

行业报告

AI迭代短期产生爆发式硬件需求， 产业整体仍处发展早期

报告日期：2026-06-08

研究部

姓名：杨森

SFC：BJO644

电话：0755-21519178

E-mail: yangsen@gyzq.com.hk

目录

- 一、AI性能快速迭代，商业落地推动AI Agent爆发
- 二、硬件产业链不得不全栈式、性能指数级进化以满足AI发展需求
- 三、AI产业仍处于早期，具有长足发展空间
- 四、AI硬件产业链的投资思路
- 五、风险提示

AI大模型迅速迭代，在某些专业领域已可以媲美人类

ChatGPT迭代历史

以ChatGPT为例。过去8年来，其运算能力呈现指数级增长：

- 参数量从1.17亿爆发式增长到了约1.8万亿，增长超过15000倍；
- 上下文窗口从约1000个词→200万token，增长超过2000倍。

性能和架构不断升级，使得ChatGPT在某些专业领域已可以媲美人类：

- 从"预测词语"到"理解意图"：RLHF技术的引入，让模型不再只是机械地预测下一个词，而是学会了理解人类真正的需求。
- 从"处理文字"到"感知世界"：GPT-4引入多模态能力，模型开始能够看、能够听，突破了纯文字的限制。
- 从"回答问题"到"独立执行"：GPT-5系列具备Agent能力，能够主动规划任务、调用工具并独立完成复杂工作。

版本	发布时间	核心参数	核心特性与突破	重要里程碑
GPT-1	2018年6月	1.17亿参数训练数据：约5GB文本	提出"预训练+微调"框架	证明大规模无监督学习技术路线可行
GPT-2	2019年2月	15亿参数训练数据：约40GB文本（800万网页）	生成质量大幅提升；首次出现"零样本学习"迹象	OpenAI因认为"太危险"拒绝完整发布
GPT-3	2020年6月	1750亿参数训练数据：约570GB文本	少样本学习能力成熟	第一次让普通人感受到AI的实用价值；确立"规模定律"（ScalingLaw）
GPT-3.5	2022年11月	-	加入RLHF（人类反馈强化学习）；学会理解人类意图	ChatGPT正式上线5天获100万用户60天获1亿用户（人类历史上增长最快的消费级产品）
GPT-4	2023年3月	约1.8万亿参数（混合专家架构MoE）	首次引入多模态能力；能"看"图片	AI首次在人类专业考试中超过大多数人；律师资格考试：前10%；医学执照考试：通过；SAT：前7%
GPT-4o	2024年5月	-	原生多模态；文字/图片/音频同时理解、同时输出	响应延迟320毫秒，接近真人对话节奏
GPT-5	2025年8月	-	引入动态路由机制；根据任务复杂度自动分配算力	具备Agent代理能力；从"回答问题"进化到"执行任务"；
GPT-5.2	2025年12月	-	ReasoningTokens机制成熟；输出前先进行内部验证	幻觉率下降30%
GPT-5.4	2026年3月	-	上下文窗口100万token（部分版本支持200万）；新增NativeComputerUse	OSWorld基准测试得分75%，首次超过人类平均水准（72.4%）；

数据来源：ChatGPT官网、百度、国元证券经纪（香港）整理

盈利模式不断验证落地，AI Agent正进入商业爆发期

智能体（AI Agent）是由大语言模型或领域模型驱动，具备环境感知、推理决策及行动能力的系统。对企业或个人而言，在当前AI的智能水平下，AI Agent是能够自主规划、记忆、调用工具，帮助人类或独立完成相对专业任务的智能助手，从而具有商业价值。

- 技术层面：AI Agent技术生态已趋于成熟，LangChain、AutoGen、CrewAI等开源框架为构建复杂Agent系统提供了坚实基础多智能体协作系统、特定领域语言模型和ComputerUse能力也有望实现重大突破；
- 商业层面：垂直行业应用将成为主战场，金融服务、医疗健康、制造业和客服领域将率先实现规模化部署。AI Agent正在大幅度提升社会生产效率，商业模式不断落地。举例而言：根据GitHub的数据，目前全球约3000万到4000万专业软件开发者，总薪资约3万亿美元。在AI辅助下，这些开发者现在产出的工作量相当于9万亿美元的生产力。

AI Agent技术演进历史

阶段	时间范围	核心特征	关键里程碑	代表产品 / 技术
技术萌芽期	2023.1-2023.6	概念验证与早期探索	首个自主 Agent 项目发布 工具调用机制确立	AutoGPT (16 万 + 星标) OpenAI Function Calling ReAct 范式
框架爆发期	2023.7-2024.6	开源框架与商业产品涌现 多模态与长上下文能力提升	主流框架相继发布 多模态模型能力升级	LangChain、AutoGen、 CrewAI BabyAGI Claude 3.5 Sonnet (100K 上下文)
商业化加速期	2024.7-2025.12	企业级产品与服务推出 标准化协议开始形成	多智能体系统发布 通用交互协议诞生	微软 Magentic-One Anthropic MCP 协议 微软 Copilot Studio Salesforce Agentforce
生态深化期	2026 年展望	Agent Swarm 范式兴起 Agent 原生模型出现可复用 Agent 市场形成	智能体经济进入深化期 AI 向 "同事" 角色演进	Anthropic Virtual Collaborator Agent 原生基础模型 企业智能中枢

数据来源：各公司官网、百度、国元证券经纪（香港）整理

目录

- 一、AI性能快速迭代，商业落地推动AI Agent爆发
- 二、硬件产业链不得不全栈式、性能指数级进化以满足AI发展需求
- 三、AI仍处于早期，具有长足发展空间
- 四、AI硬件产业链的投资思路
- 五、风险提示

AI的性能迭代和大规模应用“迫使”硬件产业链极速进化

- AI自身性能和外部应用的指数级增长要求AI产业的硬件性能和供应量做到同步的指数级增长。但物理现实世界产能升级的速度低于产业需求增长的速度，造成多个硬件供应环节或持续进入量价齐升、供不应求的局面。
- 当下，硬件不再是被动支撑软件的"底座"，而是成为决定AI能力上限的核心瓶颈。
- 整个产业链正经历从“芯片元件-服务器机柜-数据中心”的全栈式、革命性升级，硬件升级不再是单一环节的突破，而是芯片、封装、散热、电源、网络等多个环节的协同创新。

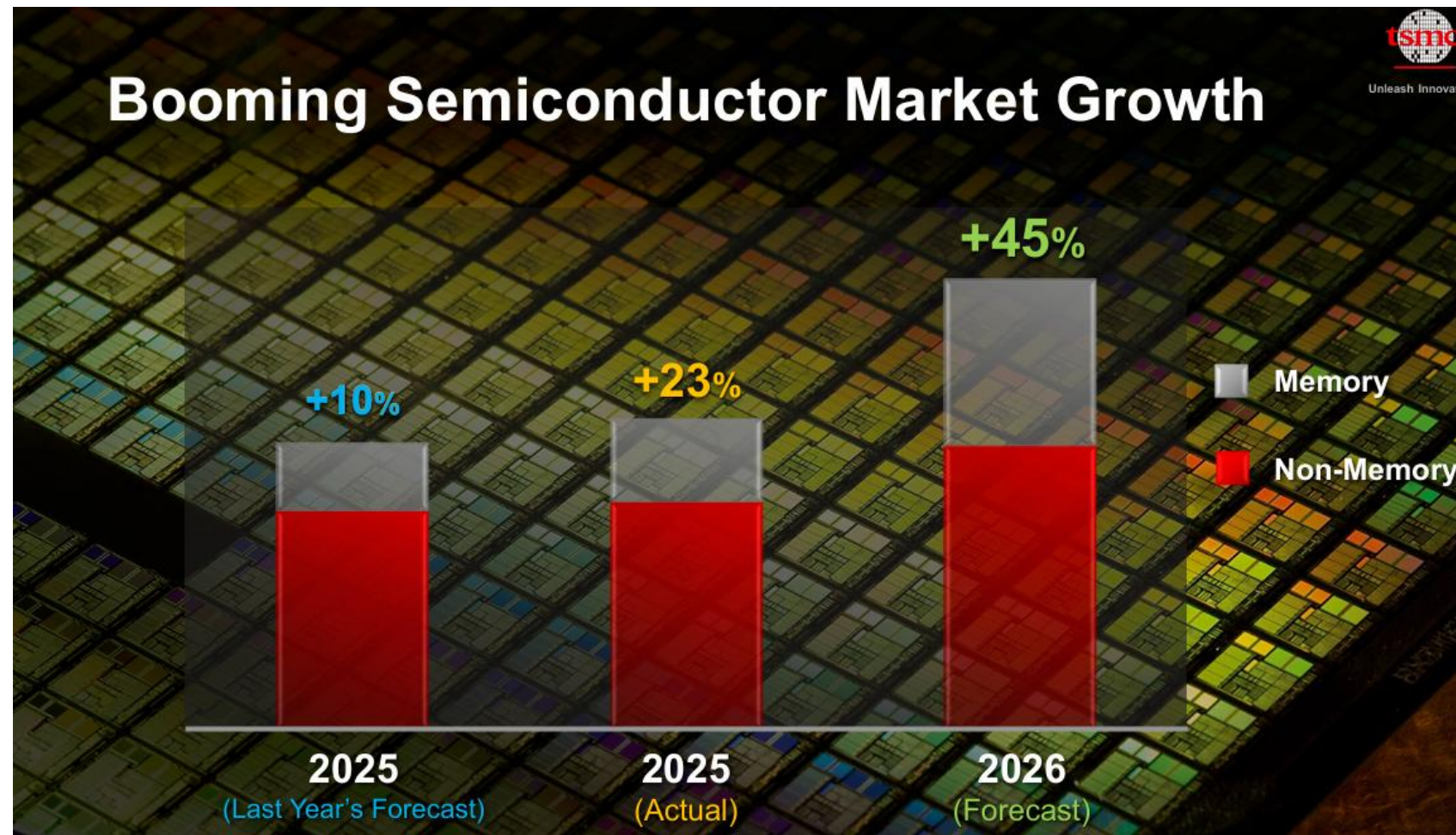
AI全产业链		
AI应用层		
分类		具体内容
模型		语言模型、视觉模型（图 / 视频）、Agent
应用（软）		AI 游戏、AI 传媒、AI 漫剧、AI 编程、AI 电商、AI 医疗
应用（硬）		消费电子、机器人、自动驾驶、AI 玩具、AI 手机、AI 眼镜、遥控、电池、灵巧手、线形关节、旋转关节、芯片 / 通信 / 触算法、视觉 / 雷达、电子皮肤、环卫智驾、矿卡智驾、端侧算力、摄像头、激光雷达、踏面单元
AI硬件基建层		
层级	模块	具体内容
芯片元件层	算力	GPU/TPU/ASIC/LPU、CPU、SoC
	存储	存储芯片（利基 / NAND/DRAM/HBM）、存储模组、设备（洁净、半导体设备）
	AI 元件	电容（钽电容）、电阻（MLCC）、电感
	PCB	高多层 / HDI / 正交、CCL、电子布（7628, Low Dk, Low CTE、Q 布）、铜箔（HVLP3/4/5）、树脂（环氧树脂、PPO、脲醛树脂）、PCB 设备、钻头（钨 & 金刚石）
服务器 / 机柜层	液冷	一次液冷（冷水机组、冷却塔等）、二次液冷（冷板、快接头、Manifold、CDU、循环泵）、浸没式液冷、金刚散热
	电源	超级电容（MLCC等）、服务器电源、BBU、PSU、PDU
	服务器 / 机柜	服务器整机、机柜系统
数据中心层	光通信	光材料（硅晶 / 磷化铟 / 钕铁单晶）、光芯片（EML/CW/VCSEL/DFB）、光模块（800G/1.6T/3.2T/LPO/NPO/CPO）、交换机（电交换机 / 光交换机 / OCS）、光钎（光棒 / 光钎 / 空芯光钎）、光设备
	电力	变压器（干式变压器、SST）、UPS、HVDC、柴油发电机、燃气轮机、SOFC
	AIDC	云服务、算力租赁

