

# Gaudi 3、MTIA v2 相继发布，AI 产业链加速升级

计算机行业

推荐

维持评级

## 摘要 | 04.05-04.12

- 休市情况：**港股4月5日到7日休市；沪、深股通（北向）4月5日到4月7日休市、港股通（南向）4月5日到4月7日休市。
- 股指动态：**美股三大指数均下跌，恒生科技指数小幅上涨。标普500指数-1.56%，纳斯达克综合指数-0.45%，费城半导体指数-1.54%；TAMAMA科技指数+1.03%；纳斯达克中国金龙指数-3.22%；恒生科技指数+0.68%；计算机-4.22%。
- 个股表现：**热门科技股表现分化。据统计，相比4月5日收盘价，4月12日盘后，苹果合计+4.11%，英伟达+0.20%，特斯拉+3.73%，谷歌+3.43%，亚马逊+0.57%，META-2.93%，微软-0.85%，ARM+1.21%，英特尔-7.80%，高通-0.14%，AMD-4.19%。
- 10年期国债及汇率：**截至12日，中国10年期国债利率上升至2.28%，较4月8日累计上升0.15bps；美国10年期国债利率上升至4.50%，累计上升8bps。4月12日，美元兑人民币中间价报7.10；较4月8日价累计调贬20个基点。
- 核心观点**

Meta推出新一代AI芯片MTIA v2，采用台积电5nm制程。与MTIA v1相比，整体性能提高了3倍。英特尔推出其最新AI芯片Gaudi 3，并发布中国特供版本。Gaudi 3使用的台积电N5工艺技术制造，可缩小其与英伟达在半导体制造技术方面的差距。此外，两款经大幅“阉割”的产品Gaudi 3 HL-328 OAM及Gaudi 3 HL-388 PCIe专供中国，以应对美国的出口管制规则。我们认为，企业持续加大AI芯片研发力度，产业链上游竞争加剧，该趋势将提升芯片效率，优化算力成本，并赋能中游大模型、AI平台、软件等多方面的开发，进而加速产业链升级。此外，部分公司针对美国出口限制开发专供芯片至中国，旨在限制前提下，尽可能维持在中国市场的业务及市场份额，遵循政策的同时，最小化其可能面临的损失。

- 风险提示**

技术迭代不及预期风险；科技巨头竞争加剧风险；法律监管风险；供应链风险；下游需求不及预期风险。

## 分析师

吴砚靖

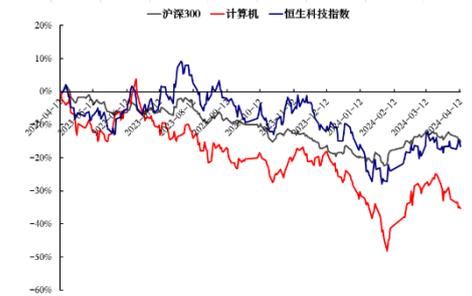
☎：(8610) 66568589

✉：wuyanqing@chinastock.com.cn

分析师登记编码：S0130519070001

## 国内表现

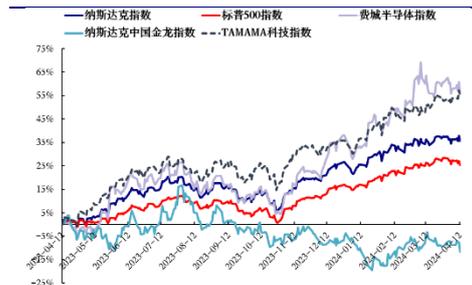
2024-04-12



资料来源：Wind，中国银河证券研究院

## 全球行情

2024-04-12



资料来源：Wind，中国银河证券研究院

## 相关研究

【银河计算机】全球科技动态追踪\_GPT-5有望今年夏季发布，多模态能力预期提升

【银河计算机】全球科技动态追踪\_Figure联合OpenAI发布人形机器人，加速AGI时代到来

【银河计算机】全球科技动态追踪\_Claude-3 AI模型发布，博通2024Q1财报AI增长强劲

## 目 录

一、全球市场表现.....	3
(一) 股市动态.....	3
(二) 债市及汇率情况.....	3
(三) 重点公司表现.....	3
二、行业要闻.....	4
(一) 算力及终端.....	4
(二) 大模型及云应用.....	8
三、风险提示.....	9

