

扬帆起航 AI浪潮下的通信投资策略演进

通信行业

通信团队 证券分析师：马军 SAC NO:S1120523090003

2025年3月13日



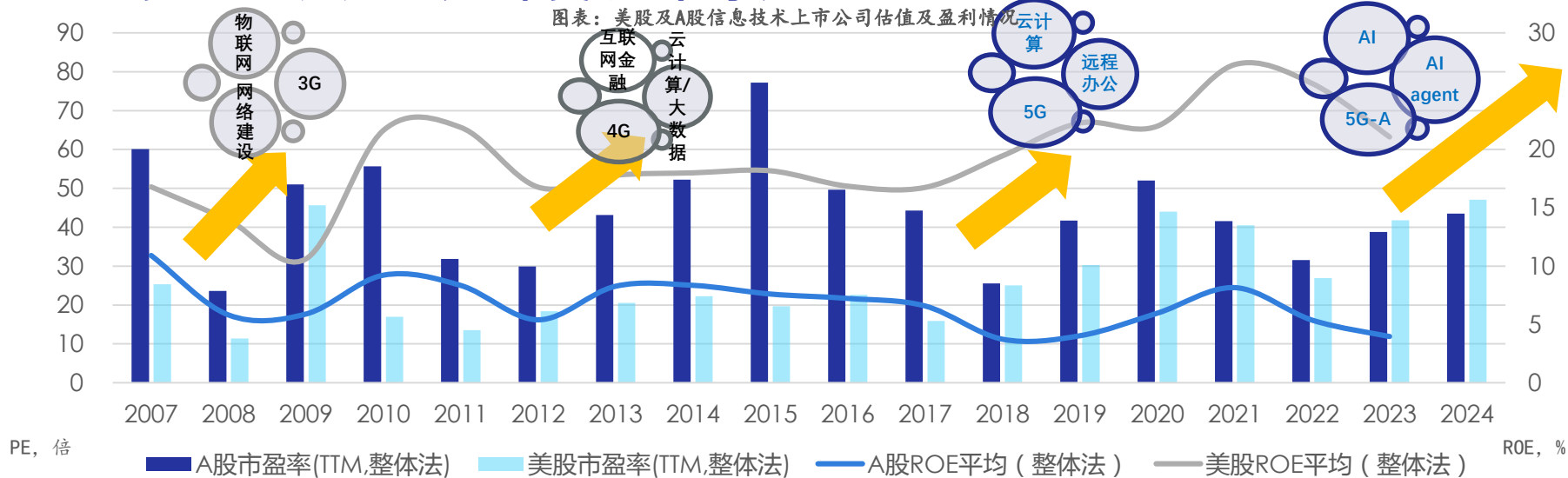


01 大势研判：科技行业大年

市值排名	1980年	1985年	1990年	1995年	2000年	2005年	2010年	2015年	2017年	2018年	2019年	2020年	2022年	2023年	2024年	2025年2月	2025年3月
1	IBM	IBM	IBM	GE	GE	GE	埃克森美孚	Apple	Apple	Microsoft	Apple	Apple	Apple	Apple	Apple	Apple	Apple
2	AT&T	埃克森	埃克森	AT&T	埃克森美孚	埃克森美孚	Apple	AlphaBet	AlphaBet	Apple	Microsoft	Microsoft	Microsoft	Microsoft	NVIDIA	NVIDIA	Microsoft
3	埃克森	GE	GE	埃克森	辉瑞	Microsoft	Microsoft	Microsoft	Microsoft	Amazon	Amazon	Amazon	AlphaBet	AlphaBet	Microsoft	Microsoft	NVIDIA
4	Indiana标准石油	AT&T	菲利普莫里斯	可口可乐	花旗	花旗	伯克希尔	伯克希尔	amazon	AlphaBet	AlphaBet	AlphaBet	Amazon	Amazon	Amazon	Amazon	Amazon
5	斯伦贝谢	GM	荷兰皇家石油	默克	思科	宝洁	GE	埃克森美孚	facebook	伯克希尔哈撒韦	腾讯	Facebook	伯克希尔哈撒韦	NVIDIA	AlphaBet	AlphaBet	AlphaBet
6	壳牌	荷兰皇家石油	百时美施贵宝	荷兰皇家石油	沃尔玛	沃尔玛	沃尔玛	Amazon	腾讯	腾讯	伯克希尔哈撒韦	腾讯	联合健康	Facebook	Facebook	Facebook	Facebook
7	美孚	杜邦	默克	菲利普莫里斯	Microsoft	美国银行	Google	Facebook	伯克希尔哈撒韦	Facebook	Facebook	TESLA	强生	TESLA	TESLA	BROADCOM	伯克希尔哈撒韦
8	California标准石油	阿克石油	沃尔玛	宝洁	AIG	强生	雪弗兰	GE	阿里巴巴	阿里巴巴	阿里巴巴	阿里巴巴	埃克森美孚	伯克希尔哈撒韦	BROADCOM	TESLA	台积电
9	阿克石油	贝尔南方	AT&T	强生	默克	AIG	IBM	强生	强生	强生	强生	台积电	沃尔玛	ELI LILLY	台积电	台积电	ELI LILLY
10	GE	西尔斯	可口可乐	Microsoft	Intel	辉瑞	宝洁	富国银行	摩根大通	摩根大通	埃克森美孚	伯克希尔哈撒韦	腾讯	台积电	伯克希尔哈撒韦	伯克希尔哈撒韦	TESLA

技术变革+叠加产业政治，美股估值攀升

图表：美股及A股信息技术上市公司估值及盈利情况

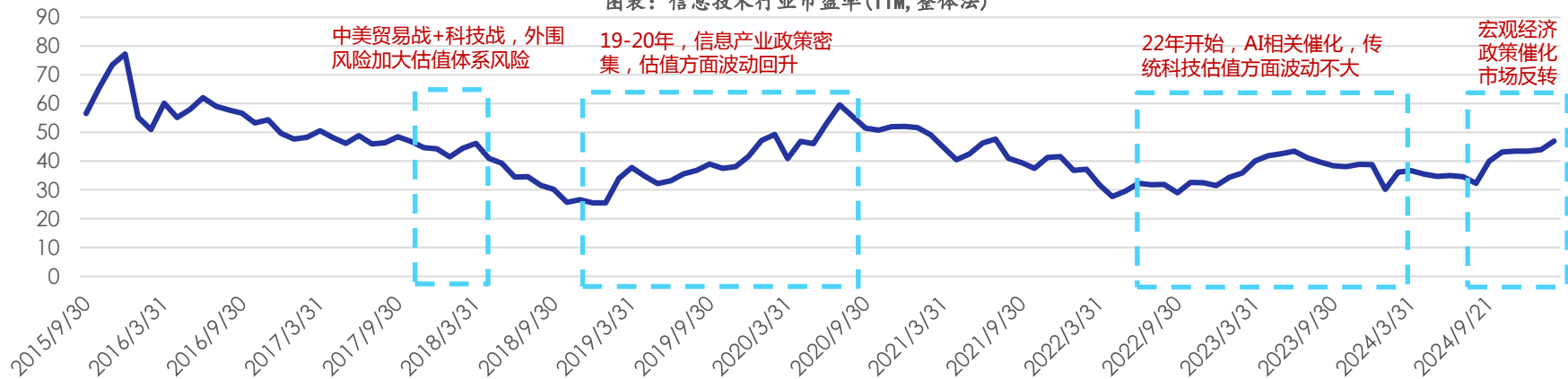


	2008-2010	2013-2015	2018-?	2023-?
通信技术	3G	4G	5G、5G-A	
终端	智能手机萌芽	智能手机高速增长、成熟	泛在智能设备、物联网	AI侧终端
应用	移动上网	移动互联网、互联网金融	高清、AR/VR、产业互联网	AI agent、低空经济

资料来源：wind，华西证券研究所整理

A股：政策催化，市场反转，行业估值拉升

图表：信息技术行业市盈率(TTM, 整体法)

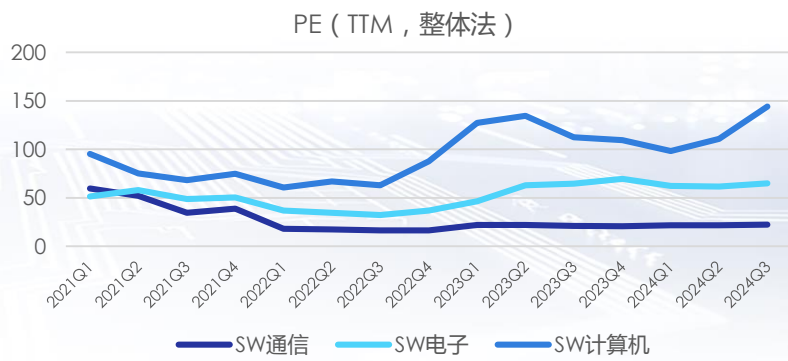
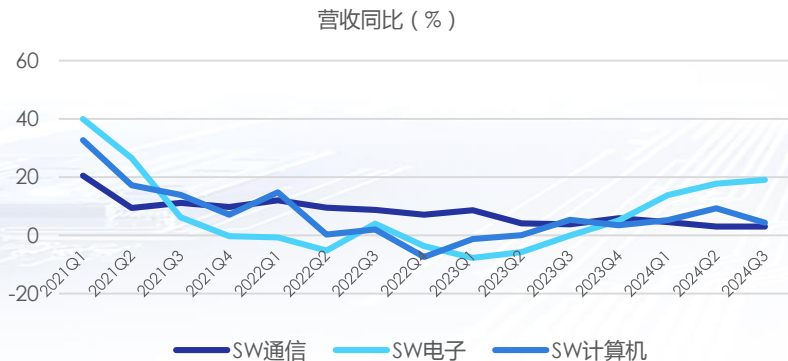
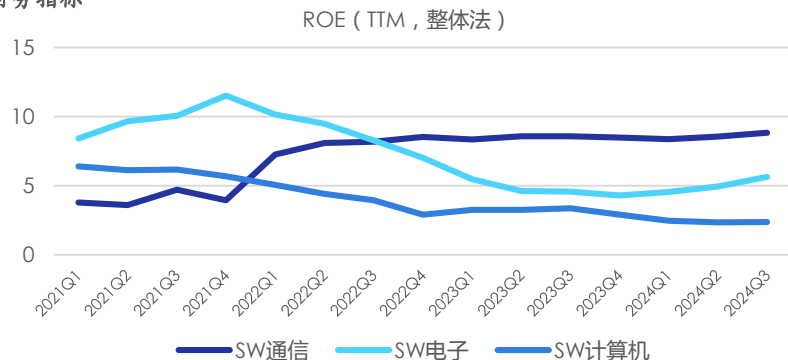
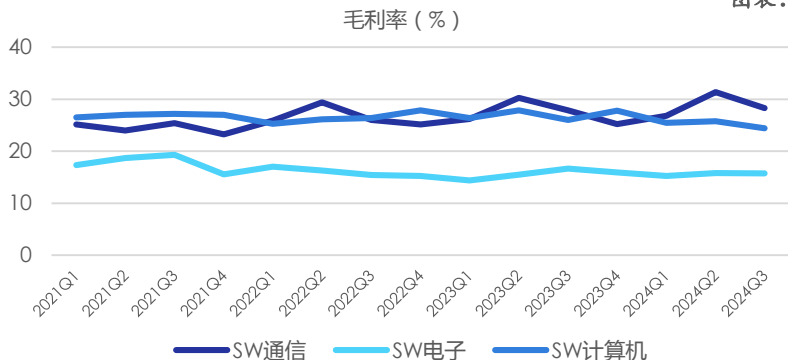