数据来源：各公司官网、百度、国元证券经纪（香港）整理

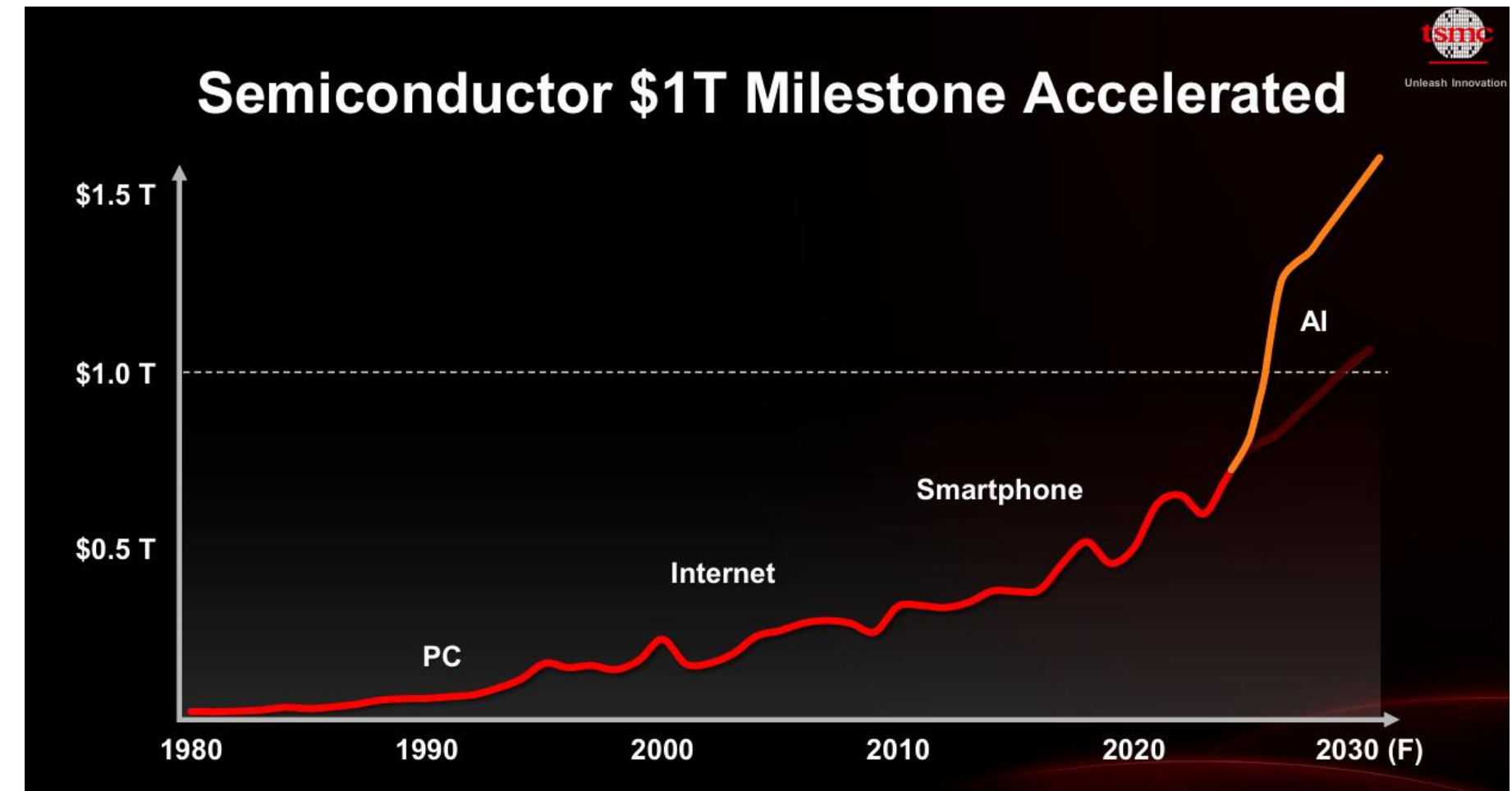
半导体：AI需求推动产业双位数增长

- 半导体市场在AI应用驱动下呈现加速成长。2025年实际成长率达23%,远超前一年预估的10%。展望2026年,台积电预测市场将成长45%,存储器与非存储器产品均贡献成长动能。此成长幅度反映AI运算需求对半导体产业需求结构的影响,持续性成长动能跨越整个价值链。全球半导体市场迈向1兆美元的时程较原先预期提前数年。
- 个人计算机、网际网络与智能型手机各自带动产业成长,但AI应用呈现更陡峭的加速曲线。台积电预估2030年全球半导体市场将超过1.5兆美元,AI为主要成长驱动力。此时程压缩反映机器学习工作负载对硅晶圆消耗量的影响规模。