## 一、全球市场表现

### (一) 股市动态

**休市情况：**港股 4 月 5 日到 7 日休市；沪、深股通（北向）4 月 5 日到 4 月 7 日休市、港股通（南向）4 月 5 日到 4 月 7 日休市。

**美股三大指数均下跌，恒生科技指数小幅上涨。**标普 500 指数-1.56%，纳斯达克综合指数-0.45%，费城半导体指数-1.54%；TAMAMA 科技指数+1.03%；纳斯达克中国金龙指数-3.22%；恒生科技指数+0.68%；计算机-4.22%。

表 1：主要股指周变动

指数代码	指数简称	涨跌幅%					市盈率 PE (TTM)
		本周	上周	本月	本年度	2023	
SPX.GI	标普500指数	-1.56	-0.95	-2.49	7.41	24.23	25.58
IXIC.GI	纳斯达克指数	-0.45	-0.80	-1.25	7.75	43.42	41.78
SOX.GI	费城半导体指数	-1.54	-1.75	-3.27	13.64	64.90	49.06
8884057.WI	TAMAMA科技指数	1.03	0.68	1.71	14.95	67.81	37.18
HXC.GI	纳斯达克中国金龙指数	-3.22	0.37	-2.86	-8.25	-3.39	19.52
HSTECH.HI	恒生科技指数	0.68	-0.77	-0.09	-7.70	-8.83	21.36
CI005027.WI	计算机	-4.22	-2.73	-6.84	-15.03	8.90	78.76

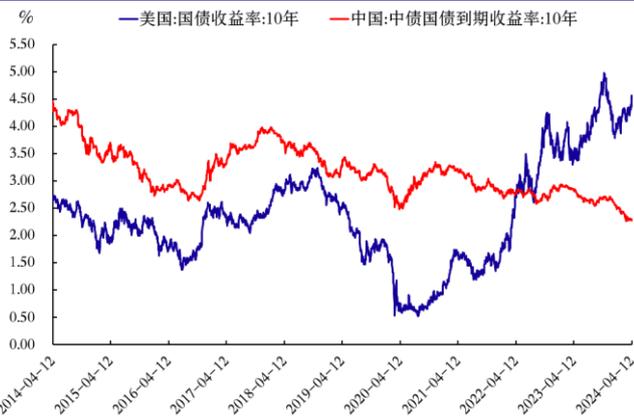
资料来源：Wind，中国银河证券研究院

### (二) 债市及汇率情况

**债市：**截至 12 日，中国 10 年期国债利率上升至 2.28%，较 4 月 8 日累计上升 0.15bps；美国 10 年期国债利率上升至 4.50%，累计上升 8bps。

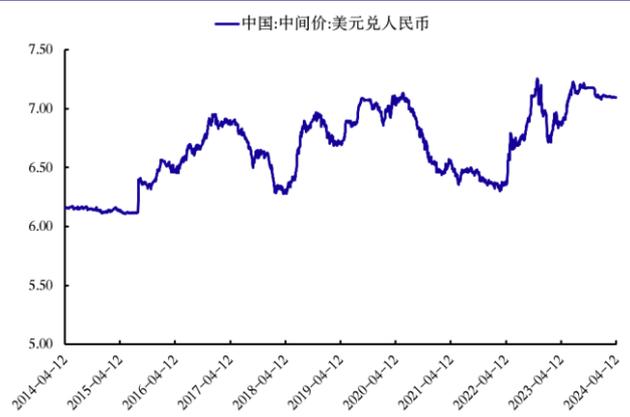
**汇率：**4 月 12 日，美元兑人民币中间价报 7.10；较 4 月 8 日价累计调贬 20 个基点。

图1：国债收益率（10年期）



资料来源：Wind，中国银河证券研究院

图2：美元兑人民币汇率(中间价)



资料来源：Wind，中国银河证券研究院

### (三) 重点公司表现

**热门科技股表现分化。**据统计，相比 4 月 5 日收盘价，4 月 12 日盘后，苹果合计+4.11%，英伟达+0.20%，特斯拉+3.73%，谷歌+3.43%，亚马逊+0.57%，META-2.93%，微软-0.85%，ARM+1.21%，英特尔-7.80%，高通-0.14%，AMD-4.19%。

