

英伟达 GTC 2025 召开，创新产品持续推出

2025 年 3 月 18 日

概览【03.07-03.14】

- **股指动态：**过去一周，全球主要科技股指普遍下跌，中概股科技小幅上涨。标普 500 指数-2.27%，纳斯达克综合指数-2.43%，费城半导体指数-0.67%；TAMAMA 科技指数-2.62%；纳斯达克中国金龙指数+0.37%；恒生科技指数-2.59%；计算机-1.20%。
- **个股表现：**过去一周，热门科技股普遍下跌。据统计，相比 3 月 7 日收盘价，3 月 14 日盘后，苹果合计-10.70%，英伟达+7.98%，特斯拉-4.83%，谷歌-4.70%，亚马逊-0.65%，META-2.80%，微软-1.21%，ARM-6.06%，英特尔+16.52%，高通-2.88%，AMD+0.66%。
- **10 年期国债及汇率：**3 月 14 日，美国 10 年期国债利率为 4.31%，相较 3 月 7 日累计下降 1bps；3 月 14 日，美元兑人民币中间价报 7.17；较 3 月 7 日价累计调贬 33 个基点。
- **核心观点：**

英伟达 GTC 2025 大会于 3 月 17 日至 21 日在美国加州圣何塞举行，大会亮点如下：1) 英伟达 CEO 黄仁勋的主题演讲预计将重点介绍 Blackwell Ultra GPU 和下一代 Vera Rubin 超级芯片。英伟达预计将发布新一代 Blackwell Ultra GPU，包括 GB300 和 B300 系列。这些芯片将采用台积电 4NP 工艺，HBM 容量提升至 288GB，采用 HBM3e 12 高堆叠技术，TDP 功耗从 1.2kW 激增至 1.4kW，并全面采用液冷方案。性能方面，B300 系列的 FP4 计算性能预计比 B200 高出 50%，有望在 2025 年三季度开始出货。同时可能会披露下一代 Vera Rubin 超级芯片架构的部分细节。2) 3 月 20 日，英伟达将首次推出“量子日”，聚焦量子运算的现状与未来发展，并邀请 D-Wave、IonQ 等业界领袖共同讨论量子技术的应用。尽管英伟达对量子计算持谨慎态度，预计实用化量子计算机仍需 10 到 20 年，但这一活动表明英伟达正在积极探索量子计算的未来应用。3) **AI 智能体与多模态大模型：**大会预计将展示 AI 在人形机器人和自动驾驶汽车领域的最新进展。例如，Jetson Thor 芯片专为人形机器人设计，支持复杂实时计算需求。此外，英伟达还将展示其在自动驾驶技术方面的最新成果。4) **英伟达计划推出支持 115.2Tbps 传输的 CPO 交换机 Quantum 3400 X800 IB，集成 144 个 MPO 接口，市场预计 2025 年 Q3 量产。**5) **液冷技术普及：**预计全系产品转向浸没式液冷设计，降低 40% 散热功耗。6) **中国人工智能日：**当地 3 月 18 日上午 9:30，英伟达计划将举行“中国人工智能日”（China AI Day）线上专场，汇聚国内多家云与互联网企业。英伟达通过“中国人工智能日”活动，进一步强调了其对中国市场的重视。尽管受到出口限制，但英伟达仍在积极与中国企业合作，推动 AI 技术在中国的落地和应用。

- **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。

计算机行业

推荐 维持评级

分析师

吴砚靖

☎：010-66568589

✉：wuyanqing@chinastock.com.cn

分析师登记编码：S0130519070001

国内表现

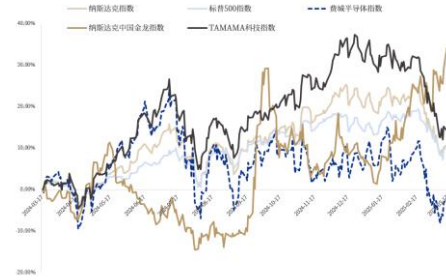
2025-03-17



资料来源：中国银河证券研究院

全球行情

2025-03-17



资料来源：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_降息改善流动性，重视港股科技投资机会

【银河计算机】全球科技动态追踪_特斯拉、谷歌等科技龙头陆续披露季报

【银河计算机】全球科技动态追踪_大选对科技的影响：特朗普与哈里斯的不同路径

目录

Catalog

一、全球市场表现	3
(一) 股市动态	3
(二) 债市及汇率情况	3
(三) 重点科技股表现	3
二、行业要闻	4
(一) 算力及终端	4
(二) 大模型及云应用	7
三、风险提示	11

一、全球市场表现

（一）股市动态

过去一周，全球主要科技股指普遍下跌，中概股科技小幅上涨。标普 500 指数-2.27%，纳斯达克综合指数-2.43%，费城半导体指数-0.67%；TAMAMA 科技指数-2.62%；纳斯达克中国金龙指数+0.37%；恒生科技指数-2.59%；计算机-1.20%。