资料来源：wind，华西证券研究所整理

A股：TMT底部修复，市场情绪关注点

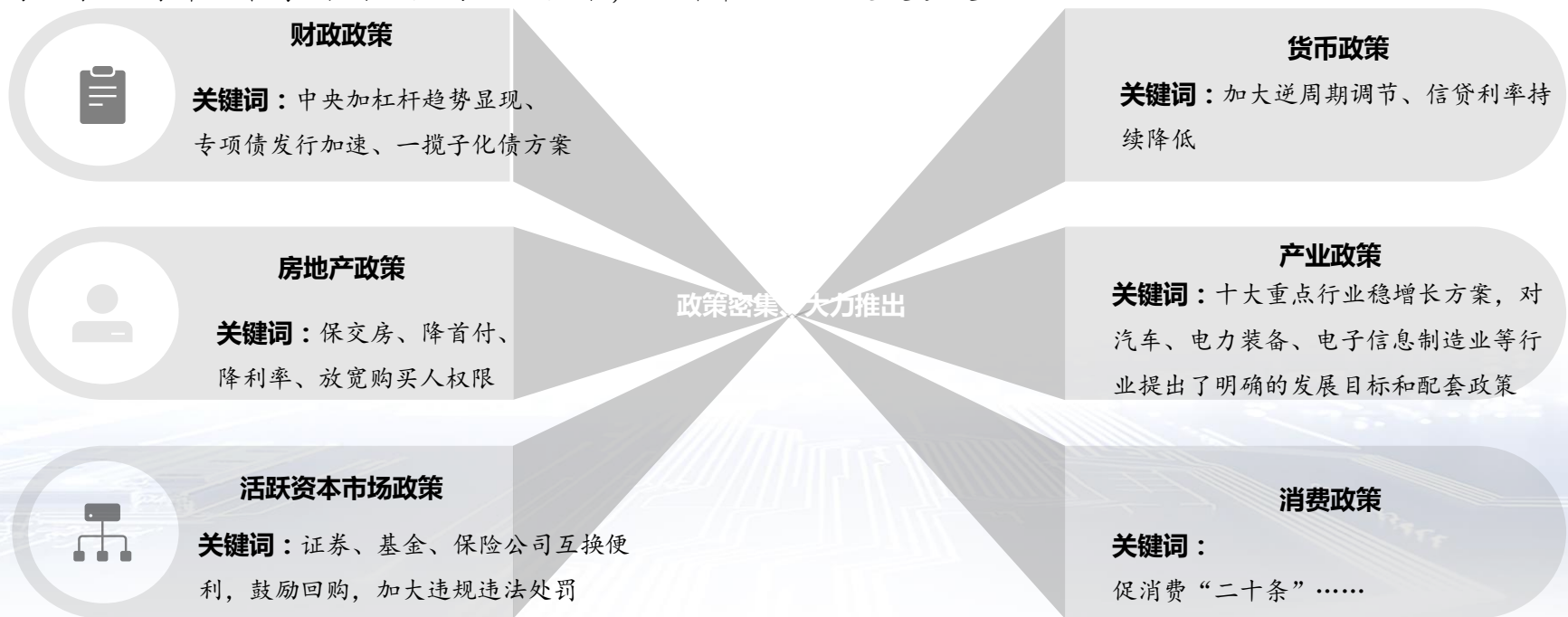
整体营收与ROE稳定回升，季度毛利率有所下滑，市场情绪拉动PE快速回升。

图表：TMT核心财务指标



资料来源：wind，华西证券研究所整理

国内宏观经济政策持续有力，预计宏观调控的政策工具包还会不断丰富完善。实际经济指标转暖及改善程度待观察。内部政策导向确定性与一致性下，预计市场信心缓慢修复。

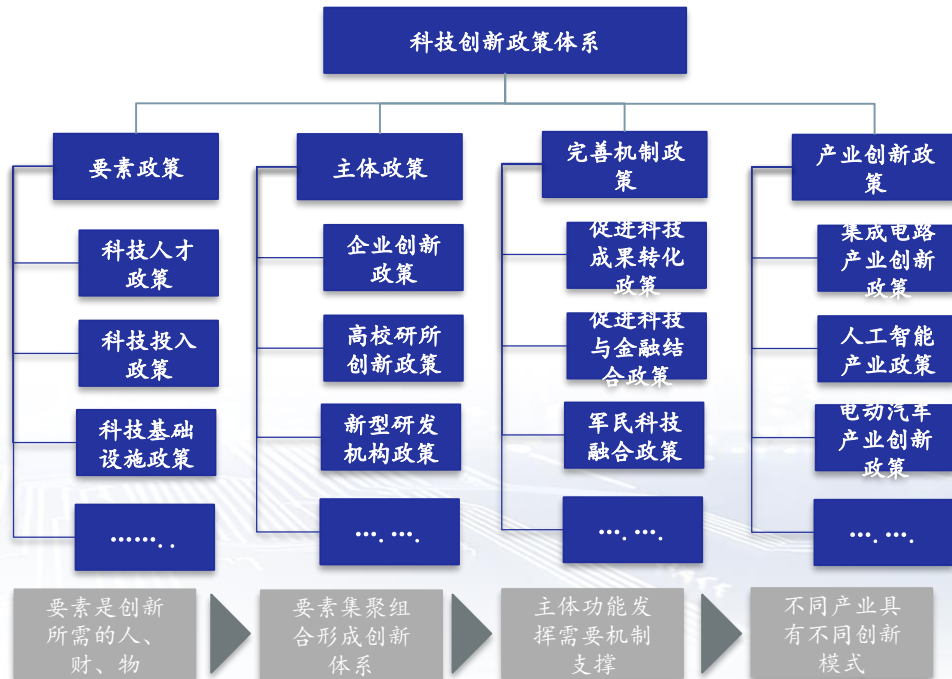


中国政府高度重视科技产业发展，科技产业政策目标明确：瞄准人工智能、量子信息、集成电路、生命健康、脑科学、生物育种、空天科技、深地深海等前沿领域，实施一批具有前瞻性、战略性的国家重大科技项目。采取基础研究与科技应用并举的方式，除了政府会加大力度实施国家重大科技项目及建设科研中心外，企业进行科研也会有更多税收等财政政策支持。

中国5年规划比较

	十二五规划 (2011-2015年)	十三五规划 (2016-2020年)	十四五规划 (2021-2025年)
政策定位	全面建设小康社会的关键时期，加快转变经济发展方式的攻坚时期	全面建成小康社会决胜阶段	
发展环境	后全球金融危机时代	全球经济增长乏力、贸易保护主义兴起	
政策理念	减持科学发展	创新、协调、绿色、开放、公用	创新、协调、绿色、开放、公用
经济增速目标	GDP年均增长+7%	经济保持中高速增长，2020年GDP相比2010年翻倍	在质量效益明显提升的基础上实现经济持续健康发展
政策重点支持产业	战略新兴产业： 1、节能环保 2、新一代信息技术 3、生物医药 4、高端装备 5、新能源 6、新材料 7、新能源汽车	战略新兴产业 1、新一代信息技术 2、生物医药 3、高端装备制造 4、新能源 5、新材料 6、新能源汽车 7、航空航天 8、中国制造2025	战略新兴产业 1、新一代信息技术 2、生物技术 3、新能源 4、新材料 5、高端装备 6、新能源汽车 7、绿色环保 8、航空航天 9、海洋装备

科技创新政策体系



积极活跃的并购

并购6条落地，并购市场成为上市公司产业整合或转型的重要途径：9月24日发布《关于深化上市公司并购重组市场改革的意见》，支持上市公司注入优质资产、提升投资价值。9月24日至10月20日，A股市场52家上市公司披露了资产重组相关公告，涉及能源、船舶、医药、半导体、设备制造等多个领域。

经济发展的驱动模式已发生转变，从过去的数量型发展进入质量型驱动关键阶段，并购重组是实现产业整合和转型升级、优化资源配置、助力上市公司实现高质量发展的重要途径。



四类重要的并购趋势值得重视

科技类企业
并购小型科
技企业

制造类企业
的出海并购

周期类企业
横向并购

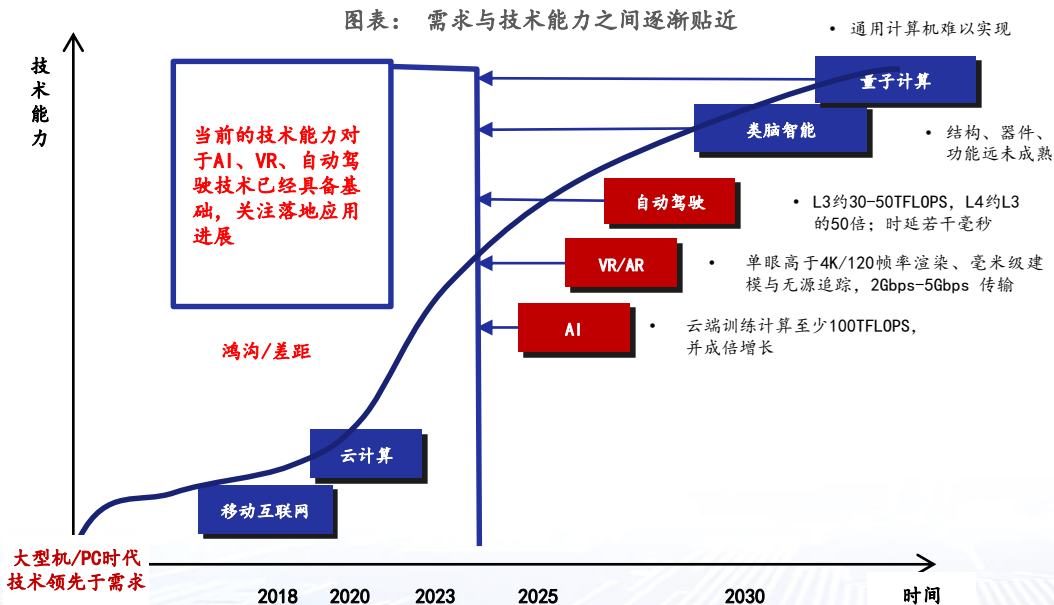
消费类企业
纵向并购



02 2025年关注点

AI泛在化、空天地一体（6G）、智能制造

产业趋势：新技术萌芽期，底层创新不断



■ 摩尔定律、多核高主频架构、软件并行加速等计算技术固有升级驱动力先后面临发展瓶颈，计算技术升级趋势逐渐放缓。

■ 后摩尔时代颠覆性技术前瞻：

- 新材料：III-V、SiGe、SOI、MEMS等
- 新架构：硅光子、CPO、RISC-V、存算一体、光计算、chiplet
- 新封装：3D封装、AiP封装
- 新装备：光刻机、检测设备等
- 新软件：EDA到PDA

■ 计算需求的增速远超摩尔定律，数据总量激增：进入Z字节时代，海量数据隐藏着巨大价值，实现前提就是具备分析计算能力。

■ 类型多样：由结构化向非结构化及不规则/定制演进。

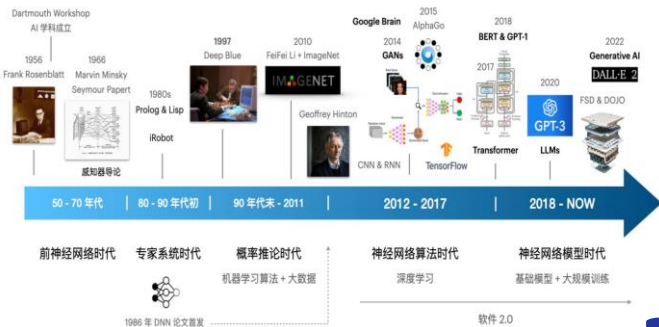
大模型能力爆发，模型算法+基础设施投资需要持续加大

- 以GPT为代表的AI大模型能力突出，AI技术的不断进步使得数据处理和分析的效率大幅提升。
- 深度学习的两大驱动力：提升硬件算力和模型，通过更多的层和非线性计算增加非线性能力的演进，持续带动模型算法及硬件研发投入加大。

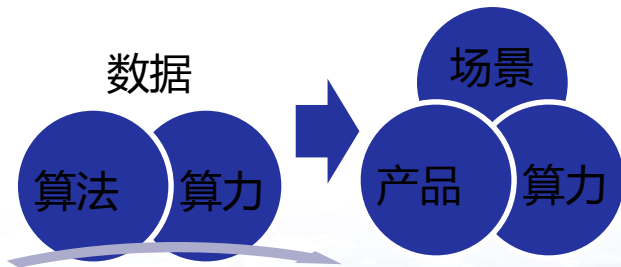
人类有望进入AGI（通用人工智能）时代

- AI大模型成为人工智能迈向通用智能的里程碑技术。基于数据的互联网时代和基于算力的云计算时代之后，我们将进入基于大模型的AI时代。
- AI大模型的落地应用使得AI的三要素由“数据、算法、算力”演变为“场景、产品、算力”。

