

## 海外AIGC持续突破，Gemini 1.5 Pro、OpenAI Sora 引爆AI视频生成赛道

计算机行业

推荐

维持评级

摘要 (2024.02.02-2024.02.16|美东) :

**股指动态：**过去两周，美股、港股及台股全线上涨。标普 500 指数合计上涨 0.95%，纳斯达克综合指数上涨 0.94%，TAMAMA 科技指数上涨 0.34%；费城半导体指数上涨 4.40%；纳斯达克中国金龙指数上涨 9.74%；恒生科技指数上涨 9.82%。

**美股科技：**热门科技股大部分上涨。据统计，相比 2 月 2 日收盘价，2 月 16 日盘后，苹果下跌 1.78%，英伟达上涨 9.75%，特斯拉上涨 6.41%，谷歌下跌 4.18%，亚马逊下跌 1.31%，META 下跌 0.35%，微软下跌 1.56%，ARM 上涨 80.76%，英特尔上涨 2.44%，高通上涨 7.76%，AMD 下跌 2.13%。

**10 年期国债及汇率：**过去两周，美国 10 年期国债利率由 4.03% 上涨至 4.30%，累计上浮 27bps；上周，中国 10 年期国债利率由 2.42% 上涨至 2.43%，累计上浮 0.58bps。2 月 9 日，美元兑人民币中间价报 7.10；较 2 月 2 日价 7.10，累计调贬 30 个基点。

### 核心观点

2 月 15 日谷歌发布 Gemini 1.5 Pro，上下文窗口容量扩展至 100 万个 tokens。据 Medium，Gemini 1.5 Pro 参数量为 175B，相比于上一代 Gemini Pro 参数量 600B 显著降低。Gemini 1.5 Pro 的参数效率提升，可减少对 LLM 微调及知识注入的需要。预计模型训练将产生协同效应，进行大规模模型推理时，微调成本有望大幅降低。**两大海外巨头加码 AI 视频赛道：**OpenAI 发布可实现文生视频的生成式 AI 模型 Sora，目前 Sora 可完成最长约 60S 的文生视频，效果惊艳；Meta 紧随其后发布非生成视频模型 V-JEPA，迈向 AMI（先进机器学习）愿景的下一步。

我们认为，在 AIGC 领域，难度排序由易到难依次是文本、图像、音频、视频。Gemini 1.5 pro 以及 Sora 发布标志着 AGI 时代正加速到来，二者均在“文生视频”领域取得里程碑式突破。未来随着模型规模增大，参数将继续放大，多模态大模型陆续推出将引发对算力需求的指数级提升，包括训练和推理，并且在影视制作、游戏开发、教育培训、广告及社交等场景应用也将带来颠覆式革命。