**表 2: 重点公司周数据**

所属板块	地区	证券代码	公司名称	股价 (美元/港元/新台币)		区间涨跌幅 (%)	总市值 (亿美元/亿港元/亿新台币) [2024-04-12]	市销率 PS (TTM)	市盈率 PE (TTM)	PE(2022)	PE(2023)	PE(2024E)
				[2024-04-05]	[2024-04-12]							
算力及终端	美股	AAPL.O	苹果(Apple)	169.58	176.55	4.11	27,262.64	7.07	27.02	20.71	30.87	26.68
		NVDA.O	英伟达(NVIDIA)	880.08	881.86	0.20	22,046.50	36.19	74.08	36.86	280.04	77.93
		TSLA.O	特斯拉(TESLA)	164.90	171.05	3.73	5,447.58	5.63	36.32	30.98	52.67	48.31
		HPQ.N	惠普(HP)	29.11	28.53	-1.99	279.16	0.53	8.17	8.24	9.14	0.00
		CSCO.O	思科(CISCO)	48.55	48.49	-0.12	1,963.45	3.43	14.61	16.57	16.28	0.00
		ASML.O	阿斯麦	979.55	961.84	-1.81	3,795.32	12.41	43.63	36.70	34.34	46.93
		AMD.O	超威半导体(AMD)	170.42	163.28	-4.19	2,638.83	11.64	309.00	79.12	278.85	62.55
		INTC.O	英特尔(INTEL)	38.71	35.69	-7.80	1,519.28	2.80	89.95	13.61	125.43	36.36
		QCOM.O	高通(QUALCOMM)	171.53	171.29	-0.14	1,911.60	5.27	24.62	9.51	22.36	19.85
		ARM.O	ARM	124.82	126.33	1.21	1,298.77	44.21	1,527.96	0.00	147.03	0.00
	ON.O	安森美半导体 (ON SEMICONDUCTOR)	69.32	66.64	-3.87	286.45	3.47	13.12	14.18	16.47	15.37	
	0909.HK	明源云	2.30	2.35	2.17	45.64	2.50	-7.06	-10.65	-8.66	-13.94	
	9698.HK	万国数据-SW	7.29	6.32	-13.31	96.34	0.88	-2.01	-21.99	-2.92	-5.20	
	1686.HK	新意网集团	2.54	2.46	-3.15	57.54	2.27	6.34	11.66	7.98	0.00	
	2330.TW	台积电	780.00	818.00	4.87	233,332.29	9.81	25.30	11.71	18.34	21.06	
2454.TW	联发科	1,160.00	1,180.00	1.72	18,875.56	4.35	24.52	8.46	21.09	0.00		
互联网	美股	GOOGL.O	谷歌(ALPHABET)-C	152.50	157.73	3.43	19,694.94	6.38	26.57	19.04	23.69	23.17
		AMZN.O	亚马逊(AMAZON)	185.07	186.13	0.57	19,352.28	3.37	63.61	-314.82	51.61	45.33
		META.O	脸书 (META PLATFORMS)	527.34	511.90	-2.93	13,050.40	9.67	33.38	13.75	23.27	26.57
		NFLX.O	奈飞(NETFLIX)	636.18	622.83	-2.10	2,695.36	7.99	49.84	29.21	39.40	37.52
		PDD.O	拼多多	117.87	115.50	-2.01	1,534.52	4.39	18.11	22.77	22.94	12.50
	港股	NTES.O	网易	96.80	92.15	-4.80	621.50	4.07	14.30	16.34	14.46	14.01
		BIDU.O	百度	106.49	97.54	-8.40	349.16	1.80	11.92	36.42	14.52	11.07
		TCOM.O	携程网	48.14	48.85	1.47	339.79	5.31	23.84	116.72	17.58	26.34
		BABA.N	阿里巴巴	71.66	71.29	-0.52	1,862.64	1.39	12.86	23.78	18.64	12.96
		9988.HK	阿里巴巴-SW	70.05	71.65	2.28	14,598.14	1.43	13.23	23.80	18.53	13.30
		0700.HK	腾讯控股	310.00	309.60	-0.13	29,242.81	4.35	23.00	15.17	21.90	16.99
		80700.HK	腾讯控股-R	287.60	287.20	-0.14	26,479.36	4.45	23.54	0.00	21.86	17.40