表1：主要股指周变动

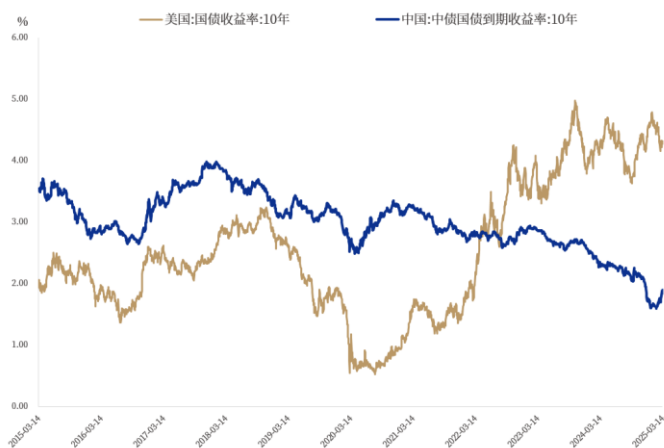
指数代码	指数简称	涨跌幅%				市盈率 PE (TTM)
		区间[03.07-03.14]	本月	本年度	2024	
SPX.GI	标普500指数	-2.27	-4.69	-3.51	23.31	25.95
IXIC.GI	纳斯达克指数	-2.43	-5.51	-7.78	28.64	38.64
SOX.GI	费城半导体指数	-0.67	-2.16	-6.34	19.27	43.45
8884057.WI	TAMAMA科技指数	-2.62	-7.36	-12.86	46.87	31.64
HXC.GI	纳斯达克中国金龙指数	0.37	9.56	24.26	4.43	24.31
HSTECH.HI	恒生科技指数	-2.59	5.47	31.44	18.70	25.43
CI005027.WI	计算机	-1.20	4.90	19.87	9.91	133.39

资料来源：WIND，中国银河证券研究院

（二）债市及汇率情况

3月14日，美国10年期国债利率为4.31%，相较3月7日累计下降1bps；3月14日，美元兑人民币中间价报7.17；较3月7日价累计调贬33个基点。

图1：美债收益率（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）重点科技股表现

过去一周，热门科技股普遍下跌。据统计，相比3月7日收盘价，3月14日盘后，苹果合

计-10.70%，英伟达+7.98%，特斯拉-4.83%，谷歌-4.70%，亚马逊-0.65%，META-2.80%，微软-1.21%，ARM-6.06%，英特尔+16.52%，高通-2.88%，AMD+0.66%。