### 风险提示

AI 技术研发不及预期风险；巨头竞争加剧风险；法律监管风险；供应链风险；消费需求不及预期风险。

分析师

吴砚靖

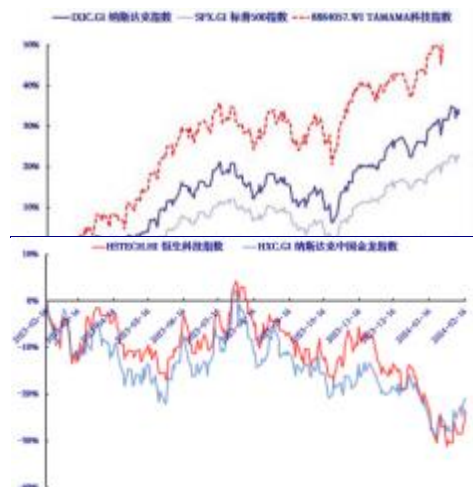
☎：(8610) 66568589

✉：wuyanqing@chinastock.com.cn

分析师登记编码：S0130519070001

美股数据

2024-02-16



港股台股数据

2024-02-16

资料来源：Wind，中国银河证券研究院

相关研究

【银河计算机】海外科技追踪周报\_Meta2024Q1 收入指 引超预期， 发布开源大模型 CodeLlama70B

【银河计算机】海外科技追踪周报\_英特尔 2024Q1 指 引低于预期，SK 海力士 2023Q4 扭亏

【银河计算机】海外科技追踪周报\_CES2024 即将开展， AIPC 以至 “AIforAll” 将成亮点

## 目 录

一、海外市场表现.....	3
(一) 海外股市动态.....	3
(二) 美债及汇率情况.....	3
(三) 主要公司财务指标.....	3
二、本周热点内容.....	4
(一) 算力及终端.....	4
(二) 软件及应用.....	7
(三) 互联网.....	11
三、海外季度财报总结.....	12
(一) 铿腾电子 2023 年业绩实现增长，2024Q1 指引未及预期.....	12
(二) Arm 2024 Q3 业绩超预期，全年利润预期大幅上调.....	13
(三) 亚马逊下季度业绩预期上调，AI 推动 AWS 保持增长.....	14
四、风险提示.....	15

## 一、海外市场表现

### (一) 海外股市动态

表 1：海外主要股市指数双周变动

指数代码	指数简称	涨跌幅%					市盈率 PE (TTM)
		本周	上周	双周	本年度	2023	
IXIC.GI	纳斯达克指数	-1.34	2.31	0.94	5.09	43.42	41.17
SPX.GI	标普500指数	-0.42	1.37	0.95	4.94	24.23	25.56
8884057.WI	TAMAMA科技指数	-2.52	2.93	0.34	8.02	67.81	35.94
HXC.GI	纳斯达克中国金龙指数	4.31	5.20	9.74	-6.03	-3.39	20.69
SOX.GI	费城半导体指数	-0.87	5.32	4.40	8.44	64.90	48.26
HSTECH.HI	恒生科技指数	6.89	2.74	9.82	-11.20	-8.83	17.96
CI005027.WI	计算机	0.00	5.92	5.92	-22.74	8.90	66.84

资料来源：Wind，中国银河证券研究院

### (二) 美债及汇率情况

过去两周，美国 10 年期国债利率由 4.03% 上涨至 4.30%，累计上浮 27bps；上周，中国 10 年期国债利率由 2.42% 上涨至 2.43%，累计上浮 0.58bps。

