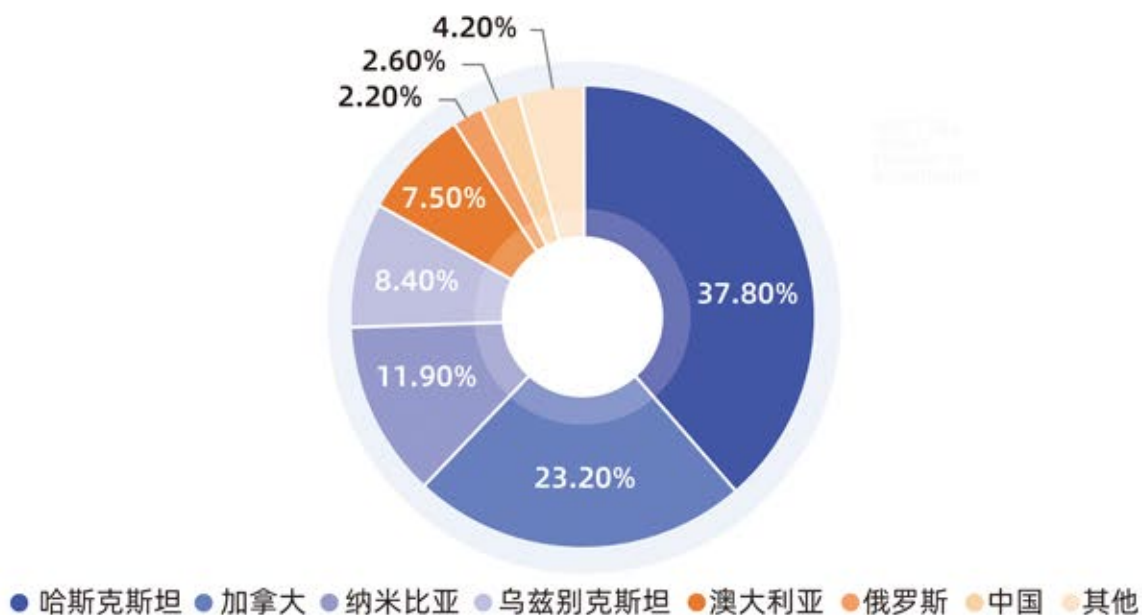
» 4.1.2 风光出力不稳定，大规模装机对电网提出较大挑战

风光出力极不稳定，对电网冲击很大。光伏发电完全依赖光照强度和天气，出力呈现倒U型，风力发电主要依赖季节性风力资源，出力都不稳定，对电网冲击较大，配储又显著提高成本，且不提供惯性动量，风光的消纳难题已经显现。

» 4.1.3 核裂变面临铀资源和乏燃料处理双重难题

当前全球核电利用主要以三代裂变堆为主，面临铀资源匮乏和乏燃料处理两大难题。全球经济可采天然铀资源储量约700万tU，按照目前强度可供使用100年。中国天然铀资源相对匮乏，储量占比仅2.6%，进口占比超过80%，大规模发展核裂变面临铀资源的约束。另一方面，当前三代堆芯中U-235仅能消耗一半，乏燃料中存在较强放射性，需要地质深埋数十万年才能降到安全水平。

图：全球经济可采天然铀资源储量分布



数据来源：公开数据整理

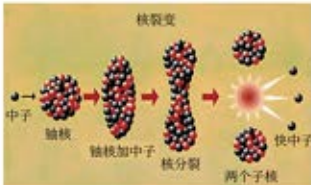
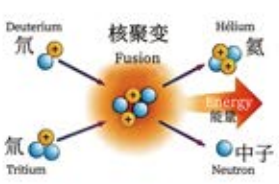
4.2 可控核聚变是未来能源的理想解决方案

4.2.1 可控核聚变技术原理

可控核聚变是在极度的高温高压下，使两个轻原子核聚合成一个较重的原子核，并释放出巨大能量的过程。

自20世纪50年代起，无数科学家与工程师投身于这项被誉为“人类迄今构想的最复杂能源系统”的探索。其核心挑战在于如何在地球上创造并约束高达上亿摄氏度的等离子体，并实现持续稳定的能量净输出。