图：AI进化史



图：高端智能制造不同领域



- **交互革命**
- 人机交互方式：GUI --- DUI/HUI
- Prompt工程价值凸显

AGI 0.1

AGI 1.0

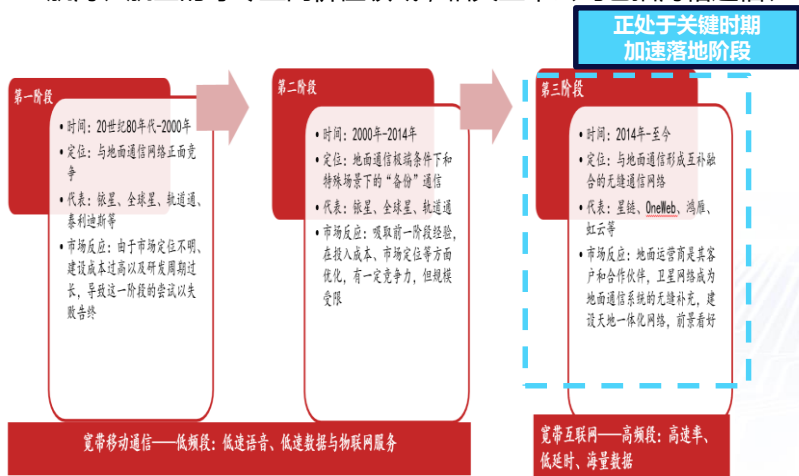
- **知识革命**
- 语言是知识的载体，未来模型人人可训、人人可用

- **思维革命**
- AI具备独立思考与逻辑判断能力
- 具备智能连接物理世界

AGI 2.0

卫星互联网，新基建之一，处于行业加速期

- **通信卫星制造**：低轨卫星数量预期发射量较大，卫星及星载核心硬件最先受益。通信卫星和有效载荷是通信转发器和天线（相控阵天线和固定多波束天线），Ka频段高通量卫星需求带动下，相控阵天线具备规模经济优势。
- **导航应用方面**：由于北斗三号北斗卫星组网已经完成，北斗三号终端有望加速渗透尤其是军工市场，相关具备芯片能力的上市公司有望受益，相关受益标的包括海格通信、华力创通、北斗星通等。
- **卫星通信应用方面**：地面接收硬件终端（卫星接收天线、手持终端等）以及应用服务。尤其是服务高通量卫星之前卫通信领域多用于应急、航海、航空的等专业高价值领域，相关上市公司包括海格通信、华力创通、盛洋科技等。



图表：中国卫星通信规模及增长预测

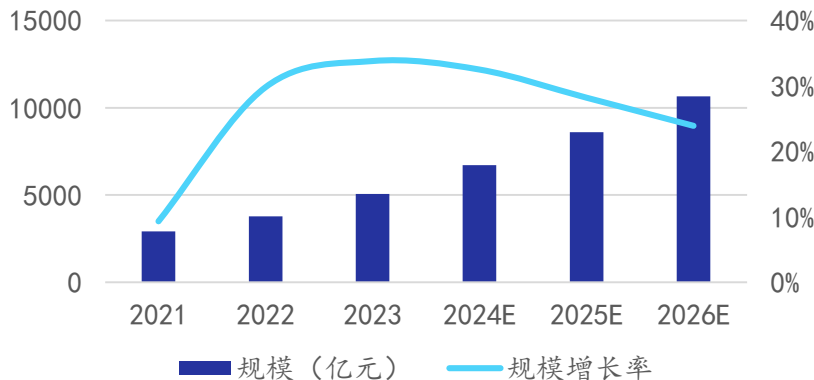


低空经济，新质生产力，政策加速落地

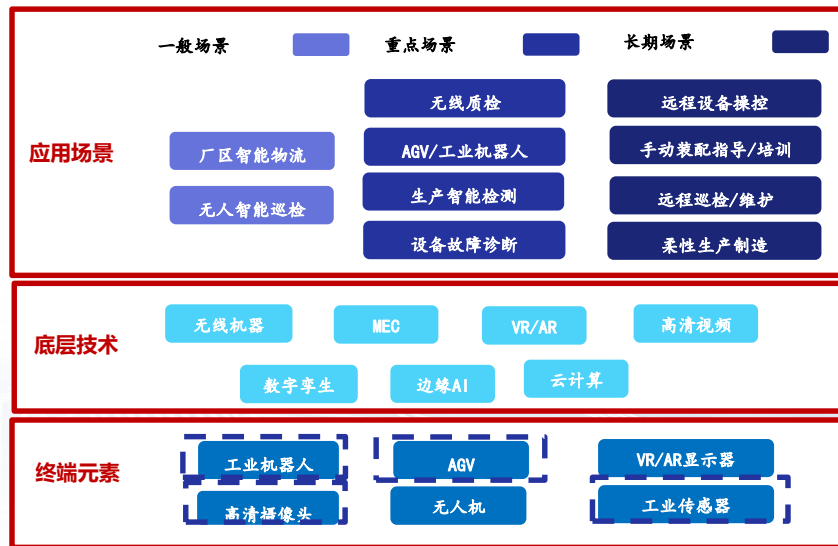
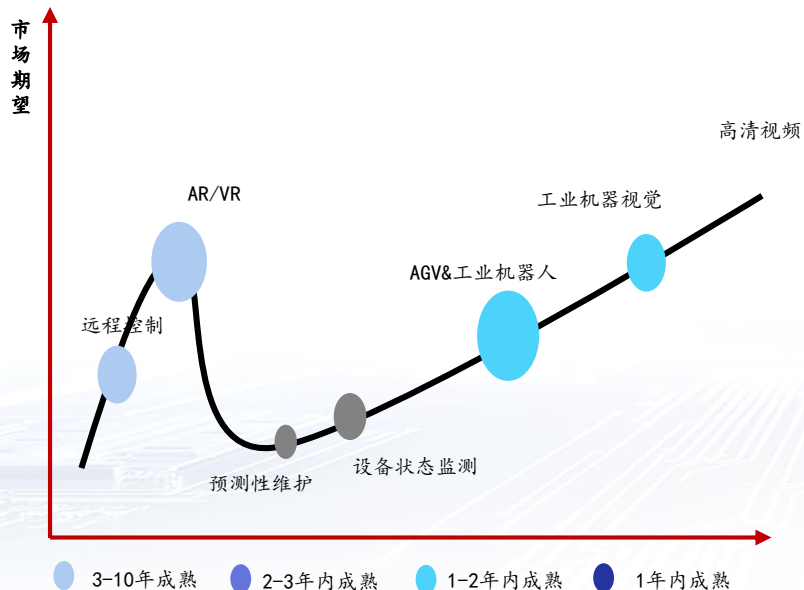
- 目前感知网络的产业链及建设进度还未明确，且覆盖成本还需要通信覆盖来率先补贴，相关产业链具备较好的发展机遇。
- 1) 5G-A通感一体化带来通信天线及毫米波产业链应用延伸；2) 低空经济的发展带来无线通信技术空间上的延伸带动短波超短波、无线中继组网等产业链；3) 卫星通信作为低空飞行器重要的补充通信方式，随着国内低轨卫星的发射组网逐渐成熟，低空经济有望成为低轨卫星通信的重要落地应用场景；4) 异构多传感器提供多维度环境信息，保障飞行安全：低空飞行器的感知域主要由摄像头、毫米波雷达、激光雷达、全球导航卫星系统（GNSS）+惯性测量单元（IMU）等模块组成，相关产业链与智能驾驶产业链高度重合。



图表：中国低空经济规模及增长预测



- 针对不同场景的需求共性进行分析归纳，输出了5G重点场景两大类解决方案，同时对各类场景规模化时间、行业整体规模进行预测。
- 重点解决方案：智能监控、工业机器视觉、云化AGV、数据采集（工业传感）、工业AR。
- 长期目标：无线云化PLC控制及云化机器人控制两个场景仍需要实验室去开展相关研究及验证工作