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） [2025-03-14]	市销率 PS (TTM)	市盈率 PE (TTM)	PE(2023)	PE(2024)	PE(2025E)
				[2025-03-07]	[2025-03-14]							
算力及终端	美股	AAPL.O	苹果(Apple)	239.07	213.49	-10.70	32,070.62	8.10	33.35	30.87	40.18	29.00
		NVDA.O	英伟达(NVIDIA)	112.69	121.67	7.98	29,687.48	22.75	40.73	280.04	110.51	41.80
		TSLA.O	特斯拉(TESLA)	262.67	249.98	-4.83	8,040.65	8.23	113.39	52.67	183.15	74.59
		HPQ.N	惠普(HP)	30.54	28.41	-6.02	267.82	0.50	9.85	9.14	11.03	0.00
		CSCO.O	思科(CISCO)	63.94	60.50	-5.38	2,406.87	4.44	26.20	16.28	22.85	26.19
		ASML.O	阿斯麦	732.22	714.00	-2.49	2,811.95	9.50	35.47	34.34	33.90	27.75
		AMD.O	超威半导体(AMD)	100.31	100.97	0.66	1,636.20	6.35	99.71	278.85	119.39	36.94
		INTC.O	英特尔(INTEL)	20.64	24.05	16.52	1,041.37	1.96	-5.55	125.43	-4.63	47.03
		QCOM.O	高通(QUALCOMM)	161.22	156.58	-2.88	1,731.77	4.26	16.41	22.36	16.75	15.77
		ARM.O	ARM	125.55	117.94	-6.06	1,243.09	33.65	154.23	147.03	424.91	146.04
	台股	ON.O	安森美半导体 (ON SEMICONDUCTOR)	45.40	43.19	-4.87	182.01	2.57	11.57	16.47	16.96	13.82
		0909.HK	明源云	4.34	3.66	-15.67	71.15	4.03	-17.19	-8.66	0.00	0.00
		9698.HK	万国数据-SW	36.30	32.75	-9.78	499.25	4.10	-11.30	-2.92	0.00	-68.69
		1686.HK	新意网集团	8.60	7.91	-8.02	186.42	6.53	19.51	7.98	11.01	27.82
互联网	美股	2330.TW	台积电	1,005.00	959.00	-4.58	294,956.49	8.59	21.20	18.05	23.76	16.46
		2454.TW	联发科	1,465.00	1,400.00	-4.44	22,423.27	4.22	21.08	21.09	21.30	0.00
		GOOGL.O	谷歌(ALPHABET)-A	173.86	165.49	-4.70	20,290.32	5.76	20.15	23.69	23.09	18.39
		AMZN.O	亚马逊(AMAZON)	199.25	197.95	-0.65	20,978.21	3.29	35.41	51.61	39.22	30.06
		META.O	脸书 (META PLATFORMS)	625.66	607.60	-2.80	15,394.51	9.36	24.69	23.27	23.79	23.44
		NFLX.O	奈飞(NETFLIX)	891.11	918.00	3.02	3,926.81	10.07	45.08	39.40	43.77	38.18
		PDD.O	拼多多	119.92	122.54	2.18	1,701.80	3.20	11.01	22.94	0.00	9.53
		NTES.O	网易	102.78	102.80	0.02	666.78	4.51	15.98	14.46	13.87	14.70
		BIDU.O	百度	94.80	93.82	-1.03	330.80	1.78	9.95	14.52	8.94	9.31
		TCOM.O	携程网	63.83	62.03	-2.82	423.03	5.72	17.86	17.58	19.77	19.04
	港股	BABA.N	阿里巴巴	140.62	141.10	0.34	3,321.30	2.45	20.02	18.64	17.91	19.22
		9988.HK	阿里巴巴-SW	140.00	135.80	-3.00	25,813.18	2.43	19.86	18.53	17.79	18.16
		0700.HK	腾讯控股	533.50	521.50	-2.25	47,867.56	6.71	25.43	21.90	0.00	21.36
		80700.HK	腾讯控股-R	497.60	487.60	-2.01	44,183.19	6.96	26.36	21.86	0.00	21.57
		9999.HK	网易-S	161.60	161.40	-0.12	5,182.20	4.56	16.16	13.96	13.86	14.56
		9888.HK	百度集团-SW	90.80	91.65	0.94	2,570.98	1.79	10.02	14.48	9.04	8.62
		89888.HK	百度集团-SWR	84.65	85.35	0.83	2,373.10	1.80	10.08	14.50	9.19	8.69
		1024.HK	快手-W	63.70	64.90	1.88	2,792.99	2.04	16.82	32.60	0.00	13.10
		81024.HK	快手-WR	59.30	60.50	2.02	2,578.02	2.11	17.39	32.58	0.00	13.22
		9626.HK	哔哩哔哩-W	181.50	166.30	-8.37	700.56	2.42	-48.17	-7.41	-41.13	96.59
		2518.HK	汽车之家-S	58.50	57.50	-1.71	292.90	3.85	16.75	12.86	14.02	5.72
		9898.HK	微博-SW	85.50	83.55	-2.28	203.29	1.49	8.71	7.52	7.79	6.98
软件及应用	美股	MSFT.O	微软(MICROSOFT)	393.31	388.56	-1.21	28,885.48	11.03	31.14	38.62	35.56	30.38
		SNOW.N	SNOWFLAKE	158.91	156.11	-1.76	515.32	14.21	-40.08	-82.25	-60.96	0.00
		ORCL.N	甲骨文(ORACLE)	155.16	149.27	-3.80	4,185.88	7.50	34.42	34.08	44.53	36.30
		CRM.N	赛富时(SALESFORCE)	282.89	279.40	-1.23	2,685.03	7.09	43.33	1224.61	77.36	0.00
		ADBE.O	奥多比(ADOBE)	449.40	394.74	-12.16	1,716.71	7.79	25.43	50.01	35.27	24.99
		INTU.O	财捷(INTUIT)	609.61	598.80	-1.77	1,674.02	9.75	55.12	73.39	59.40	47.22
		SNPS.O	新思科技(SYNOPSYS)	450.88	446.23	-1.03	689.96	11.36	32.70	63.63	33.15	0.00
		CDNS.O	铿腾电子(CADENCE)	243.21	247.43	1.74	678.23	14.61	64.26	71.17	77.96	0.00
		ADSK.O	欧特克(AUTODESK)	259.75	254.43	-2.05	541.94	8.84	48.74	63.29	70.14	0.00
		U.N	Unity	24.19	20.77	-14.14	85.09	4.69	-12.81	-18.90	-13.85	-18.54
	港股	0020.HK	商汤-W	1.75	1.71	-2.29	632.83	15.55	-10.00	-5.46	0.00	-20.05
		80020.HK	商汤-WR	1.63	1.60	-1.84	584.12	15.94	-10.25	-5.46	0.00	-20.31
		3888.HK	金山软件	45.50	41.15	-9.56	549.36	5.06	38.20	61.69	0.00	26.12
		0268.HK	金蝶国际	17.06	15.58	-8.68	552.71	8.32	-349.81	-178.41	-199.39	279.31
		9878.HK	汇通达网络	17.14	16.50	-3.73	92.82	0.12	25.81	34.57	0.00	12.58
		3650.HK	KEEP	6.29	5.28	-16.06	27.76	1.16	-10.03	6.00	0.00	148.50
		0354.HK	中国软件国际	6.43	6.06	-5.75	165.56	0.90	23.32	22.39	0.00	17.45
		1357.HK	美图公司	5.54	6.00	8.30	273.51	8.21	54.97	38.62	0.00	31.99
		3896.HK	金山云	9.02	8.80	-2.44	334.87	4.15	-14.70	-3.18	0.00	-53.59
		2013.HK	微盟集团	2.47	2.36	-4.45	85.33	4.12	-9.09	-9.62	0.00	-75.71
		1675.HK	亚信科技	11.34	9.82	-13.40	91.98	1.28	15.63	13.69	10.05	12.13
		2121.HK	创新奇智	5.77	3.99	-30.85	22.55	1.47	-3.53	-8.15	0.00	-4.45
		2400.HK	心动公司	38.60	38.45	-0.39	188.83	4.47	540.77	-53.48	0.00	16.24
		0777.HK	网龙	12.68	12.26	-3.31	65.13	0.88	13.21	10.84	0.00	0.00