全球半导体市场蓬勃发展



半导体市场提前数年达到1万亿美元规模



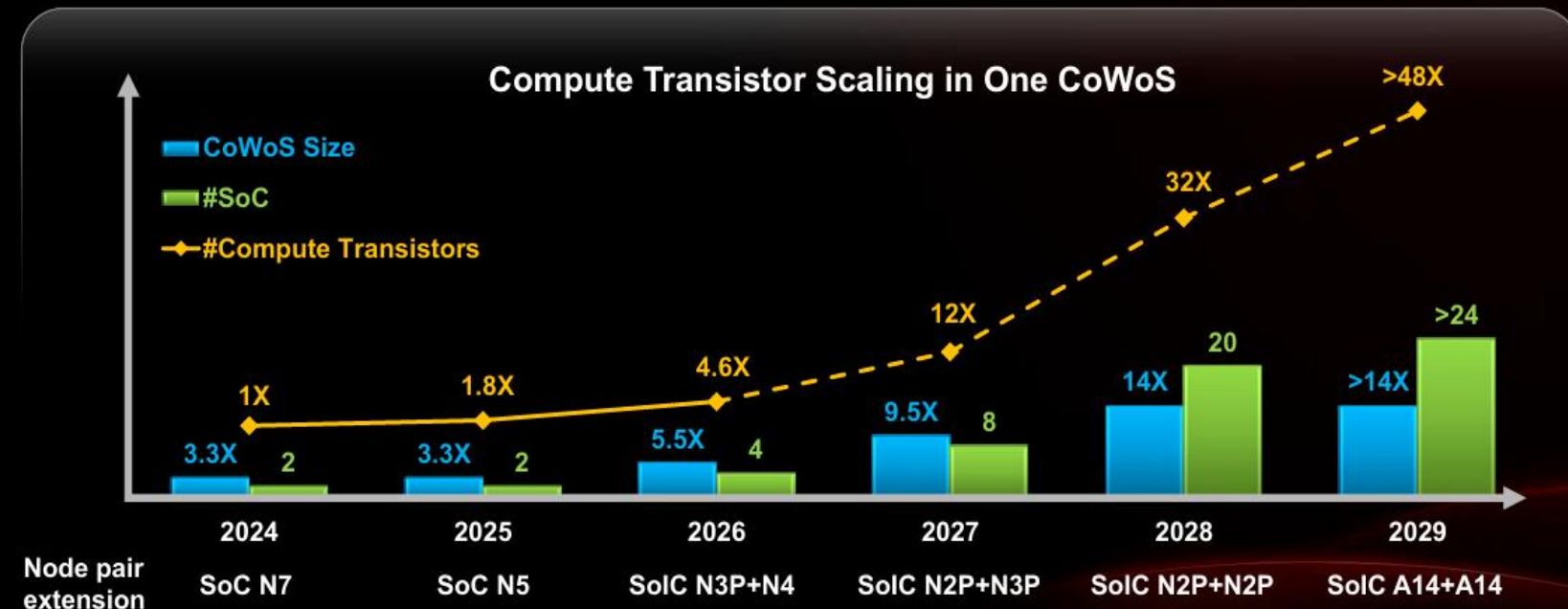
先进工艺节点保障AI性能持续提升

- 个人计算机、网际网络与智能型手机各自带动产业成长,但AI应用呈现更陡峭的加速先进逻辑节点、SoIC3D堆栈与CoWoS整合的协同组合驱动AI运算的指数扩展。从2024年至2029年,单一CoWoS封装内的运算晶体管总数将增加48倍。此扩展源自三个并行矢量:CoWoS Interposer尺寸从3.3倍扩展至超过14倍reticle尺寸、透过SoIC堆栈使系统单晶片数量从2颗增加至超过24颗,以及个别晶粒晶体管密度透过逻辑节点从N7进展至A14而改善。
- 台积电计划2026年导入N3P-on-N4SoIC堆栈于更大的5.5-reticle Interposer达成4.6倍扩展。2027年世代运用N2P-on-N3PSoIC于9.5-reticle Interposer达到12倍扩展。2028年平台采用N2P-on-N2PSoIC于14-reticleInterposer达成32倍扩展。最终,2029年系统整合A14-on-A14SoIC于超过14-reticle Interposer,相对于2024年基准线超越48倍运算晶体管扩展。

系统集成拓展AI算力

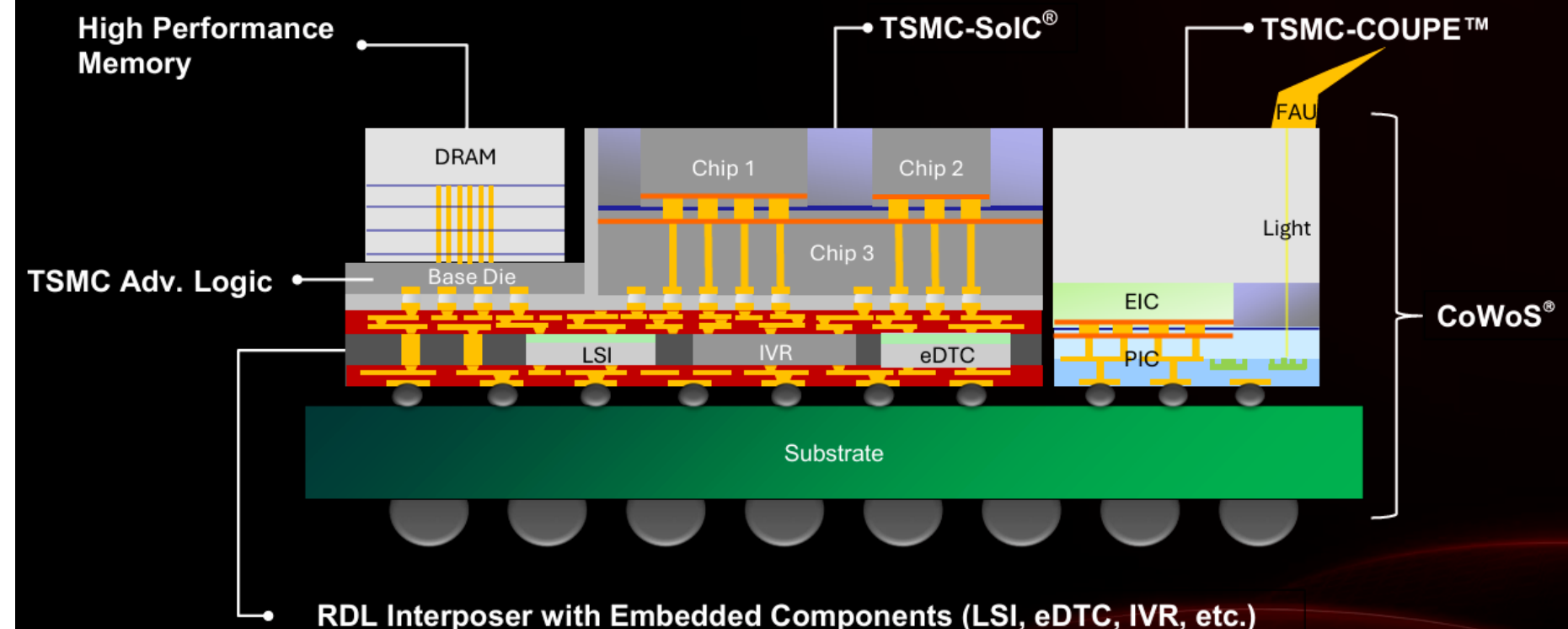
System Integration Scales AI Compute

- AI compute scaling is driven by the combination of advanced logic, SoIC 3D stacking, and CoWoS technologies
- From 2024 to 2029, AI compute transistors in one CoWoS will increase by 48X



台积电HPCAI平台

TSMC HPC Platform for AI



芯片：从Blackwell到Rubin，英伟达GPU向Useful AI持续进化

國元國際
GUOYUAN INTERNATIONAL

- 英伟达的GPU从初代迭代至今，已有二十六年历史。从最初主要面向图形处理，逐步拓展至通用计算、人工智能、光线追踪以及高性能计算等多个前沿领域。英伟达每一代的新架构都在不断推进计算能力的极限，也推动了整个AI产业链的发展；
- 2017年，英伟达推出Volta架构，首次将Tensor Core（张量核心）引入消费级GPU，开启了“AI原生架构”的新时代。每一代Tensor Core都代表了一次重大的架构升级，在支持的数据类型、计算吞吐量、能效比和专用功能上都有显著提升。至2024年推出的Blackwell架构，已迭代至第五代，伴随的是各项性能的指数级增长。
- 2026年6月试产启动下一代AI平台Vera Rubin，首批产品将于7月向微软、谷歌、亚马逊、Meta、甲骨文等北美云厂商交付，第三季度进入批量生产阶段，第四季度实现产能爬坡。该平台采用台积电3nm制程工艺，集成72颗Rubin GPU和36颗Vera CPU，搭载新一代HBM4显存（高端版本支持1025GBHBM4e），推理性能较前代Blackwell架构提升35倍，单机架算力密度较Blackwell架构提升4倍。该平台支持高度复杂的分布式、异构计算，可以说是为了迎接“有用的AI”时代而专为AI Agent量身打造的平台。

英伟达历代GPU性能演进

架构	代表产品	发布时间	Tensor Core 代数	支持精度格式	FP16/BF16Dense 算力	FP8Dense 算力	显存规格	显存带宽	NVLink 版本 (双向带宽)	标志性技术创新
Volta	V100	2017	第一代	FP16/FP32	125 TFLOPS	-	16GB/32GB HBM2	900 GB/s	2.0(300 GB/s)	首次引入 Tensor Core混合精度训练硬件基础
Turing	T4	2018	第二代	FP16/INT8/INT4	65 TFLOPS	-	16GB GDDR6	320 GB/s		INT8/INT4 推理加速 70W 低功耗设计
Ampere	A100	2020	第三代	FP32/TF32/FP16/BF16/INT8/INT4	312 TFLOPS	-	40GB/80GB HBM2e	2.0 TB/s	3.0(600 GB/s)	TF32 精度(无需改代码加速)MIG 多实例 GPUBF16 成为训练主流
Hopper	H100	2022	第四代	FP32/TF32/FP16/BF16/FP8/INT8/INT4	989 TFLOPS	1979 TFLOPS	80GB HBM3	3.35 TB/s	4.0(900 GB/s)	FP8 训练支持 Transformer EngineTMA 张量内存加速器
Blackwell	B200	2024	第五代	FP32/TF32/FP16/BF16/FP8/FP4/INT8	~1125 TFLOPS	2250 TFLOPS	192GB HBM3e	8.0 TB/s	5.0(1800 GB/s)	双 die 封装FP4 硬件级支持支持 72 卡 NVLink 互联硬件级机密计算

数据来源：英伟达官网、国元证券经纪（香港）整理

AI Agent时代的算力重构，CPU会越来越重要

- 未来的AI算力竞争，不再是单一硬件的比拼，而是CPU、GPU、内存、网络、先进封装协同发力的全栈式竞争。
- 传统数据中心配比：过去，在AI训练和大语言模型为主的场景下，数据中心中CPU与GPU的配比通常为1:4甚至1:8，即每1颗CPU对应4颗或8颗GPU，此时GPU承担主要的计算负载，CPU主要负责调度和数据搬运。
- 当前及未来趋势：随着AI进入“智能体时代”，智能体需要自主拆解任务、规划步骤、调用工具、处理数据等，这些工作主要由CPU完成，GPU仅负责模型推理环节。因此，CPU与GPU的配比正在从1:4或1:8向1:1靠拢，甚至在部分场景下CPU需求可能超过GPU。