产业链供应链安全：

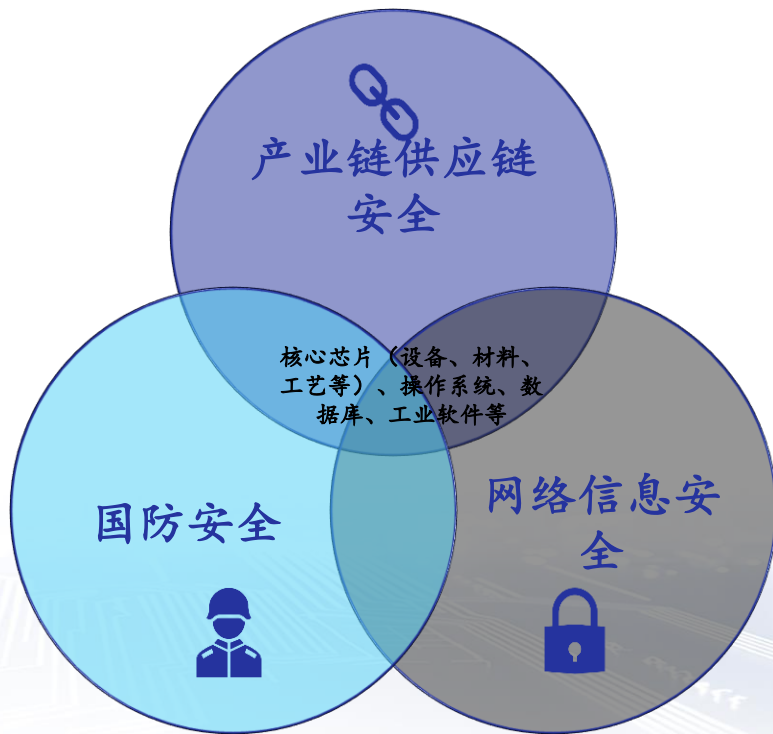
关键核心技术和关键零部件的自主研发，
技术自立自强。

国防军事安全：

打造强大战略威慑力量体系，增加新域新
质作战力量比重，军工行业将向着更加全
面、更多维度方向发展。

网络信息安全：

网络安全作为网络强国、数字中国的底
座，将在未来的发展中承担托底的重担，
是现代化产业体系中不可或缺的部分。



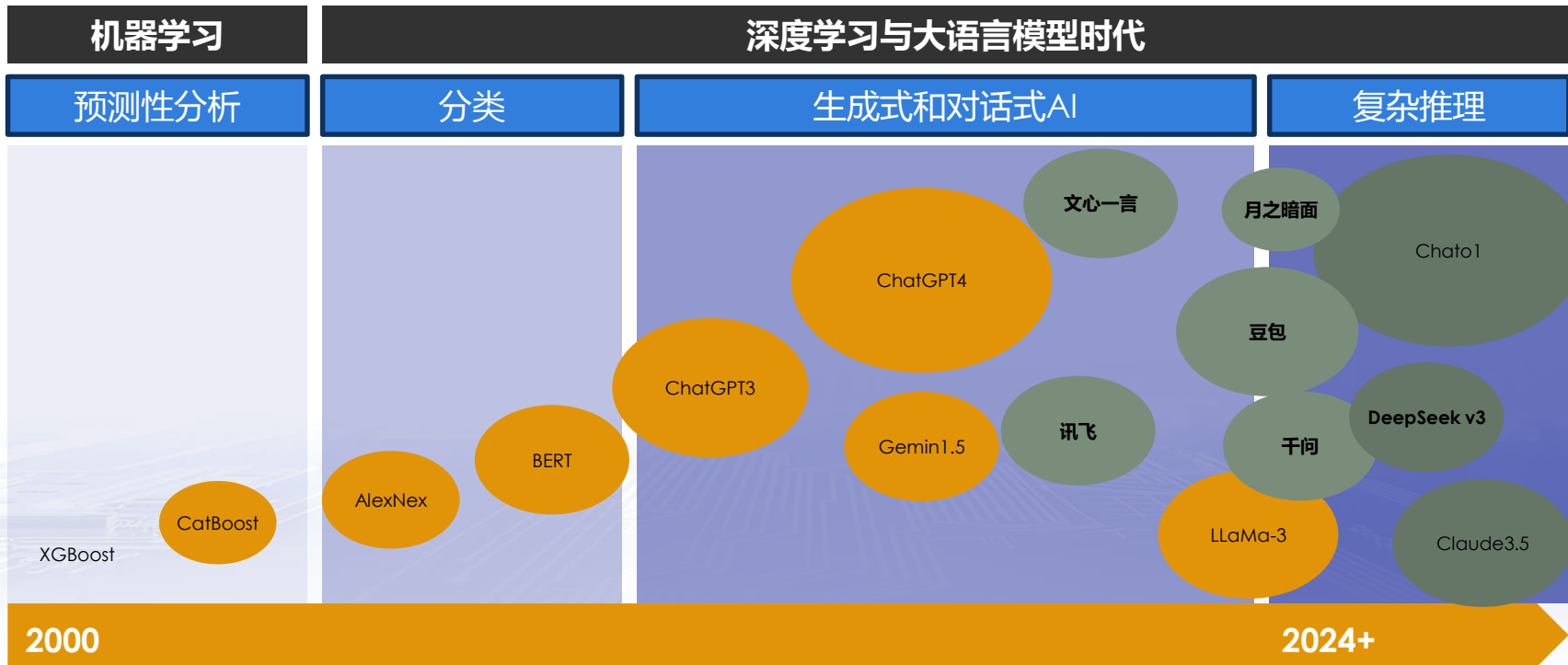


03 AI平权，泛在化与端侧化演进

边缘AI+端侧AI

GPT3引发大模型涌现，百模大战

由OpenAI发布的GPT3开始，大语言模型的理解和生成能力、通用和泛化能力大幅提升，引爆AGI大模型喷薄发展，开源模型与闭源模型并驾齐驱。

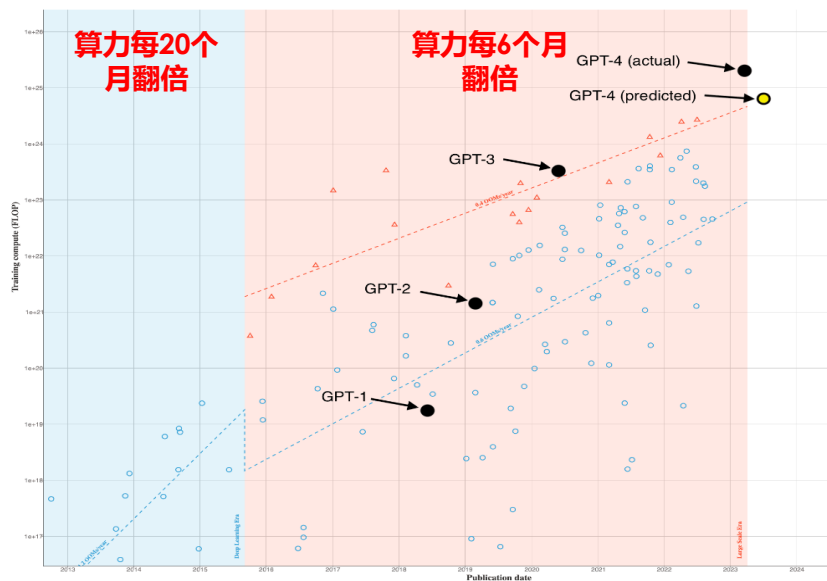


AI大模型出现后，算力及网络需求倍数提升

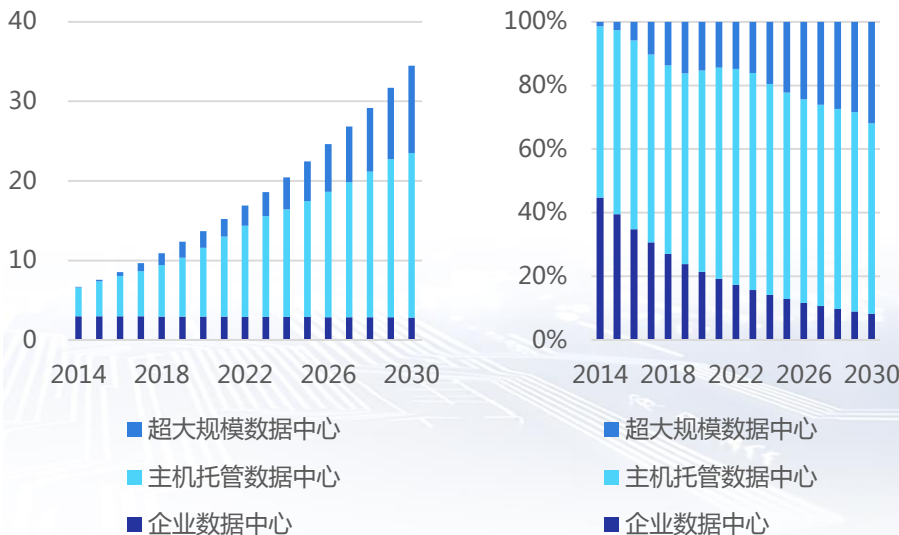
大模型和AIGC高速演进，高性能计算行业面临迫切的更高效率的算力需求

大模型训练的算力当量正在进一步增大，从 GPT-3 到 GPT-4 算力当量增长了 68 倍。算力当量越大，交叉熵越小，模型能力越强。数据中心为训练AI模型产生的能耗将为常规云工作的三倍，美国数据中心电力需求以10%复合增速增长。

图表： AI需求带来算力增速爆发



图表：数据中心算力需求及占比（千兆瓦）



产业趋势：需求驱动IDC向AIDC演进

在生成式AI及大模型的需求驱动下，客户算力需求由传统CPU通用算力向GPU、TPU等智能算力演进。根据客户和智算需求可大致分为三类：

纯智算

主要面向大模型的通用训练，比如华为盘古大模型等，机柜功率普遍在35KW以上，平均功耗较高通常采用全液冷形式；

训推一体

主要面向公有云和互联网，比如运营商公有云、字节跳动等，机柜功率在8KW-15KW或35KW以上，考虑成本通常采用以液冷为主的风液混合形式；

普智一体

主要面向私有云和极低时延的端侧推理，主要应用于央国企、政府部门和中小型产业数字化客户，包括烽火等，机柜功率在2KW-8KW或35KW以上，通常以风冷为主，少许配置液冷形式。

对于数据中心运营方来说，多种不同客户对应不同的智能算力需求，以及同一客户训推重点也对算力结构有所要求，因此，其运营的数据中心灵活部署要求也在不断提升，具备标准化建造及运营能力的中大型数据中心厂商具备更好的服务基础。

- 对于互联网大厂，一般都集中以训推一体需求为主，大部分自建数据中心，注重大网络的集群化部署，对于机电配套设备要求高，定制化以及低容错的高要求运维能力至关重要，该类大数据中心集群通常以高功率机柜为主，且基本具备一定规模，网络配置优异，通常选用核心城市周边地区以获得更高的能耗容量，伴随智算占比不断提升，通常采用冷板式液冷模式为主，预计在未来需求中仍将伴随客户Capex不断提升。主要关注阿里、腾讯、字节等相关主流供应商；
- 对于政府、金融客户，通常以私有部署或中低功率机柜需求为主，对于安全等级、灾备能力要求更高，该部分客户也主要采取定制化、自有自建部署为主；
- 对于企业客户，目前其功率需求仍相对较低，或通常直接租用公有云服务，其客户流动更加频繁，但对于价格敏感度更高。

产业趋势：龙头公司开支持续高涨

AI应用催化，互联网大厂资本开支加速增长，重心向算力方向转移。其中，国内互联网大厂包括腾讯、阿里等季度资本开支均出现同比大幅增长，且结构重心向AI领域倾斜。

- **字节**：12月18日，字节跳动正式发布豆包视觉理解模型，并接入豆包App和PC端产品，豆包大模型全面升级。伴随字节在抖音等App流量加速上涨过程中，字节对数据中心、AI硬件等资本开支不断加大，豆包App的全面升级有望推动公司持续加大AI相关领域投资。
- **腾讯**：2023年全年经营性资本开支172亿元，同比增长69%。2024年Q1-Q3经营性资本支出分别为66/72/147亿元，同比分别增长557%/144%/122%，主要受GPU服务投资驱动。
- **阿里**：2024Q1-Q3购置物业及设备资本开支约102/119/170亿元，同比分别增长305%/99%/313%，主要系围绕AI基础设施投入，在2024财年报告中，公司强调将持续对AI基础模型和AI创新进行投入。

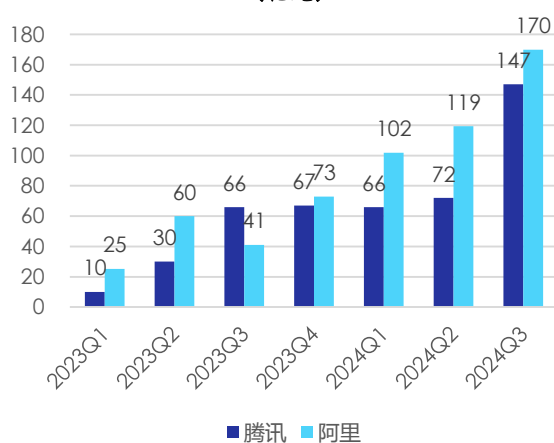
图表：美国科技龙头公司资本开支
(未特殊标注单位均为亿美元)

公司	2024年 资本性支出	2023年 资本性支出	同比增长 (%)
谷歌 (ALPHABET) -A	525.35	322.51	62.9%
微软 (MICROSOFT)	444.77	281.07	58.2%
Meta Platforms	372.56	272.66	36.6%

图表：三大运营商资本开支
(未特殊标注单位均为亿元)

单位：(亿元)	中国 移动	中国 电信	中国 联通
2023年资本开支	1803	988.4	739
2024年预计资本开支	<1730	960	650
2024H1资本开支	640	472	239
2024H1智算规模 (EFLOPS)	19.6	21	10
2024H1新增智算规模 (EFLOPS)	9.5	10	
2024H1累计5G中频基站	229		131
2024H1低频基站	70.5		79

图表：腾讯及阿里季度资本开支情况
(亿元)



产业趋势：Scaling Law逐渐递减

Scaling Law规律带来百模大战：伴随OpenAI2020年提出的Scaling Law，众多互联网企业加入模型大战，不断提升资本支出，加大用于训练的计算资源以及数据，从而提升自身模型能力。

- Scaling Law指出，对于模型，模型规模（如参数数量N）、训练数据集大小（D，token数）和计算量（C）之间存在一定可预测的关系，三者满足 $C \approx 6ND$ ；而模型的最终性能主要与计算量C、模型参数量N和数据大小D三者相关，而与模型的具体结构（层数/深度/宽度）基本无关。
- 随着模型规模的增大，每增加相同数量的参数或计算资源，获得的性能提升逐渐减少。这对模型设计及其部署策略有重要指导意义，提示研究人员和工程师在扩大模型规模时，需要综合考虑成本和收益。

伴随数据资源耗尽，Scaling Law边际效益逐渐递减：

- （1）人类创作的文本资料（D）几乎耗尽；
- （2）在计算量（C）没有大幅提升和数据量（D）耗尽下，仅提升参数（N）无助于模型效能；
- （3）计算量（C）短期无法显著提升，主要是Blackwell系统尚未量产和电力供应限制等因素。

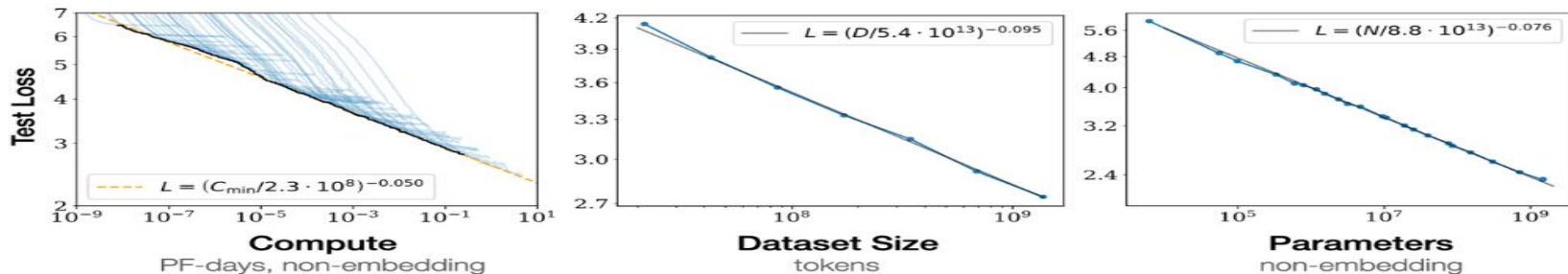


Figure 1 Language modeling performance improves smoothly as we increase the model size, dataset size, and amount of compute used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