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

【英伟达 GTC 2025 大会于 3 月 17 日至 21 日在美国加州圣何塞举行，大会亮点如下】

英伟达将于 GTC2025 发布新一代 Blackwell Ultra GPU

在 GTC 2025 大会上，英伟达 CEO 黄仁勋将于当地时间 3 月 18 日上午十点（北京时间 3 月 19 日凌晨一点）带来主题演讲，预计重点介绍新一代 Blackwell Ultra GPU，包括 GB300 和 B300 系列。Blackwell 以极高的算力密度和能效比为核心，旨在支持大规模 AI 模型训练和推理需求。本系列芯片将采用台积电 4NP 工艺，HBM 容量提升至 288GB，采用 HBM3e 12 高堆叠技术，TDP 功耗从 1.2kW 增加至 1.4kW，并全面采用液冷方案。性能方面，B300 系列的 FP4 计算性能预计比 B200 高出 50%，计划于 2025 年第三季度开始出货。此外，黄仁勋还将披露下一代 Vera Rubin 超级芯片架构的部分细节，展示了英伟达在高性能计算领域的持续创新。

英伟达首次举办“量子日”，探讨量子计算未来

今年的 GTC 大会上，英伟达首次设立“量子日”，定于当地时间 3 月 20 日举行，聚焦量子计算的现状与未来发展。在“量子日”活动中，英伟达 CEO 黄仁勋将与 D-Wave Quantum 和 Rigetti Computing 等十余家量子计算行业领军企业的高管同台，讨论量子计算的技术现状、潜力以及未来发展方向。目前，市场对“量子日”活动寄予厚望，认为其可能会对量子计算相关产业产生积极的影响。对此，英伟达表示，将在 GTC 大会上宣布该公司在量子计算领域取得的进展，这些进展将“缩短量子应用实用化的时间”。尽管英伟达对量子计算持谨慎态度，预计量子计算机的实用化仍需 10 到 20 年，但这一活动表明英伟达正在积极探索量子计算的未来应用，致力于推动这一前沿领域的发展。

英伟达预计分享 AI 智能体与多模态大模型领域的最新进展

在本次 GTC 大会上，英伟达有望展示下一代 Rubin 平台技术创新，该平台预计采用台积电 N3 工艺，配置 8 个 HBM4 立方体，总容量 384GB，比 Blackwell Ultra 增加 33%。Rubin Ultra 也有望于 2027 年上市，可能会进一步突破 GPU 设计的界限，可能包含 12 个 HBM4E 内存堆栈，可能提供高达 576GB 的总内存。大会期间，英伟达或将展示了 AI 在人形机器人和自动驾驶汽车领域的最新进展。例如，专为人形机器人设计的 Jetson Thor 芯片，支持复杂的实时计算需求，提升了机器人在动态环境中的适应能力。此外，英伟达还展示了其在自动驾驶技术方面的最新成果，强调了多模态大模型在感知和决策中的应用，推动了智能交通的发展。

英伟达计划推出 CPO 交换机，助力 AI 计算发展

英伟达计划推出三款 CPO(共封装光学)交换机，总交换容量分别达到 115.2T、204.8T 和 409.6T，集成 144 个 MPO 接口，进一步降低信号延迟与能耗。CPO 技术通过将光模块与芯片封装集成，显著提升带宽密度并降低功耗和延迟，满足 AI 集群对高速互联的需求。这一技术有望加速超大规模数据中心的网络架构升级。英伟达或将于 GTC 上展示 Blackwell Ultra 如何利用 CPO 技术优化其 NVLink 或 Infiniband 互连系统，从而支持超大规模 AI 计算的需求。该产品预计于 2025 年第三季度量产，标志着英伟达在高性能网络设备领域的又一重要突破。

英伟达全面普及液冷技术，降低散热功耗

高性能计算设备的散热挑战日益突出，例如本次预计推出的 B300GPU 系列产品，其热设计功率（TDP）从上一代的 1.2kW 提升至 1.4kW，对散热系统提出更高要求。为此，英伟达计划在 GB300 全系列产品中全面应用液冷技术，其中冷板式方案成为当前主流，而浸没式液冷技术被视为长期发展方向，预计可将散热功耗降低 40%。这一技术升级不仅解决高功耗芯片的散热难题，也为数据中心能效优化提供新路径，同时还体现了英伟达对可持续发展的承诺，对全球追求绿色科技的响应。