		9999.HK	网易-S	149.80	151.10	0.87	4,870.90	4.27	15.01	16.52	13.96	14.66
		9888.HK	百度集团-SW	105.10	97.55	-7.18	2,736.49	1.84	12.21	36.49	14.48	11.30
		89888.HK	百度集团-SWR	97.35	90.45	-7.09	2,477.89	1.89	12.49	0.00	14.50	11.57
1024.HK	快手-W	49.65	49.15	-1.01	2,134.84	1.69	30.25	-19.96	32.60	16.11		
81024.HK	快手-WR	46.00	45.50	-1.09	1,933.09	1.73	30.90	0.00	32.58	16.47		
9626.HK	哔哩哔哩-W	86.20	94.60	9.74	398.51	1.60	-7.49	0.00	0.00	-19.81		
2518.HK	汽车之家-S	51.35	53.05	3.31	270.23	3.41	13.03	14.79	12.86	12.33		
9898.HK	微博-SW	72.00	64.00	-2.54	150.57	1.09	5.62	52.13	7.52	5.67		
软件及应用	美股	MSFT.O	微软(MICROSOFT)	425.52	421.90	-0.85	31,349.01	13.77	37.98	24.58	38.62	35.66
		SNOW.N	SNOWFLAKE	153.86	158.56	3.05	529.91	18.88	-63.38	-67.89	-82.25	0.00
		ORCL.N	甲骨文(ORACLE)	124.90	121.11	-2.72	3,328.73	6.34	31.28	32.81	34.08	25.15
		CRM.N	赛富时(SALESFORCE)	301.91	294.32	-2.51	2,854.90	8.19	69.03	91.82	1224.61	0.00
		ADBE.O	奥多比(ADOBE)	485.12	474.09	-2.27	2,123.92	10.65	44.24	32.69	50.01	37.44
	港股	INTU.O	财捷(INTUIT)	635.74	621.11	-2.16	1,738.98	11.52	62.78	52.92	73.39	62.61
		SNPS.O	新思科技(SYNOPSYS)	575.45	557.16	-3.18	849.91	13.86	60.39	49.43	63.63	54.73
		CDNS.O	铿腾电子(CADENCE)	308.02	305.85	-0.70	833.75	20.39	80.08	51.91	71.17	0.00
		ADSK.O	欧特克(AUTODESK)	244.74	236.98	-3.17	506.94	9.22	55.95	81.13	63.29	0.00
		U.N	Unity	25.36	25.31	-0.20	97.68	4.47	-11.82	-9.35	-18.80	0.00
		0020.HK	商汤-W	0.65	0.66	1.54	220.89	5.41	-3.11	-10.98	-5.46	-5.02
		80020.HK	商汤-WR	0.60	0.62	3.33	200.02	5.61	-3.22	0.00	-5.46	-5.21
		3888.HK	金山软件	23.75	25.70	8.21	344.17	3.50	64.51	-5.25	61.69	28.36
		0268.HK	金蝶国际	8.69	8.24	-5.18	296.12	4.64	-127.85	-133.49	-178.41	-944.47
		9878.HK	汇通达网络	30.80	30.30	-1.62	170.46	0.19	34.46	70.68	34.57	25.99
3650.HK	KEEP	4.30	6.33	47.21	33.28	1.38	2.73	0.00	6.00	-12.37		
0354.HK	中国软件国际	4.65	4.63	-0.43	134.60	0.69	17.10	24.18	22.39	12.06		
1357.HK	美图公司	3.10	3.18	2.58	144.21	4.73	34.55	57.75	38.62	25.13		
3896.HK	金山云	1.65	1.49	-9.70	56.70	0.73	-2.36	-2.69	-3.18	3.94		
2013.HK	微盟集团	1.44	1.37	-4.86	38.29	1.56	-4.58	-8.29	-9.62	-74.28		
1675.HK	亚信科技	7.18	7.20	0.28	67.34	0.77	11.45	13.73	13.69	6.78		
2121.HK	创新奇智	5.81	4.85	-16.52	27.40	1.42	-4.26	-31.65	-8.15	-6.85		
2400.HK	心动公司	13.96	14.20	1.72	68.24	1.82	-74.47	-16.75	-53.48	18.94		
0777.HK	网龙	10.82	11.12	2.77	59.08	0.73	9.73	9.71	10.84	5.56		

