

# 浪潮之巅: AI+产业链从技术底座到应用全解析

杨林	资格证书编号: S0850517080008;	余伟民	资格证书编号: S0850517090006
张晓飞	资格证书编号: S0850523030002;	毛云聪	资格证书编号: S0850518080001
陈星光	资格证书编号: S0850519070002;	崔冰睿	资格证书编号: S0850523100002

联系人: 杨昊翊

2024年3月28日

**1.计算机：**我们认为，2023年是国产大模型快速发展的元年，在过去一年多的发展时间中，部分领先的国产大模型目前已经接近世界第一梯队，国产大模型目前仍在持续发展的过程中，对于算力、数据的需求仍将不断扩大，而伴随大模型的升级迭代，应用的落地也有望全面铺开，我们判断，2024年有望成为国产大模型全面商业落地的元年，随着国产模型性能的全面升级，先进的大模型能力也将逐渐转化为落地实际场景的产品应用，从而带动大模型在办公、教育、医疗、工业制造等众多垂直行业的持续落地，AI2B与AI2C的需求天花板也有望全面打开。AI的投资也从映射和跟随，变为引领。

**2.通信：**AI的大模型训练及应用提升通信能力需求。1) GPU：英伟达带动下，新的AI芯片计算能力不断提升，芯片间互联速度、内存容量也快速增长。2) 交换机：服务器与交换机、交换机与交换机之间接口速率从100G、400G快速向800G、1.6T等演进。3) 光器件：光模块作为光电转换的核心器件，速率升级同时，产品上也衍生出LPO、CPO、硅光、薄膜铌酸锂等方向；并带来光电芯片新需求。

**3.电子：**1) 存储为半导体核心受益板块。未来在AI模型逐渐复杂化的趋势下，将刺激更多的存储器用量，并同步带动服务器DRAM、SSD以及HBM的需求成长。2) AI端侧数字芯片是不可或缺一环。到2024年，边缘AI芯片的出货量将增至15亿颗，年增长至少20%，远高于全球半导体整体增长率(大约9%)。3) AI PC、AIoT能见度增强。生成式AI模型的爆发式增长，带来将专用的AI加速硬件集成到PC的需求。随AI PC持续演绎，头部厂商积极布局；2024年出货的PC中，AI PC占比将会接近20%。此外，多模态加持，智能眼镜需求端迎来催化，以Micro LED+光波导的光显方案为代表的AR眼镜，有望成为AI大模型的极佳载体。

**4.传媒：**随着国内大模型陆续上线，AI应用的商业化变现将逐渐拉开帷幕。我们认为，2023年是大模型向多模态化进化的一年，与此同时AI提供了更多商业化变现的途径，2024年生成式AI技术有望在IP开发、互动陪伴、游戏、营销、电商、教育等方向获得广泛应用。且随着AI产品的逐步落地，2024年将主要是去伪存真的逻辑验证阶段，检验AI技术的应用是否能够很高效的产出，多模态进化之后是否在更复杂的视频、游戏等领域有生产力的提升和用户需求的解决，一旦实现AI赋能，渗透率从0~1将会带来收入和利润的巨大增量和弹性。

**风险提示：**国内AI大模型进展进度不及预期，相关公司相关内容产品开发进度不及预期，政策监管风险等。

## 1.算力产业链分析:

1.1算力规模持续增长,智能算力增长迅速

1.2 AI服务器市场前景广阔,生成式AI服务器占比将不断提升

## 2.人工智能产业链产业链分析

2.1 大模型:百花齐放

2.2 应用场景:千行百业AI落地不断加深

## 3.投资逻辑与标的推荐

## 4.风险提示

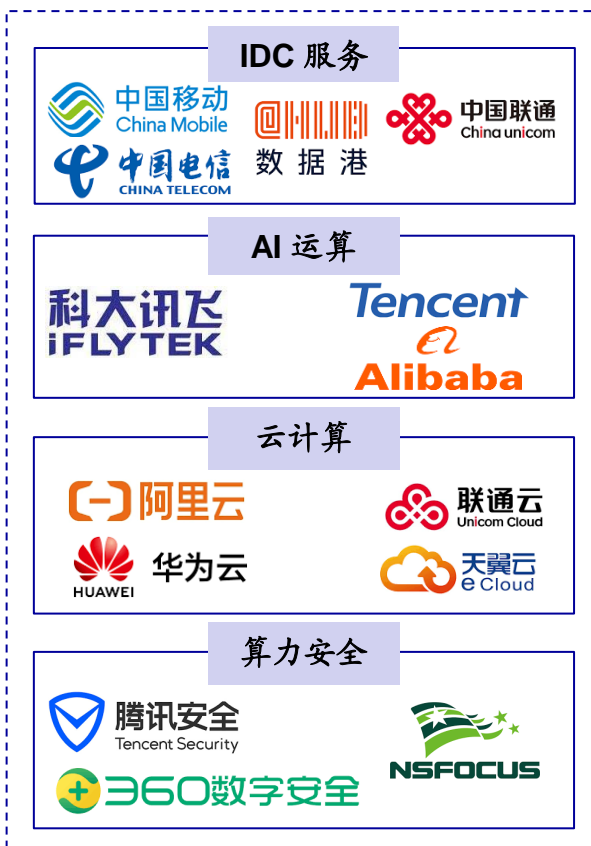


➤ AI算力产业链上游为算力基础硬件设施，包括芯片、服务器、光模块、交换机；中游为算力网络平台，包括数据中心、云运算等，下游则是各类用户以及细分的应用场景。

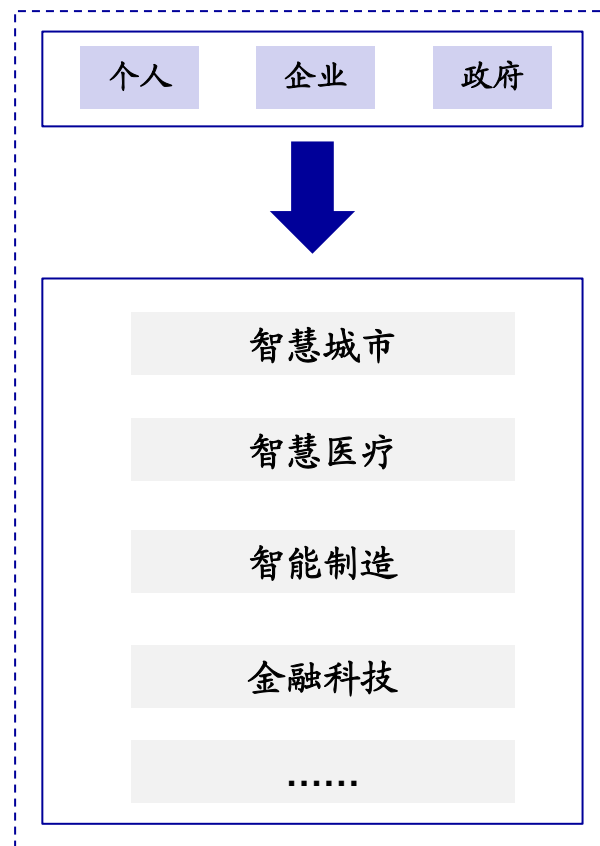
## 上游：算力基础硬件设施



## 中游：算力网络平台



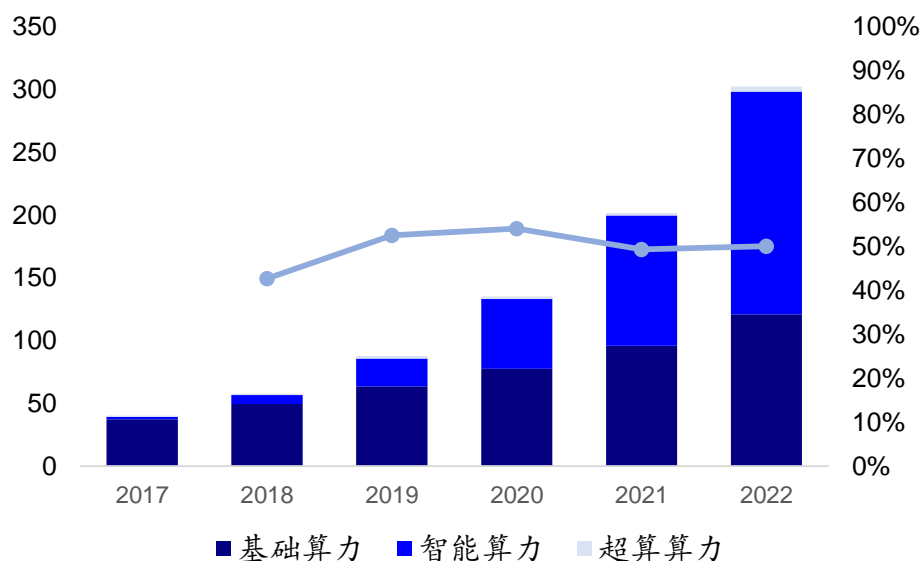
## 下游：应用场景与用户



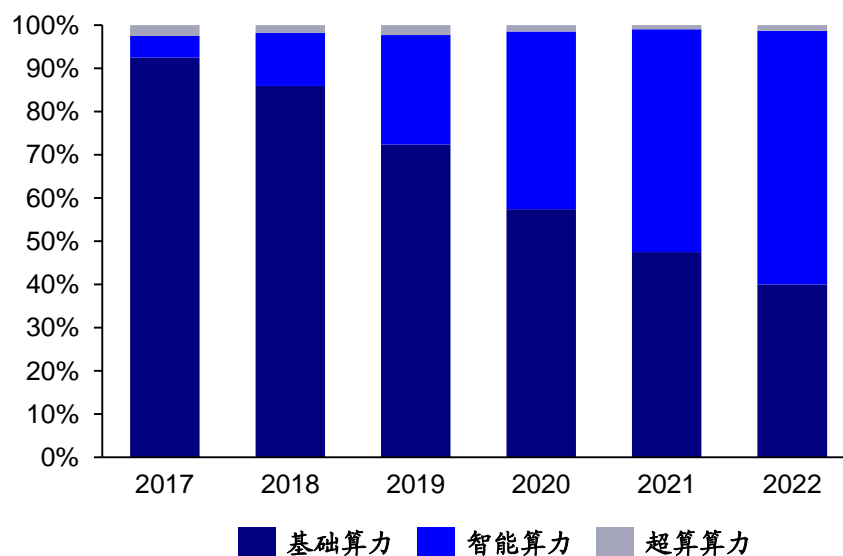
## 算力规模持续增长，智能算力增长迅速

- 经中国信息通信研究院测算，2022年我国计算设备算力总规模达到302EFlops，全球占比约为33%，连续两年增速超过50%，高于全球增速。基础算力稳定增长，基础算力规模为120EFlops，增速为26%，在我国算力占比为40%，其中2022年通用服务器出货量达到384.6万台，同比增长3%，六年累计出货量达到2091万台。
- 智能算力增长迅速，2022年智能算力规模达到178.5EFlops，增速为 72%，在我国算力占比达 59%，成为算力快速增长的驱动力。