英伟达举办“中国人工智能日”，尽显对中国市场的重视

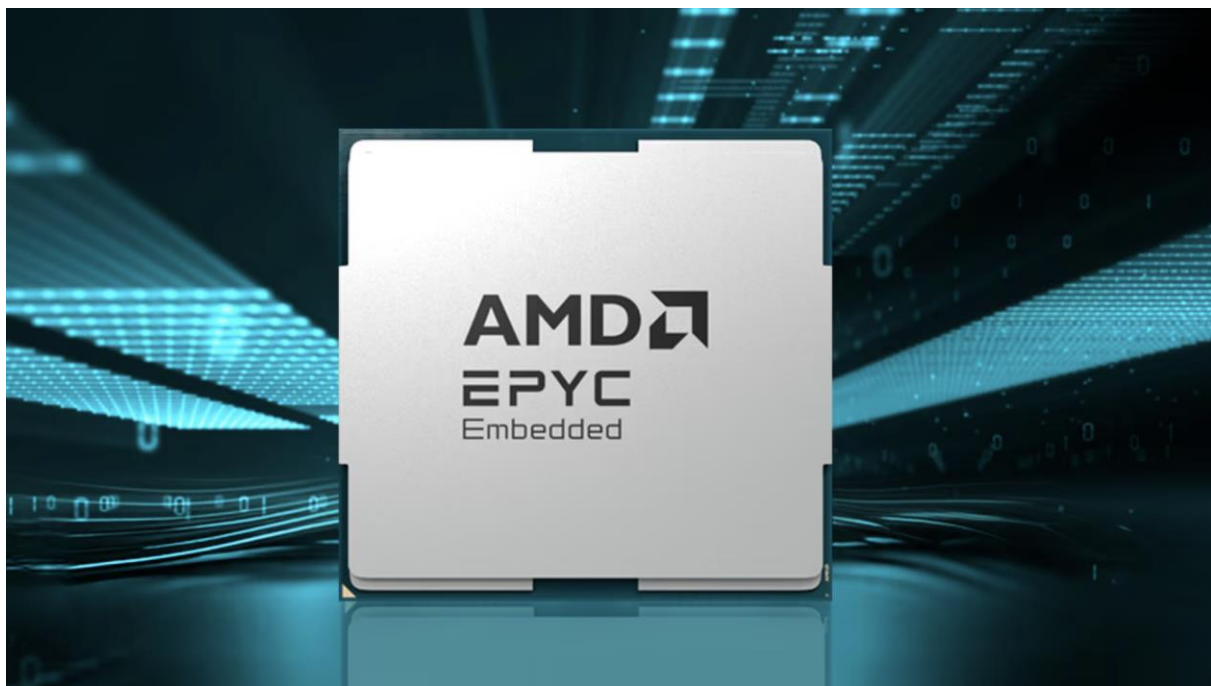
今年的 GTC 大会上，英伟达首次举行了“中国人工智能日”（China AI Day）线上专场，于北京时间 3 月 18 日上午 9:30 开始直播。本次专场汇聚了包括字节跳动、阿里云、百度等在内的多家中国云与互联网企业。与会者分享了在大语言模型、多模态大语言模型、数据科学等领域的前沿进展，展示了如何通过软硬件协同优化实现生产级 AI 的性能和效率提升。英伟达 CEO 黄仁勋表示，中国市场营收占比稳定在出口管制前的一半，未来有望随技术合作深化实现环比增长。通过这一活动，英伟达进一步强调了其对中国市场的重视，推动 AI 技术在中国的落地和应用。

【AMD 推出第五代 EPYC 嵌入式处理器】

3 月 11 日，AMD 正式宣布推出第五代 AMD EPYC（霄龙）嵌入式处理器，扩展其 x86 嵌入式处理器产品组合。本系列针对嵌入式市场进行了优化，在尖端计算能力与专用嵌入式功能之间实现了平衡，从而提高了产品寿命、可靠性、系统弹性和嵌入式应用程序开发的简易性。

本系列处理器采用“Zen 5”架构，可提供更高的吞吐量和更高的能效。与其他产品对比，插槽吞吐量提高 1.3 倍，每瓦性能提高 1.3 倍。且每个插槽的 DDR5 内存容量高达 6TB，I/O 连接性得到扩展，支持多达 160 个 PCIe Gen5 通道，可实现存储容量扩展和高速数据传输。本系列处理器在单个插槽中可支持 8 到 192 个核心，将网络和存储工作负载的数据处理吞吐量提高 1.3 倍和 1.6 倍。此外，AMD 承诺，本系列产品将延长供应支持至七年，以满足客户的长期需求。该系列产品还配备 NTB（非透明桥接）技术，在两个处理器包之间实现跨不同内存域的高性能数据传输。

图3：第五代 AMD EPYC 嵌入式处理器



资料来源：AMD 官网，中国银河证券研究院

【英特尔扩展 Tiber 信任服务组合，积极防范安全风险】

3 月 12 日，英特尔宣布扩展 Intel Tiber 信任服务组合，旨在帮助企业在日益复杂的现代应用架构中提升安全性，并满足不断变化的监管需求。此次扩展增强了对平台生命周期完整性、联邦 AI 训练、零信任安全和供应链安全的支持，为企业提供更加完善的数据和系统保护。

新功能包括平台生命周期监控，帮助企业识别和减轻系统生命周期内的风险；基于 OpenFL 的安全联邦 AI 训练技术，确保 AI 模型训练过程中数据隐私和合规性；此外，还强化了零信任安全验证，提高硬件、软件 and 数据的完整性，使企业能够更可靠地管理自身系统。英特尔表示，将始终致力

于提供从芯片级别开始的安全防护，并与行业领先的技术合作伙伴携手，在不断变化的网络安全环境中不断提升安全防护能力。此次扩展在 AI 发展和数据安全性方面创造了新的机遇，确保了系统和数据的长期可靠性。

图4：英特尔 Tiber 信任服务组合概念图



资料来源：英特尔官网，中国银河证券研究院

【ASML 与 imec 签署五年战略合作协议，共同推动欧洲半导体技术创新】

3 月 11 日，ASML 宣布与比利时微电子研究中心（imec）签署为期五年的战略合作协议，旨在推进半导体行业的发展，并推动可持续创新。

此次合作将汇集双方在先进芯片制造领域的专业知识，重点开发高端节点技术，包括 0.55 NA EUV（极紫外光刻）、0.33 NA EUV、DUV（深紫外光刻）浸没、YieldStar 光学计量以及 HMI 单束和多束技术。以上技术将安装在 imec 最先进的试验线中，并纳入由欧盟及比利时政府政府资助的 NanoIC 试验生产线，为国际半导体生态系统提供最先进的 2nm 以下工艺的研发基础设施。此外，合作还将拓展至硅光子学、存储器和先进封装等领域，为人工智能及其他高科技市场提供创新解决方案。本次合作涉及多个欧洲国家的联合资助，致力于推动欧洲在全球芯片制造领域的领先地位。