目前行业巨头纷纷发布适应AI Agent时代的CPU，包括：

英伟达发布VeraCPU：该架构放弃了传统CPU的多核租赁模式，转而追求极致的单线程性能和带宽：

- IPC：拥有全球最高的指令周期（IPC）
- 带宽：内部织网带宽达到3.6TB/s，内存带宽是高性能x86的2到3倍
- 效率：运行SQL的速度提升3倍，实时流处理性能提升6倍

AMD也发布了搭载Zen5架构的Venice系列CPU，已采用台积电2纳米工艺完成量产，主打高性能与高能效，卡位高端服务器市场。

存储：功能升级，供给或许“永远”跟不上需求

- 从H100的80GB再到Rubin Ultra的384GB，英伟达每一代AI芯片的HBM搭载容量都在大幅提升。这一容量升级意味着即便AI芯片的出货量保持稳定，单芯片HBM容量的提升也会推动HBM总需求的持续增长。AI系统正在改变存储需求的结构：HBM是训练和高端推理的带宽核心，DDR/LPDDR承担CPU侧控制平面和通用内存，eSSD支撑推理数据、缓存、向量检索和高吞吐存取，HDD则继续承接冷数据和超大规模对象存储。也就是说，存储不再只是简单功能品类，而是AI系统每一层都要消耗的基础资源。
- 此外，存储需求被用户数、使用时长、任务复杂度、推理token消耗等多个变量相乘放大。AI并不是只活跃在云端，手术室、工厂、汽车，凡是AI业务发生的场景，算力就必须在场。未来数以千亿计的AI Agent将时刻不停且极速地调用工具、访问长期记忆并进行推理规划，这需要依托顶级CPU构筑的底层设施来承载，使模型大脑真正转化为能执行任务的数字机器人。未来每个AI Agent都需要专属算力与存储，这意味着需求规模将有指数级增长空间。
- 2026年6月初，SK海力士宣布大规模扩产计划，将在未来五年内将晶圆产能翻倍，以应对全球人工智能基础设施建设带来的内存芯片持续短缺。SK集团董事长崔泰源表示：全球存储芯片供应缺口可能持续至2030年，公司正加大资本支出以弥合供需失衡。

英伟达历代GPU的HBM容量

指标	H100(2022)	H200(2023)	GB200(2024)	GB300(2025)	VR100(2026)	Rubin Ultra(2027)
HBM 类型	HBM3	HBM3e	HBM3e	HBM3e	HBM4	HBM4e
HBM 容量 (GB)	80(基准)	141(+76%)	192(+36%)	288(+50%)	288(相同容量)	1,000(+247%)
容量对比 H100	1 倍	1.8×	2.4 倍	3.6 倍	3.6 倍	12.5 倍
内存带宽 (TB/s)	3.35(基准值)	4.8(+43%)	7.7(+60%)	8.0(+4%)	22(+175%)	32(+45%)
带宽与 H100	1 倍	1.4 倍	2.3 倍	2.4 倍	6.6×	9.6 倍
每 GPU 总 DRAM 容量 (GB)*	320	380	430	530	1050	2500

数据来源：英伟达官网、百度、国元证券经纪（香港）整理。注：Rubin和Rubin Ultra的参数为市场预测数，与实际发布或有差异。

存储：结构性短缺或导致存储价格持续处于高位

- 根据Trend Force集邦咨询最新存储器产业研究，AI发展从大型模型训练转向以推理为核心的Agentic AI（代理式AI）应用，驱动存储器需求结构性扩张，由于供给缺口短期无法补足，推升价格上涨。因此，Trend Force集邦咨询大幅上调全球存储器产值预估，将2026年产值从前一版的5,516亿美元提高至8,893亿美元，2027年则预计由8,427亿美元上修至逾1.28万亿美元，年增率约44%。
- 其中：由于HBM对晶圆耗用量的增加，压缩Conventional DRAM可用产能。在需求扩张的情况下，更强化供应商对合约价的话语权，预估可支持涨势至2027年。因此，Trend Force集邦咨询上调2026年DRAM产值至6,187亿美元，年增率达303%；预计2027年产值将进一步成长至9,033亿美元，年增46%。
- 由于需求强劲但供给紧张，NAND Flash价格获得支撑，Trend Force集邦咨询上修2026年全球NAND Flash产值达2,706亿美元，年增高达280.7%，2027年更将进一步扩大至近3,794亿美元，持续取得40.2%的年成长。

2023-2027年DRAM与NAND Flash产值预估

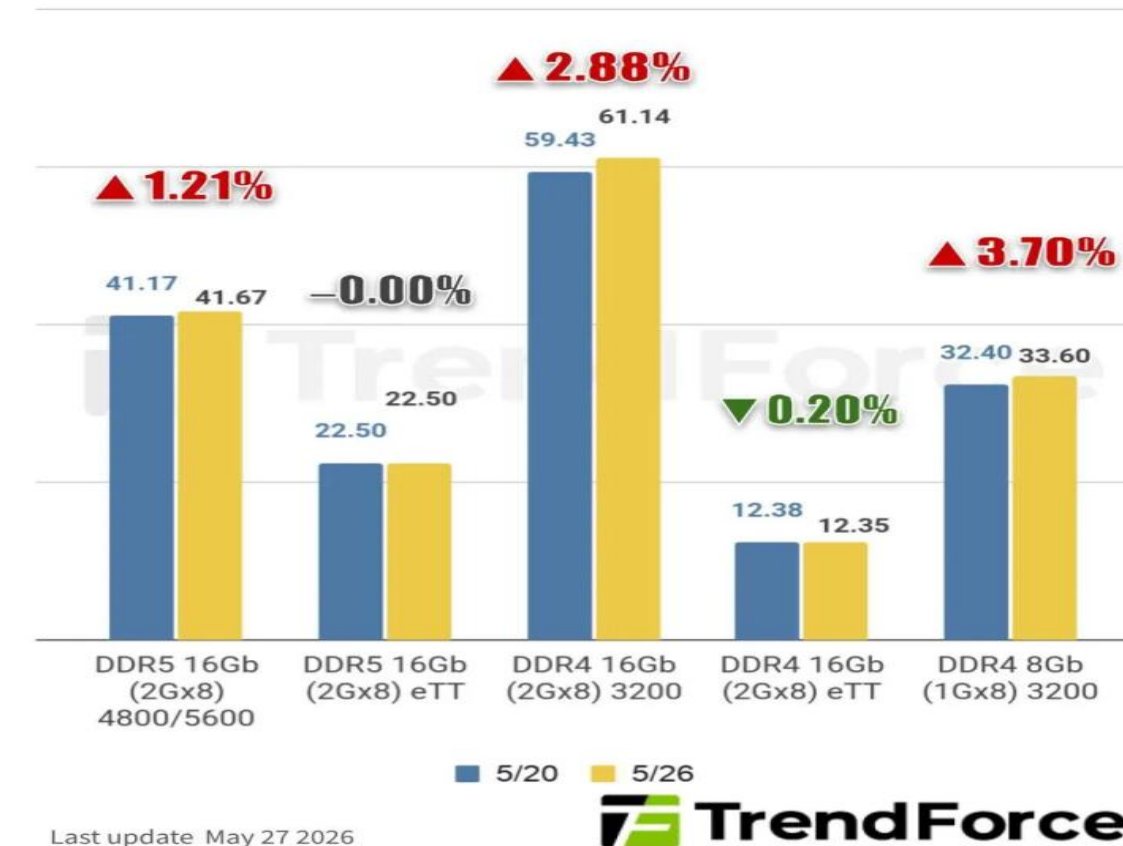
2023-2027年DRAM与NAND Flash产值预估 (unit: US\$ billion)