资料来源: Wind, 中国银河证券研究院

## 二、行业要闻

### (一) 算力及终端

#### 【Meta 推出新一代 AI 芯片 MTIA v2】

Meta 于 2024 年 4 月 10 日公布了其新一代 AI 芯片: MTIA v2, 并公布其技术细节。Meta 称: 与其上一代芯片 MTIA v1 相比, 新发布的芯片显著提高了性能, 该芯片从架构上追求计算、内存带宽和内存容量的适当平衡, 目的是在 Facebook 这类社交媒

体上用于广告排名和推荐广告。Meta 的下一代大规模基础设施正在以人工智能为中心构建，包括支持生成式人工智能的产品和服务、先进的人工智能研究。Meta 预计，随着支持 AI 模型的算力需求随着模型的复杂性而增加，这项投资将在未来几年增长。

**Meta 表示：MTIA v2 采用台积电 5nm 制程，相对 MTIA v1 的 7nm 制程有所升级。**与 MTIA v1 相比，整体性能提高了 3 倍，原因在于新一代 MTIA 采用更大的物理尺寸设计，比其上一代芯片拥有更多的处理内核。虽然它消耗更多的功率（90W 对 25W），但这也拥有更多的内部存储器（128M 64MB），并以更高的平均时钟速度运行（从 800MHz 上升到 1.35GHz）。

**新一代 AI 芯片对 Meta 的重要性：**1) 解决在 Facebook、Instagram 和 WhatsApp 中运行 AI 产品算力膨胀的问题；2) 帮助 Meta 减少对英伟达的依赖，并降低其整体成本。

MTIA v2 目前已在 16 个数据中心区域上线。

图3: MTIA v2 和 MTIA v1 的技术规格对比图

First Gen MTIA	Next Gen MTIA
<b>Technology</b> TSMC 7nm	<b>Technology</b> TSMC 5nm
<b>Frequency</b> 800MHz	<b>Frequency</b> 1.35GHz
<b>Instances</b> 1.12B gates, 65M flops	<b>Instances</b> 2.35B gates, 103M flops
<b>Area</b> 19.34mm x 19.1mm, 373mm <sup>2</sup>	<b>Area</b> 25.6mm x 16.4mm, 421mm <sup>2</sup>
<b>Package</b> 43mm x 43mm	<b>Package</b> 50mm x 40mm
<b>Voltage</b> 0.67V logic, 0.75V memory	<b>Voltage</b> 0.85V
<b>TDP</b> 25W	<b>TDP</b> 90W
<b>Host Connection</b> 8x PCIe Gen4 (16 GB/s)	<b>Host Connection</b> 8x PCIe Gen5 (32 GB/s)
<b>GEMM TOPS</b> 102.4 TFLOPS/s (INT8) 51.2 TFLOPS/s (FP16/BF16)	<b>GEMM TOPS</b> 708 TFLOPS/s (INT8) (sparsity) 354 TFLOPS/s (INT8) 354 TFLOPS/s (FP16/BF16) (sparsity) 177 TFLOPS/s (FP16/BF16)
<b>SIMD TOPS</b> Vector core: 3.2 TFLOPS/s (INT8), 1.6 TFLOPS/s (FP16/BF16), 0.8 TFLOPS/s (FP32) SIMD: 3.2 TFLOPS/s (INT8/FP16/BF16), 1.6 TFLOPS/s (FP32)	<b>SIMD TOPS</b> Vector core: 11.06 TFLOPS/s (INT8), 5.53 TFLOPS/s (FP16/BF16), 2.76 TFLOPS/s (FP32) SIMD: 5.53 TFLOPS/s (INT8/FP16/BF16), 2.76 TFLOPS/s (FP32)
<b>Memory Capacity</b> Local memory: 128 KB per PE On-chip memory: 128 MB Off-chip LPDDR5: 64 GB	<b>Memory Capacity</b> Local memory: 384 KB per PE On-chip memory: 256 MB Off-chip LPDDR5: 128 GB
<b>Memory Bandwidth</b> Local memory: 400 GB/s per PE On-chip memory: 800 GB/s Off-chip LPDDR5: 176 GB/s	<b>Memory Bandwidth</b> Local memory: 1 TB/s per PE On-chip memory: 2.7 TB/s Off-chip LPDDR5: 204.8 GB/s

资料来源: Meta 官网, 中国银河证券研究院

**【英特尔推出其最新 AI 芯片 Gaudi 3，并发布中国特供版本】**

英特尔在 2024 年 4 月 19 日推出其最新人工智能芯片 Gaudi 3，并公布其技术细节。

与英伟达 H100 芯片相比，英特尔 Gaudi 3 在具有 7B 和 13B 参数的 Llama2 模型以及 GPT-3 175B 参数模型中的训练时间快 50%。此外，在 Llama 7B 和 70B 参数模型以及 Falcon 180B 参数模型中，英特尔 Gaudi 3 加速器的推理吞吐量比 H100 平均高出 50%，推理能效平均高出 40%。

据 IEEE Spectrum 的数据，Gaudi 3 正在使用台积电的 N5 工艺技术制造，缩小了英特尔和英伟达在半导体制造技术方面的差距。

英特尔 Gaudi 3 将于 2024 年第二季度提供给原始设备制造商，包括戴尔科技、惠普企业、联想和超微。

图4: 英特尔在发布会上展示的 Gaudi 3 技术细节



资料来源: Intel 官网, 中国银河证券研究院

此外，英特尔准备了两款经过大幅“阉割”的 Gaudi 3 产品，以满足美国对中国的出口管制规则。一款是 Gaudi 3 HL-328 OAM，另一款是 Gaudi 3 HL-388 PCIe。英特尔没有透露这两款处理器的性能水平，但两者都具有 450W 的热设计功率（低于非中国版版本的 900W 和 600W），这表明性能从根本上降低。面向中国的 Gaudi 3 系列处理器具有 128 GB 的 HBM2E 内存（使用带有 1024 位接口的八个堆栈）和 3.7 TB/s 的带宽。中国特定的 Gaudi 3 的横向扩展能力与其他 Gaudi 3 加速器相似。