图：我国算力规模及增速



图：我国算力内部结构

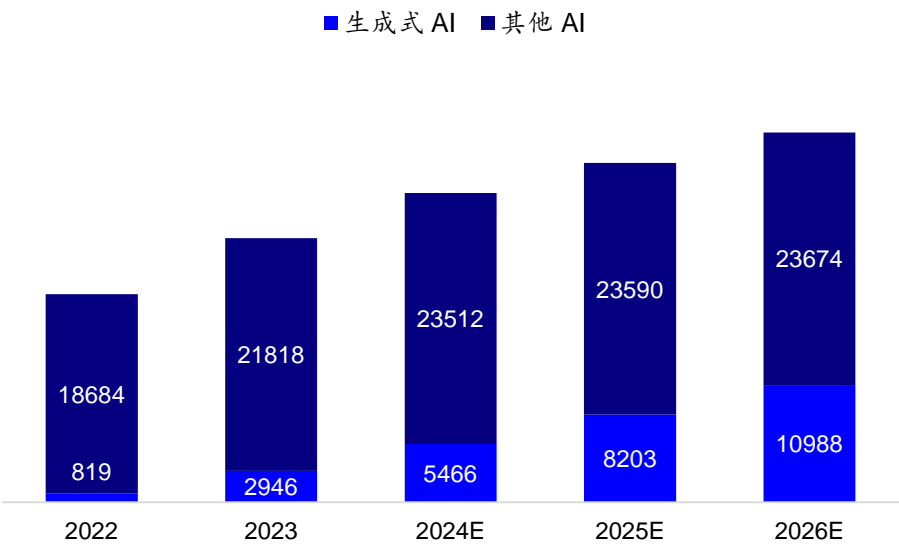


# AI服务器市场前景广阔，生成式AI服务器占比将不断提升

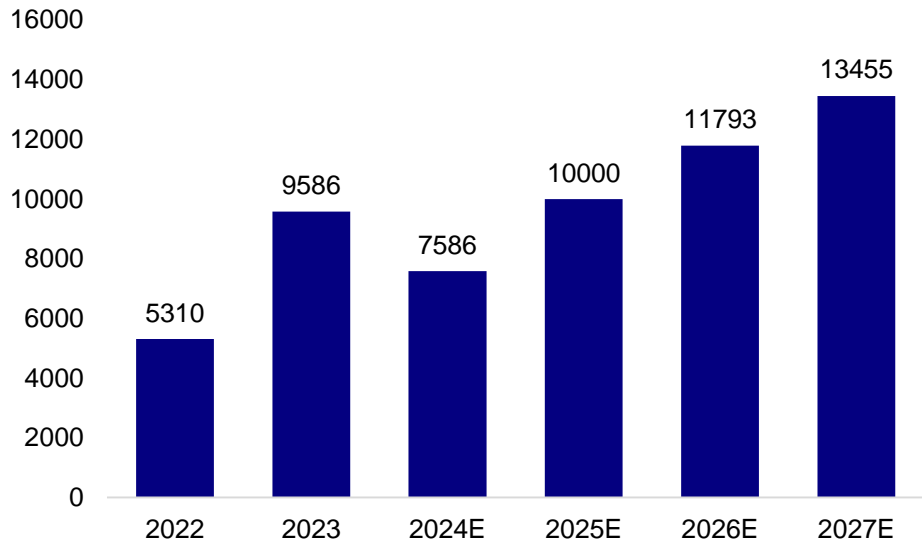


- 根据IDC测算，全球AI硬件市场（服务器），将从2022年的195亿美元增长到2026年的347亿美元，五年CAGR达17.3%；其中，用于运行生成式AI的服务器市场规模在整体AI服务器市场的占比将从2023年的11.9%增长至2026年的31.7%。
- 随着国内数字基础设施建设不断加快，以及AI应用场景的逐步落地，算力需求激增，带来AI服务器的需求的快速增长。IDC 预测，未来市场需求量也将会实现大幅度上升，预计2023年，中国AI服务器市场规模将达9586亿美元，同比增长80.5%，2027年将达到13455亿美元，五年CAGR为21.8%。随着生成式AI任务的不断增加，市场对于高性能和高能效的人工智能服务器需求将持续增长。未来的人工智能服务器将注重提高计算能力和处理效率，以适应更复杂、更大规模的人工智能应用。

图：全球 AI 服务器市场规模预测（百万美元）



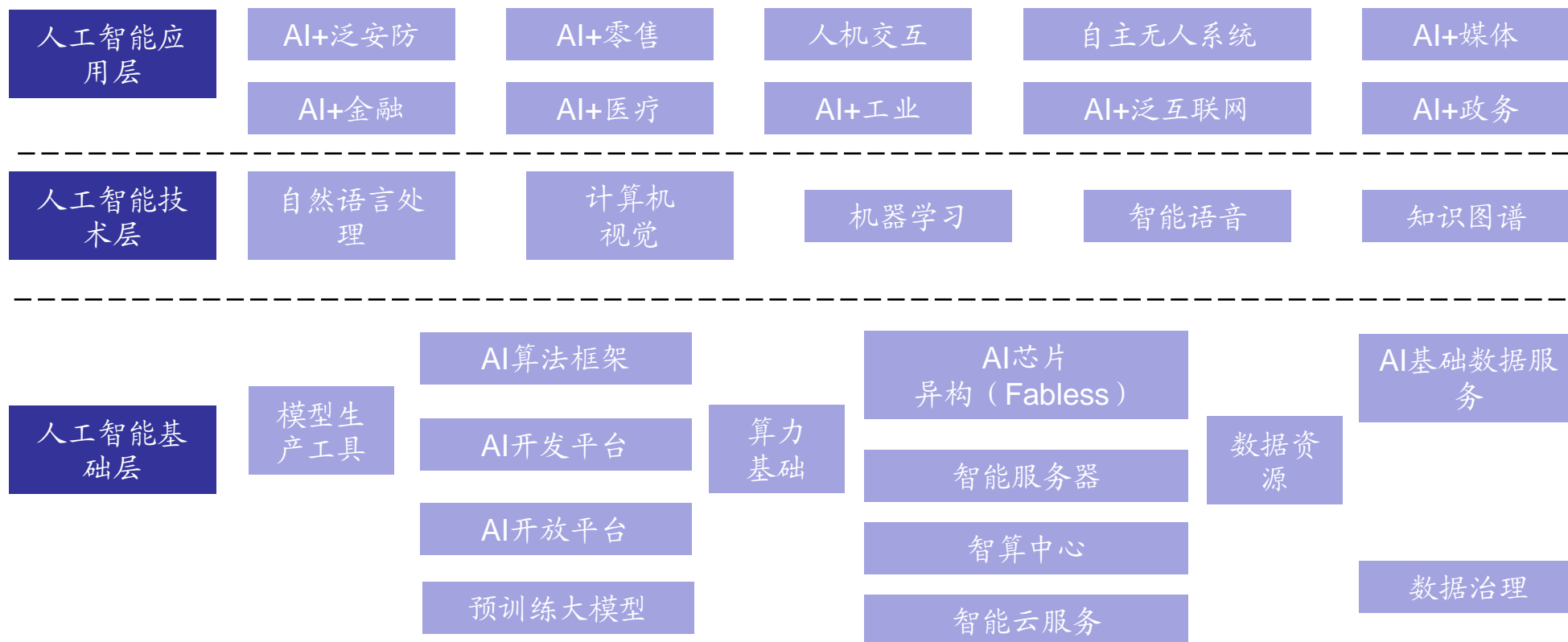
图：中国 AI 服务器市场规模预测（百万美元）



资料来源：IDC，海通证券研究所

- 人工智能产业链主要分为基础层、技术层和应用层三大层级，其中基础层主要包括AI模型生产工具、AI算力基础和AI数据资源三大要素。技术层包括计算机视觉、智能语音、自然语言处理以及机器学习和知识图谱。应用层涵盖AI+泛安防、AI+零售、人机交互、自主无人系统、AI+媒体、AI+金融、AI+医疗、AI+工业、AI+泛互联网、AI+政务等多领域。

图 人工智能产业链



► 国外的大模型主要厂商有OpenAI、谷歌、Meta 等。

图：OpenAI 、谷歌、 Meta主要大模型情况

厂商	发布时间	模型名称	底座模型	参数规模 (B)	预训练数据规模
Open AI	2022.5	GPT-3	-	175	300B tokens
	2021.12	WebGPT	GPT-3		
	2022.3	InstructGPT	GPT-4	175	
	2022.11	ChatGPT	GPT-3		
	2023.1	GPT-4	-	-	
	2024.1.25	GPT-4 Turbo		1760	16T tokens
谷歌	2019.1	T5	-	11	1T tokens
	2020.6	Gshard	-	600	1T tokens
	2021.9	FLAN	LaMDA-PT	137	
	2021.12	GLaM	-	1200	280B tokens
	2022.1	LaMDA	-	137	768B tokens
	2022.4	PaLM	-	540	
	2022.7	Flan-T5	T5	11	
	2022.10	Flan-PaLM	-	540	
	2023.1	Bard	-		
	2023.5	PaLM2	-	16	100B tokens
	2023.12	Gemini			
	2024.2	Gemini 1.5	-		
Meta	2022.5	OPT	-	175	180B Tokens
	2022.6	NLLB	-	54.5	
	2022.11	OPT-IML	OPT	175	
	2023.1	LLaMA	-	65	1.4T Tokens
	2023.6	LLaMA2	-	70	2T Tokens