2 月 9 日，美元兑人民币中间价报 7.10；较 2 月 2 日价 7.10，累计调贬 30 个基点。

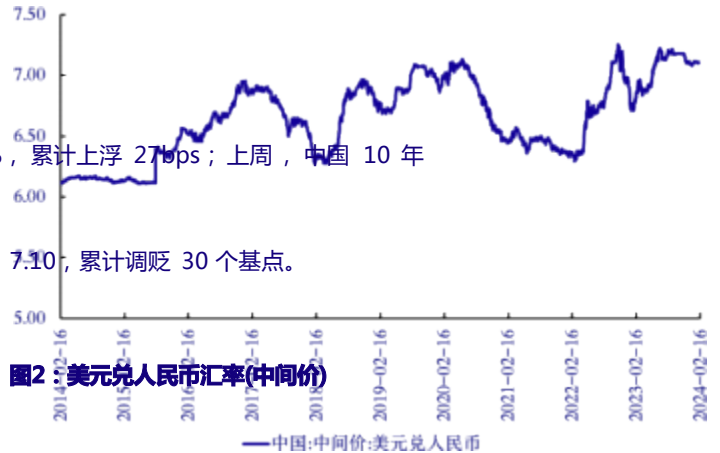


图 1：国债收益率（10 年期）



资料来源：Wind，中国银河证券研究院

资料来源：Wind，中国银河证券研究院

### (三) 主要公司财务指标

请务必阅读正文最后的中国银河证券股份有限公司免责声明。

**表 2：美股、港股、台股主要科技公司双周数据**

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) [2024.02.16]	市净率 PS (TTM)	市盈率 PE (TTM)	PE(2022)	PE(2023)	PE(2024E)
				[2024.02.02]	[2024.02.16]							
算力及终端	美股	AAPLO	苹果(Apple)	185.85	182.31	-1.78	28,152.09	7.30	27.90	20.71	30.87	27.64
		NVDA.O	英伟达(NVIDIA)	661.60	726.13	9.75	17,935.41	39.97	94.95	82.30	0.00	69.53
		TSLA.O	特斯拉(TESLA)	187.91	199.95	6.41	6,367.99	6.58	42.46	30.98	52.67	53.70
		HPQ.N	惠普(HP)	28.68	28.58	-0.35	283.20	0.53	8.68	8.24	9.14	0.00
		CSCO.O	思科(CISCO)	50.18	48.44	-3.47	1,968.35	3.44	14.64	16.57	16.28	0.00
		ASML.O	阿斯麦	890.66	928.94	4.48	3,665.50	11.99	42.14	36.70	34.34	45.26
		AMD.O	超威半导体(AMD)	177.66	173.87	-2.13	2,809.37	12.39	328.97	79.12	278.85	70.12
		INTC.O	英特尔(INTEL)	42.60	43.51	2.44	1,839.60	3.39	108.92	13.61	125.43	48.98
		QCOM.O	高通(QUALCOMM)	141.69	152.69	7.76	1,704.02	4.70	21.95	9.51	22.36	16.99
		ARM.O	ARM	71.00	128.34	80.76	1,319.43	44.91	1,552.27	0.00	147.03	0.00
		ON.O	安森美半导体 (ON SEMICONDUCTOR)	70.83	78.64	11.03	336.05	4.07	15.39	14.18	16.47	15.84
港股	0909.HK	明源云	2.04	2.33	14.22	45.26	2.43	-4.56	-10.65	0.00	-15.07	
	9698.HK	万国数据-SW	5.24	7.06	34.73	107.62	1.01	-7.27	-21.99	0.00	-6.35	
	1686.HK	新意网集团	2.72	2.75	1.10	64.32	2.74	7.10	11.66	7.98	0.00	
台股	2330.TW	台积电	635.00	683.00	7.56	203,454.03	8.19	21.12	11.71	18.34	17.49	
	2454.TW	联发科	936.00	970.00	3.63	15,516.35	3.72	22.23	8.46	0.00	0.00	
互联网	美股	GOOGL.O	谷歌(ALPHABET)-C	142.38	140.52	-1.31	17,542.58	5.68	23.67	19.04	23.69	20.60
		AMZN.O	亚马逊(AMAZON)	171.81	169.51	-1.34	17,607.65	3.06	57.87	-314.82	51.61	40.81
		META.O	脸书 (META PLATFORMS)	474.99	473.32	-0.35	12,066.84	8.94	30.86	13.75	23.27	25.55
		NFLX.O	奈飞(NETFLIX)	564.64	583.95	3.42	2,527.10	7.49	46.73	29.21	39.40	35.91
		PDD.O	拼多多	124.59	135.26	8.56	1,797.05	6.50	27.93	22.77	0.00	18.04
		NTE.S.O	网易	98.83	107.76	9.04	706.82	4.91	18.62	16.34	0.00	16.39
		BIDU.O	百度	102.79	107.12	4.21	382.24	2.03	11.86	36.42	0.00	12.24
		TCOM.O	携程网	37.00	41.25	11.49	284.01	5.16	18.96	116.72	0.00	19.21
		BABA.N	阿里巴巴	71.85	73.91	2.87	1,901.53	1.44	13.33	23.78	18.64	11.90
港股	9988.HK	阿里巴巴-SW	70.70	73.00	3.25	14,873.19	1.45	13.48	23.80	18.53	12.03	
	0700.HK	腾讯控股	279.40	291.80	4.44	27,521.94	4.22	12.99	15.17	0.00	15.30	
	80700.HK	腾讯控股-R	257.00	268.40	4.44	24,995.15	4.23	13.02	0.00	0.00	15.49	
	9999.HK	网易-S	155.80	171.50	10.08	5,528.51	4.99	18.94	16.52	0.00	16.67	
	9888.HK	百度集团-SW	102.00	106.90	4.80	2,989.76	2.07	12.10	36.49	0.00	12.50	
	89888.HK	百度集团-SWR	93.50	98.65	5.51	2,715.27	2.08	12.17	0.00	0.00	12.70	
	1024.HK	快手-W	39.85	45.10	13.17	1,966.42	1.66	145.41	-19.96	0.00	16.17	
	81024.HK	快手-WR	36.55	41.70	14.09	1,785.89	1.67	146.52	0.00	0.00	16.46	
	9626.HK	哔哩哔哩-W	71.90	81.85	13.84	344.80	1.42	-6.30	-8.90	0.00	-15.88	
	2518.HK	汽车之家-S	45.80	53.50	16.81	272.52	3.44	13.14	14.79	12.86	12.82	
	9898.HK	微博-SW	65.35	75.20	15.07	176.92	1.30	5.63	52.13	0.00	6.24	
		MSFT.O	微软(MICROSOFT)	411.22	404.06	-1.56	30,023.42	13.19	36.37	24.58	38.62	33.97