我们认为，企业持续加大 AI 芯片研发力度，产业链上游竞争加剧，该趋势将提升芯片效率，优化算力成本，并赋能中游大模型、AI 平台、软件等多方面的开发，进而加速产业链升级。此外，部分公司针对美国出口限制开发专供芯片至中国，旨在限制前提下，尽可能维持在中国市场的业务及市场份额。遵循政策的同时，最小化其可能面临的损失。

图5: 面向中国的两款处理器技术规格

Family	Intel® Gaudi® 3-OAM			Intel® Gaudi® 3-PCIe	
Part Number	HL-325L	HL-328	HL-335	HL-338	HL-388
Export Market	Non-PRC	PRC (*)	Non-PRC	Non-PRC	PRC (*)
Availability	Mar '24	Jun '24	Oct '24	Sep '24	Sep '24
Form Factor	OAM	OAM	OAM	Dual-slot PCIe, Full Height, 10.5" Length	
Cooling	Air	Air	Liquid (1P or 2P)	Air	Air
TDP [W]	900	450	900	600	450
Scale-up support	via x8 Baseboard (HLB-325)			x4 Top-Board (HLTB-304)	
Scale-out support	via Baseboard (6x OSFP800) or via Host-Nic			via Host-Nic or via dual QSFP-112 on PCIe card	
HBM Capacity	128GB				
Peak HBM Bandwidth	3.7TB/sec				
HBM Interface and Type	1024-bit x 8 stacks HBM2e				
Last-Level Cache Capacity	96MB				
Host Interface	PCIe Gen5 x16				
Media Decoders	HEVC/H.265, AVC/H.264, VP9, JPEG				

资料来源: Intel 官网, 中国银河证券研究院

### 【美国希望 ASML 停止在中国维修深紫外（DUV）光刻机】

由于美国和荷兰政府的限制，中芯国际等中国芯片制造商无法从 ASML 采购深紫外（DUV）光刻机，但他们仍然可以使用现有的 DUV 工具生产相当先进的芯片。据路透社报道，美国政府对此并不完全理解，因此下周，美国将试图敦促荷兰政府禁止 ASML 为中国工厂已经安装的光刻机提供维修服务。

美国负责出口政策的高级官员艾伦·埃斯特维兹（Alan Estevez）将于下周与荷兰官员和 ASML 代表会面并讨论此事。报道称，讨论还可能包括扩大禁止购买 ASML 设备的中国芯片工厂名单。如果美国和荷兰同意限制 ASML 机器在中国的维修，这可能会损害甚至破坏中国使用其芯片制造商已经拥有的光刻设备生产芯片的能力。

如果 ASML 被限制在中国提供服务，这可能会影响其财务业绩，并且还可能面临客户的诉讼。

### 【欧洲汽车制造商麦格纳将试用 Sanctuary AI 人形机器人】

Sanctuary AI 在 2024 年 4 月 11 日宣布与麦格纳建立战略合作伙伴关系。此次合作包括：将 Sanctuary AI 开发通用 AI 机器人用于麦格纳的制造业务；利用麦格纳的工程和制造能力，提高 Sanctuary AI 机器人可的扩展性。

麦格纳是全球最大的汽车供应商之一，也是工业机器人的重要买家。该战略合作伙伴关系可以促进 Sanctuary AI 机器人的规模化使用，同时使该技术成熟，以便为麦格纳和其他汽车工业客户在复杂的制造环境中使用。

Sanctuary AI 于 2023 年 5 月推出了自己的人形机器人 Phoenix。该机器人 5 英尺 7 英寸，重 155 磅。Sanctuary AI 开发了许多人形机器人的相关技术，包括其用于机器人的类人手，和为其人形通用机器人 Phoenix 开发的 AI 控制系统 Carbon。Sanctuary AI 拥有涵盖视觉和触觉伺服、语言模型集成和 AI 训练的重要专利，为能够执行数百种不同任务的人形通用机器人的商业化铺平了道路。