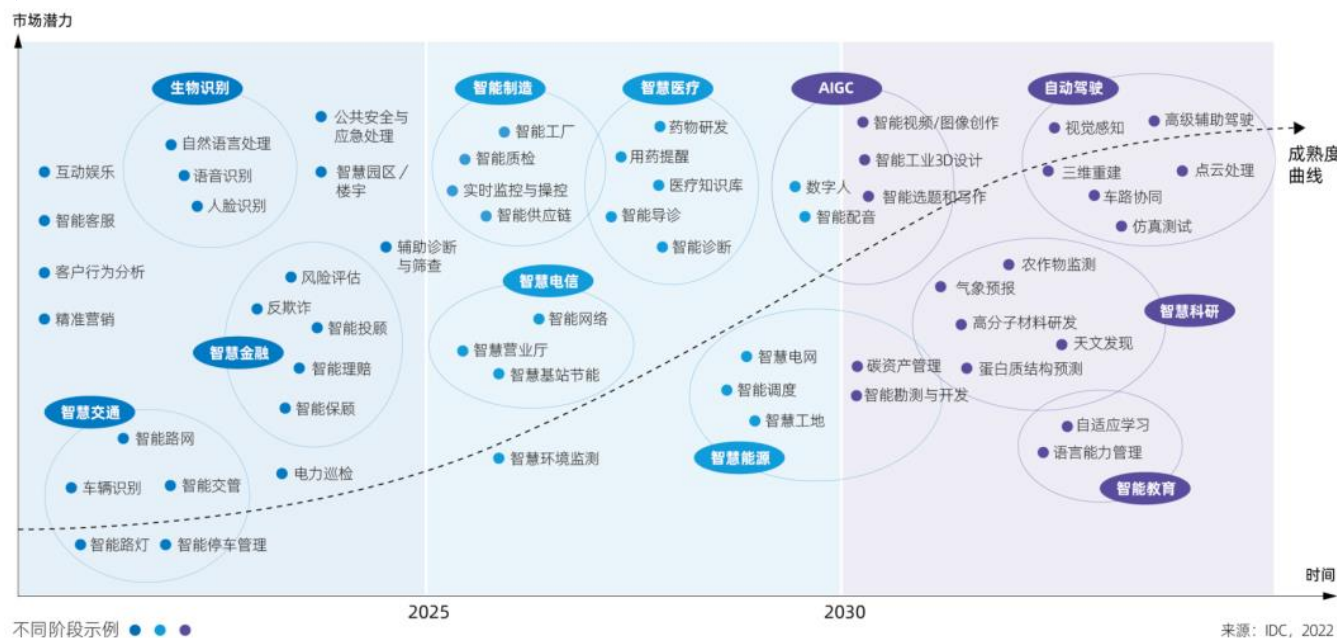


- **百度**于2023年发布新一代知识增强大语言模型**文心一言**，该模型是在**ERNIE**及**PLATO**系列模型的基础上研发的。它的关键技术包括有监督精调、人类反馈的强化学习、提示、知识增强、检索增强和对话增强。
- **阿里**2023年4月推出**通义千问**大模型，10月发布**通义千问2.0**。在10个权威测评中，**通义千问2.0**多方面性能超过**GPT-3.5**，正在加速追赶**GPT-4**。
- **华为**于2021年4月发布**华为云盘古大模型**。其中**盘古NLP大模型**是业界首个千亿参数的中文预训练大模型。在不断迭代中，目前已迭代至**盘古3.0**，包含五大基础模型：**NLP大模型**、**CV大模型**、**多模态大模型**、**预测大模型**、**科学计算大模型**，运用于商务、金融、预报等多个场景。
- **腾讯**于2023年9月发布其自研通用大模型**混元大模型**。拥有超千亿参数规模，预训练语料超2万亿 tokens，具备强大的中文创作能力，复杂语境下的逻辑推理能力，以及可靠的任务执行能力。
- **科大讯飞**于2023年5月6日，发布**星火认知大模型**，经过持续迭代，先后推出**V1.5**和**V2.0**版本，目前已迭代至**V3.5**。最新的模型文本生成、语言理解等七大能力全面升级，其中数学、中文超过**GPT-4 Turbo**，代码达到**96%**水平，多模态达到**91%**水平。
- **商汤科技**于2023年3月14日发布**多模态多任务通用大模型“书生（INTERN）2.5”**，在多模态多任务处理能力方面实现多项全新突破。2024年1月17日，商汤科技与上海AI实验室联合香港中文大学和复旦大学正式发布新一代大语言模型**书生·浦语2.0（InternLM2）**，该模型在**2.6万亿token**的高质量语料上训练得到的，包含**7B**及**20B**两种参数规格及基座、对话等版本，满足不同复杂应用场景需求。

## 应用场景：千行百业AI落地不断加深

- 从场景应用维度看，智能化场景在行业的落地随着时间的推移，正呈现出更深入、更广泛的趋势。人工智能持续为提升用户体验做出贡献，当前诸如智能客服、智能推荐、精准营销等场景深入落地到各行业；人工智能也在精准科学防疫，加强公共卫生安全体系建设中承担重要角色，在病毒演变预测、疫苗药物研发、辅助诊断等维度实现广泛应用；长期来看，企业通过在数字人等数字化营销内容创作领域布局，创造差异化的营销体验，升级品牌形象；另外，科学家们越来越多地利用人工智能技术和方法，从数据中建立模型，重点围绕新药创制、基因研究、新材料研发等领域加速对前沿科学问题的探究。

图：中国人工智能应用场景发展



## 应用场景：千行百业AI落地不断加深

- **AI+医疗**：基于迭代优化的大模型技术，讯飞医疗全面升级医疗诊后康复管理平台，将专业的诊后管理和康复指导延伸到院外。根据患者健康画像自动分析，平台可为患者智能生成个性化康复计划，包括重点关注、用药指导、康复运动、出院随访、健康知识、患者咨询等，并督促患者按计划执行。讯飞诊后康复管理平台还可以通过外呼机器人和小程序、APP为康复过程中病患提供及时应答，回复开放性和交叉性的问题。
- **AI+金融**：在信贷领域，征信数据一直是非结构化数据的典范。由于它的复杂性和多样性，很难使用传统的数据处理方式进行分析。为了破解这一难题，度小满智能征信中台将大型语言模型LLM、图算法应用在征信报告的解读上，能够将报告解读出40万维的风险变量，将银行风控模型的风险区分度提升了26%。

图：讯飞星火医疗大模型正式发布





## 应用场景：千行百业AI落地不断加深

- **AI+工业：**在星火认知大模型的基础上，羚羊平台结合工业产业的发展现状，推出了工业大模型——“**羚机一动**”。中小企业在羚羊平台上自由发布需求，羚机一动针对企业需求给出专业化建议策略，智能匹配方案、服务商、专家等资源。星火认知大模型还可在企业内部知识库和工业知识库之上构建企业知识大脑，在研发、生产、服务营销各个环节上，精准地定位问题、得到有效解决方案。
- **AI+办公：**在软件领域，人工智能大模型逐步应用于自然语言处理、计算机视觉、语音识别等领域，相关软件产品的智能化水平得到提升。2023年4月，WPS AI对外亮相，并陆续应用于文字、表格、演示文稿等常用办公组件。除了文本生成等功能外，WPS AI可以通过自然对话的人机交互生成复杂的函数公式，进一步降低办公软件的操作门槛。2023年11月，WPS AI公测发布，正式对用户开放使用。

图：羚羊工业大模型



图：WPS AI





领域	公司名称	2024 营业收入 (百万元)	2024营业收入 增长率 (%)	2024净利润 (百万元)	2024净利润 增长率 (%)	2024 PE (倍)	2024 PS (倍)
芯片	寒武纪-U	1528	115.35	-557	33.38	-126.60	46.13
	海光信息	8421	40.07	1677	32.86	104.25	20.77
	景嘉微	1292	60.86	254	266.45	131.22	25.84
	龙芯中科	1127	121.59	53	116.16	645.14	30.48
	英伟达	111132	82.42	56833	90.97	40.72	20.82
服务器	浪潮信息	79833	13.97	2197	32.00	27.02	0.74
	中科曙光	17179	19.68	2427	32.09	28.37	4.01
	神州数码	130723	6.48	1469	18.62	13.67	0.15
	软通动力	23331	16.02	1127	30.27	38.07	1.84
	中国长城	18442	21.66	360	231.82	85.87	1.68
	拓维信息	4022	32.68	302	107.78	60.06	4.51
	协创数据	5924	27.19	484	68.47	28.86	2.36

领域	公司名称	2024 营业收入 (百万元)	2024营业收入 增长率 (%)	2024净利润 (百万元)	2024净利润 增长率 (%)	2024 PE (倍)	2024 PS (倍)
大模型	科大讯飞	25584	24.44	1250	54.26	84.90	4.15
	拓尔思	1266	26.05	282	154.82	43.52	9.69
	中控技术	11642	35.08	1349	22.71	25.82	2.99
	商汤-W	5165	51.66	-3357	47.87	-6.98	4.54
应用	海康威视	100059	11.98	17019	20.56	17.79	3.03
	金山办公	5880	29.05	1709	29.66	79.19	23.01
	新国都	5586	46.96	959	27.07	12.45	2.14
	赛意信息	3157	23.12	357	29.28	21.05	2.38
	大华股份	37851	15.39	4442	-10.54	13.44	1.58
	万兴科技	1922	26.58	145	52.60	88.06	6.66
	虹软科技	861	28.50	200	123.62	62.64	14.58
	新致软件	2199	30.29	167	136.75	27.29	2.07
	微软	239184	12.87	87964	21.56	35.62	13.10

- 1.人工智能发展不及预期;
- 2.政策不及预期;
- 3.AI商业落地不及预期。

## 1. 硬件分析

1.1 NVIDIA H100 架构分析

1.2 英伟达发布GH200芯片

1.3 谷歌OCS架构

## 2. 产业格局及市场分析

2.1 芯片工艺壁垒高，下游应用分布广泛

2.2 市场份额中国厂商登顶，光芯片空间可期

2.3 数通市场：技术迭代升级，800G等高速模块未来将成主流

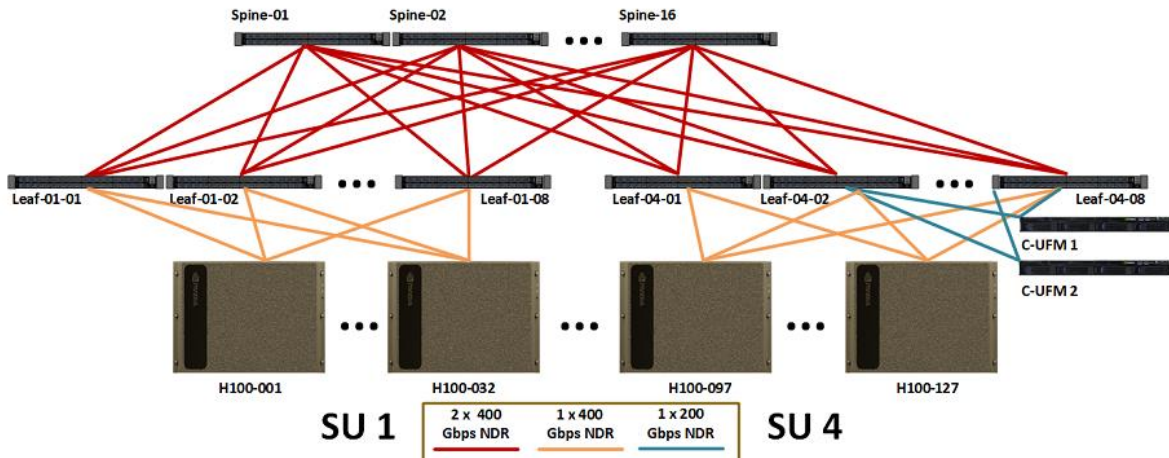
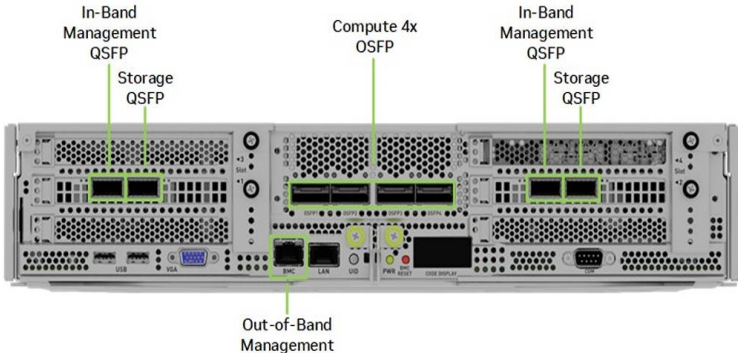
2.4 LPO、CPO等新技术未来可期

## 3. 标的推荐

## 4. 风险提示



图：NVIDIA H100 产品及架构情况



Node Count	GPU Count	SU Count	Leaf SW Count	Spine SW Count	Core SW Count	Node-Leaf Cable Count	Leaf-Spine Cable Count	Spine-Core Count
128	1024	4	32	16	--	1024	1024	1024
256	2048	8	64	32	--	2048	2048	2048
512	4096	16	128	128	64	4096	4096	4096
1024	8192	32	256	256	128	8192	8192	8192
2048	16384	64	512	512	256	16384	16384	16384