美股												
软件及应用	SNOW.N	SNOWFLAKE	218.76	230.31	5.28	758.41	28.94	-86.78	-67.89	-82.25	0.00	
	ORCL.N	甲骨文(ORACLE)	115.79	111.31	-3.87	3,059.83	5.93	30.18	32.81	34.08	23.12	
	CRM.N	赛富时(SALESFORCE)	285.66	289.72	1.42	2,804.49	8.26	108.20	91.82	1224.61	71.02	
	ADBE.O	奥多比(ADOBE)	634.76	546.66	-13.88	2,470.90	12.73	45.52	32.69	50.01	39.41	
	INTU.O	财捷(INTUIT)	639.58	655.74	2.53	1,835.65	12.45	71.01	52.92	73.39	65.57	
	SNPS.O	新思科技(SYNOPSYS)	552.05	552.91	0.16	843.42	14.38	68.33	49.43	63.63	0.00	
	CDNS.O	铿腾电子(CADENCE)	296.55	290.30	-2.11	790.31	19.32	75.91	51.91	71.17	0.00	
	ADSK.O	欧特克(AUTODESK)	256.45	258.30	0.72	552.54	10.34	60.26	81.13	63.29	64.58	
	U.N	Unity	32.84	33.90	3.23	128.83	6.35	-14.98	-9.35	0.00	0.00	
	港股	0020.HK	商汤-W	0.78	0.85	8.97	284.49	6.85	-4.36	-10.98	0.00	-9.20
80020.HK		商汤-WR	0.71	0.79	11.27	258.37	6.91	-4.40	0.00	0.00	-9.42	
3888.HK		金山软件	19.24	22.00	14.35	300.42	3.31	84.89	-5.25	0.00	23.59	
0268.HK		金蝶国际	7.63	8.13	6.55	295.26	5.11	-86.08	-133.49	0.00	-476.36	
9878.HK		汇通达网络	30.10	29.60	-1.66	166.52	0.18	36.42	70.68	0.00	23.40	
3650.HK		KEEP	4.69	5.14	9.59	27.02	1.14	3.01	0.00	0.00	-10.48	
0354.HK		中国软件国际	4.38	4.72	7.76	137.21	0.68	23.47	24.18	0.00	11.10	
1357.HK		美图公司	2.29	2.54	10.92	113.74	4.49	17.83	57.75	0.00	20.60	
3896.HK		金山云	1.31	1.48	12.98	56.32	0.69	-2.15	-2.69	0.00	-3.75	
2013.HK		微盟集团	1.71	1.92	12.28	53.66	2.30	-2.96	-8.29	0.00	-117.03	
1675.HK		亚信科技	7.62	8.15	6.96	76.23	0.89	8.21	13.73	0.00	6.70	
2121.HK		创新奇智	6.11	6.52	6.71	36.84	1.85	-10.44	-31.65	0.00	-37.22	
2400.HK		心动公司	8.94	10.92	22.15	52.47	1.35	-62.62	-16.75	0.00	12.98	
0777.HK		网龙	10.54	11.40	8.16	60.56	0.76	7.26	9.71	0.00	5.44	