数据来源：Trend Force集邦咨询、国元证券经纪（香港）整理

DRAM Avg. Spot Price

DRAM Avg. Spot Price

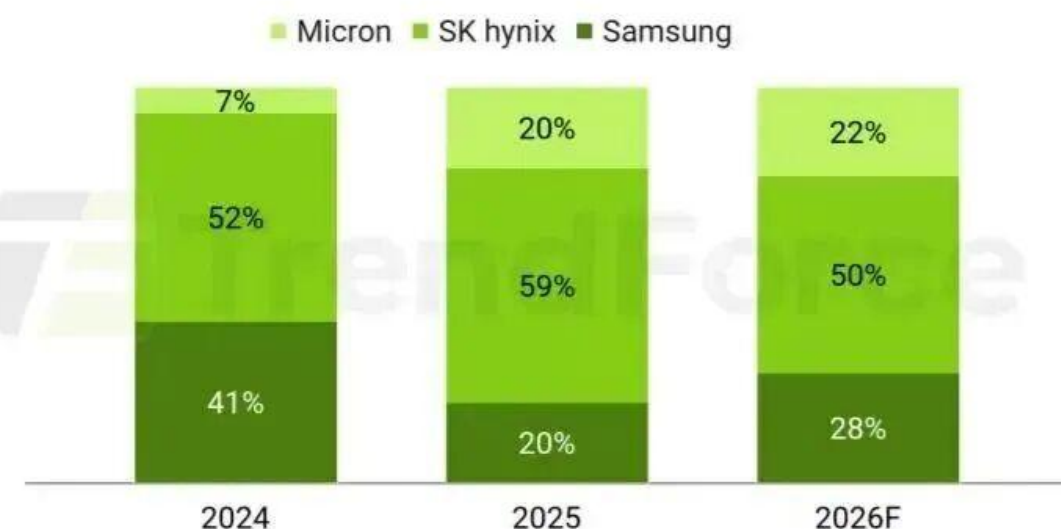


存储：结构性短缺或导致存储价格持续处于高位

- 2026年6月市场研究机构集邦咨询Trend Force发布最新研究指出，三大存储原厂三星电子、SK海力士、美光将于2027年大幅调高HBM高带宽存储器的报价。这一判断的核心依据是:DRAM供不应求的市况、新旧世代HBM的高制造难度及高成本，以及HBM对普通DRAM产能的挤占效应在2027年将更加严重。集邦咨询指出自2025年下半年以来，普通型DRAM价格大涨反映出供不应求的形势。但由于三大原厂的HBM采用年度议价机制，导致HBM合约价无法及时反映市场的季度涨价趋势。
- 据市场估算,HBM总可寻址市场(TAM)规模在2023–2027年间快速扩张,从2023年的约30亿美元升至2027年的约850亿美元。其中GPU用HBM2024–2027年复合增长率为73%,ASIC用HBM同期复合增长率为88%,ASICHBM增速高于GPUHBM。随着HBM芯片尺寸的扩大和需求的增加，HBM消耗的晶圆产能也在增加。由于HBM每GB消耗的晶圆产能约是普通DRAM的三倍，HBM产能的扩大会直接挤占普通DRAM的产能。

2024-2026年HBM生产位元占比

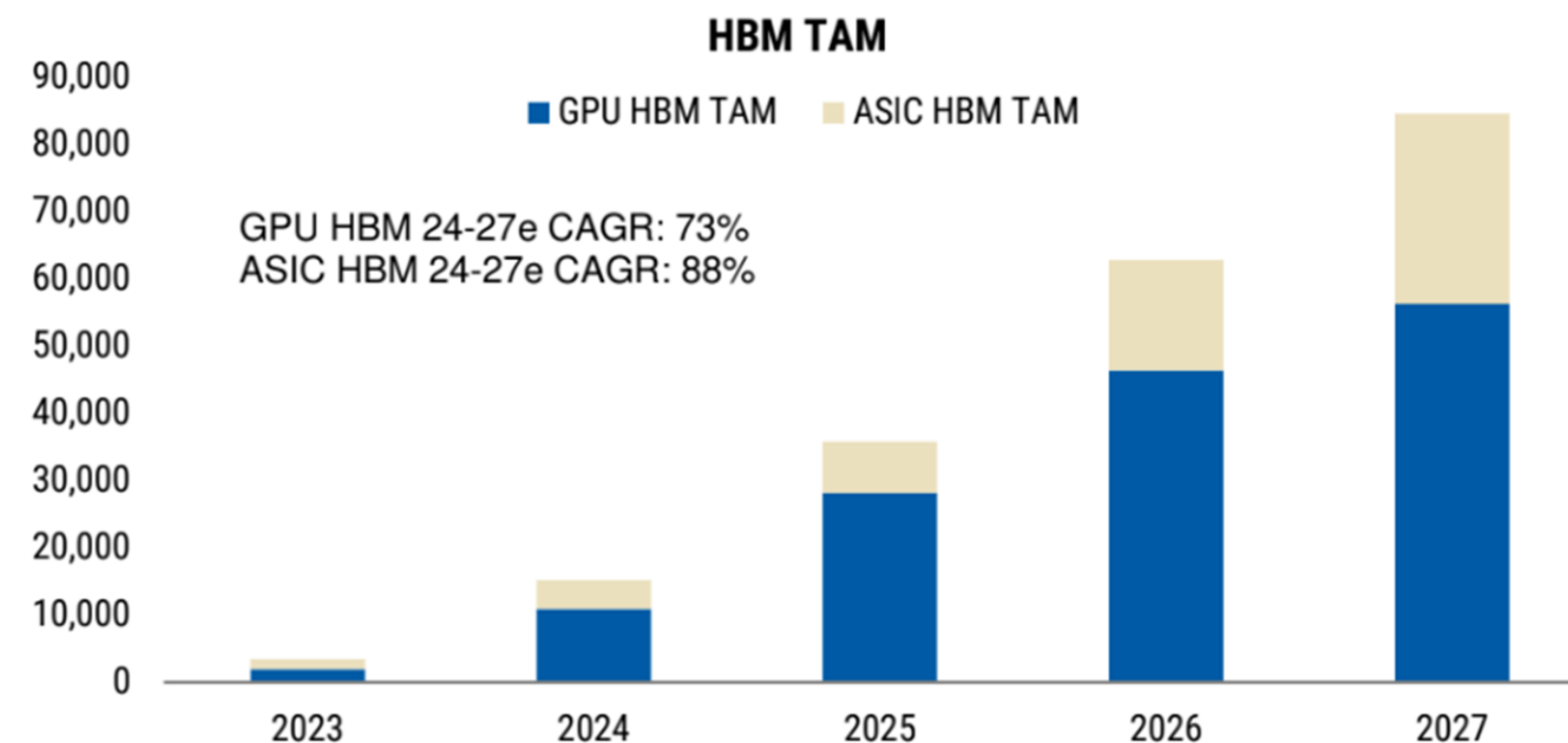
2024-2026年HBM生产位元占比



Source: TrendForce, Feb. 2026

TrendForce

HBM TAM



AI服务器：受益于AI浪潮，处于快速发展期

- AI服务器是适配大模型训练、推理任务的专用高性能算力硬件，是人工智能产业的核心算力底座。区别于传统通用服务器，其核心架构为“CPU+多GPU加速卡”，具备超高算力、高速互联、大容量存储、高效散热四大核心特性，伴随AI大模型迭代，产业链规模持续快速增长；
- AI服务器的核心价值集中在芯片、高速互联和散热电源三大高附加值环节。高速互联环节，800G、1.6T光模块、高速连接器、PCB高频板实现算力节点高速互通，是多GPU协同工作的关键，国内厂商已实现全球领先供货能力。电源与散热是AI服务器刚需，高算力芯片功耗激增，推动电源模块、液冷散热、CDU冷却单元需求爆发，单台AI服务器电源价值量远超普通服务器，液冷技术逐步替代风冷，成为高密度数据中心主流方案。另外，AI服务器的高功耗、高频率、高密度特性使得MLCC从传统的辅助元件转变为保障服务器稳定运行的核心基础设施，其作用在AI时代显著增强。
- 面对汹涌的市场需求，戴尔将2027财年全年的AI服务器收入预期从今年2月给出的500亿美元大幅上调至600亿美元，隐含约144%的同比增长。

AI服务器区别于传统服务器的发展趋势

设计维度	发展趋势
机架设计	在万卡级以上大模型训练集群中，机架设计正从传统的独立服务器堆叠，向芯片-服务器-机架-网络-供电-液冷全栈协同优化的系统级计算机架构演进；中低端训练集群和绝大多数推理集群仍采用"服务器+交换机"的传统堆叠模式。领先厂商（如NVIDIA、华为）已推出集成72-144颗AI芯片的机架级一体化计算单元，通过跨服务器的高速互联技术，使整个机架对外呈现为单一逻辑计算资源。
散热设计	单机柜功率超过40kW的AI训练集群，液冷技术已从可选项转变为刚需；单机柜功率低于40kW的推理集群和边缘AI节点，风冷仍为主流技术路线。当前冷板式液冷在AI训练集群中的渗透率约为74%，浸没式液冷主要应用于150kW以上的超高密度场景。液冷与风冷将长期按功率密度分层共存，不会出现液冷全面取代风冷的情况。
电力设计	为应对AI集群高功率密度和剧烈功率波动的挑战，动态功率调度技术已成为训练集群的标准配置；800V高压直流供电技术目前处于商用元年（全球渗透率约3-5%），预计2027-2028年进入大规模部署阶段。当前AI数据中心主流供电架构仍为48V直流，800V直流主要用于面向未来600kW以上超高密度机架的前瞻性建设。
网络连接	AI数据中心网络架构分化为**卡间互联（Scale-up）和集群组网（Scale-out）**两层：训练集群普遍采用NVLink/HCCS等专用低延迟卡间互联技术，以及InfiniBand无损集群组网技术；推理集群则主要采用PCIe卡间互联和200G/400G以太网组网。AI训练集群对网络的零丢包要求远高于传统数据中心，万分之一的丢包率即可导致30%以上的算力损失。

数据来源：戴尔官网、百度、国元证券经纪（香港）整理

PCB板：层数不断增加，价值量和需求量有望长期增长

- AI服务器、GPU等算力设备对信号传输密度要求越来越高，推动PCB层数从传统10-16层向20-64层甚至更高层数发展，板厚也相应增加，以满足复杂电路布局和电源分配需求。
- 与单层或双层PCB相比，高多层PCB面临着更高的复杂性和性能挑战。主要体现在三方面：一是高频信号传输特性要求。高多层PCB常用于需要高频信号传输的应用，如通信设备和高速数字信号处理。这要求PCB必须有更好的介电性能和信号传输速度，以减少信号损耗和噪音。二是热管理需求。高多层PCB通常用于高功率电子设备，需要有效的热管理措施保证元件正常工作温度。这就要求PCB有更好的导热性能和热膨胀系数控制，以优化热传导和分布，避免热点和热应力问题。三是机械强度和稳定性要求。高多层PCB通常有着较大的尺寸和较高的层数，因此需要更高的机械强度和稳定性。板材的刚性和耐久性对于抵抗振动、冲击和弯曲应力至关重要，以确保高多层PCB的可靠性和寿命。
- 更强的性能需求导致PCB板的加工难度和材料价值不断提升，PCB行业有望保持长期增长。