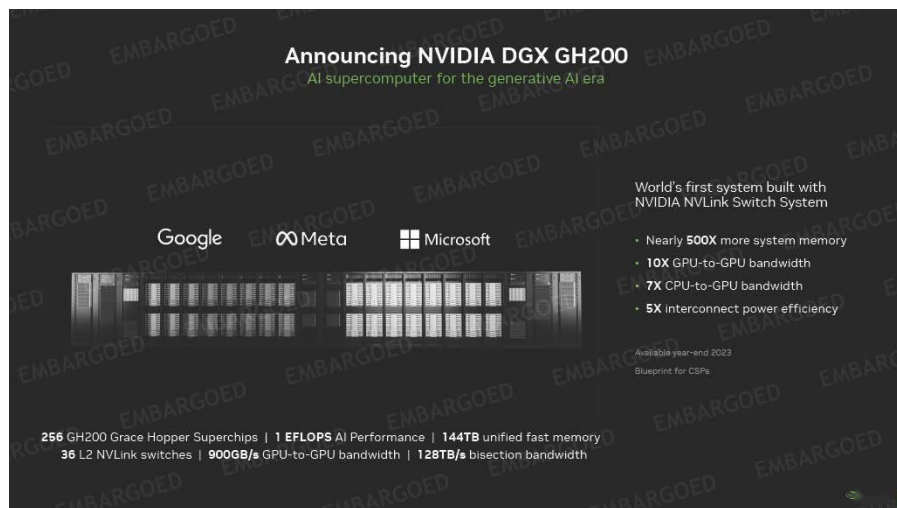
资料来源：英伟达官网，海通证券研究所

请务必阅读正文之后的信息披露和法律声明

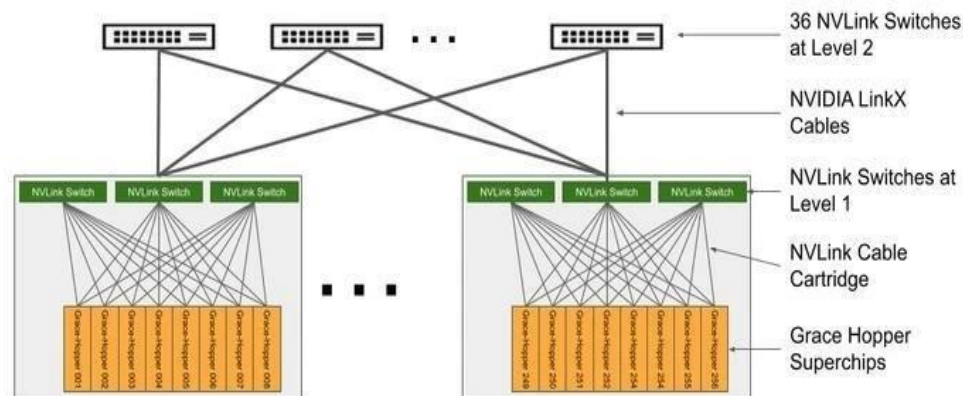
# 英伟达发布GH200芯片

2023年5月29日，AI芯片大厂英伟达（NVIDIA）CEO黄仁勋在COMPUTEX 2023展前发布会上，正式发布了一款新型大内存AI超级计算机——由NVIDIA GH200 Grace Hopper超级芯片和NVIDIA NVLink Switch System 驱动的NVIDIA DGX超级计算机，旨在助力开发面向生成式AI语言应用、推荐系统和数据分析工作负载的巨型、下一代模型。

图：GH200产品及架构



## Fully Connected NVLink across 256 GPUs



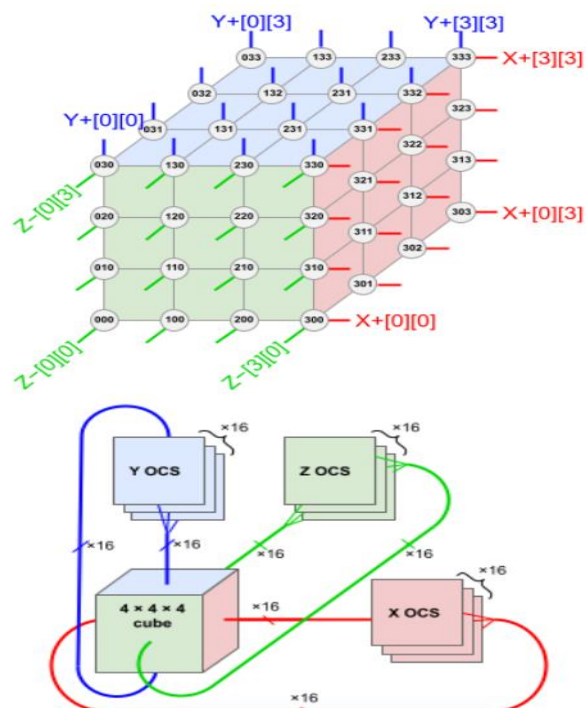
资料来源：芯智讯、腾讯网、IT之家，海通证券研究所

请务必阅读正文之后的信息披露和法律声明

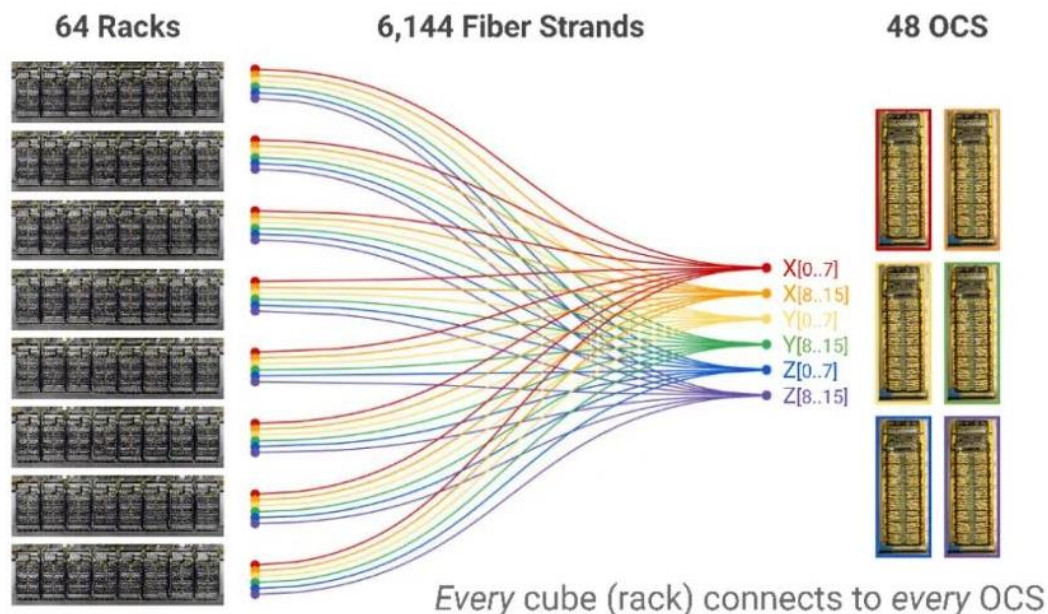
## 谷歌OCS架构

根据谷歌公布了其训练语言大模型的超级计算机的细节，基于 TPU 的超算系统已经可以比英伟达的同类更加快速、节能。如今 TPU 已经发展到了第四代。TPU v4 的性能比 TPU v3 高 2.1 倍，性能功耗比提高 2.7 倍。基于 TPU v4 的超级计算机拥有 4096 块芯片，整体速度提高了约 10 倍。

图：谷歌OCS架构



## The Physical System



资料来源：电子发烧友、系统范畴论微信公众号，海通证券研究所

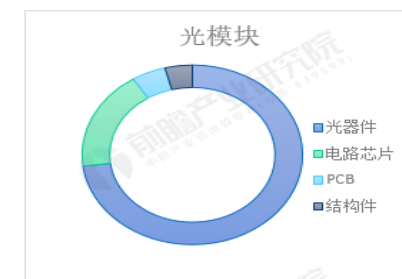
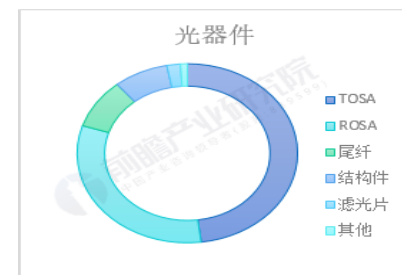
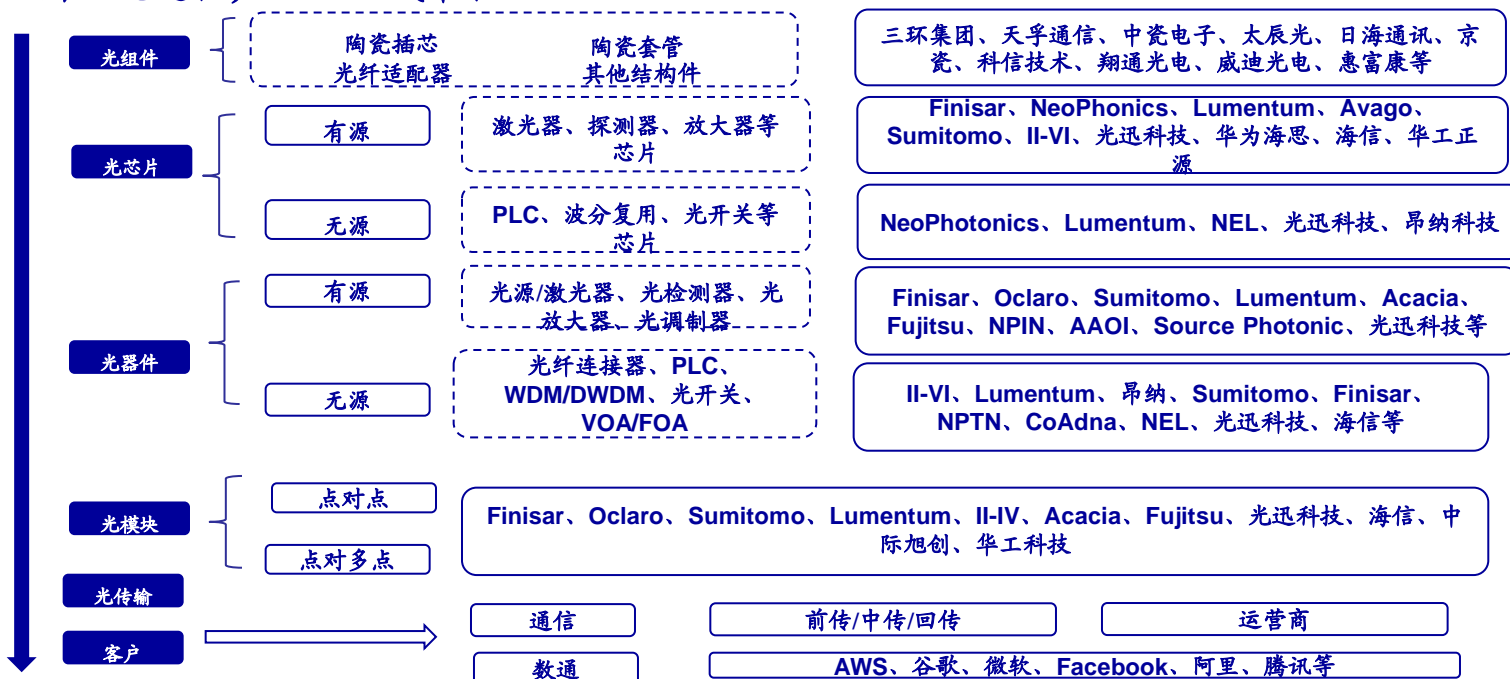
请务必阅读正文之后的信息披露和法律声明