资料来源：Wind，中国银河证券研究院

\*2月2日至16日，ARM.O 累计涨幅达 80.76%，总市值由 727.92 亿美元升至 1319.43 亿美元，主要系 2024Q3 财季业绩超预期增长，全年业绩预期上调所致。

## 二、本周热点内容

### (一) 算力及终端

#### •英伟达公开展示超级计算机 Eos。

英伟达在 2 月 15 日发布了一段视频，首次公开展示其最新的数据中心规模超级计算机 Eos，为人们提供了一窥高级 AI 工厂背后架构的机会。

Eos 是一种超大型 NVIDIA DGX SuperPOD，NVIDIA 开发人员利用加速计算基础设施和全面优化的软件实现 AI 突破。Eos 配备了 576 套 NVIDIA DGX H100 系统、NVIDIA Quantum-2 InfiniBand 网络和软件，提供总共 18.4 exaflops 的 FP8 AI 性能。Eos 在 2023 年的超级计算贸易展览会上公布，以希腊女神的名字命名，这位女神据说每天打开黎明的大门，这反映了 NVIDIA 对推进 AI 技术的承诺。Eos 超级计算机每个 DGX H100 系统均配备八个 NVIDIA H100 Tensor Core GPU。Eos 总共配备了 4,608 个 H100 GPU。因此，Eos 可以处理最大的人工智能工作负载，以训练大型语言模型、推

荐系统、量子模拟等。它展示了 NVIDIA 技术在大规模应用时的能力。Eos 的到来恰逢其时，人们正在利用生成式人工智能改变世界，从药物发现到聊天机器人，再到自主机器等等。

为了实现这些突破，他们需要的不仅仅是人工智能专业知识和开发技能。他们需要一个人工智能工厂——一个始终可用的专用人工智能引擎，可以帮助提高他们大规模构建人工智能模型的能力。

Eos 在全球最快超级计算机 TOP500 排行榜中排名第 9，突破了人工智能技术和基础设施的界限。它包括 NVIDIA 先进的加速计算和网络，以及 NVIDIA Base Command 和 NVIDIA AI Enterprise 等先进的软件产品。

### **•英伟达 NVIDIA RTX 2000 Ada Generation GPU 的推出为 AI 加速设计和可视带来更好的性能和多功能性。**

2 月 12 日，英伟达推出新的 NVIDIA RTX 2000 Ada Generation GPU。全新 NVIDIA RTX 2000 Ada Generation GPU 为紧凑型工作站提供了最新的 AI、图形和计算技术，在专业工作流程中提供的性能是上一代 RTX A2000 12GB 的 1.5 倍。

从制作令人惊叹的 3D 环境到简化复杂的设计审查，再到完善工业设计，该卡的功能为 AI 加速的未来铺平了道路，使专业人士能够在不影响性能或功能的情况下实现更多目标。

现代多应用工作流程，例如人工智能工具、多显示设置和高分辨率内容，对 GPU 内存提出了很高的要求。借助 RTX 2000 Ada 的 16GB 内存，专业人士可以利用最新的技术和工具，更快、更好地处理数据。

在 NVIDIA RTX 技术的支持下，新 GPU 通过 NVIDIA DLSS 提供令人印象深刻的图形逼真度，提供超高质量、逼真的光线追踪图像，速度比以前快 3 倍以上。此外，RTX 2000 Ada 可为企业虚拟现实工作流程（例如产品设计和工程设计审核）提供身临其境的体验。

RTX 2000 Ada 集性能、多功能性和 AI 功能于一身，可帮助各行各业的专业人士提高效率。建筑师和城市规划师可以使用它来加速可视化工作流程和结构分析，从而提高设计精度。使用工业 PC 的产品设计师和工程师可以通过快速、逼真的渲染和人工智能驱动的生成设计来快速迭代产品设计。内容创作者可以无缝编辑高分辨率视频和图像，并使用人工智能实现逼真的视觉效果和内容创作辅助。