（二）大模型及云应用

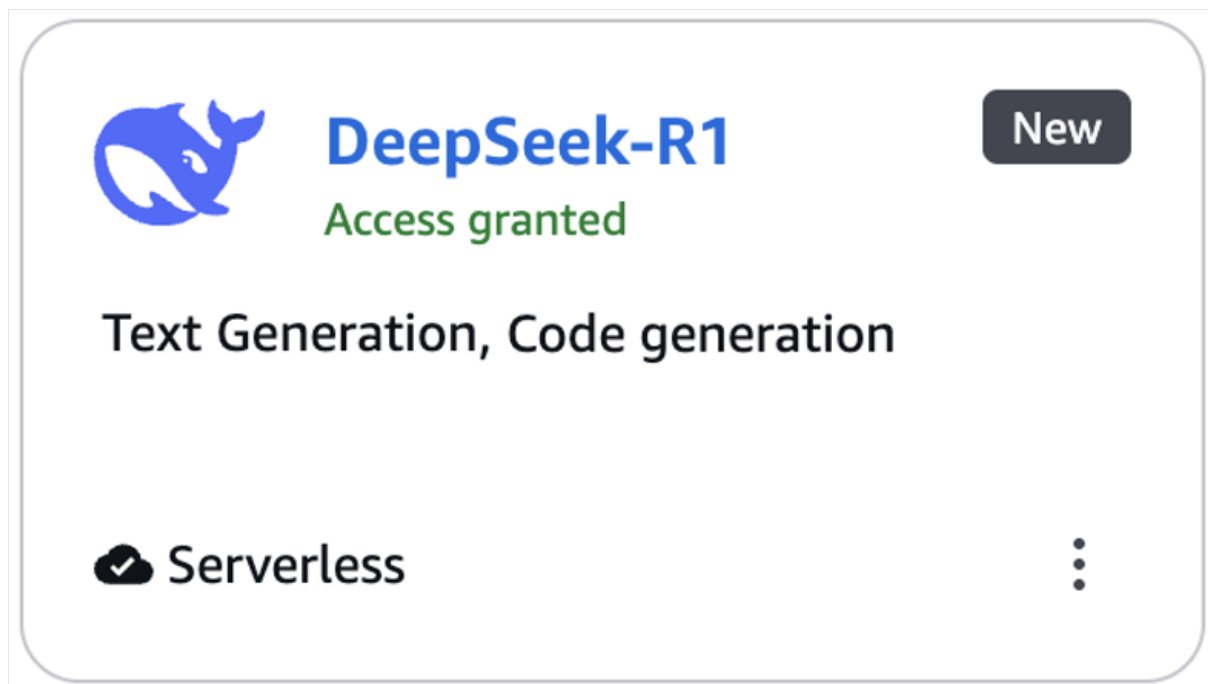
【AWS 正式上线 DeepSeek-R1，提供全面托管 AI 模型服务】

3 月 10 日，亚马逊云服务（AWS）宣布，DeepSeek-R1 模型已正式在 Amazon Bedrock 上全面上线，成为首家提供完全托管模型的云服务提供商。DeepSeek-R1 由人工智能初创公司 DeepSeek 开发，以其卓越的推理能力和深度理解能力闻名，能高效处理复杂任务，并在成本效益方面远优于同类模型。

通过 Amazon Bedrock，AWS 客户能够轻松访问并部署 DeepSeek-R1 及其精简版本。无需关心任何技术维护问题，即可用于解决复杂问题、编写代码、处理数据等。除此之外，AWS 还为客户提供了使用自定义模型的机会。客户可以上传经过自己调试的 DeepSeek-R1 Distilled Llama 变体版本，并通过 Amazon Bedrock Custom Model Import 以完全托管的方式运行它们。该功能使客

户能够通过单一 API 导入并使用自定义模型，实现与现有模型的一同运行。AWS 表示，DeepSeek-R1 的加入不仅降低了推理成本，还提升了客户在生成式 AI 领域的应用能力。