# 光器件产业格局：芯片工艺壁垒高，下游应用分布广泛

- ▶ 光器件位于光通信产业链中游，上游包括光芯片、电芯片、光组件等，产业链下游是光通信设备商，最终客户方面，传统客户包括了 2B 侧电信市场的大型运营商和数通市场的云计算巨头，近年来光器件厂商开始逐渐向 2B 侧的非通信领域（如医疗检测等）和 2C 侧消费级应用场景（如 AR，激光雷达等）延伸，以寻求更大的发展空间。
- ▶ 光模块产品所需原材料主要为光器件、电路芯片、PCB 以及结构件等。其中，光器件的成本占比最高，在 73% 左右。光器件主要由 TOSA (以激光器为主的发射组件)、ROSA (以探测器为主的接收组件)、尾纤等组成，其中 TOSA 占到了光器件总成本的 48%; ROSA 占到了光器件总成本的 32%。

图：光通信产业链及成本占比



资料来源：路透社微信公众号，未来智库微信公众号，讯石光通讯微信公众号，海通证券研究所

请务必阅读正文之后的信息披露和法律声明



## 光器件：市场份额中国厂商登顶，光芯片空间可期

根据Ofweek光通讯网援引LightCounting最新版2022年全球光模块TOP10榜单，中国光模块厂商2010年仅有武汉电信器件有限公司（WTD，后与光迅科技合并）入围。在2022年全球光模块市场，旭创科技与Coherent并列TOP10榜首，Cisco（Acacia）排名第三，华为（海思）排名第四，光迅科技排名第五，海信宽带排名第六，新易盛排名第七，华工正源排名第八，Intel排名第九，索尔思光电排名第十。

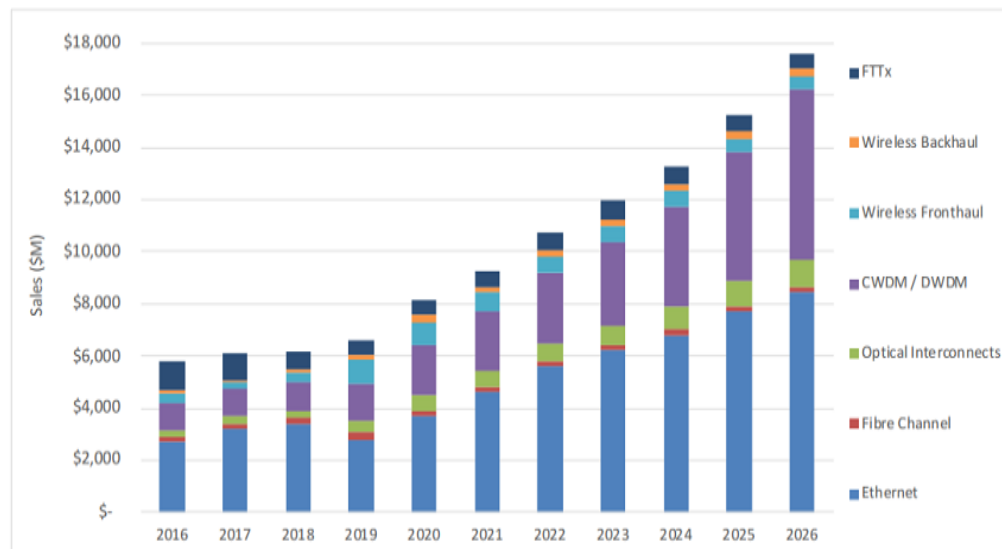
据中际旭创2021年年报援引Lightcounting预测，光模块的市场规模在未来5年将以 CAGR14%保持增长，2026年预计达到 176亿美元。受益于数据中心建设、5G 网络深入布局，中国光模块市场有望进一步增长。

图：全球光模块供应商TOP10

Ranking of Top 10 Transceiver Suppliers			
2010	2016	2018	2022
Finisar	Finisar	1 Finisar	Innolight & Coherent (tie)
Opnext	Hisense	2 Innolight	
Sumitomo	Accelink	3 Hisense	Cisco (Acacia)
Avago	Acacia	4 Accelink	Huawei (HiSilicon)
Source Photonics	FOIT (Avago)	5 FOIT (Avago)	Accelink
Fujitsu	Oclaro	6 Lumentum/Oclaro	Hisense
JDSU	Innolight	7 Acacia	Eoptolink
Emcore	Sumitomo	8 Intel	HGG
WTD	Lumentum	9 AOi	Intel
NeoPhotonics	Source Photonics	10 Sumitomo	Source Photonics

Source: LightCounting

图：全球光模块市场规模

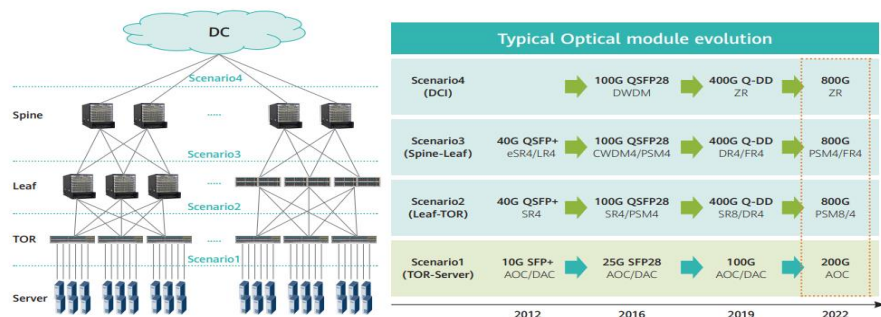


资料来源：OFweek光通讯网、ofweek、Wind、Lightcounting、中际旭创2021年年报，海通证券研究所 请务必阅读正文之后的信息披露和法律声明

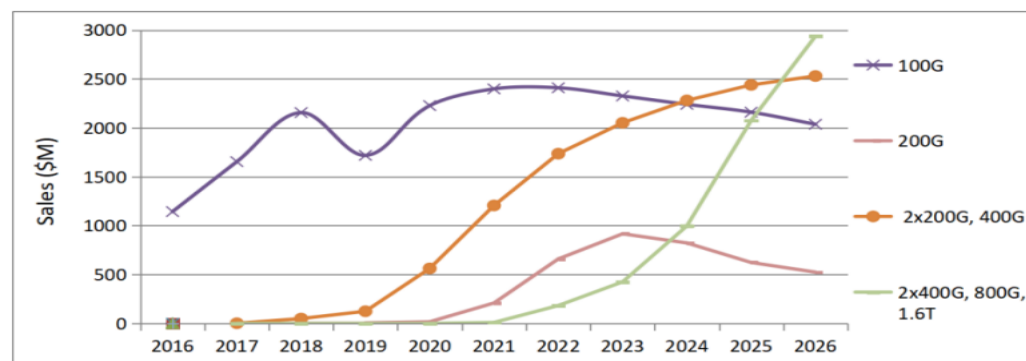
# 数通市场：技术迭代升级，800G等高速模块未来将成主流

LightCounting 最新数据显示，未来随着 AI、元宇宙等新技术不断发展，以及网络流量长期保持持续增长，以太网光模块销售额也将保持较快增长并不断迭代升级。

图：典型的超大规模数据中心互连路线图



图：2016-2026以太网光模块销售情况及预测



资料来源：中际旭创2021年报援引Lightcounting，中兴通讯《800G可插拔MSA白皮书》，海通证券研究所

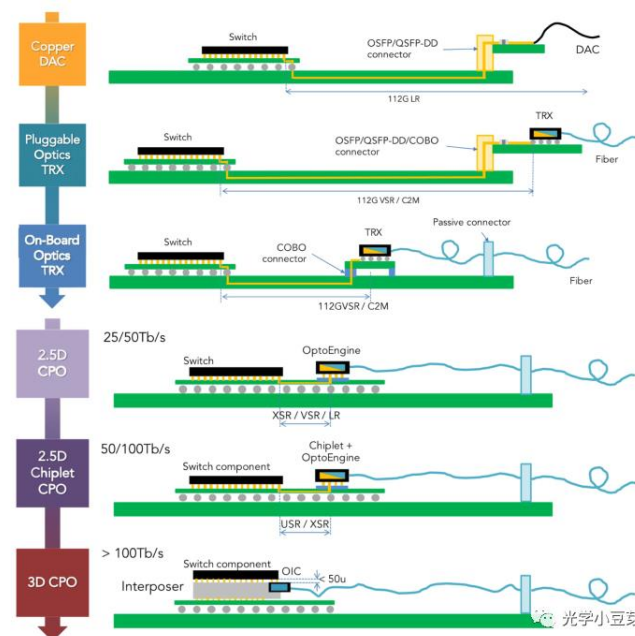
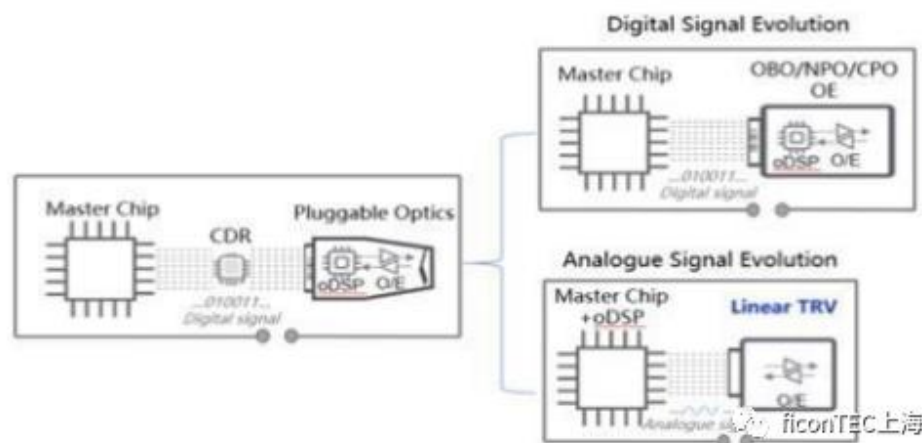
请务必阅读正文之后的信息披露和法律声明

# AI驱动，LPO、CPO等新技术未来可期

LPO (linear drive pluggable optics, 线性驱动可插拔光模块) 是指基于Linear Driver芯片技术实现的可插拔光模块, 线性驱动方案的直接表征在于去DSP化, 在数据链路中只使用线性模拟元件, 无CDR或DSP的设计方案。该方案主要运用于数据中心等短距离传输场景, 通过线性直驱技术替换传统的DSP, 实现系统降低功耗和延迟的优势。

Co-packaged optics(以下简称CPO)技术的出发点主要是为了取代传统的可插拔光模块, 将光芯片与交换芯片在基板上封装在一起, 进一步降低功耗, 提高带宽, 其主要用于数据中心的以太网网络。从可插拔光模块到CPO的演化图如下图所示, 光引擎与switch芯片的距离越来越近。

图：LPO及CPO方案



资料来源：Lightcounting、ficonTEC上海、光学小豆芽微信公众号，海通证券研究所

请务必阅读正文之后的信息披露和法律声明

## 建议关注公司：光模块、通信设备、PCB等

子领域	运营商	公司名称	股价（元）	市值（亿元）	24年PE（倍）	24年净利润（亿元）	年初至今涨幅（%）
光器件	300308.SZ	中际旭创	154.20	1238	29	42.42	36.57
光器件	300502.SZ	新易盛	67.01	476	38	12.58	35.87
光器件	002281.SZ	光迅科技	35.94	285	40	7.16	26.11
光器件	300620.SZ	光库科技	48.62	121	78	1.56	7.33
光器件	300394.SZ	天孚通信	151.00	596	51	11.61	64.99
光器件	000988.SZ	华工科技	31.16	313	23	13.58	4.70
PCB	002916.SZ	深南电路	87.33	448	26	17.07	23.02
PCB	002463.SZ	沪电股份	30.08	575	27	21.65	35.99
网络设备	000063.SZ	中兴通讯	26.15	1156	11	108.20	-1.25
网络设备	000938.SZ	紫光股份	20.63	590	21	28.08	6.61
网络设备	301165.SZ	锐捷网络	35.41	201	27	7.34	-6.55
网络设备	600498.SH	烽火通信	16.87	200	27	7.41	1.38
网络设备	301191.SZ	菲菱科思	81.36	56	19	2.95	-11.76
激光芯片	688498.SH	源杰科技	122.17	104	114	0.91	-18.00
激光芯片	688048.SH	长光华芯	41.70	74	63	1.16	-33.46
通信芯片	600702.SH	盛科通信	77.41	258	12	20.78	-19.95