产业趋势：开源与闭源并驾齐驱

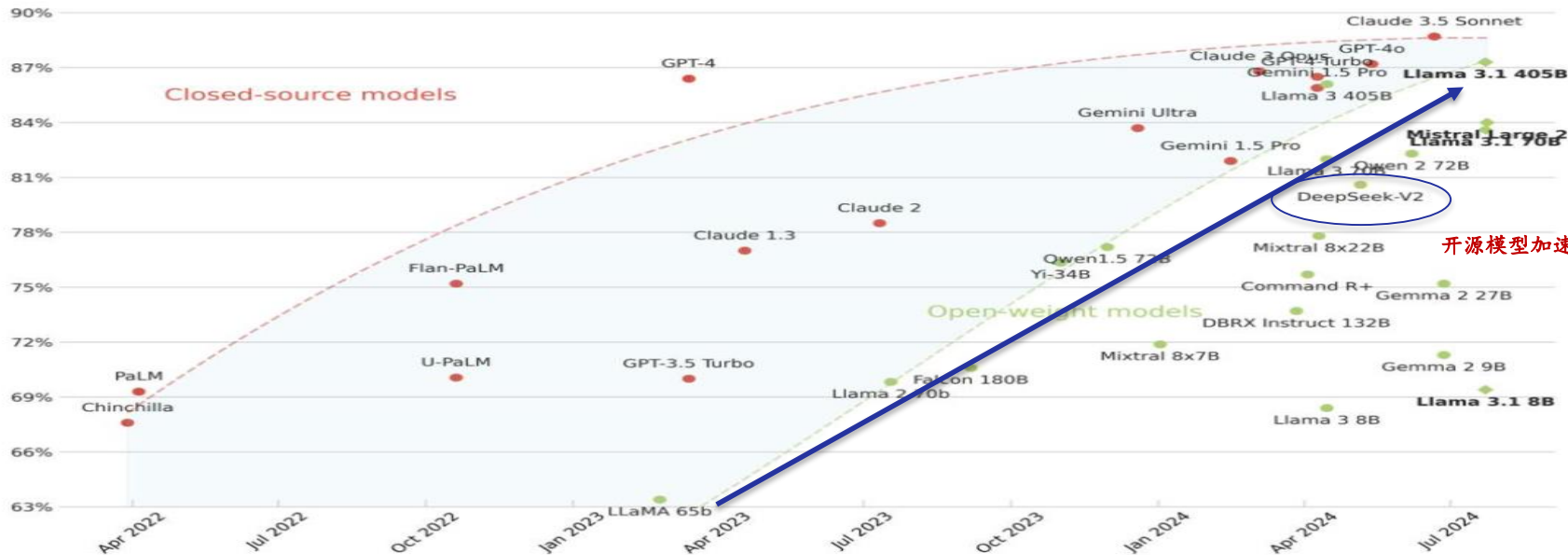
- 基础大模型普遍进入性能提升瓶颈期，开源模型吸引大量开发展，开源模型与闭源模型性能差距逐渐缩小。
- 在Scaling Law边际效益递减规律下，开源模型通过提升优化训练方式来提升模型效能。

Closed-source vs. open-weight models

Llama 3.1 405B closes the gap with closed-source models for the first time in history.

@maximelabonne

MMLU (5-shot)

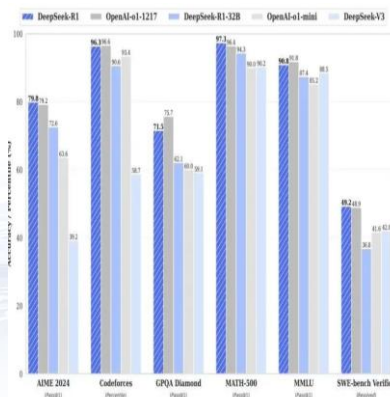


开源模型加速追赶

国产开源模型引发巨震，有望加速完善产业生态

- 在OpenAI发布O推理模型后，国内大模型厂商紧随其后，陆续推出自己的推理模型。这些模型在数学、代码、推理谜题等多种复杂推理任务上取得显著进步。**主要模型机构包括DeepSeek、月之暗面、阿里的QwQ和Marco-o1等。**
- **DeepSeek开源模型引动巨震，有望加速完善产业生态**：2024年12月，DeepSeekV3模型发布，2025年1月，DeepSeek正式发布R1系列模型并开源，根据AI产品榜统计，DeepSeek仅上线18天日活1500万，成为全球增速最快的AI应用。其模型能力比肩OpenAI，部署成本显著下降；同时，包括华为、腾讯、阿里、百度以及微软、AWS等陆续宣布支持DeepSeek模型，芯片领域包括英伟达、AMD、英特尔以及国产GPU天数智芯、摩尔线程沐曦等陆续宣布支持DeepSeek模型。
- API/token价格的显著下降，开源生态加速国产GPU适配，助力下游软件应用和端侧AI加速落地，产业链商业闭环有望实现。

图：DeepSeek模型与OpenAI模型主要性能对比



Benchmark (Metric)	DeepSeek V3	Qwen2.5 72B-Inst.	Llama 3.1 405B-Inst.	Claude 3.5 Sonnet-1022	GPT-4o 0513
Architecture	MoE	Dense	Dense	-	-
# Activated Params	37B	72B	405B	-	-
# Total Params	671B	72B	405B	-	-
MMLU (EM)	88.5	85.6	88.6	88.3	87.2
MMLU-Pro (EM)	89.1	85.6	86.2	88.9	88
MMLU-Pro (EM)	75.9	71.6	73.3	78	72.6
DROP (3-shot F1)	91.6	76.7	88.7	88.3	83.7
English IF-Eval (Prompt Select)	86.1	84.1	86	86.5	84.3
GPQA-Diamond (Pass@1)	59.1	49	51.1	65	49.9
SimpleQA (Correct)	24.9	9.1	17.1	28.4	38.2
FRAMES (Acc)	73.3	69.8	70	72.5	80.5
LongBench v2 (Acc)	48.7	39.4	36.1	41	48.1
HumanEval-Mul (Pass@1)	82.6	77.5	77.2	81.7	80.5
LiveCodeBench(Pass@1-corr)	40.5	31.1	28.4	36.3	33.4
LiveCodeBench (Pass@1)	37.6	28.7	30.1	32.6	34.2
Code Codeforces (Proseval)	51.6	24.8	25.3	20.3	23.6
SWE Verified (Resolved)	42	23.8	24.5	50.8	38.8
AIdo-Eval (Acc)	79.7	65.4	63.9	84.2	72.9
AIdo-Polygot (Acc)	49.6	7.6	5.8	45.3	16
AIME 2024 (Pass@1)	39.2	23.3	23.3	16	9.3
Math MATH-500 (EM)	90.2	80	73.8	78.3	74.6
CNMO 2024 (Pass@1)	43.2	15.9	6.8	13.1	10.8
CLUEWSC (EM)	90.9	91.4	84.7	85.4	87.9
Chinese C-Eval (EM)	86.5	86.1	61.5	76.7	76
C-SimpleQA (Correct)	64.1	48.4	50.4	51.3	59.3

图：DeepSeek模型与其他模型价格对比

模型	输入费用 (每百万 tokens)	输出费用 (每百万 tokens)	最大上下文
DeepSeek v3	0.14 美元 (缓存未命中) / 0.014 美元 (缓存命中)	0.28 美元	64k
OpenAI gpt-4o	2.5 美元 (缓存未命中) / 1.25 美元 (缓存命中)	10 美元	128k
Claude 3.5 sonnet	3 美元	15 美元	200k
OpenAI gpt-o1	15 美元 (缓存未命中) / 7.5 美元 (缓存命中)	60 美元	200k
Claude 3 Opus	15 美元	75 美元	200K

资料来源：AI在线，观察者网，华西证券研究所整理

DeepSeek一体机推动私有云、政企云部署加速

目前DeepSeek大模型一体机分为推理一体机和训推一体机。DeepSeek推理一体机内置DeepSeek-R1 32B、70B、满血版671B等不同尺寸模型，价格在几十万元到数百万元不等，主要面向对数据安全、数据隐私较为敏感的企业用户。而训推一体机的售价更高，用于DeepSeek-R1 32B模型的预训练和微调的一体机价格就达到数百万元。

根据财联社信息，硅基流动联合北京昇腾人工智能计算中心推出支持基于昇腾算力的DeepSeek-R1&V3系列模型、Qwen、FLUX.1等上百种主流模型的私有化集群部署，覆盖多场景应用需求。根据《电子工程专辑》，昇腾DeepSeek一体机支持2倍于业界的多用户并发数，最多可达192个用户同时在线，单用户时延仅50ms，满足企业多部门高效协同工作需求，同时支持DeepSeek V3/R1以及参数量1.5B-70B蒸馏模型全系列灵活部署，为企业提供极大选择空间。

中长期看好中大型客户定制化数据中心或公有云服务商：

- 包括光环新网、奥飞数据、数据港等，相关受益标的包括润泽科技、万国数据等；对于液冷等确定性架构迭代，持续推荐英维克，相关受益标的包括高澜股份等。

部署难度降低，推动私有云、政企云业务加速：

- 我们认为，目前C端AI功能大部分仍主要取代搜索引擎，相比之下，企业端的垂类应用落地方案相比C端AI应用可能更快，比如AI办公、AI医疗等应用方案，相关受益标的包括相关国资云、运营商以及数据要素相关厂商等，包含中国移动、中国电信、中国联通、以及数据要素厂商。
- 其二，昇腾DeepSeek一体机对政企客户AI渗透进程加速推动，有望为政企、金融等对数据安全较高的客户提供可靠的解决方案。政企算力有望加速部署，相关受益标的包括烽火通信、神州数码、拓维信息、华工科技等。

边缘计算：AI应用新增云安全和算力分发需求

得益于云原生技术应用、物联网产业成熟以及AI大模型催化，边缘计算市场保持快速增长。根据MarketsandMarkets报告，全球边缘计算市场规模将从2024年的600亿美元增长至2029年的1106亿美元，年复合增长率预计达到13%。

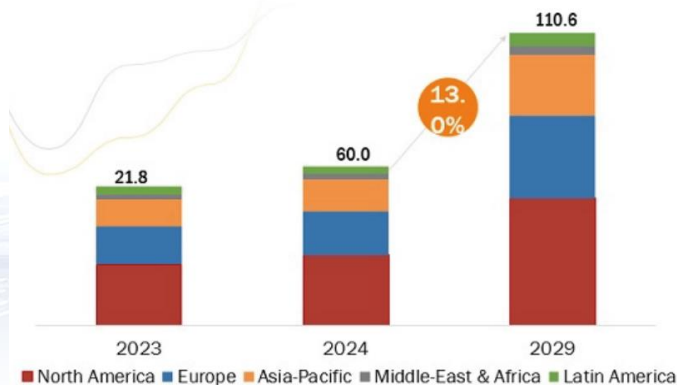
根据IDC最新发布《中国边缘云市场跟踪研究，2023H2》，2023年下半年，中国边缘云市场规模总计62.6亿元，同比增速36.1%。

云原生与AI技术驱动，行业加速智能化转型，带来边缘计算刚需：伴随企业数字化转型进程，云原生技术应用规模不断扩展，带来边缘云刚需。根据IDC《中国云原生基础设施技术能力评估报告（2023）》预测，到2025年，云原生工作负载将会占据服务器上工作负载的1/3。

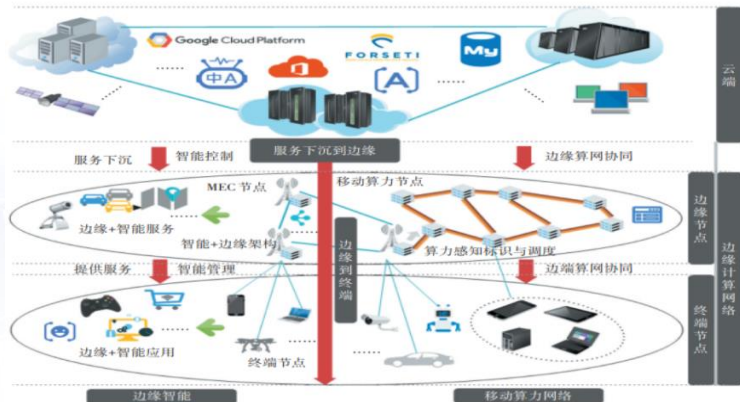
在AI技术发展进程中，AI算力与技术迭代加速，先进的AI模型也优先以云服务形式对外提供服务，这需要在基础设施层面全面彻底云原生化，以此为基础构建AI原生的应用架构，并面向AI研发进行MLOps/LMOps的研发流程升级。

边缘计算为云原生和AI大模型推理等提供诸多领域应用创新。比如在智慧城市中，边缘节点处理交通摄像头数据，利用AI生成交通流量预测和优化方案，有效减少拥堵。在娱乐媒体行业，边缘节点处理直播视频流，生成实时字幕、翻译和视频特效，增强观众的观看体验。在智能零售中，边缘设备处理顾客行为数据，通过AI生成个性化推荐，可以提升购物体验等。