图：可控核聚变技术原理

类别	原理	能量	清洁程度	原理示意图
核裂变	用中子轰击重原子核（如铀-235），使其分裂并释放出更多中子与巨大能量，从而引发持续链式反应的过程。	1g铀 ⁻²³⁵ 完全裂变≈2.6吨标准煤。	较低：核裂变会产生难以处理的长寿命放射性废物，如 ¹³⁷ Cs（半衰期30年），处理不当会造成长期污染。	
核聚变	在极高温（如1亿℃）和高压下，使两个轻原子核（如氘和氚）聚合成一个较重的原子核，并释放巨大能量。	1g氘氚混合物聚变≈11.2吨标准煤，是核裂变的约4倍。	核聚变本身非常清洁，其有限的放射性主要来自中子活化的堆壁材料以及半衰期较短的氚（12.43年），危害远小于核裂变。	

资料来源：公开资料整理

核聚变技术目前呈现百花齐放的格局，全球主流的装置以托卡马克为主，使用超导磁体进行磁约束，利用氢的同位素氘氚进行聚合反应，生成氦并释放能量。

图：可控核聚变主流装置

主流装置	托卡马克	仿星器	磁镜
原理示意图			
核心原理	利用等离子体自身感生的大电流与外部线圈共同产生螺旋磁场。	完全通过外部复杂扭曲的线圈直接“雕刻”出所需的螺旋磁场，无需或仅需很小的等离子体电流。	利用两端强、中间弱的“磁镜场”（瓶颈效应）约束直线型磁场中的等离子体。
磁场产生	内外结合：环向场由外部线圈产生；极向场由等离子体电流和外部线圈共同产生。	完全外部：由特殊设计的复杂三维螺旋线圈产生。	简单对称：通过电流强度不同的环形线圈在直管上排列，形成轴对称磁场。
等离子体电流	必需且巨大（如ITER达15兆安培）。	不需要或很小。	不需要。
最大优势	性能最高，研究最成熟，是ITER选择的路线，聚变参数最优。	无大电流导致的破裂风险，原则上可实现稳态连续运行。	结构概念简单，易于维护。
主要挑战	大电流导致不稳定性与脉冲式运行；控制系统复杂。	工程制造难度极高，线圈加工精度要求苛刻；粒子逃逸。	存在终端损失，粒子从两端逃逸，难以达到聚变条件。
代表装置	TER(国际)、EAST(中国)、KSTAR(韩国)	Wendelstein7-X(德国)、LHD(日本)	WHAM(美国)：传统磁镜已非主流，其改进型由商业公司继续研究。

资料来源：公开资料整理

核聚变相较于核裂变具备显著的优势。

能量密度高：1g氘氚反应释放11吨标准煤的能量，是裂变4倍。

清洁：不产生温室气体，不产生放射性乏燃料。

安全性高：不预置燃料，原料一旦中断，很快停止反应。

燃料充足：海水中氘资源十分丰富，富集重水电解即可得到大量氘气，中子辐照Li-6就能持续生成氚，可供人类使用3000万年。

» 4.2.2可控核聚变已成大国战略竞争的新高地

可控核聚变领域的全球竞争，其意义早已超越单纯的科学探索或商业角逐，它是一场重塑全球能源与地缘政治格局、并深刻考验国家综合实力的世纪长征。可控核聚变是人类迄今构想的最复杂能源系统，涉及等离子体物理、超导材料、精密制造、人工智能等多学科尖端领域，其探索极大地推动了高温超导材料、人工智能、高性能计算、先进制造等前沿领域的突破，若能成功将彻底颠覆现有的全球能源贸易体系和地缘政治格局。当前中国在全球可控核聚变领域处于世界领先地位，合肥科学岛聚集了EAST、BEST（2027年建成）、CRAFT等世界级装置，形成全球唯一研究集群。此外上海、成都、南昌等地也有实验装置在建，形成国家队牵头，民营企业积极参与的良好格局。