图5: AWS 正式上线 DeepSeek-R1



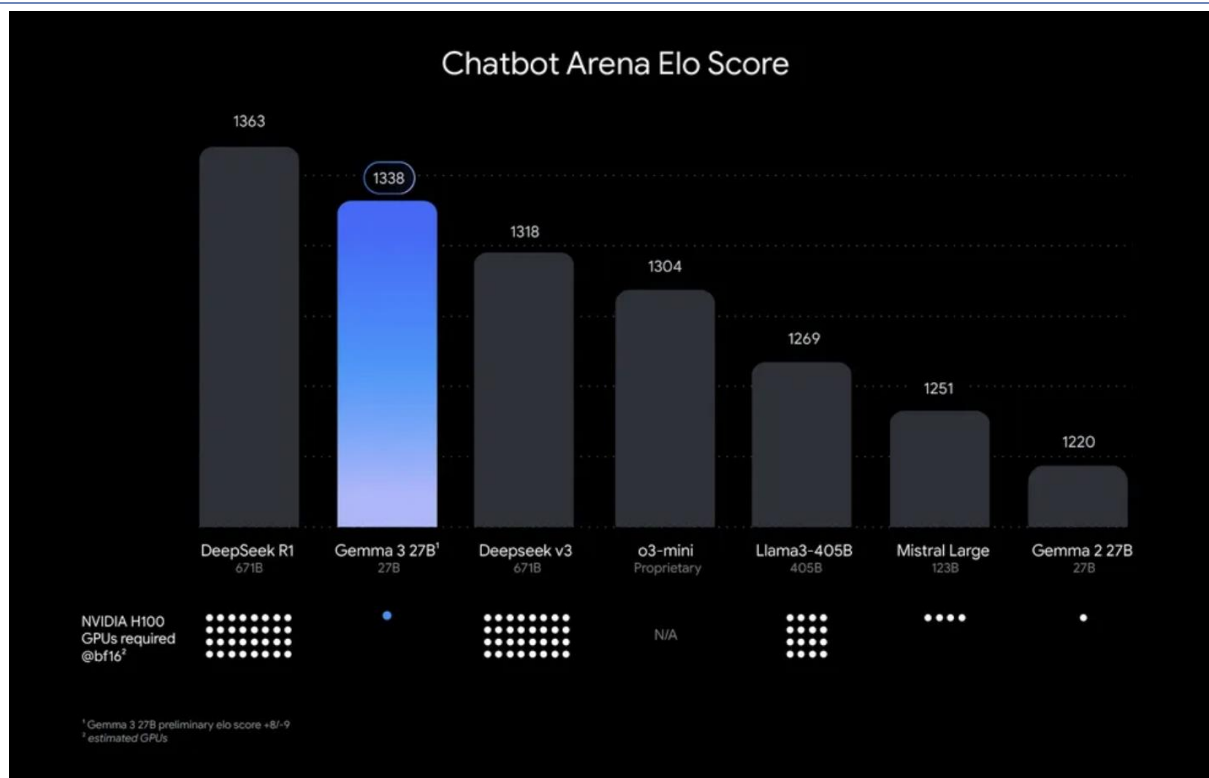
资料来源: AWS 官网, 中国银河证券研究院

【谷歌推出最新 AI 模型 Gemma 3，为用户提供更强大的开放模型】

3 月 12 日，谷歌宣布正式推出 Gemma 3。这是一款最新的轻量级开放模型，基于与 Gemini 2.0 相同的研究技术构建而成，旨在实现直接在各种设备上高速运行 AI，被称作单 GPU/TPU 运行的最强模型。Gemma 3 具备更强的计算性能、更长的上下文窗口，并且提供多个不同规模的模型（1B、4B、12B 和 27B），使用户可以根据不同需求选择最合适的模型。Gemma 3 主要特点如下：拥有卓越的单加速器性能，在基准测试中优于 DeepSeek-V3 等已推出的 AI 模型；支持 140 多种语言，增强全球适用性；提供 128k-token 上下文窗口，可处理更长文本和复杂推理；支持函数调用和结构化输出，提升 AI 任务自动化能力。此外，谷歌还推出了官方量化版本，确保模型在低计算资源环境下依然保持高性能。

为增强 AI 安全性，谷歌同时发布了 ShieldGemma 2。其基于 Gemma 3 构建而成，是一款拥有 40 亿参数的图像安全检测模型，可识别有害、暴力等内容，为用户在使用中进行更安全的图像输入与输出。