图表：边缘计算市场全球规模预测（十亿美元）



图表：面向边缘智能的移动算力网络架构



边缘计算：AI应用新增云安全和算力分发需求

DeepSeek模型价格远低于其他大模型，同时其蒸馏模型（包括1.5B、7B、8B）也持续推动轻量化模型部署。

我们认为，DeepSeek蒸馏模型的出现，将进一步促使物联网产业链向智能化、高端化方向演进，芯片企业针对边缘端AI推理芯片也有望满足DeepSeek蒸馏模型在物联网设备上的高效运行能力，从而持续推动边缘计算需求。

比如摩尔线程就基于Ollama开源框架，完成了DeepSeek 7B蒸馏模型在MTT S80和MTT S4000的部署，并在多种中文任务中展现优异性能，验证了摩尔线程自研全功能GPU的通用性与CUDA兼容性。同时包括移远通信、美格智能等相机发布的AI算力模组产品也有望在端侧AI加速渗透过程中拉动需求。

边缘计算需求持续增长，主要得益于三方面，其一是企业数字化转型过程中云原生技术的广泛应用；其二是IoT产业不断成熟，推动边缘计算应用广泛渗透；其三是在AI大模型快速发展进程中，为边缘计算打开新的成长机遇。

当前时点，伴随DeepSeek轻量化低成本部署方案的推动，包括边缘计算以及端侧AI有望迎来新一轮发展机遇。相关受益标的包括加速网络以及边缘安全厂商网宿科技，边缘节点厂商优刻得、首都在线，混合云青青科技，边缘端侧AI模组厂商移远通信、美格智能等和边缘计算高算力SoC类芯片厂商。

图表：边缘AI相关标的

公司名称	股票代码	收盘价 (元)	总市值 (亿元)	EPS				PE			
				2023	2024E	2025E	2026E	2023	2024E	2025E	2026E
网宿科技	300017.SZ	13.2	323.7	0.3	0.3	0.3	0.4	53.0	53.0	42.7	35.8
首都在线*	300846.SZ	28.2	141.1	-0.7	0.2	0.4		-38.7	156.7	74.0	
移远通信*	603236.SH	81.1	212.2	0.3	2.0	2.8	3.6	238.5	40.2	29.0	22.8
美格智能*	002881.SZ	52.8	138.3	0.2	0.5	0.7	0.9	213.0	112.4	77.4	57.9

注：*为wind一致性预测

资料来源：wind，华西证券研究所

物联网：端侧AI持续渗透，AIOT模组增长动能强劲

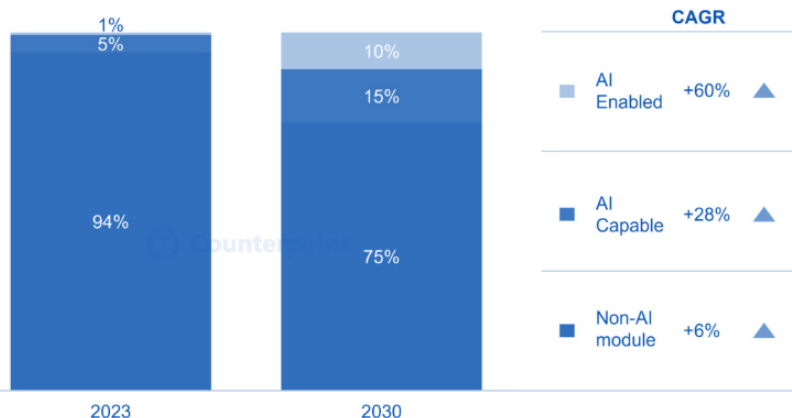
伴随AI发展的加速，尤其在端侧设备端的不断渗透，AI嵌入式蜂窝物联网模组增长潜能充足。而AI嵌入式蜂窝模组又可以细分为智能模组和AI使能型模组，后者按照 AI 能力的高低又可进一步细分为低级 AI 模组和高级 AI 模组。

— **智能模组**：配有 CPU 和 GPU 以及连接基带，主要侧重于连接和进行比较基础的数据处理。

— **AI 使能型模组**：这类模组集成了 CPU、GPU、NPU、TPU 等先进处理器，或者配有专用的 AI 引擎，以增强人工智能功能。其中低级 AI 模组提供中等 AI 能力，通常以低于 8 TOPS 的速度进行 AI 推理；高级 AI 模组支持 8 TOPS 以上的 AI 推理，可以处理更复杂的任务。

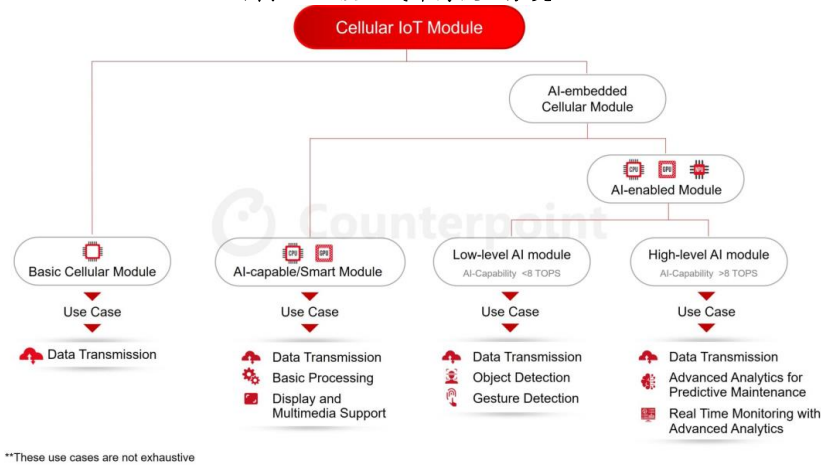
根据 Counterpoint 的 AI Module Tracker and Forecast 的数据，到 2030 年，AI 嵌入式蜂窝模组预计将占有物联网模组出货量的 25%，高于 2023 年的 6%，复合年增长率为 35%。

图表：不同AI功能模组2023年出货量及2030年预计出货量占比



资料来源：Counterpoint，华西证券研究所

图表：AI嵌入式蜂窝模组分类



物联网：端侧AI持续渗透，AIOT模组增长动能强劲

我们认为，当前AI发展正处于由模型训练转向推理的关键阶段，各种终端和应用在逐渐明确AI的使用方向，与此同时，国产算力需求也正在培育推动相关配套硬件及软件生态发展。

而物联网模组行业市场格局维持头部集中，整体销量仍在稳定抬升，ARPU也在伴随网络制式升级逐渐企稳。

当前，AI端侧的渗透逐渐推动下游包括机器人、AI眼镜&可穿戴、工控等应用方向有望爆发，也给其中的模组厂商带来结构性市场机会，相关受益标的包括移远通信、美格智能等。

图表：物联网厂商的AI相关模组产品

厂商	AI相关模组产品
移远通信	近期宣布推出针对玩具市场的AI智能解决方案。该方案利用移远通信和全链路的音频算法，并深度融合豆包等大模型的先进AI技术和丰富数据资源，实现智能玩具与云端AI能力的无缝对接，使得用户与玩具能够进行实时、流畅的对话，无论是趣味畅聊、百科问答还是信息查询，智能玩具都能迅速响应。
广和通	深度布局视觉视觉与听觉计算，推出高算力和轻量化机器视觉解决方案，在高端智能影像终端、工业视觉终端、车载终端、机器人均具有广泛应用场景。
美格智能	合作伙伴阿加犀联合高通在2025CES展会上面向全球发布人形机器人，搭载了2颗美格智能基于高通QCS8550计算平台开发的高算力AI模组SNM970，单颗算力48T，累计AI算力接近100T，为人形机器人的整体控制系统提供计算能力。根据公司公众号文章，其发布的终端侧AI产品矩阵为包括机器人、AR/AI眼镜、无人机智控、AI+医疗等多领域提供落地方案。



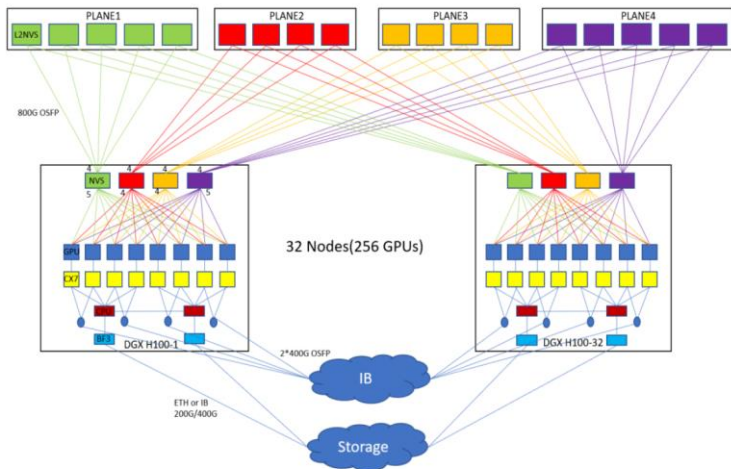
03 AI对算力网络的改造

交换机、CPO、液冷、电源

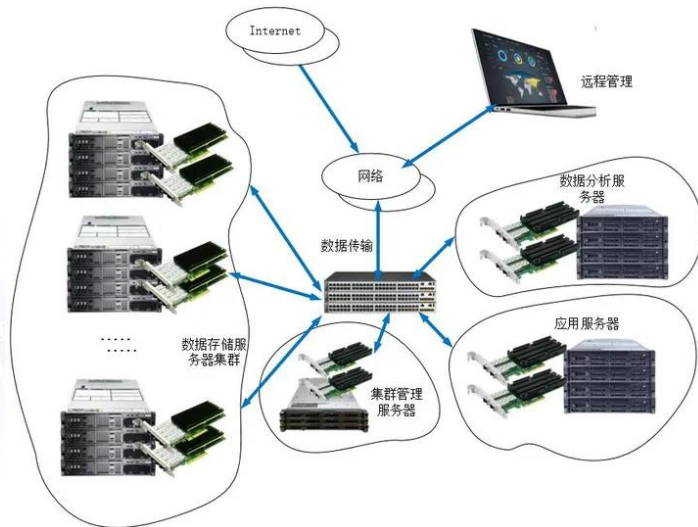
大规模的无阻塞网络是构建AI大模型的基础设施

- 大规模的无阻塞网络是构建AI大模型的基础设施：面对千亿、万亿参数规模的大模型训练，仅仅是单次计算迭代内梯度同步需要的通信量就达到了百 GB 量级。此外还有各种并行模式、加速框架引入的通信需求，使得传统低速网络的带宽远远无法支撑 GPU 集群的高效计算。因此要充分发挥 GPU 计算资源的强大算力，必须构建一个全新的高性能网络底座，用高速网络的大带宽来助推整个集群计算的高效率。
- 大规模 AI 训练集群架构中，GPU 之间的通信实际上由多种形式的网络来承载的：机间网络（网卡 + 交换机）与机内网络（NVLink/NVSwitch 网络、PCIe 总线网络）。