历代英伟达GPU平台PCB层数

GPU平台	量产时间	覆铜板材料	PCB层数	核心变化
H100/H200	2023-2024	M7-M8	40-60层	AI服务器PCB需求井喷
Blackwell B200/GB200	2024-2025	M8	50-70层	算力密度翻倍，PCB面积增大
Vera Rubin	2026下半年	M9	70-78层	正交背板替代铜缆，单板价值量翻倍
Rubin Ultra	2027	M10	78+层	信号损耗降低30%以上
Feynman	2028	M10+	80+层	下一代架构，更高集成度

数据来源：英伟达官网、百度、国元证券经纪（香港）整理

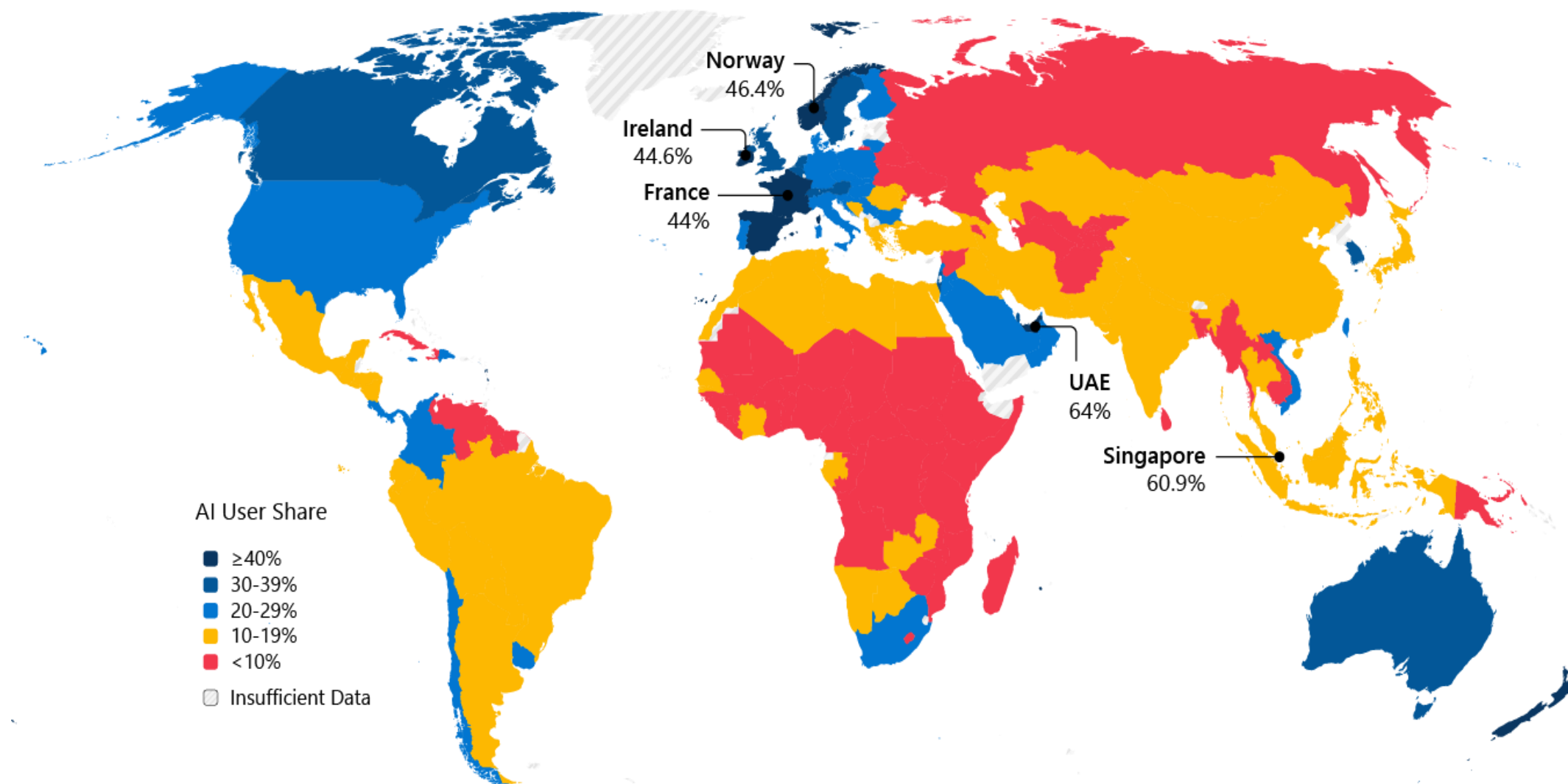
目录

- 一、AI性能快速迭代，商业落地推动AI Agent爆发
- 二、硬件产业链不得不全栈式、性能指数级进化以满足AI发展需求
- 三、AI仍处于早期，具有长足发展空间
- 四、AI硬件产业链的投资思路
- 五、风险提示

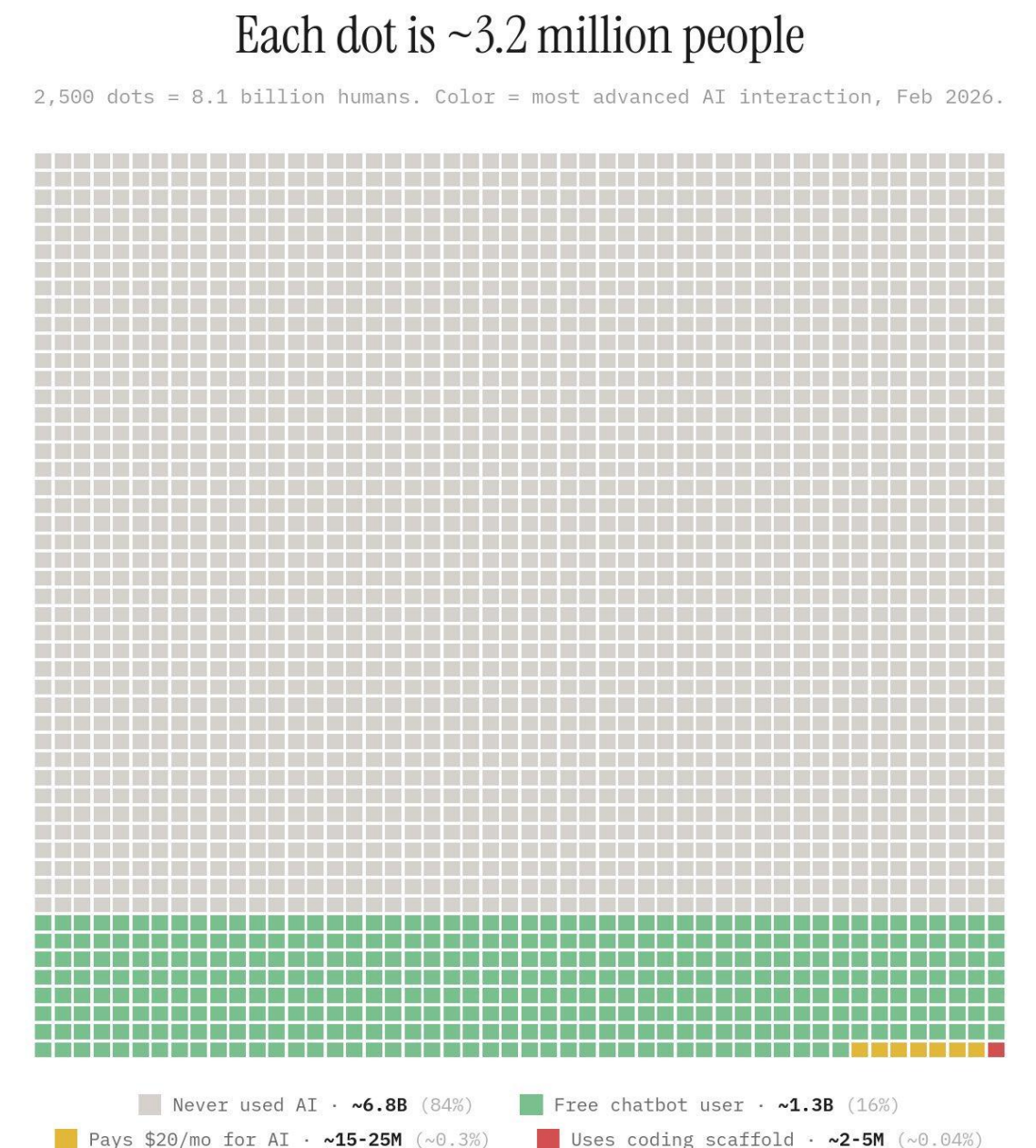
全球深度使用AI的渗透率仍很低，具有广阔提升空间

- 根据微软研究院数据，2025年下半年，人工智能的全球普及率持续上升，较上半年提升1.2个百分点。目前全球约每六人中就有一人使用生成式人工智能工具；
- 但根据AI Native Dev Studio的补充数据，目前使用生成式AI的人群中，大部分是免费使用。真正为AI工具付费、深度使用AI的人群只有1500~2500万人，占全球人口比例仅仅约0.3%。