其他的人形机器人公司在此前也与汽车制造商有类似的合作。Figure 和 Apptronic 这两家公司分别与宝马和梅赛德斯合作，在制造过程中试运行他们的人形机器人。

图6: Phoenix 机器人的基本信息



资料来源: Sanctuary AI 官网, 中国银河证券研究院

### 【苹果将应用 M4 芯片于新款 Mac 中】

据彭博新闻社和路透社在 2024 年 4 月 11 日的报道：苹果即将生产具有 AI 处理能力的 M4 芯片，并计划将其应用在每款 Mac 型号中。预计至少有三个主要芯片型号：低端型号代号为 Donan、中端型号代号为 Brava、高端型号代号为 Hidra。报告称，该公司的目标是在今年年底和明年初发布新款电脑，其中包括 iMac，14 英寸 MacBook Pro，14 英寸和 16 英寸 MacBook Pro 以及 Mac mini。

另外，自新冠疫情时期居家办公的销售热潮结束以来，PC 制造商一直处于低迷状态。该行业复兴的希望将寄托在新一代具有更强大处理功能的笔记本电脑和台式机上，使其可以处理人工智能任务，并无需将数据发送到云端。

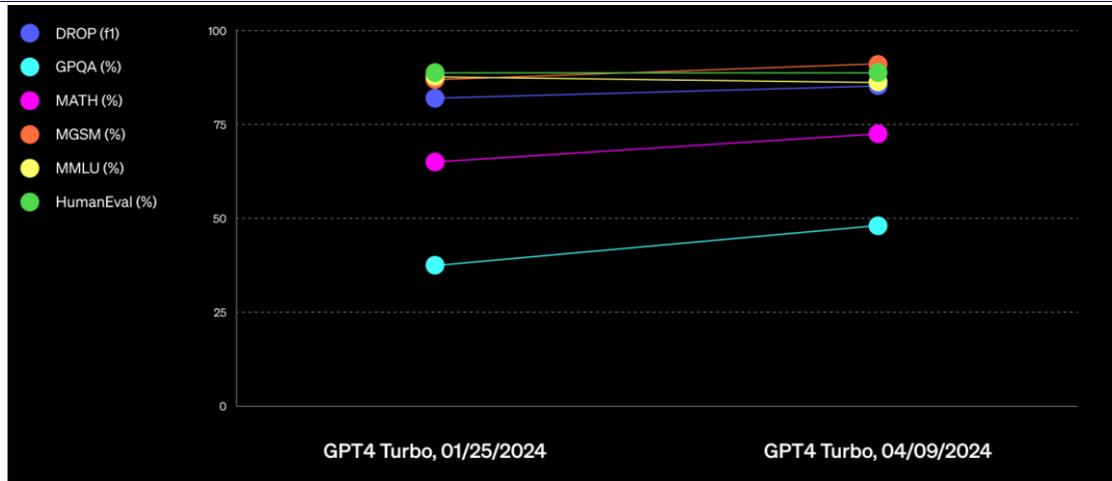
## （二）大模型及云应用

### 【ChatGPT-4 Turbo 现在已向付费用户通过 API 调用方式】

OpenAI 在 2024 年在 X 的一篇推文中宣布：配备视觉能力的 GPT-4 Turbo 现在已经可以通过 API 调用，此服务现在只向高级 ChatGPT 付费用户（ChatGPT Plus、Team 或 Enterprise 客户）开放。

OpenAI 在推文中写道：该新模型（“gpt-4-turbo-2024-04-09”）在写作、数学、逻辑推理和编码方面进行了改进，以及采用了最新更新的数据库。该模型是根据截至 2023 年 12 月的公开数据进行训练的，这与 ChatGPT 中提供的上一版 GPT-4 Turbo 形成鲜明对比，后者的截止日期为 2023 年 4 月。此外，当用户使用新版 GPT-4 Turbo 进行写作时，回答将更直接、更精简。

图7：GPT-4 Turbo 在写作、数学、逻辑推理等方面较之前版本均有提升



资料来源：X: OpenAI 推文，中国银河证券研究院

### 【Meta 确认其 Llama 3 开源大语言模型将于下个月推出】

根据 The Information 的报道。2024 年 4 月 9 日在伦敦举行的一次活动中，Meta 声称计划在下个月发布 Llama 3 的初始版本。据 Meta 首席产品官 Chris Cox，将通过 Llama 3 为 Meta 的多种产品提供支持。然而，Meta 并没有公布 Llama 3 的参数规模，预计它将拥有 1400 亿个参数，而现有的 Llama 2 只有 700 亿个参数。