表：中国可控核聚变格局

项目	企业/机构	位置	投资金额 (元)	建设周期*	预估 2026-2030 资本开支*	运作情况
东方超环EAST	中科院等离子体所	合肥	128亿	/	/	已建运行
中国环流三号(HL-3)	核工业西南物理研究所	成都	/	/	未知	改造
BEST (紧凑型聚变实现装置)	聚变新能/ 中科院等离子体所	合肥	85亿	2025-2027	65亿左右	在建
CRAFT(聚变堆主机 关键系统综合研究设施)	中科院等离子体所	合肥	/	/	/	在建
CFETR (中国聚变工程实验堆)	中科院等离子体所	合肥	预估 1000亿左右	2035年 左右建成	500亿左右	已规划 并预研究
环流四号	中国聚变能源	上海	注册资本 150亿	预计 2030年建成	预估150亿	已规划
星火一号	江西聚变新能	南昌	计划投资 200亿	预计2029年 年底建成	预估200亿	在建
Z箍缩裂变聚变混合堆	先觉聚能	成都	50亿	2035-2040	/	已规划
玄龙-50U/和龙2	新奥集团	廊坊	40亿	/	/	已建运作
洪荒70/170	能量奇点	上海	8亿	/	/	已建运作
CTRFR-1	星环聚能	西安	10亿	2024-2027	5亿左右	在建
KMAX-U/Xe onova-1	星能玄光	合肥	计划成本 10亿	/	/	已规划
HHMAX-901	瀚海聚能	成都	/	/	/	在建

资料来源：公开资料整理

海外聚变产业亦在蓬勃发展。根据国际原子能机构（IAEA）正式发布的《2025年世界聚变展望》，全球近40个国家正积极推进聚变计划，其中美国相对领先。

表：各国聚变产业发展情况

	示范堆名称	目标	预期时间
中国	CFETR	填补ITER与商业堆技术空白，目标Q>20、验证聚变堆超导技术，发电功率百万千瓦级	2050年前示范发电； 2050-2059年(2050年代)商业发电
欧盟	EU-DEMO	验证聚变电站经济性、可靠性与持续性，验证超导技术，要求氦增殖率>1.1、连续发电500MW	2050年前示范发电； 2050年代商业发电
美国	FNSF	验证氦燃料自持、材料辐照耐受性及聚变系统集成技术	2040年前示范发电； 2050年代商业发电
日本	JA-DEMO	紧凑型托卡马克，验证高温超导、氦增殖及燃料循环技术，实现500MW稳态净电力输出	2040年前示范发电； 2050年代商业发电
英国	STEP	全球首个商用球形托卡马克，验证氦自持、高温超导技术，实现100MW电力输出	2040年前示范发电； 2050年代商业发电
韩国	K-DEMO	验证稳态高约束模式(H-Mode)运行、氦自持及高温超导技术，目标百万千瓦级发电	2040年前示范发电； 2050年代商业发电
俄罗斯	TRT	发电功率500MW、目标Q>10	2030年前示范发电； 2050年代商业发电