2025年下半年各经济体AI普及度



全球使用AI人口结构（按使用程度）



数据来源：微软研究院、AI Native Dev Studio/Agent Integrator、国元证券经纪（香港）整理

从虚拟到现实：物理AI是新的万亿美元市场

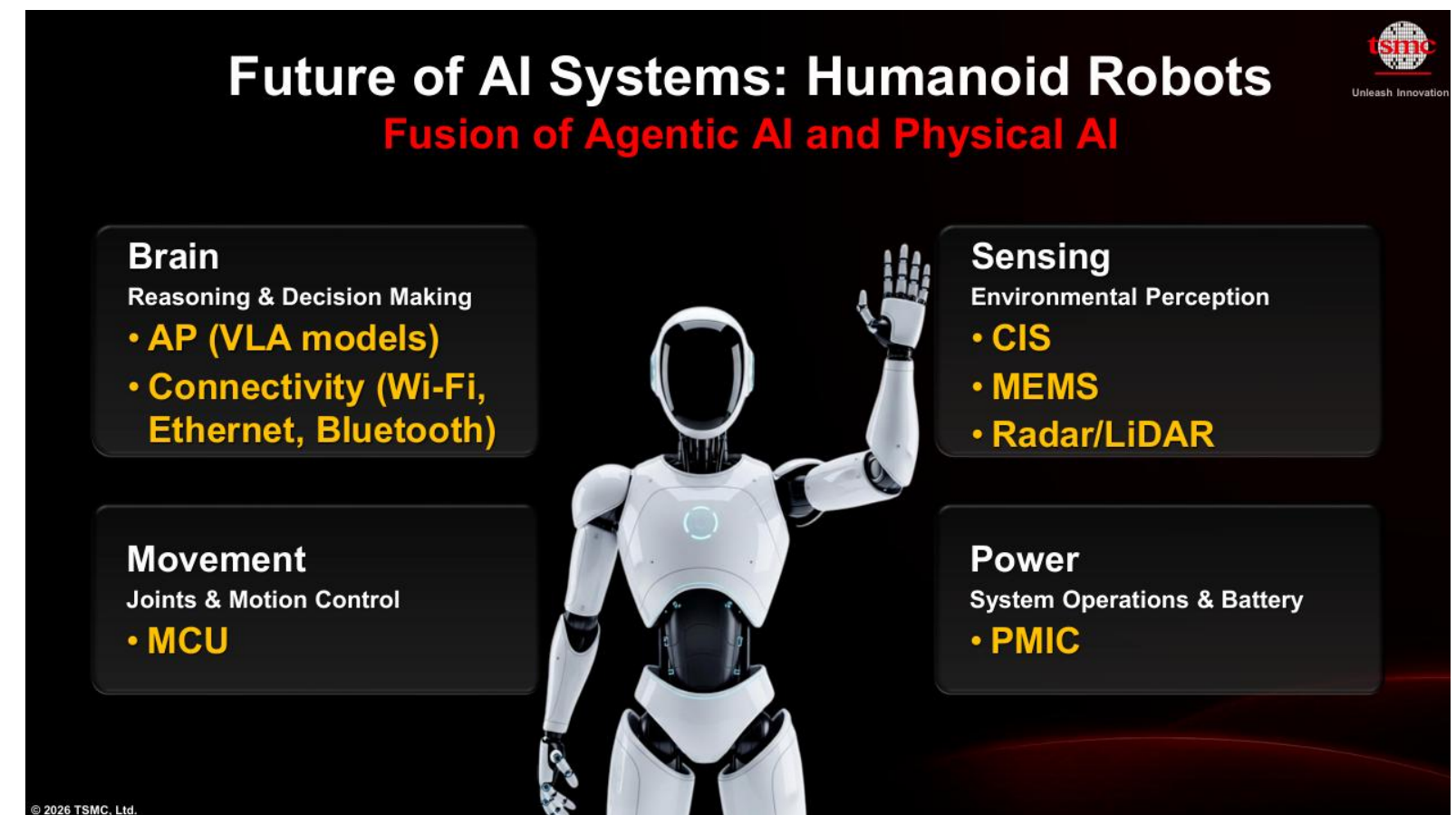
- 物理AI (Physical AI) 是指能够理解、推理并与物理世界交互的人工智能系统，其核心在于让AI不仅具备对虚拟信息的处理能力，还能通过物理载体（如机器人、自动驾驶汽车等）在真实世界中感知、决策和执行任务。
- 相对于当下“对话式”AI，物理AI是AI发展的大趋势。市场观点认为，未来物理AI的市场规模将达到万亿美元规模。
- NVIDIA已推出NVIDIACosmos™3，这是一款面向物理AI的开放世界基础模型，基于突破性的混合Transformer(Mixture-of-Transformers)架构所构建，将视觉推理、世界生成和动作预测集成至单一系统中。Cosmos3解决了物理AI的一项核心难题：使机器人、智能汽车或视觉智能体能够在训练数据有限和仿真堆栈碎片化的情况下，在现实世界中进行泛化。
- NVIDIA创始人兼首席执行官黄仁勋表示：“得益于多模态推理语言、视觉和世界模型领域的多项突破，物理AI爆发的时代已近在眼前。Cosmos3系列开放前沿全模态模型，助力开发者在构建机器人、智能汽车和视觉AI方面实现了代际飞跃，使其能够在物理世界中完成感知、推理、规划并采取行动。”

Cosmos3欲打造物理AI算力底座



数据来源：英伟达、国元证券经纪（香港）整理

未来AI系统：具身机器人



数据来源：台积电、国元证券经纪（香港）整理

目录

- 一、AI性能快速迭代，商业落地推动AI Agent爆发
- 二、硬件产业链不得不全栈式、性能指数级进化以满足AI发展需求
- 三、AI仍处于早期，具有长足发展空间
- 四、AI硬件产业链的投资思路
- 五、风险提示

AI硬件：短期有业绩，中长期有持续升级空间

- **整体估值相对合理：**根据Wind一致预期，截至2026年6月5日，全球AI硬件龙头在饱满订单的支持下预期未来3年（2026~2028年）均有望取得净利润的较快增长。很多龙头企业的PEG维持在1附近；
- **龙头公司业绩短期有超预期的可能性：**当下，存储、MLCC和PCB等核心元件由于供求关系非常紧张，产品价格存在进一步上调的可能性，进而导致公司净利润上升和PE下降；
- **产业链中长期增长有空间：**随着AI渗透率提升和物理AI的发展，整个产业的规模有望持续扩大，相关上市公司的业绩也有望持续增长。
- **建议积极关注AI硬件产业链。**按产业链各环节看：
 - 晶圆代工：台积电（TSM.N）、中芯国际（0981.HK）、华虹半导体（1347.HK）；
 - 算力芯片：英伟达（NVDA.O）、英特尔（INTC.O）、AMD（AMD.O）；
 - 存储：SK海力士（000660.KS）或XL二南方海力士（7709.HK）、三星电子（005930.KS）或XL二南三星（7747.HK）、美光科技（MU.O）、兆易创新（3986.HK）；
 - 服务器：联想集团（0992.HK）、戴尔科技（DELL.N）
 - 高速连接：迈威尔科技（MRVL.O）、剑桥科技（6166.HK）、长飞光纤光缆（6869.HK）、舜宇光学科技（2382.HK）、环联连讯（1473.HK）；
 - 先进封装：胜宏科技（2476.HK）、建滔积层板（1888.HK）；
 - MLCC制造和销售：天利控股集团（0117.HK）。
 - 其中，平台型龙头公司的技术领先性和业绩增长稳定性更强，值得优先关注。

AI硬件产业链关注标的（表一）

模块	公司简称	证券代码	总市值 (亿美元)	今年涨幅%	净利润（亿美元）				PE				PEG(至 2028年)
					2025	2026E	2027E	2028E	2025A	2026E	2027E	2028E	
算力	英伟达	NVDA.O	49,634	10.1	728.8	1,148.5	2,131.5	2,898.6	68.1	43.6	23.4	17.2	0.7
	超威半导体	AMD.O	7,605	117.8	43.4	88.8	180.2	243.7	175.4	85.6	42.5	31.7	1.1
	英特尔	INTC.O	4,981	168.8	-2.7	-1.5	64.4	103.4	-1,865.7	-3,373.1	83.1	51.3	7.7
	博通	AVGO.O	18,263	11.7	231.3	427.9	731.2	1,002.5	79.0	42.7	25.0	18.5	0.7
	高通	QCOM.O	2,276	27.5	55.4	93.2	91.6	100.9	41.1	24.8	25.0	22.5	1.1
存储	美光科技	MU.O	9,744	202.9	85.4	582.0	930.6	1,025.5	114.1	16.8	10.5	9.5	0.1
	西部数据	WDC.O	1,764	197.3	18.9	45.1	57.5	77.1	93.4	40.1	31.8	23.3	0.7
	SK海力士	000660.KS	9,565	218.6	296.8	781.7	963.2	970.4	34.4	12.2	9.9	9.9	0.3
	三星电子	005930.KS	12,471	175.0	306.0	1,068.3	1,268.5	1,536.1	43.5	11.7	9.8	8.1	0.2
	铠侠	285A.T	2,663	648.8	18.3	27.1	105.7	130.2	156.8	98.3	25.2	20.4	1.1
	阿斯麦	ASML.O	6,466	54.0	112.6	142.5	190.1	222.9	57.4	45.4	34.0	29.0	1.8
AI电器元件	村田制作所	6981.T	1,187	201.1	15.8	14.6	18.7	22.0	81.4	81.3	63.6	54.1	6.9
	TDK电子	6762.T	498	87.5	11.3	12.2	14.9	16.8	47.8	40.8	33.5	29.7	2.9
	舜宇光学科技	2382.HK	115	27.0	6.6	6.1	7.2	8.6	17.5	18.9	16.0	13.3	2.0
	比亚迪电子	0285.HK	82	-15.0	5.0	5.4	6.8	7.9	16.5	15.2	12.1	10.4	0.9
	天利控股集团	0117.HK	6	1,830.3	0.2	0.0	0.0	0.0	32.1	0.0	0.0	0.0	0.0
PCB	胜宏科技	2476.HK	486	17.0	6.1	0.0	0.0	0.0	75.9	0.0	0.0	0.0	0.0
	建滔积层板	1888.HK	200	278.5	3.1	6.4	8.1	10.0	64.1	31.2	24.7	20.0	0.7
芯片制造	台积电	TSM.N	21,533	37.0	538.8	742.0	911.0	1,080.1	40.0	29.3	23.9	20.4	1.1
	中芯国际	0981.HK	955	5.9	6.9	12.2	17.2	22.5	113.7	63.2	44.9	34.4	1.3
	华虹宏力	1347.HK	382	95.6	0.5	1.8	2.7	3.7	591.2	174.3	121.0	86.2	1.9
	德州仪器	TXN.O	2,594	66.2	49.7	63.1	74.4	86.6	51.9	41.2	34.9	29.9	2.0
	意法半导体	STM.N	629	173.4	1.7	0.0	0.0	0.0	378.6	0.0	0.0	0.0	0.0
	恩智浦半导体	NXPI.O	747	37.1	20.2	35.6	34.8	42.1	37.0	21.1	21.4	17.8	0.8