资料来源：Wind，海通证券研究所（数据截至20240327，盈利预测采用WIND一致预期）



- 1.人工智能发展不及预期;
- 2.芯片短缺及供应风险;
- 3.AI商业落地不及预期的风险。

## 1. 半导体产业分析：大模型迭代利好**DRAM**需求向好

## 2. **AI+硬件**持续演绎，头部厂商积极布局

2.1 端侧数字芯片：不可或缺一环，**AI**时代芯片最后一公里

2.2 **AI PC**：Copilot稳步推进，端侧**AI**能见度最高

2.3 **AI+AR**眼镜：多模态加持，智能眼镜需求端迎来催化

2.4 **AI 手机**：关注三星产业链

## 3. 风险提示

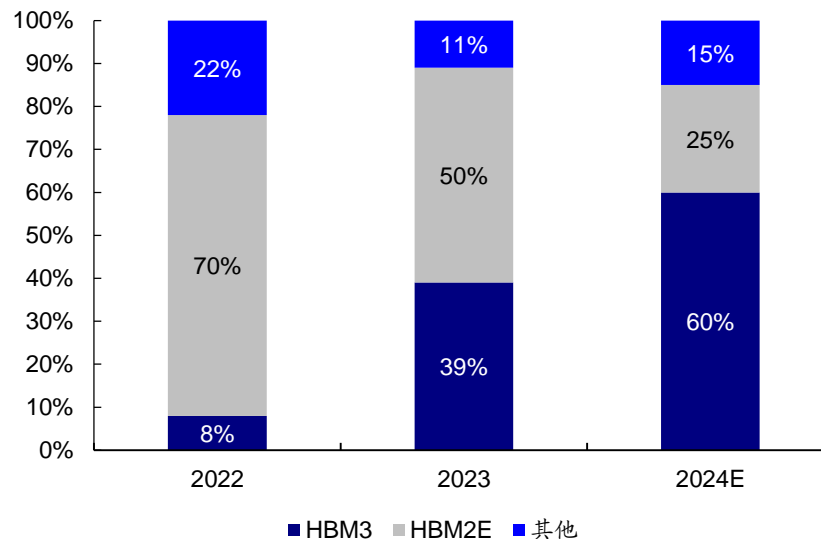
# 存储为半导体核心受益板块，大模型迭代利好DRAM需求向好

根据TrendForce集邦咨询微信公众号，AI服务器可望带动存储器需求成长，目前通用服务器的DRAM普遍配置约为500-600GB，而AI服务器在单条模组上则多采64~128GB，平均容量可达1.2~1.7TB之间。相较于一般服务器而言，AI服务器多增加GPGPU的使用，因此以A100 80GB配置4或8张计算，HBM用量约为320~640GB。随着使用HBM3的加速芯片陆续放量，2024年市场需求将大幅转往HBM3，而2024年将直接超越HBM2e，比重预估达60%，且受惠于其更高的平均销售单价，将带动HBM营收显著成长。未来在AI模型逐渐复杂化的趋势下，将刺激更多的存储器用量，并同步带动服务器DRAM、SSD以及HBM的需求成长。

表：通用服务器及AI服务器平均存储容量差异

	通用服务器	AI服务器	AI服务器（E）
服务器DRAM容量	500-600GB	1.2-1.7TB	2.2-2.7TB
服务器NAND容量	4.1TB	4.1TB	8TB
HBM容量	-	320-640GB	512-1024GB

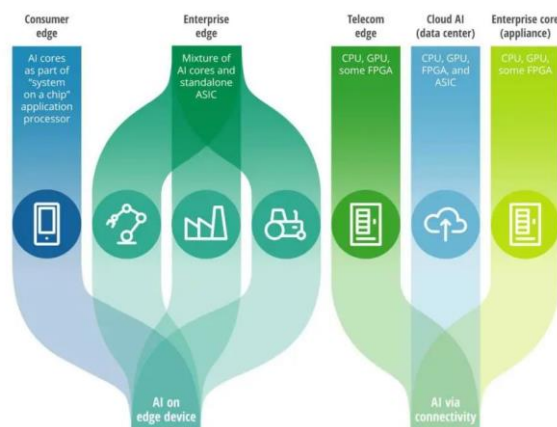
图：2022-2024E HBM结构



# AI端侧数字芯片：不可或缺一环，AI时代芯片最后一公里

- AI生态应是分布式的，分布在不同的终端设备上，以适应人工智能的感知、决策和执行环节。
- 据电子工程专辑公众号援引Deloitte分析，AI芯片(包括边缘和云端)的市场规模将从2018年的约60亿美元增长到2025年的900亿美元，这期间的年复合增长率高达45%。到2024年，边缘AI芯片的出货量将增至15亿颗，年增长至少20%，远高于全球半导体整体增长率(大约9%)。
- 目前的边缘AI芯片主要出现在消费类电子设备，其中高性能手机占据了较大的消费应用边缘AI芯片市场。然而，边缘AI芯片正越来越多地应用在非消费类设备和场合，比如智能安防、ADAS/自动驾驶、智能家居、可穿戴智能设备，以及商业和工业场合的AI应用(智能交通、智慧城市、工厂机器视觉、机器人和AGV等)。

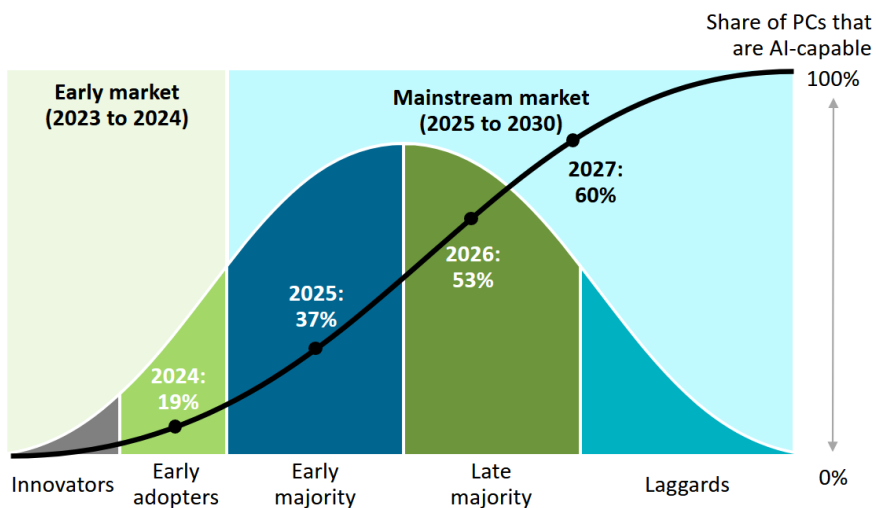
图：AI边缘芯片潜在下游领域



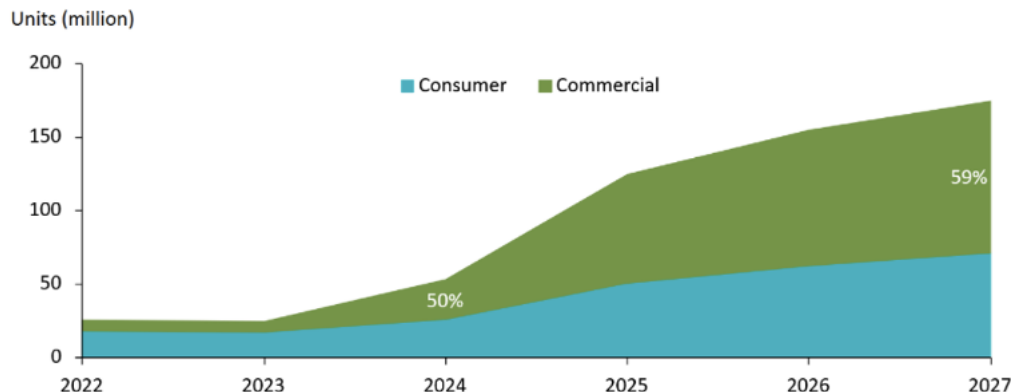
## AI PC: Copilot稳步推进，端侧AI能见度最高

- 生成式AI模型的爆发式增长，带来将专用的AI加速硬件集成到PC的需求。参考Canalys围绕硬件的初步定义，我们认为AI PC需要具备专用芯片组/块以承载端侧的AI运行负载，如配备NPU处理器，从而使得其能够在本地而非云端运行AI模型。在此定义下，根据Canalys数据，2024年出货的PC中，AI PC占比将会接近20%；到2027年，得益于换机动能和全新本地体验，AI PC占比将达60%以上。
- 随AI PC持续演绎，头部厂商积极布局。2023年底，英特尔及高通分别推出可集成NPU单元的处理器酷睿Ultra和骁龙X Elite。随后，联想、戴尔、惠普、三星、华硕、荣耀等陆续上线AI PC产品。此外，微软有望通过新增Copilot键以赋能终端体验。

图：AI PC全球接受度曲线



图：2022-2027年AI PC出货量预计



资料来源：Canalys, 36kr, IT之家, 澎湃新闻, 机器之心, 微软, 海通证券研究所



## AI+AR眼镜：多模态加持，智能眼镜需求端迎来催化

- 我们判断，针对AR眼镜的光显方案，Micro OLED+Birdbath的方案更适配影游娱乐场景，而Micro LED+光波导的方案更适合全天候佩戴场景。其中，多模态赋能下，以Micro LED+光波导为代表的AR眼镜，有望成为AI大模型的极佳载体。
- 目前，海外Meta Ray-ban多模态AI功能测试中，眼镜可以准确描述所看到物体衬衫并提供数款搭配建议。国内MYVU基于Flyme AI大模型也能够带来旅游攻略、学习计划、商业分析等等智能助手功能，和提词器、实时翻译（中英双语）、骑行导航等场景功能。