资料来源：国际原子能机构（IAEA）《2025年世界聚变展望》

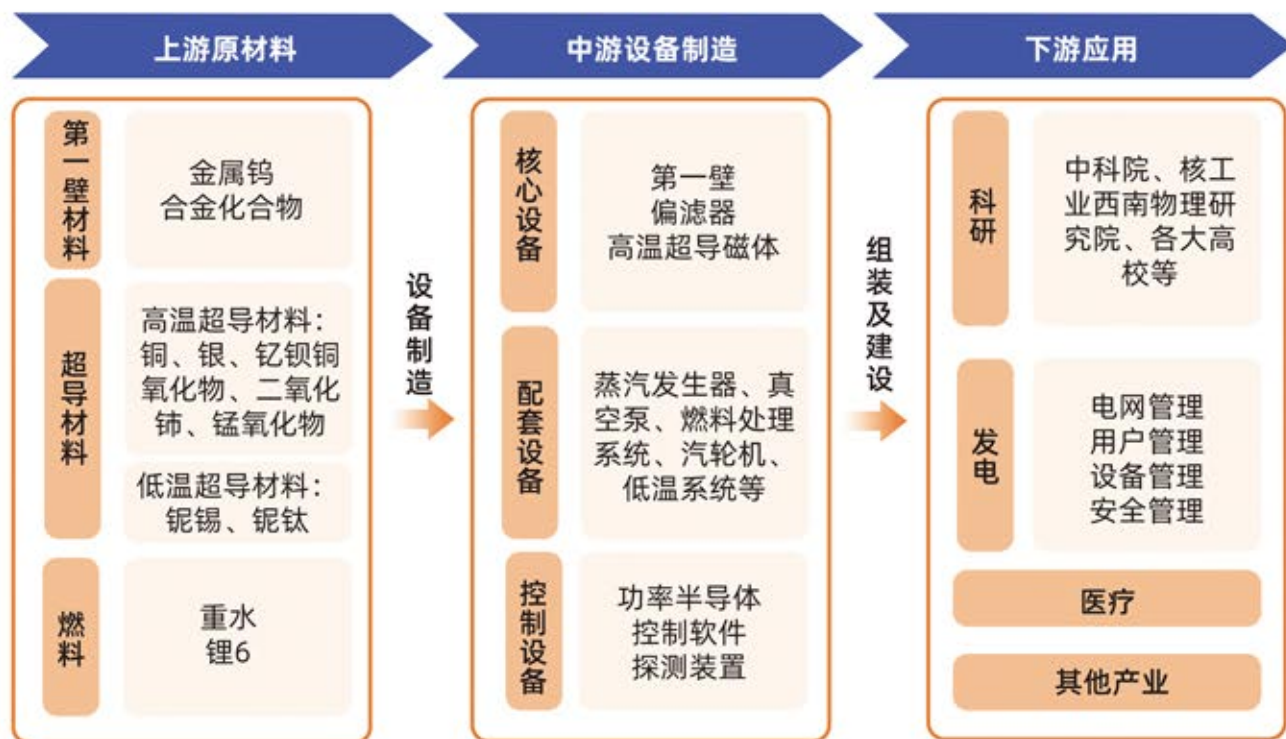
4.3 政策和资本双轮驱动，产业浪潮澎湃向前

» 4.3.1 聚变正引爆一个前所未有的全新生态链

可控核聚变正在强力锻造一条从上游核心材料（超导材料、第一壁材料）、中游关键设备（磁体系统、真空室），到下游聚变发电及多元化应用的全新产业链。上游材料主要涉及重水、Li6、高温超导磁体、特种钢材合金。中游装备主要涉及真空腔室、偏滤器、燃料处理系统、蒸汽发生器等特种装备。下游应用目前主要在发电。我国产学研一体化推进，已经形成了全球领先的供应链体系。

价值量上，当前国内每年裂变堆核准约10个，照此估算市场规模在2000亿，其中上游材料占比20%，中游装备占比50%，未来随着电力需求增长，聚变将是数千亿级别的市场。

图：可控核聚变产业链



资料来源：公开资料整理

4.3.2 政策和资本双轮驱动，中国企业奋勇争先

中国已构建起一个由顶层立法、国家战略、地方配套、国际合作及监管创新共同支撑的综合性政策体系。可控核聚变更被正式纳入“十五五”规划，明确为前瞻布局的未来产业与经济增长新引擎！资本力量全面进发，除了传统科研院所财政拨款，中核集团牵头组建了注册资本150亿元的中国聚变能源有限公司，星环聚能、奇点能源也陆续在一级市场实现融资，配套的超导材料等中上游企业也在积极融资，上海超导即将IPO——资本巨浪正加速推动产业链成熟与壮大。

我们正迈向激动人心的里程碑，预计2027年前我国将建成聚变能实验装置BEST，完成聚变发电演示；2030年突破核心技术并“点亮第一盏灯”，2035年实现CFETR工程堆发电；2045~2050年建成商用示范堆，实现大规模并网。在这场决定未来的全球科技竞赛中，国内企业正持续取得技术进展，并逐步拓展国际合作。相关上市公司必将深度受益，共同拥抱这场波澜壮阔的产业浪潮！

中国创新药 “燃烧的远征”

5.1 激荡十年——政策、生态与管线爆发（2015-2025）

这十年是中国创新药“把欧美半个世纪的功课压缩成‘速成班’”的十年。

» 5.1.1 政策顶层设计：从“审评积压”到“全链条支持”