### 【xAI 发布 Grok-1.5V，xAI 首款多模态大模型】

xAI 于 2024 年 4 月 12 日发布其首款多模态大模型 Grok-1.5V。Grok 除了文本功能外，现在还可以处理各种各样的视觉信息，包括文档、图表、图表、屏幕截图、照片，并能进行多学科推理。

xAI 在官网展示了 Grok-1.5V 的 7 个示例，涵盖以下内容：把手绘图表变成 Python 代码、通过食品标签来计算卡路里、从孩子画作中构思睡前故事、解释网络梗图、把表格数据转换为 csv 文件、针对日常生活问题提供解答、以及解决编程相关问题。

与 GPT-4V、Claude 3 Sonnet、Claude 3 Opus、Gemini Pro 1.5 等同类大模型相比，Grok-1.5V 的表现不相上下，在一些基准测试中优于同类。

Grok-1.5V 将很快提供给其早期测试者和现有的 Grok 用户。

图8: Grok-1.5V 与同类大模型的比较图

Benchmark	Grok-1.5V	GPT-4V	Claude 3 Sonnet	Claude 3 Opus	Gemini Pro 1.5
<b>MMMU</b> Multi-discipline	53.6%	56.8%	53.1%	59.4%	58.5%
<b>Mathvista</b> Math	52.8%	49.9%	47.9%	50.5%	52.1%
<b>AI2D</b>  Diagrams	88.3%	78.2%	88.7%	88.1%	80.3%
<b>TextVQA</b> Text reading	78.1%	78.0%	-	-	73.5%
<b>ChartQA</b> Charts	76.1%	78.5%	81.1%	80.8%	81.3%
<b>DocVQA</b> Documents	85.6%	88.4%	89.5%	89.3%	86.5%
<b>RealWorldQA</b> Real-world understanding	68.7%	61.4%	51.9%	49.8%	67.5%

资料来源: xAI 官网, 中国银河证券研究院

### 三、风险提示

技术迭代不及预期风险；科技巨头竞争加剧风险；法律监管风险；供应链风险；下游需求不及预期风险。

## 图表目录

图 1: 国债收益率 (10 年期) .....	3
图 2: 美元兑人民币汇率(中间价).....	3
图 3: MTIA v2 和 MTIA v1 的技术规格对比图.....	5
图 4: 英特尔在发布会上展示的 Gaudi 3 技术细节 .....	6
图 5: 面向中国的两款处理器技术规格.....	6
图 6: Phoenix 机器人的基本信息.....	7
图 7: GPT-4 Turbo 在写作、数学、逻辑推理等方面较之前版本均有提升.....	8
图 8: Grok-1.5V 与同类大模型的比较图 .....	9

## 表格目录

表 1: 主要股指周变动.....	3
表 2: 重点公司周数据 .....	4

## 分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

**吴砚靖**，TMT/科创板研究负责人。北京大学软件项目管理硕士，10年证券分析从业经验，历任中银国际证券首席分析师，国内大型知名PE机构研究部执行总经理。具备一二级市场经验，长期专注科技公司研究。

## 免责声明

本报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险、应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。银河证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

## 评级标准

评级标准	评级	说明
评级标准为报告发布日后的6到12个月行业指数（或公司股价）相对市场表现，其中：A股市场以沪深300指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准，北交所市场以北证50指数为基准，香港市场以摩根士丹利中国指数为基准。	行业评级	推荐：相对基准指数涨幅10%以上
		中性：相对基准指数涨幅在-5%~10%之间
		回避：相对基准指数跌幅5%以上
	公司评级	推荐：相对基准指数涨幅20%以上
		谨慎推荐：相对基准指数涨幅在5%~20%之间
		中性：相对基准指数涨幅在-5%~5%之间
	回避：相对基准指数跌幅5%以上	

## 联系

### 中国银河证券股份有限公司研究院

深圳市福田区金田路3088号中洲大厦20层

上海浦东新区富城路99号震旦大厦31层

北京市丰台区西营街8号院1号楼青海金融大厦

公司网址：www.chinastock.com.cn

### 机构请致电：

深广地区：程曦 0755-83471683chengxi\_yj@chinastock.com.cn

苏一耘 0755-83479312suyiyun\_yj@chinastock.com.cn

上海地区：陆韵如 021-60387901luyunru\_yj@chinastock.com.cn

李洋洋 021-20252671liyongyang\_yj@chinastock.com.cn

北京地区：田薇 010-80927721tianwei@chinastock.com.cn

唐嫚羚 010-80927722tangmanling\_bj@chinastock.com.cn