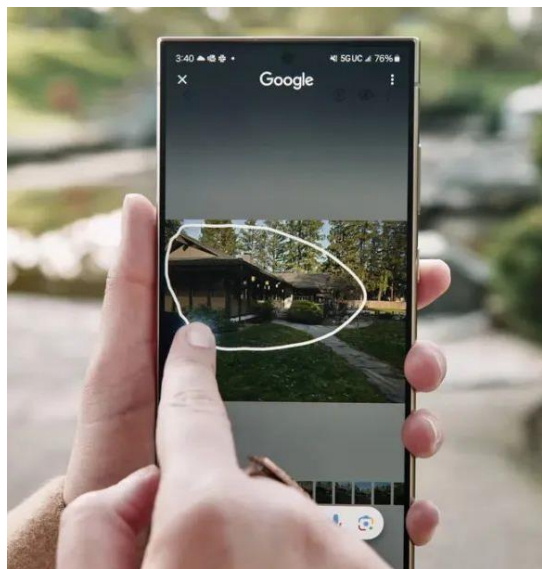
图：MYVU Discovery探索版配置



## AI 手机：关注三星产业链

- ▶ 三星首推AI手机，有望加速行业换机潮。根据韩联社，2024年1月，三星推出首款AI手机Galaxy S24，28天在韩国销量突破100万部，刷新该系列销量纪录。Galaxy S24 AI功能主要应用在翻译、搜索、影像3大方面，对应手机用户通话（信息）、搜索和拍照3大高频使用场景。
- ▶ 在国内，Galaxy AI在中国的本土化合作供应商包括百度、WPS、美图等。其中，百度提供的是云服务、美图提供的是自研视觉大模型MiracleVision与三星相册合作、WPS提供用户文档生成服务。我们认为，尽管当前终端并未搭载生成式AI功能，但其可以通过OTA方式实现快速部署。整体而言，AI加持对换机潮的影响值得关注。

图：Galaxy S24即圈即搜



图：Galaxy S24智能填充



1. 人工智能发展不及预期;
2. 政策不及预期的风险;
3. AI消费电子领域落地不及预期。

## 1.AI应用发展趋势分析

## 2.AI应用端产品分析

**2.1 IP开发、影视：AI增能提效，打破产能瓶颈，加快IP多元开发**

**2.2 互动、陪伴：AI定制个性化服务**

**2.3 游戏：AI“生产力价值”有望全面赋能游戏产能**

**2.4 营销：垂类模型持续赋能产业链**

**2.5 电商：AI有望重塑未来购物模式**

**2.6 教育：AI覆盖多层次教学环节，推进个性化教育**

## 3.投资逻辑与标的推荐

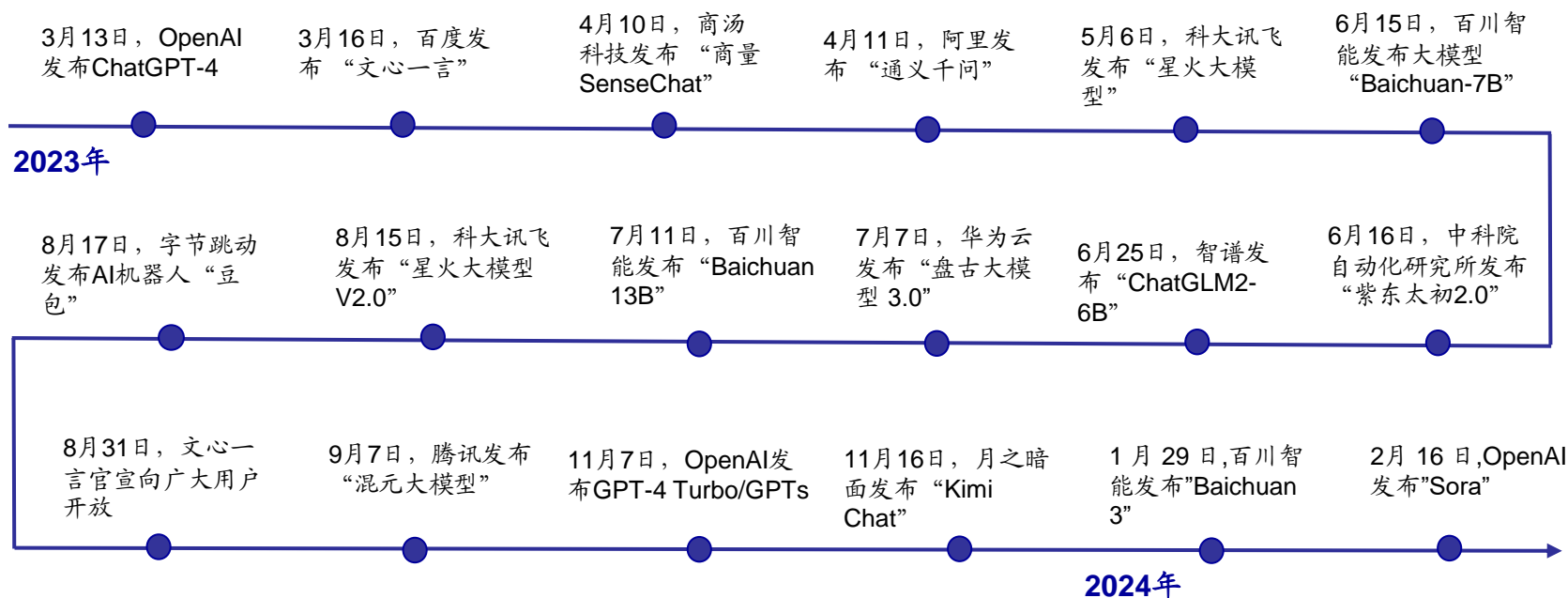
## 4.风险提示



## 发展趋势：大模型陆续上线，商业化变现拉开帷幕

2023年3月，OpenAI发布了ChatGPT-4，引起了全球市场广泛关注。2024年2月，OpenAI发布了文生视频大模型Sora，推动大模型向多模态跃迁。2023年，文心一言等首批国产AI大模型向公众开放，拉开中国大模型商业化变现帷幕。我们认为，大模型的全面开放，将使其直面用户和市场，并藉由更多的用户反馈推进自身的迭代。

图 2023.3-2024.3国内外AI重要进展



资料来源：公司官网，公司官方微信公众号，凤凰网微信公众号，海通证券研究所

# 应用端：多模态进化更迭，关注AI应用落地



我们认为，2023年是大模型向多模态化进化的一年，与此同时AI提供了更多商业化变现的途径。2024年生成式AI技术有望在IP开发、互动陪伴、游戏、营销、电商、教育等方向获得广泛应用。随着AI产品的逐步落地，2024年将主要是去伪存真的逻辑验证阶段，检验AI技术的应用是否能够很高效的产出，多模态进化之后是否在更复杂的视频、游戏等领域有生产力的提升和用户需求的解决，一旦实现AI赋能，渗透率从0~1将会带来收入和利润的巨大增量和弹性。

图 国内AI应用方向及进展

AI应用方向	国内相关公司	AI应用进展
IP开发、影视	上海电影	宣布开展“iNEW”战略，以“iPAi星球计划”为抓手，结合AI应用，主攻IP内容焕新和IP商业开发两大方向
互动、陪伴	掌阅科技	研发了AI智慧阅读产品“阅爱聊”，融合了阅读的乐趣和对情感的需求，可以对话系统中既定的角色形象
游戏	巨人网络	利用AIGC实现了游戏玩法的创新，基于大模型能力，创造性地打造出了一个AI推理剧场
营销	蓝色光标	2023年9月发布垂类BLUE AI模型，并宣布在2024年将在已有实践应用的基础上，向AI Agent（智能体）进化
	易点天下	数字营销创作平台KreadoAI以AIGC赋能创意营销并帮助广告主降本提效
电商	值得买	2024年2月值得买科技自研AI购物助手“小值”正式在“什么值得买”App上线，为存在不同决策难点的消费者提供个性化的建议，从而提升消费决策的质量和效率
教育	世纪天鸿	研发专注于服务老师的AI智能助手“小鸿助教”，涵盖教师在教学知识、教学方法、学生管理和工作事务等多方面的个性化需求

资料来源：公司官网，公司公告，公司微信公众号，海通  
证券研究所

# IP开发、影视：AI增能提效，打破产能瓶颈，加快IP多元开发



我们认为，在Sora发布后，多模态模型带来的内容生成能力继续往最核心的视频、3D内容等方向延伸，这也是图文之后价值量更大、影响范围更大的应用领域；且Sora的生成视频能力已经被评价为“世界模拟器”，生成时长、显示效果、物理逻辑都有很不错的表现，未来对IP开发领域预期会有很大的帮助。

IP是内容源头，AIGC技术能极大降低视频内容生成的中间成本。我们认为，AI加持下视频素材的版权价值量将会明显提升，在AI模型的演进之下，优质内容和应用大爆发的时代即将到来，而专业的IP开发方（PGC）具备明显的先发优势。

面对Sora掀起的文生视频新浪潮，上海电影宣布开展“iNEW”战略，以“iPAi星球计划”为抓手，结合AI应用，主攻IP内容焕新和IP商业开发两大方向。在IP商业开发方面，2023年，上影IP授权商品GMV总量已超10亿元，未来三年，上影力争实现IP合作产品涉足5大领域、30个行业、500个以上品牌，落地全国30个省份，合作商品GMV超百亿元目标。

图：上海电影“iPAi星球计划”



表：上影集团IP数字资产化布局

旗下公司	主要任务
上美影	探索“AI+动画”创作方向
上译厂	探索“AI+有声”创作方向
上海电影技术厂	探索“AI+影视基地”一站式解决方案
上海影视乐园	探索VR、AR+文旅实景体验方式
上美设计公司	探索AI数字展陈+IP行业先例

我们认为，在AI技术加持下，相关厂商可以针对服务内容进行优化训练，通过机器学习的方式，为用户打造更逼真、更人性化的交互体验，提高产品用户粘性。

在AI+互动领域，掌阅科技研发了AI智慧阅读产品“阅爱聊”，融合了阅读的乐趣和对情感的需求，可以对话系统中既定的角色形象，帮助用户阅读各种类型的文本内容，用聊天对话的形式把名著或小说轻松读一遍，让阅读成为一种更加愉悦、有趣的体验，同时也有许多角色基于虚构人物或中外历史名人，甚至还有完全虚拟的角色。

图：阅爱聊内测版



表：“阅爱聊”产品特点

产品特点	表现
多风格支持	支持多种风格，例如幽默、正式、友好、教育等。会根据不同的风格，调整生成内容的语气、态度和情感，让内容更加符合预期和效果
双模式解读	支持角色和读书双重模式。既可以从书中找到一个角色对话，也可以从一个角色中找到一本书阅读，让内容更加完整和有深度
交互式体验	不仅可以为用户生成内容，还可以和用户进行互动。用户也可以随时修改、编辑或者保存生成的内容
沉浸式阅读	阅读前可与本书作者对话全面了解全书梗概与精髓，甚至经典语录。阅读中可通过对话了解后续章节，阅读后可与本书作者交流心得与感想。未来甚至可以推书、催更

资料来源：掌阅科技官方公众号，阅爱聊微信小程序，海通证券研究所

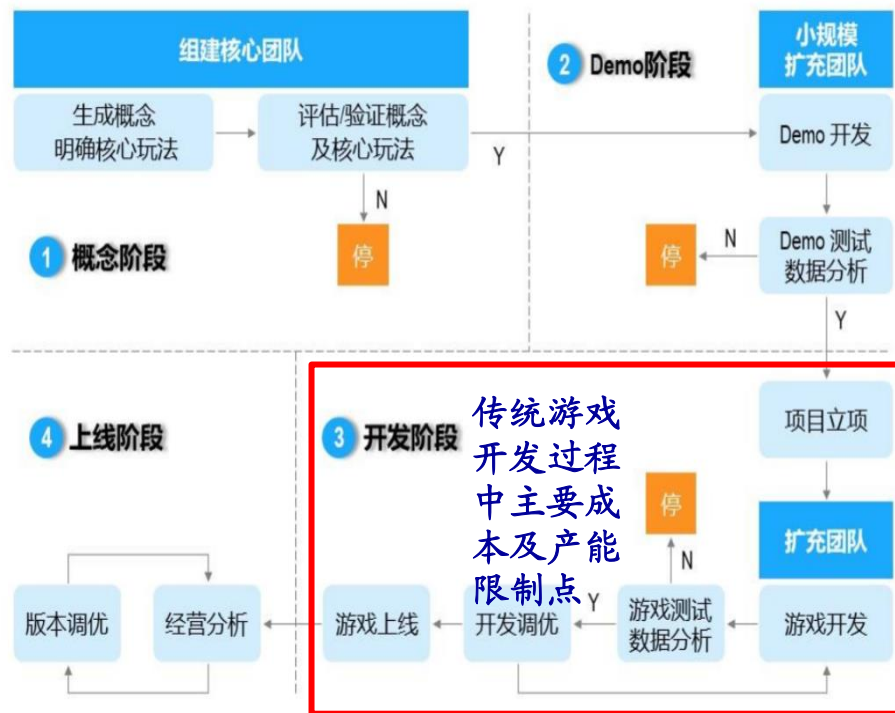


# 游戏：AI“生产力价值”有望全面赋能游戏产能

我们认为，AI技术可以解决游戏行业众多痛点。AI技术可以全面提高游戏开发团队的生产力，一方面可以克服传统游戏开发过程中成本产能限制点，起到降本增效的作用；另一方面可作为UGC（用户生成内容）工具，降低玩家生产创作的门槛。