- **起点（2015-2017）**：以“722核查”为标志，中国药监局大刀阔斧改革，清理积压，并加入ICH（国际人用药品注册技术协调会），审评标准全面对标国际。
- **加速（2018-2021）**：医保谈判常态化，创新药纳入医保速度大幅提升；资本市场改革（港股18A、科创板）允许未盈利生物科技公司上市，打通退出通道。
- **深化（2024-2025）**：政策不仅仅是单点补强，更是研发、定价、准入、支付的“全链条支持”。

图：2015年以来的政策变化



资料来源：公开资料整理

» 5.1.2 “VC+IP+CXO” 的生态不断强化

• **投融资变奏**：2015-2021年是一级市场融资的黄金期。虽然2022-2024年经历了资本寒冬，但2025年出现了结构性转折——License-out（对外授权）首付款（约70亿美元）首次超过同期一级市场融资总额（52.19亿美元），标志着产业“自我造血”能力的形成。

• **生态完善**：中国建立了全球第二大的医药研发外包（CXO）集群，并在长三角、大湾区形成了完整的生命科学服务供应链，极大地降低了Biotech的创业门槛。

图：对外授权首付款与一级市场融资额趋势



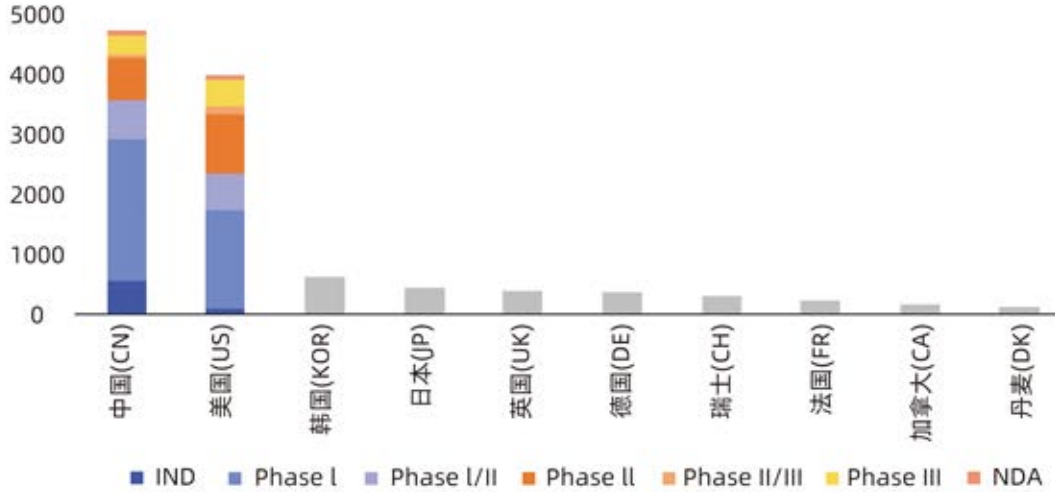
图：创新药生态体系



数据来源：医药魔方，公开资料整理。

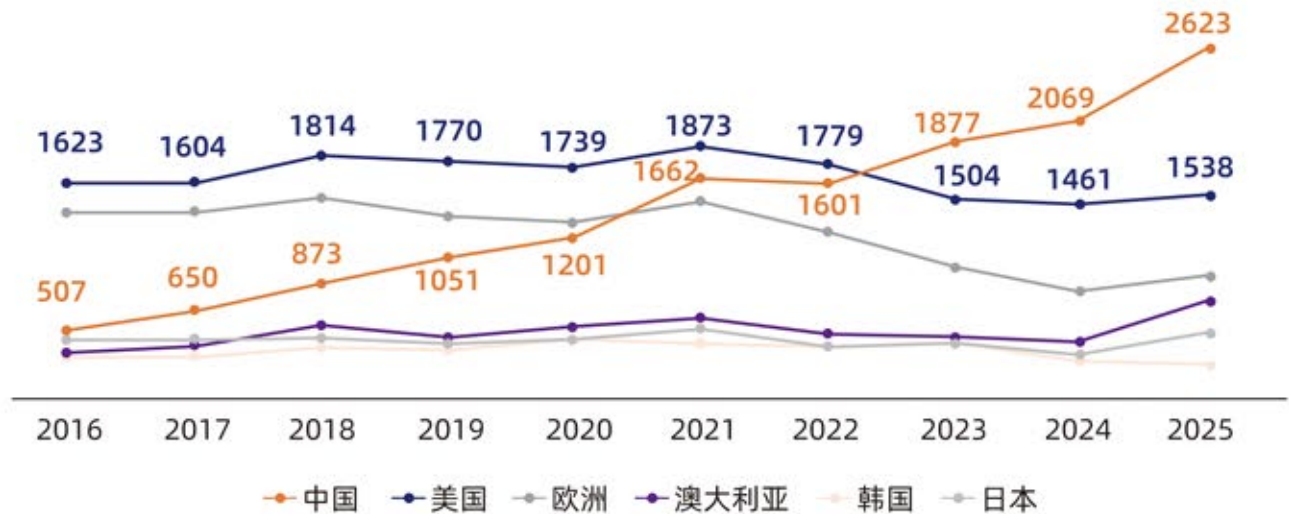
» 5.1.3 管线爆发：数量与质量的双重跃迁

图：2025年创新药管线数量：全球Top10国家/地区（中美按研发阶段拆分）



数据来源：医药魔方

图：2016-2025年创新药临床试验数量趋势



数据来源：医药魔方

» 5.1.4 中国资产的全球定价

• **交易井喷**：2015-2025年，中国相关医药交易总金额从31亿美元提升至1357亿美元，约占全球交易总金额的49%。中国已成为全球TOP 20 MNC（跨国药企）的第二大创新项目来源地。