图6: AI 模型 Chatbot Arena Elo 得分排名: Gemma 3 27B 排名靠前, 且只需要一个 GPU



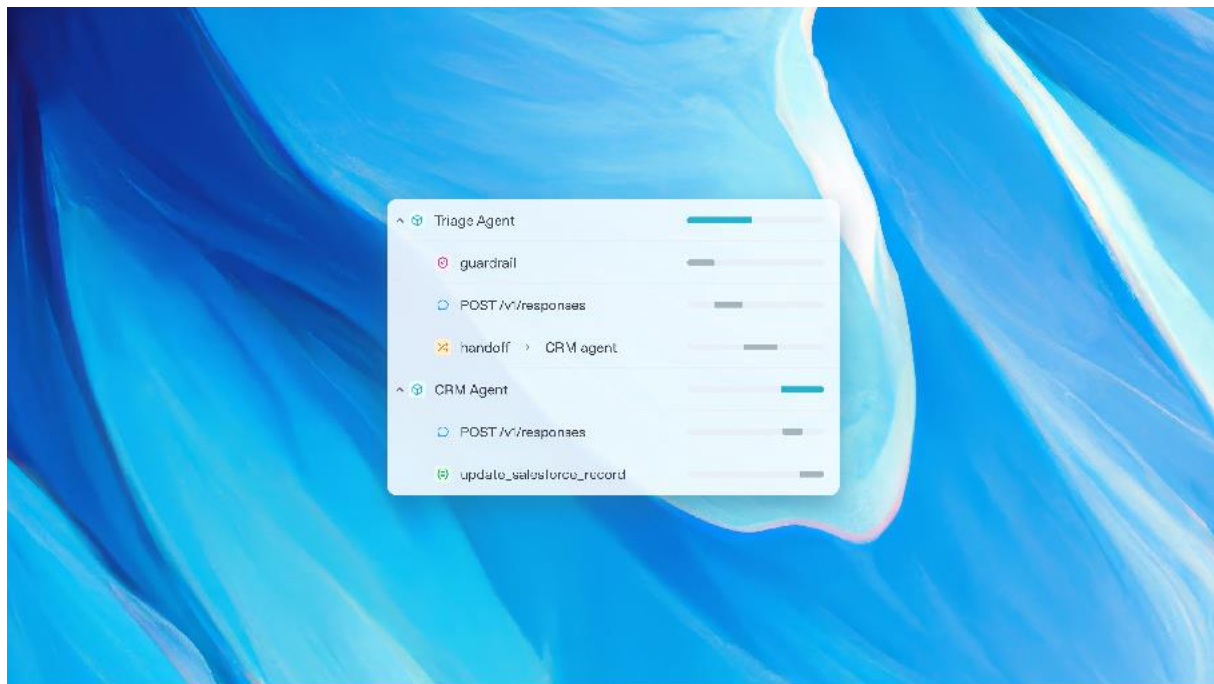
资料来源: 谷歌官网, 中国银河证券研究院

【OpenAI 发布全新 Agent 工具包, 为 AI Agent 注入新活力】

近日, OpenAI 推出了一系列专为 AI Agents(智能体)设计的全新工具和 API, 旨在优化 AI Agent 的核心逻辑、编排和交互, 使开发者更轻松构建和部署 AI Agent。主要更新包括全新的 Responses API、内置工具支持、Agents SDK 与集成可观测性工具。其中, Responses API 专为 Agent 应用设计, 是结合了 Chat Completions API 的简便性和 Assistants API 的工具调用能力的新型 API 接口。内置工具支持包括网页搜索、文件搜索和计算机操作, 增强 AI Agent 的信息检索和数据处理能力, 让 AI 做到像人一样操作电脑完成任务。Agents SDK 基于 Swarm 框架升级而成, 支持单代理和多代理协作, 能够高效执行复杂任务。集成可观测性工具用于跟踪和监测代理工作流, 帮助开发者优化 AI 代理的执行过程。

这些新工具大幅降低了 AI Agent 的开发门槛, 简化了从逻辑构建到任务编排的全过程。未来几周和几个月内, OpenAI 计划推出更多工具和功能, 以进一步加速和优化平台上的 AI Agent 的开发。

图7: Open AI 发布的一系列新功能



资料来源: Open AI 官网, 中国银河证券研究院

【三星携手英伟达推进 AI-RAN 技术发展, 助力移动网络智能化】

近日, 三星电子宣布与英伟达 (NVIDIA) 合作, 推进 AI-RAN (人工智能无线接入网络) 技术发展。此次合作旨在通过扩展 CPU 和 GPU 计算平台, 加速 AI 在移动网络中的应用, 提高网络效率和智能化水平。

自 2024 年初以来, 三星利用其无线通信和 AI 技术, 在 AI-RAN 领域取得突破。其中, 三星符合 O-RAN 标准的 vRAN (虚拟化无线接入网络) 已成功与英伟达加速计算平台进行互操作测试, 证实了 AI 在软件网络中的应用潜力。基于这一技术基础, 三星能够无缝交付 AI-RAN 解决方案, 即将其 vRAN 分布式单元 (vDU) 与英伟达的加速计算技术相结合, 并集成至安装了三星 vRAN 软件的商用服务器 (COTS), 从而实现高效的 AI-RAN 部署。三星表示, 三星正积极打造 AI 驱动的智能网络, 与英伟达的合作将进一步推动 AI 赋能下一代网络发展, 助力移动通信向智能化、自动化迈进。

【英伟达与微软合作开发神经着色技术, 将 AI 融入游戏渲染】

3 月 13 日, 英伟达在游戏开发者大会召开前夕, 宣布对 RTX 神经渲染技术进行突破性升级, 并与微软合作, 为 Microsoft DirectX 引入神经着色 (Neural Shading) 支持, 将 AI 深度融入游戏图形渲染。该功能将于 4 月发布的 DirectX 预览版中推出, 使游戏开发者能够利用 NVIDIA GeForce RTX GPU 的 AI Tensor Cores, 在游戏的图形渲染管线内加速神经网络的运行。

神经着色技术将 AI 与传统渲染结合, 实现帧率大幅提升、画质增强和系统资源占用减少, 开启图形编程的新篇章。微软表示, 将在 DirectX 和 HLSL 中引入协作向量支持, 让神经渲染成为游戏行业的重要组成部分, 同时解锁 Tensor Cores 的潜力, 为 Windows 平台带来更沉浸的游戏体验。此外, 英伟达宣布 RTX Kit 迎来重大更新, 为游戏开发者提供完整的神经渲染技术方案, 并引入对 Unreal Engine 5 的支持, 同时, NVIDIA RTX Remix 开源平台正式发布, 使 Mod 制作者能够对经典游戏进行高质量重制。

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。

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，并不构成对客户的投资咨询建议，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，香港市场以恒生指数为基准。	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 0755-83471683 chengxi_yj@chinastock.com.cn

苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn

上海地区：陆韵如 021-60387901 luyunru_yj@chinastock.com.cn

李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn

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

褚 颖 010-80927755 chuying_yj@chinastock.com.cn