图表：NV SuperPOD架构



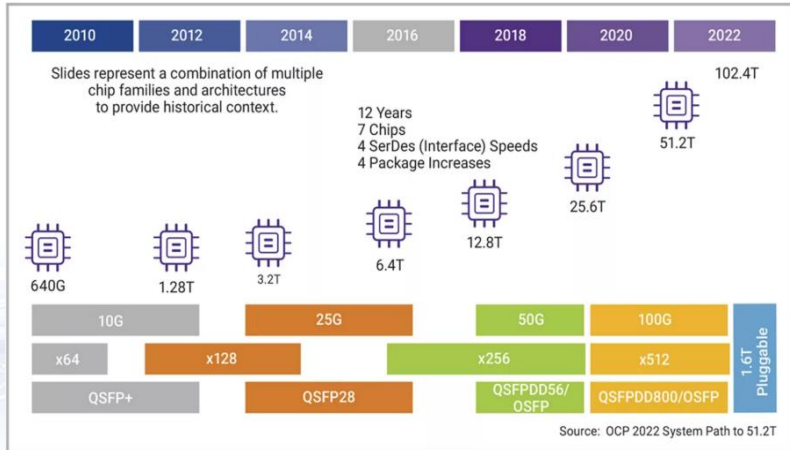
图表：网卡+交换机架构



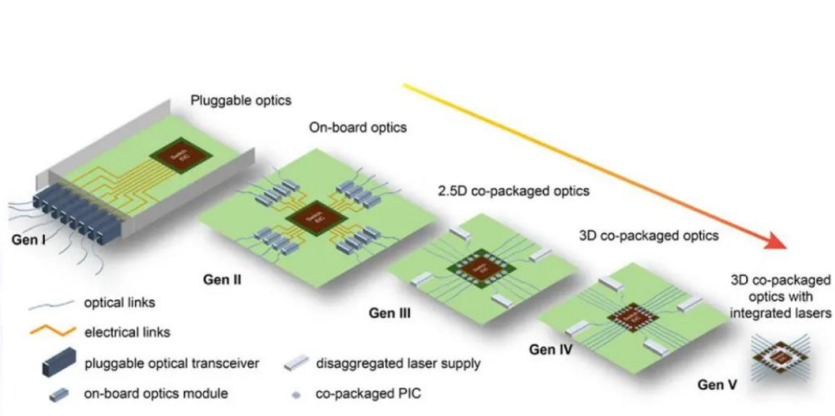
大模型驱动大容量交换芯片以前所未有的速度迭代

- **大模型训练使得算力网络往超大规模、超高带宽演进，催生高速率交换芯片快速演进及商用化：** AI 训练除了需要更高算力之外，高性能网络也是必不可少。对于大模型，单次的参数同步量一般都在百MB~GB 的量级，参数量巨大。
- 传统电交换芯片直接决定整机的交换容量、端口速率等核心性能 指标，AI 驱动高速率交换芯片逐渐成为主流。
- **大模型快速发展背景下，传统交换机芯片以前所未有的速度商用加速商用：** 2022年8月16日，博通发布了业内首款51.2T容量的交换芯片 Tomahawk 5。
- 2023年3月15日宣布其Tomahawk 5系列现已批量出货，从最初的样品到量产部署仅花七个月时间，对于博通交换芯片产品线来说是前所未有的，证明AI大模型正在快速推进相关技术迭代。

图表：电交换机芯片数据处理能力每两年翻一番



图表：以太网交换机向光电融合方案逐渐演进（3D CPO封装）

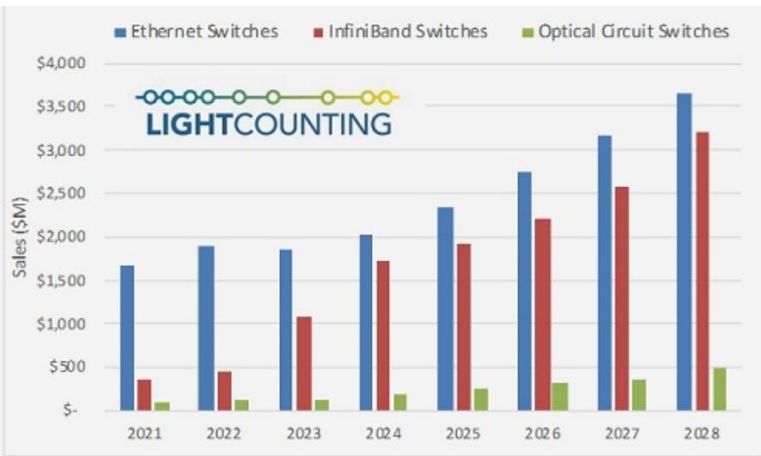
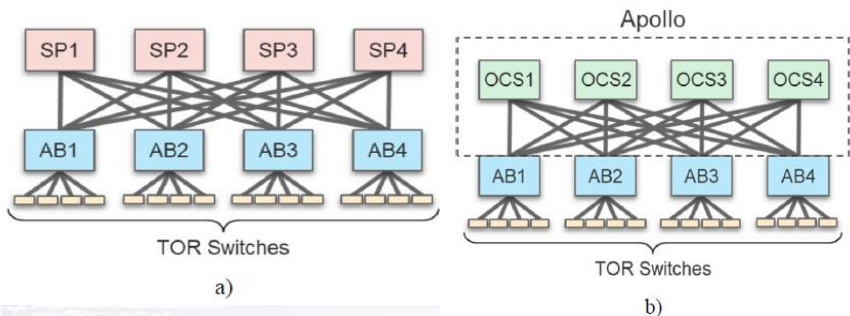


高速互联——全光交换方案OCS

- 谷歌大模型中使用的自研TPU和光电路交换机OCS进入大众视野。谷歌很早就开始在其数据中心中部署OCS。2022年谷歌发表论文“Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale”，文章详细介绍了OCS和基于自研OCS实现的Apollo网络架构。
- 与EPS相比, OCS带来了数据中心网络结构的革新, 具备能耗与成本优势, 同时不需要解码数据包, 具备低延迟特性。
- 我们认为, 两大方案仍在很长时间段内共存, 但是谷歌的Apollo结构对于脊交换机 (Spine层) 等的替换, 使得相关高速率光模块需求是不如NV IB形式的, 长期看伴随制造及微镜调配的工艺升级, 通过增加更多端口数的系统, 相关OCS架构可能会向更下一层渗透, 从而降低光模块需求。
- 根据LightCounting最新预测, AI硬件需求推动2023-2024年InfiniBand交换机销量, 但会在2025-2028年放缓, 2023-2028 CAGR 24%, 对于以太网交换机 CAGR 14%, 而谷歌的OCS方案, 预计其2023-2028 CAGR将达到32%。

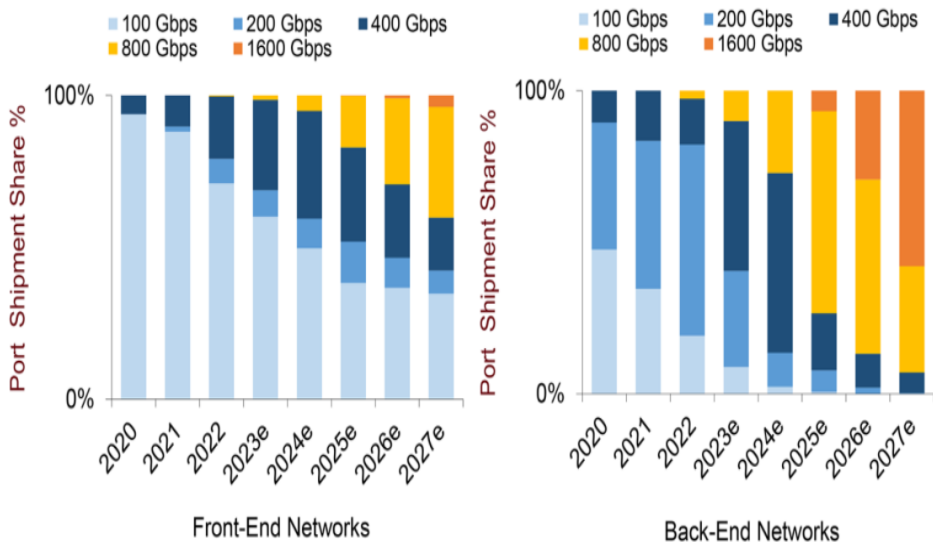
图表：传统Clos拓扑架构与谷歌的OCS全光交换机方案

图表：交换机2023-2028年市场销量预测（百万美元）

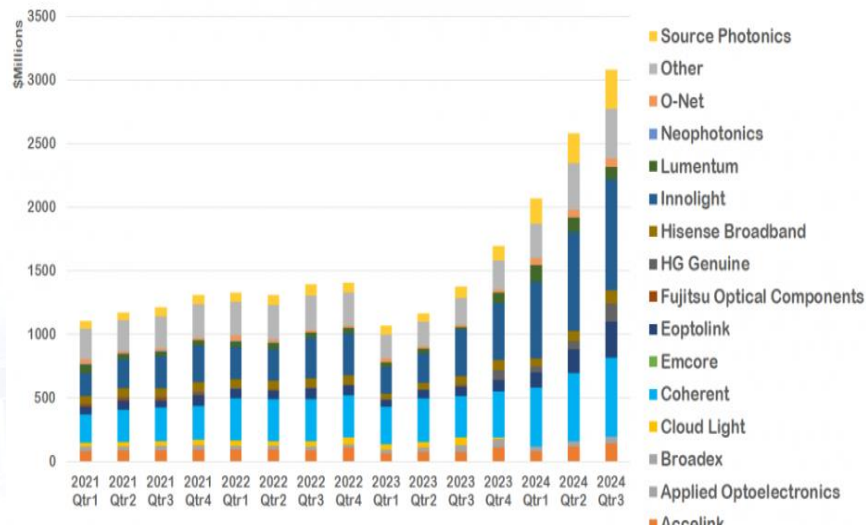


- 根据市场研究机构Cignal AI报告，AI部署为数通市场带来前所未有的发展机遇。预计2024年高速数通光模块的市场规模将超过90亿美元。400G和800G光模块的出货量在过去12个月中增长了近四倍，预计2024年将超过2000万只。
- 随着800G增长达到顶峰，云服务商正向单通道200G的1.6T方案过渡，高速数通光模块的市场规模预计将从2024年的约90亿美元扩大到2026年的近120亿美元。

图表：数据中心交换机端口速率（前端&后端）



图表：光模块速率演进图



资料来源：Dell'Oro Group, Cignal AI, C114通信网, 华西证券研究所

高速互联——NVLink与PCIe

PCIe与NVLink作为当前高性能计算领域中的两种关键互连技术，各自承载着不同的技术理念和应用定位。

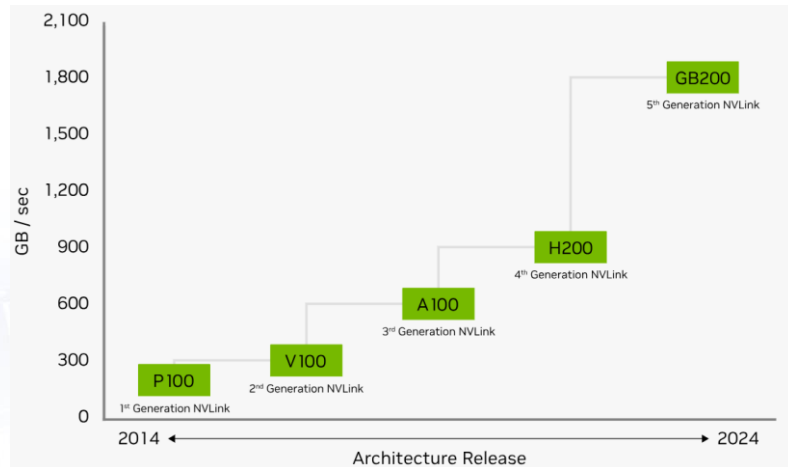
NVLink是NVIDIA为其GPU产品线专门设计的高速互联解决方案，形成了一个相对封闭的技术生态。这种封闭性使得NVLink能够针对NVIDIA的GPU进行深度优化，提供最大化的性能表现，但它的应用范围受限于NVIDIA的硬件平台，不具备PCIe的广泛兼容性。

PCIe以其开放性和通用性在多样化的计算环境中占据重要地位，随着技术的不断进步，PCIe 6.0和NVLink都将继续在各自的领域内推动高性能计算的发展。随着PCIe 6.0的推出和PCIe 7.0的即将到来，将继续朝着更高的传输速率、更低的功耗、更强的兼容性和互操作性方向发展。我们认为PCIe 7.0规范将重点关注800G以太网连接传输，并包括人工智能、机器学习、云计算和量子计算等内容，并加速在大型智算中心中的应用。

图表：NVLink技术最大化提升系统吞吐量

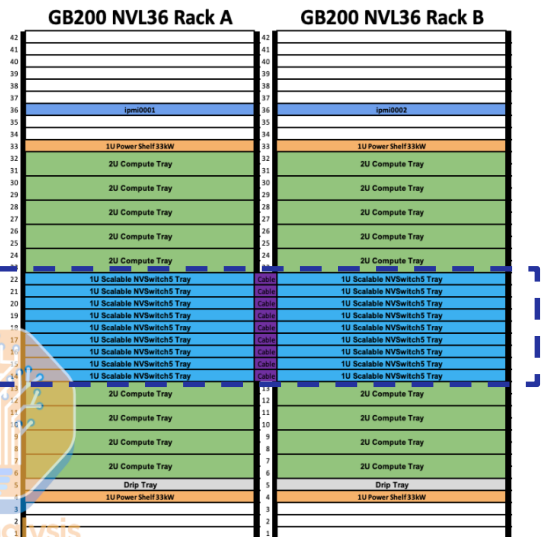
	PCIe	NVLink
技术架构生态定位	以其开放标准促进了广泛的技术融合和不同制造商和设备间的高效通信	专为NVIDIA GPU打造，形成了一个封闭但优化的性能高地
性能指标	PCIe 7.0单向速率可达128GT/s，在x16双向传输模式下，速率高达512GB/s	双向互连速度达1.8 TB/s，可扩展服务器内的多GPU 输入/输出(I/O)