图：中国创新药BD交易爆发



数据来源：医药魔方

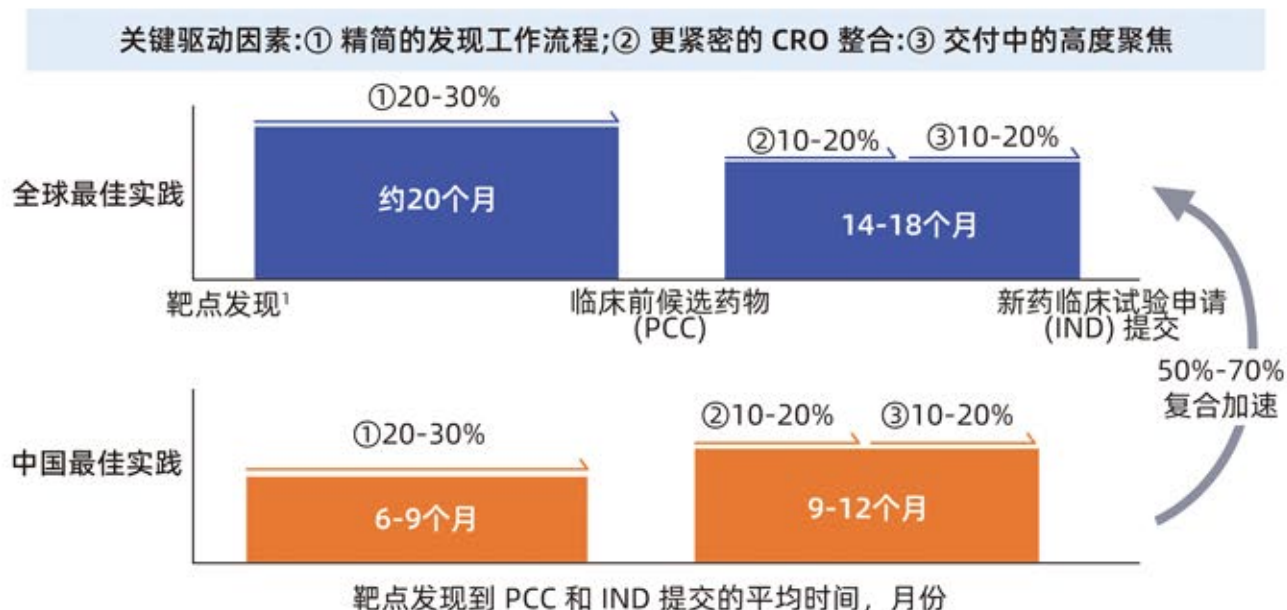
5.2 中国药企的“比较优势”与战略机遇

在“内卷”中磨练出的工程化能力，正在转化为全球竞争优势。

5.2.1 核心能力：效率与成本的极致

- 临床效率：中国在临床试验招募速度、成本、效率上具有显著优势。

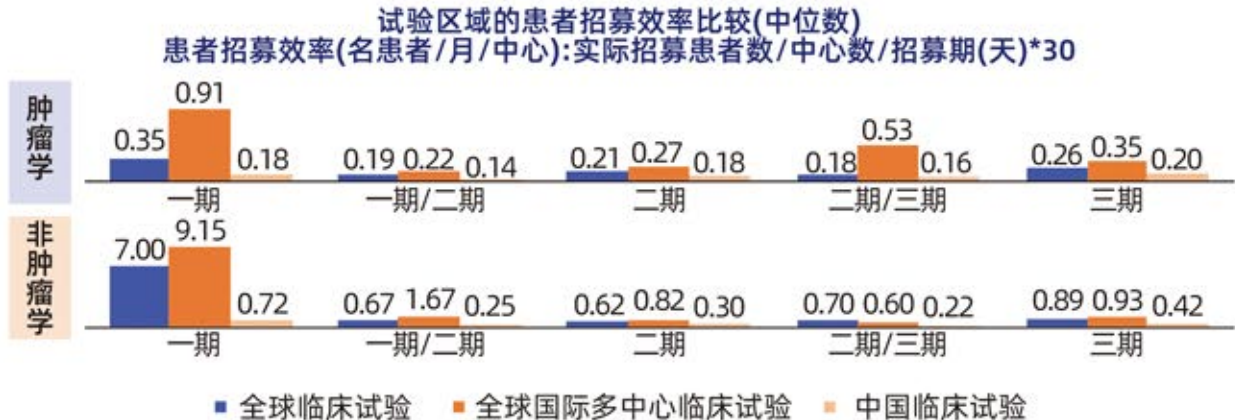
图：临床前开发效率



数据来源：医药魔方

图：试验区域的患者招募效率比较

- ▶ 招募效率反映了每个研究中心每月平均招募的患者人数。
- ▶ 肿瘤学的招募效率通常低于非肿瘤学领域。
- ▶ 总体而言，中国的招募效率通常高于全球平均水平，而全球多中心试验的效率低于全球平均水平。



数据来源：医药魔方

- **工程化创新 (1→N)**：不同于美国的“0-1”源头生物学发现，中国企业擅长“1-N”的工程化改造。在ADC（连接子、偶联技术）、双抗（结构设计）和小核酸（递送系统）领域，中国企业具备做出Best-in-Class（同类最佳）产品的能力。

» 5.2.2 历史机遇：填补MNC的专利悬崖

- **收入缺口**：全球大药企在2026-2035年面临巨大的专利悬崖，预计产生接近1200亿美元的收入缺口。肿瘤、免疫和心血管代谢领域占缺口的80%。
- **中国角色**：MNC自身研发回报率下降，中国的高质量、高性价比资产完美契合了这一需求，成为了MNC的“管线超市”。

5.3 未来十年——全球化兑现与大国竞合

中国创新药将从“借船出海”（BD）走向“造船出海”（全球商业化）。

» 5.3.1 重磅炸弹的全球上市潮

- **关键节点**：预计2028年后，中国自主研发的PD-1/VEGF双抗、多款ADC药物有望在欧美获批上市。从BD迈向里程碑兑现或深度合作分成，真正实现海外收入的规模化兑现。