在一款轻度休闲手游《太空杀》中，巨人网络的开发团队借助了AIGC 实现了游戏玩法的创新。团队基于大模型能力，创造性地打造出了一个AI推理剧场，作为《太空杀》的内置玩法。玩家需在给定的一个案件背景的情况下，和AI进行人机对话。AI的介入让玩家的想象力变得更加丰富。

图：游戏开发流程



图：巨人网络旗下游戏《太空杀》



我们认为，随着大模型基座的快速迭代和发展，通过营销领域大量数据训练建立的行业垂类大模型，能够为中小商家提供高效解决方案和个性化服务，持续赋能产业链。

2023年9月，蓝色光标发布垂类模型BLUE AI，2024年Blue AI将在已有实践应用的基础上，持续向AI Agent（智能体）进化。2023年，公司AI驱动业务提效在30%到1000%之间，突出业务场景提效十倍以上；公司在2023年里通过生成式AI赋能创造的案例达300多个，其中深度驱动并带来规模化收入的案例约在1/3以上，AI驱动的收入在1个亿以上。

2023年7月，易点天下首个AIGC数字营销创作平台KreadoAI面向全球创作者正式发布。自发布以来，KreadoAI以易点天下长期积累的丰富全面的营销大数据及AI算法模型为驱动，以AIGC赋能创意营销并帮助广告主降本提效。KreadoAI 可为全球用户提供「AI+」的多场景解决方案，已应用在商旅推荐、电商购物、应用下载、教育培训、企业服务等领域。

表：蓝色光标2023年度AI营销案例

品牌客户	AI营销作品
vivo	为vivo s18手机创造AI还原红楼梦广告
京东	用AI技术为京东11.11符号进行全面视觉升级
佳能	借助AIGC技术打造一部“次元世界的动漫作品”
伊利	打造乳品行业首款AI包装
广汽丰田	为MPV塞纳打造了数字生命的AI价值观
宁德时代	全球首创的用户自行选择剧情的AIGC广告片

资料来源：蓝色光标官方微信公众号，易点天下官方微信公众号，海通证券研究所

图：易点天下AI数字营销创作平台KreadoAI



## 电商：AI有望重塑未来购物模式

我们认为，AI在电商领域具备广阔的应用前景，有望重塑未来购物模式。卖家侧，能够打通直播带货全流程，实现降本增效；买家侧，能够基于大模型记忆推理能力，高效便捷地满足复杂个性化购物需求，辅助买家完成决策。

在卖家侧，2023年7月京东正式推出言犀大模型。京东云言犀AI开发计算平台，可为客户的大模型开发和行业应用，提供一站式的解决方案，平台精选了京东技术团队多年来开发出的100多种训练和推理优化工具，可提供更加高效的大模型开发环境，让用户可以快速地吧通用模型，转化成适合自身场景的专业模型。

在买家侧，2024年2月值得买科技自研AI购物助手“小值”正式在“什么值得买”App上线。“小值”是值得买科技基于值得买消费大模型所研发的Agent产品，能通过对话深度理解用户需求，基于全网实时消费经验、价格信息进行快速总结，提供口碑总结、商品对比、商品推荐、全网比价等服务，为存在不同决策难点的消费者提供个性化的建议，从而提升消费决策的质量和效率。

图 京东云言犀AI数字人带货



资料来源：京东云微信公众号，什么值得买微信公众号，公司公告，海通证券研究所

图 小值AI购物助手





# 教育：AI覆盖多层次教学环节，推进个性化教育

我们认为，在AIGC时代，大语言模型和教育适配程度高，其优秀的对话能力、语言理解能力和表达能力等促进AI技术在教育领域落地。此外，AI+教育市场发展空间大，客户端付费意愿强烈，商业发展逻辑清晰。

2023年，多家企业布局AI+教育并取得进展。南方传媒通过发明专利授权“基于AI深度学习的自动化数字教材建模系统”；城市传媒通过AI技术实现了精准、多维度的评价分析，并应用到了不同的学科教育中，例如AI作文批阅；中南传媒子公司中南迅智着力开展教育质量监测考试服务，以试卷、教辅等纸媒为流量入口，重点打造考试阅卷系统、考试测评系统等产品。

在教师端，世纪天鸿研发专注于服务老师的AI智能助手“小鸿助教”，涵盖教师在教学知识、教学方法、学生管理和工作事务等多方面的个性化需求，通过对话的方式，在包括教案生成、作文批改、教学活动策划、思维导图设计、教师评语编写等多种应用场景帮助老师提升工作效率。

图 南方传媒数字教材应用平台“粤教翔云”

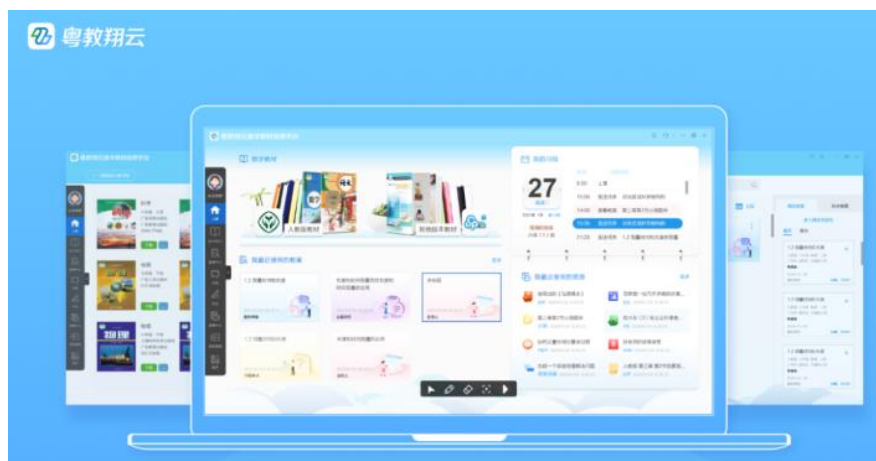


图 世纪天鸿的AI教师助手“小鸿助教”



资料来源：南方传媒官方微信公众号，小鸿助教官方微信公众号，海通证券研究所



我们建议持续关注AI+传媒方向投资机会。国内AI大模型进展迅速，我们预计2024年是国内AI应用产品的密集落地期。建议重点关注出版行业估值提升机会。国内出版公司主营稳健，现金储备丰富，股息率高，未来政策引导下，有望通过资源整合，实现战略转型。国有出版公司由于大多主营涉及教辅、教材的出版和发行，因此和学校、学生等群体联系紧密，部分出版发行国企深入学校，开展课后服务业务，促进教育产业链不断延长，推动业务转型升级。

建议关注：

- 1) IP资源和影视方面：上海电影、中文在线、掌阅科技、奥飞娱乐、捷成股份、博纳影业、百纳千成；
- 2) TOC应用方面：昆仑万维、紫天科技、焦点科技、万兴科技；
- 3) 教育方面：南方传媒、皖新传媒、时代出版、凤凰传媒、中南传媒、中文传媒；
- 4) 游戏方面：恺英网络、三七互娱、神州泰岳、吉比特、世纪华通、完美世界、掌趣科技、巨人网络。

1. 国内AI大模型进展进度不及预期;
2. 传媒公司相关内容产品开发进度不及预期;
3. 短剧、小游戏等内容监管风险。

## 分析师声明

杨林、余伟民、张晓飞、毛云聪、陈星光、崔冰睿

本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度，独立、客观地出具本报告。本报告所采用的数据和信息均来自市场公开信息，本人不保证该等信息的准确性或完整性。分析逻辑基于作者的职业理解，清晰地准确地反映了作者的研究观点，结论不受任何第三方的授意或影响，特此声明。

## 研究团队:

计算机行业首席分析师: 杨林  
资格证书编号: S0850517080008  
电 话: (021) 23154174  
Email: yl11036@haitong.com

电子行业首席分析师: 张晓飞  
资格证书编号: S0850523030002  
Email: zxf15282@haitong.com

通信行业首席分析师: 余伟民  
资格证书编号: S0850517090006  
电 话: (010) 50949926  
Email: ywm11574@haitong.com

传媒行业首席分析师: 毛云聪  
资格证书编号: S0850518080001  
电 话: (010)58067907  
Email: myc11153@haitong.com

传媒行业分析师: 陈星光  
资格证书编号: S0850519070002  
电 话: (021)23219104  
Email: cxg11774@haitong.com

电子行业分析师: 崔冰睿  
资格证书编号: S0850523100002  
电 话: (021)23185690  
Email: cbr14043@haitong.com

联系人: 杨昊翊  
Tel: (021)23185620  
Email: yhy15080@haitong.com

## 投资评级说明

	类 别	评 级	说 明
<b>1. 投资评级的比较和评级标准:</b> 以报告发布后的 6 个月内的市场表现为比较标准, 报告发布日后 6 个月内的公司股价 (或行业指数) 的涨跌幅相对同期市场基准指数的涨跌幅; <b>2. 市场基准指数的比较标准:</b> A 股市场以海通综指为基准; 香港市场以恒生指数为基准; 美国市场以标普 500 或纳斯达克综合指数为基准。	股票投资评级	优于大市	预期个股相对基准指数涨幅在 10%以上;
		中性	预期个股相对基准指数涨幅介于-10%与 10%之间;
		弱于大市	预期个股相对基准指数涨幅低于-10%及以下;
		无评级	对于个股未来 6 个月市场表现与基准指数相比无明确观点。
	行业投资评级	优于大市	预期行业整体回报高于基准指数整体水平 10%以上;
		中性	预期行业整体回报介于基准指数整体水平-10%与 10%之间;
		弱于大市	预期行业整体回报低于基准指数整体水平-10%以下。

## 法律声明

本报告仅供海通证券股份有限公司 (以下简称“本公司”) 的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下, 本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下, 本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断, 本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期, 本公司可发出与本报告所载资料、意见及推测不一致的报告。

市场有风险, 投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考, 不构成投资建议, 也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。在法律许可的情况下, 海通证券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并进行交易, 还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送, 未经海通证券研究所书面授权, 本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品, 或再次分发给任何其他人, 或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容, 务必联络海通证券研究所并获得许可, 并需注明出处为海通证券研究所, 且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可, 海通证券股份有限公司的经营经营范围包括证券投资咨询业务。