图表：NVLink与PCIe连接方式对比



- 在2024GTC大会上，英伟达发布的GB200采用铜缆连接替代传统高速背板连接中的光纤，英伟达表示DAC铜缆是InfiniBand交换网络和NVIDIA GPU加速的端到端系统中创建高速、低延迟100G/EDR、200G/EDR和400G/EDR链路的成本最低方式，高速铜连接市场被进一步打开。
- 根据Semianalysis测算，一个NVL36*2系统，需要162条 1.6T双端口ACC实现互联。

图表：GB200 NVL72架构中背板由高速铜缆连接



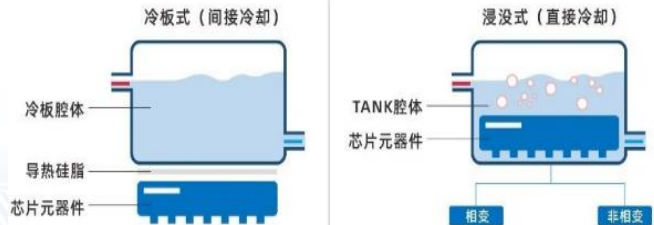
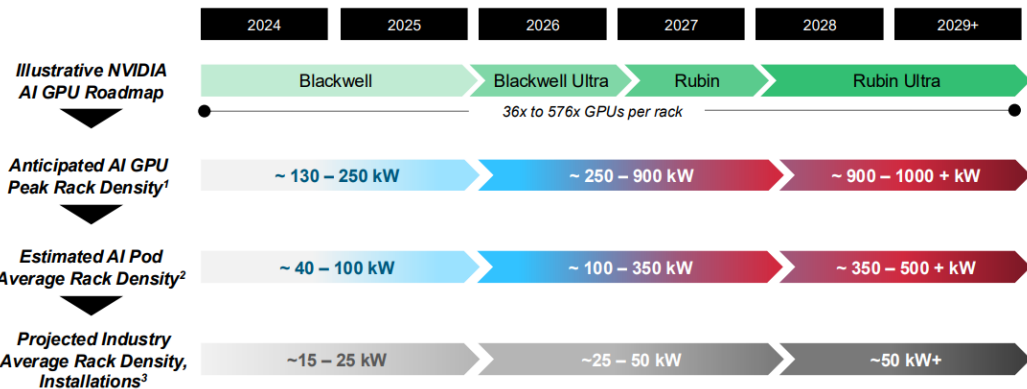
图表：NVL机柜及内部架构



- 伴随单机柜功率的不断提升，传统风冷系统也逐渐无法满足散热需求，或相应更加节能高效的散热技术方案逐渐铺开：从早期的封闭冷/热通道、机柜循环制冷、热风抽取冷却、到背板热交换、间接蒸发冷却，再到液冷方案中的喷淋液冷、冷板液冷以及浸没液冷。冷服务器生态初步形成，大多以间接冷板为主。浸没式液冷考虑其对服务器主板等技术革新需求以及下游厂商机房业务部署的差异性，规模化起量仍需要时间。
- 目前国内液冷服务器厂商主要包括华为、曙光、神威蓝光、浪潮等，海外液冷服务器厂商包括戴尔、惠普、思科、SGI、BULL、Cray、富士康、超微、Nortech（浪潮合作）、Iceotop等。此外，相关产业链中，海外产业化液冷方案制造商包括Coolit和Asetek（生产液冷配件厂）、Dynatron、K-computer；直接式液冷方案包括Green Revolution Cooling和3M（制冷液）等。国内相关企业包括英维克、维谛技术、网宿科技等。

图表：单机柜功率密度快速增加，散热技术不断更迭

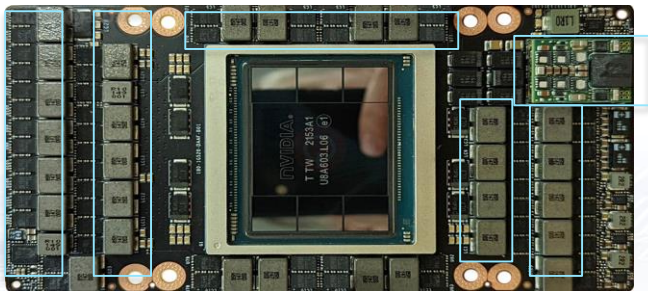
图表：冷板液冷与浸没液冷技术区别示意图



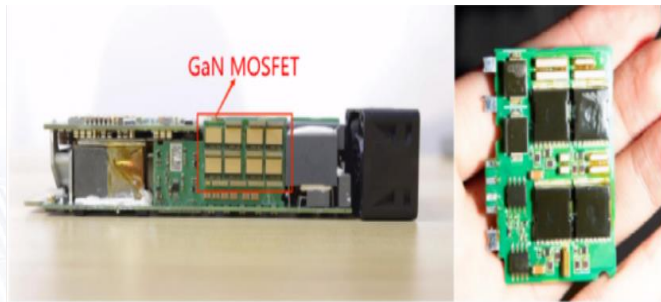
资料来源：Vertiv 2024年，开放数据中心委员会，华西证券研究所

- **智能算力推动，服务器电源需求及技术要求愈来愈高**：人工智能和云应用正在推动先进硬件的采用，能耗的不断增加。英特尔的“Sky Lake”和AMD的“Rome”等高级处理器功耗升至230-300W，英伟达的GPU功耗更将攀升到600W左右。
- **电源架构从12v提升到48v**：OCP试图通过定义电源架构的新标准来应对此类挑战，将中间总线电压从传统的12V提高至48V。这一举措可显著降低传输损耗，将电力更有效地传输到有效负载，比如AI ASIC/GPU/CPU或SOC上。
- **数据中心电源小尺寸、高功率等要求，对电源研发与生产提出更高的技术要求，服务器电源市场国产替代空间大**：服务器电源认证周期较长，从研发立项到获取认证通常需要2-3年时间，一般公司难以进入竞争，目前服务器电源市场份额主要由台达电子、光宝科技、艾默生、中国长城等企业长期占据。
- 近年来国内大陆企业包括欧陆通和高斯宝电气等规模逐渐扩大。根据高斯宝电气官网，服务器电源出货量全球前8。

图表：NVIDIA H100算力卡



图表：华为钛金级3000W氮化镓服务器电源拆解



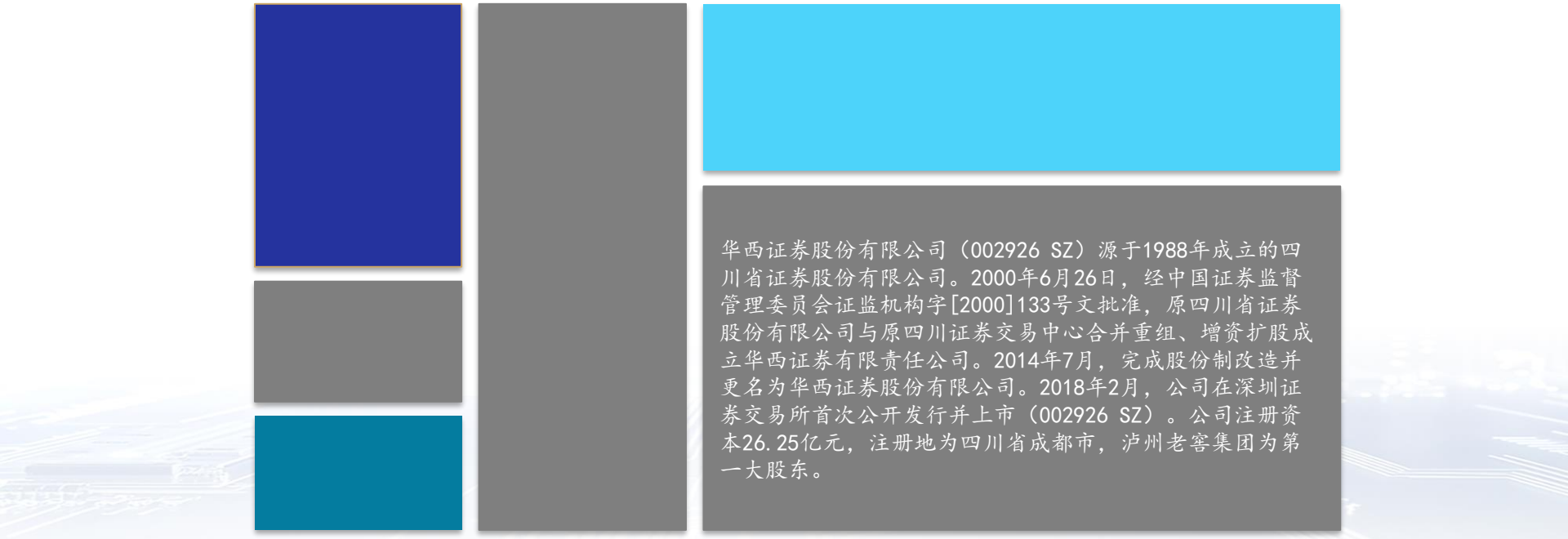
	代码	公司名称	最新收盘价 (元)	市值 (亿元)	PE		
					2024E	2025E	2026E
AI基础设施	601728.SH	中国电信	8.0	7,005.3	22.3	21.2	20.1
	600941.SH	中国移动	107.2	16,731.0	16.6	15.9	15.1
	600050.SH	中国联通	6.2	1,942.9	22.2	20.0	17.7
	000938.SZ	紫光股份	29.2	833.7	35.1	23.9	21.0
	000063.SZ	中兴通讯	37.1	1,687.5	17.6	16.0	14.4
	002837.SZ	英维克	47.9	356.6	53.9	41.0	32.6
	300593.SZ	新雷能	16.4	89.0	15.3	11.1	
	300383.SZ	光环新网	21.0	377.1	50.0	38.1	
	300738.SZ	奥飞数据	26.1	254.8	74.5	55.5	
	300442.SZ	润泽科技	67.6	1,163.7	52.8	36.2	28.6
	603881.SH	数据港	40.2	240.8	171.4	134.7	113.3
	300394.SZ	天孚通信	89.6	496.1	51.8	39.5	
	300308.SZ	中际旭创	101.8	1,124.4	44.4	36.4	
	300502.SZ	新易盛	94.3	668.4	41.5	34.9	
	688205.SH	德科立	71.5	86.4	70.4	49.2	35.6
	688498.SH	源杰科技	128.2	109.5	269.9	82.1	53.1
	688048.SH	长光华芯	63.0	111.0	-189.3	493.6	182.2
	300570.SZ	太辰光	92.6	210.4	82.6	52.5	38.2
002335.SZ	科华数据	54.5	251.4	51.2	35.7	29.0	
边缘AI相关	300017.SZ	网宿科技	13.2	323.7	53.0	42.7	35.8
	603236.SH	移远通信	81.1	212.2	40.2	29.0	22.8
	002881.SZ	美格智能	52.8	138.3	112.4	77.4	57.9

市场竞争加剧：近两年来，AI产业竞争者逐渐增多，若竞争加剧可能会造成产品同质化严重，进而对企业创新造成不良影响，影响产业发展；

产品研发不及预期：AI属于技术密集型产业，若产品研发不及预期，会影响产业进程；

Capex投入不及预期：AI硬件设备主要受下游运营商及各大互联网厂商投入周期影响，相应Capex结构性差异可能会影响上游设备需求；

下游应用推广不及预期：AI端侧设备及AI模组设备受宏观经济周期和产业备货周期双重影响，下游应用推广存在不确定性预期。



华西证券股份有限公司（002926 SZ）源于1988年成立的四川省证券股份有限公司。2000年6月26日，经中国证券监督管理委员会证监机构字[2000]133号文批准，原四川省证券股份有限公司与原四川证券交易中心合并重组、增资扩股成立华西证券有限责任公司。2014年7月，完成股份制改造并更名为华西证券股份有限公司。2018年2月，公司在深圳证券交易所首次公开发行并上市（002926 SZ）。公司注册资本26.25亿元，注册地为四川省成都市，泸州老窖集团为第一大股东。

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：<http://www.hx168.com.cn/hxzq/hxindex.html>

免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但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。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，如需引用、刊发或转载本报告，需注明出处为华西证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。