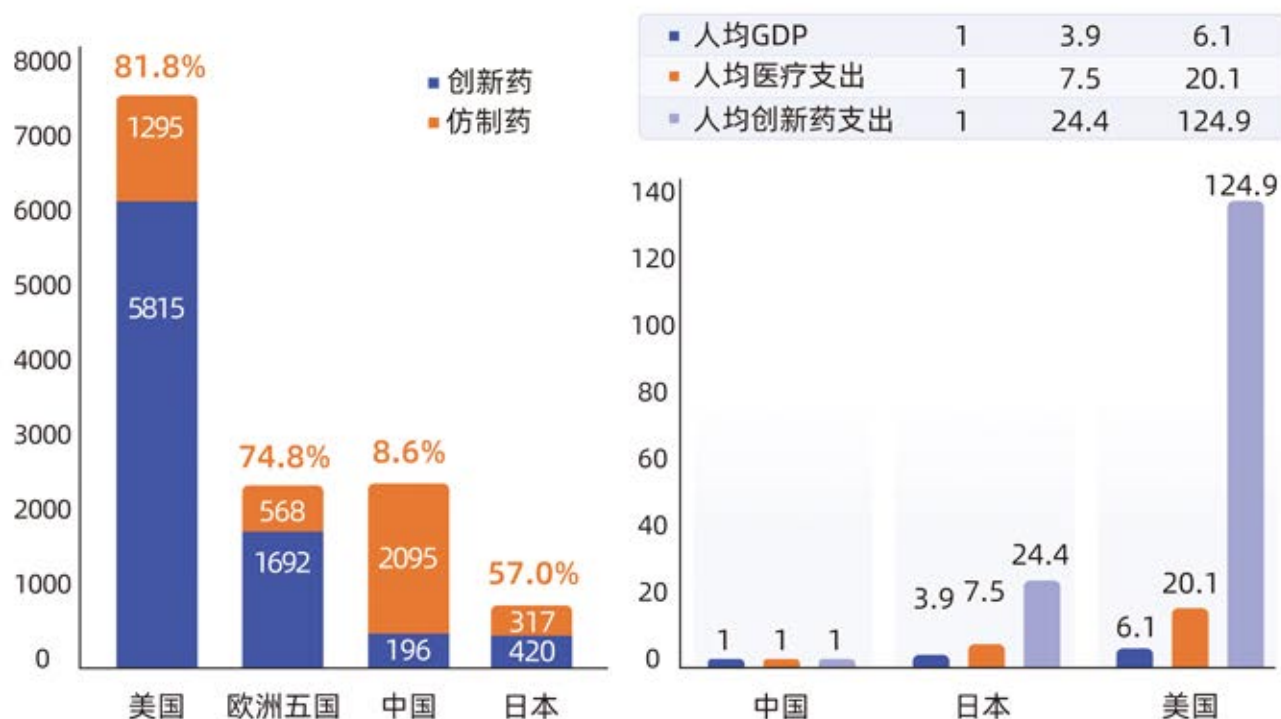
» 5.3.2 2040年市场空间展望

- **全球份额**：摩根士丹利预测，到2040年，中国源头创新的资产将占据美国FDA新药批准数量的35%。
- **销售占比**：届时，来自中国研发的创新药在海外市场的年销售额有望达到2200亿美元，中国将成为仅次于美国的全球第二大药物创新发源地。

» 5.3.3 中国商保大发展与多层次支付

- **支付破局**：随着国家医保局“1+3+N”多层次医疗保障体系的建立，商业健康险将成为高价创新药的重要支付方。未来十年，中国本土市场的创新药支出占比、人均创新药支出有望大幅提高，在这个过程中，商保将承担起类似美国商业保险支付角色。

图：美、欧、中、日药品市场规模(亿美元) 图：中美日三国人均支出(以中国为1计算)



数据来源：中国医药创新促进会

» 5.3.4 中美共同引领前沿技术

- **技术并跑**：在干细胞、基因治疗（CGT）、小核酸、双抗及ADC领域，中美差距已缩小至2-3年甚至同步。未来十年，中美将在合成生物学、AI制药等前沿领域形成“竞争与合作”并存的双极格局，共同推动人类攻克癌症、神经退行性疾病等难关。

图：中美在创新药领域形成双极格局并保持要素流动



资料来源: 公开资料整理

过去十年, 是中国创新药行业的一场激情燃烧的奋斗之旅, 我们以技术突破为火种, 以政策支持为加速器, 迎来了一个又一个里程碑。每一个突破背后, 都凝聚着无数科研人员和企业家的智慧与汗水。未来十年将是中国创新药全球发展“燃烧的远征”, 中美两国的竞争与合作, 将深刻影响全球医药格局, 我们不仅要保持前瞻视野, 更要在不断变化的环境中寻找机遇, 勇敢迎接未来的每一次挑战和跨越, 书写属于我们的“药谷”传奇。



投资中国科技
诺安一直都在