数据来源：Wind一致预期、国元证券经纪（香港）整理

AI硬件产业链关注标的（表二）

模块	公司简称	证券代码	总市值 (亿美元)	今年涨幅%	净利润（亿美元）				PE				PEG(至 2028年)
					2025	2026E	2027E	2028E	2025A	2026E	2027E	2028E	
服务器/机柜	超微电脑	SMCI.O	250	42.3	10.5	12.4	16.8	22.0	23.9	22.4	16.9	12.9	0.8
	戴尔科技	DELL.N	2,547	215.7	45.9	47.7	101.3	110.5	55.5	54.8	25.2	23.0	1.6
	惠普	HPQ.N	234	16.7	25.3	21.2	23.1	26.1	9.3	11.0	10.1	9.0	11.0
	国际商业机器	IBM.N	2,677	-2.6	105.9	96.9	105.3	113.8	25.3	27.6	25.4	23.5	11.4
	联想集团	0992.HK	394	168.7	13.8	0.00	23.0	28.6	28.7	0.0	17.5	14.1	0.0
光通信/高速连接	迈威尔科技	MRVL.O	2,305	210.4	-8.9	25.5	18.7	34.2	-260.4	90.5	123.1	67.4	-0.4
	LUMENTUM	LITE.O	672	134.3	0.3	4.7	12.8	21.9	2,594.3	153.9	59.4	35.9	0.5
	康宁	GLW.N	1,528	103.5	16.0	20.0	28.3	34.3	95.8	80.1	55.9	46.0	2.8
	环联连讯	1473.HK	2	313.0	0.0	0.0	0.0	0.0	62.2	0.0	0.0	0.0	0.0
	中兴通讯	0763.HK	260	9.3	8.0	10.2	12.8	11.9	22.8	17.8	14.2	15.2	1.3
	长飞光纤光缆	6869.HK	392	362.7	1.2	8.6	11.3	16.7	219.8	29.5	22.4	15.2	0.2
电力	赛晶科技	0580.HK	5	17.2	0.2	0.2	0.2	0.3	23.6	20.5	18.7	16.9	1.8
	重庆机电	2722.HK	12	39.0	1.1	1.5	1.7	2.0	11.0	8.1	6.9	6.0	0.4
机器人和自动驾驶相关标的	地平线机器人-W	9660.HK	94	-42.1	-14.9	-4.5	-1.9	0.8	-6.3	-20.8	-50.5	113.8	0.2
	优必选	9880.HK	71	-12.7	-1.0	-0.4	0.2	1.0	-71.3	-164.4	454.8	73.8	0.8
	越疆	2432.HK	17	-20.3	-0.1	-0.1	0.0	0.1	-143.7	-151.6	-561.4	198.5	0.8
	卧安机器人	6600.HK	23	0.4	0.0	0.1	0.2	0.3	-668.1	388.1	107.3	77.4	-1.3
	比亚迪股份	1211.HK	1,163	-5.9	46.0	62.4	78.0	90.9	22.6	16.7	13.4	11.5	0.7
	速腾聚创	2498.HK	20	-12.0	-0.2	0.1	0.5	0.8	-96.6	137.9	41.3	23.9	-0.5
	禾赛-W	2525.HK	31	-13.1	0.6	0.8	1.4	1.9	50.3	37.6	22.2	16.2	0.8
	丘钛科技	1478.HK	14	14.4	2.1	1.3	1.6	1.9	6.7	10.5	8.6	7.3	-3.2
	瑞声科技	2018.HK	70	18.9	3.6	4.2	4.9	5.6	19.8	16.8	14.4	12.5	1.0
	特斯拉	TSLA.O	14,685	-13.1	37.9	48.7	71.5	97.3	387.1	301.8	205.4	150.9	8.2
	文远知行-W	0800.HK	24	-21.8	-2.4	-2.3	-1.9	-1.3	-10.1	-10.4	-12.3	-18.8	0.6
	小马智行-W	2026.HK	41	-36.7	-1.3	-2.5	-2.3	-1.5	-30.9	-16.7	-18.0	-26.5	-3.4
	驭势科技	1511.HK	17	42.6	-0.3	0.0	0.0	0.0	-53.1	0.0	0.0	0.0	0.0
	黑芝麻智能	2533.HK	13	-23.8	-2.0	-1.4	-0.6	-0.1	-6.7	-9.8	-21.5	-144.5	0.2

数据来源：Wind一致预期、国元证券经纪（香港）整理

目录

- 一、AI性能快速迭代，商业落地推动AI Agent爆发
- 二、硬件产业链不得不全栈式、性能指数级进化以满足AI发展需求
- 三、AI仍处于早期，具有长足发展空间
- 四、AI硬件产业链的投资思路
- 五、风险提示

- **短期风险：** AI硬件行业整体估值可能受全球宏观因素（如：美元利率、地缘动荡等）以及交易集中度等产生波动；
- **中长期风险：** 如果AI的使用未能在中长期产生符合资本预期的收益率，可能导致AI Capex下降，进而使得相关公司的订单和利润减少，同时也可能造成其估值中枢下移。

免责声明



一般声明

本报告由国元证券经纪（香港）有限公司（简称“国元证券经纪（香港）”）制作，国元证券经纪（香港）为国元国际控股有限公司的全资子公司。本报告中的信息均来源于我们认为可靠的已公开资料，但国元证券经纪（香港）及其关联机构对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供投资者参考之用，不构成对买卖任何证券或其他金融工具的出价或征价或提供任何投资决策建议的服务。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐或投资操作性建议。投资者应当对本报告中的信息和意见进行独立评估，自主审慎做出决策并自行承担风险。投资者在依据本报告涉及的内容进行任何决策前，应同时考虑各自的投资目的、财务状况和特定需求，并就相关决策咨询专业顾问的意见对依据或者使用本报告所造成的一切后果，国元证券经纪（香港）及/或其关联人员均不承担任何责任。

本报告署名分析师与本报告中提及公司无财务权益关系。本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。在不同时期，国元证券经纪（香港）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。

本报告署名分析师可能会不时与国元证券经纪（香港）的客户、销售交易人员、其他业务人员或在本报告中针对可能对本报告所涉及的标的证券或其他金融工具的市场价格产生短期影响的催化剂或事件进行交易策略的讨论。这种短期影响的分析可能与分析师已发布的关于相关证券或其他金融工具的目标价、评级、估值、预测等观点相反或不一致，相关的交易策略不同于且也不影响分析师关于其所研究标的证券或其他金融工具的基本面评级或评分。

国元证券经纪（香港）的销售人员、交易人员以及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。国元证券经纪（香港）没有将此意见及建议向报告所有接收者进行更新的义务。国元国际控股有限公司的资产管理部门、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不一致的投资决策。

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。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所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地影响所预测的回报。

本报告提供给某接收人是基于该接收人被认为有能力独立评估投资风险并就投资决策能行使独立判断。投资的独立判断是指，投资决策是投资者自身基于对潜在投资的目标、需求、机会、风险、市场因素及其他投资考虑而独立做出的。

特别声明

在法律许可的情况下，国元证券经纪（香港）可能与本报告中提及公司正在建立或争取建立业务关系或服务关系。因此，投资者应当考虑到国元证券经纪（香港）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。

本报告的版权仅为国元证券经纪（香港）所有，未经书面许可任何机构和个人不得以任何形式转发、翻版、复制、刊登、发表或引用。

分析员声明

本人具备香港证监会授予的第四类牌照——就证券提供意见。本人以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

国元国际控股有限公司

香港中环康乐广场8号交易广场三期17楼

电话：(852)37696888

传真：(852)37696999

服务热线：400-888-1313

公司网址：<http://www.gyzq.com.hk>