在重要的嵌入式应用和边缘计算中，RTX 2000 Ada 可以为医疗设备提供实时数据处理，通过预测性维护优化制造流程，并在零售环境中实现 AI 驱动的智能。

### **•三星 Galaxy Watch 上的睡眠呼吸暂停功能首次获得美国 FDA 授权。**

三星 2 月 10 日宣布，三星 Samsung Health Monitor 应用程序上的睡眠呼吸暂停功能已获得美国食品和药物管理局 (FDA) 的 De Novo 授权。该功能使用兼容的三星 Galaxy 手表和手机来检测睡眠呼吸暂停的迹象，继去年 10 月在韩国宣布获得韩国食品药品安全部 (MFDS) 的批准后，这是 FDA 首次授权的此类功能。

睡眠呼吸暂停功能使 22 岁以上尚未被诊断患有睡眠呼吸暂停的用户能够在两个晚上的监测期内检测中度至重度阻塞性睡眠呼吸暂停 (OSA) 的迹象，这是一种常见的慢性睡眠状况，通常未被诊

请务必阅读正文最后的中国银河证券股份有限公司免责声明。

断出来且未经治疗。要利用该功能，用户只需在十天内两次跟踪自己的睡眠情况，每次睡眠时间超过四个小时。

OSA 会导致人在睡眠时停止呼吸，通常会导致供氧中断、睡眠质量下降和白天疲劳加剧。未经治疗的睡眠呼吸暂停会增加患高血压、冠状动脉疾病、心力衰竭、心律失常和中风等心血管疾病的风险。根据美国国家睡眠基金会(NSF)的数据，美国大约 25%的男性和 10%的女性患有 OSA。三星健康监测应用程序上的新功能预计将帮助更多的人主动检测中度或重度 OSA，并根据检测结果寻求医疗护理，以减少与健康相关的并发症的可能性。

借助睡眠呼吸暂停功能，三星进一步履行其为 Galaxy 用户提供最佳睡眠工具以改善睡眠健康习惯的承诺。该功能将于第三季度通过 Samsung Health Monitor 应用程序在美国的 Galaxy Watch 系列上提供。

#### **•思科和英伟达帮助企业快速轻松地部署和管理安全的 AI 基础设施。**

2 月 6 日，思科和英伟达宣布计划为数据中心提供易于部署和管理的人工智能基础设施解决方案，为企业在 AI 时代取得成功提供强大的计算能力。

思科凭借其在以太网网络领域的行业领先专业知识和广泛的合作伙伴生态系统，以及 NVIDIA 作为推动 AI 热潮的 GPU 的发明者，共享帮助客户以高度安全的以太网基础设施导航 AI 过渡的愿景和承诺。

思科和英伟达在过去几年中提供了一系列集成产品解决方案，跨越 Webex 协作设备和数据中心计算环境，以支持具有灵活工作空间、AI 驱动会议和虚拟桌面基础设施的混合工作力量。现在，公司正在深化在数据中心的合作伙伴关系，以使企业客户能够获得可扩展和自动化的 AI 集群管理、自动故障排除、最佳客户体验等。亮点包括：

英伟达最新的 Tensor Core GPU 可用于思科 M7 代 UCS 机架和刀片服务器，包括 Cisco UCS X-Series 和 UCS X-Series Direct，可在数据中心和边缘提供广泛的 AI 和数据密集型工作负载的最佳性能。

NVIDIA AI Enterprise 包括软件框架、预训练模型和开发工具，可实现更安全、稳定和受支持的生产 AI，现已在思科的全球价目表上提供。

通过思科验证设计(CVD)进行联合验证的参考架构，可以在跨越虚拟化和容器化环境的各种用例中轻松部署和管理任何规模的 AI 集群，并提供融合和超融合选项。用于使用 NVIDIA AI Enterprise 进行生成式 AI 推理的 FlexPod 和 FlashStack 的 CVD 将于本月推出，后续还会推出更多产品。

支持思科网络云：思科通过 Cisco Nexus Dashboard 和 Cisco Intersight 进行本地和基于云的管理，简化了 AI 基础设施管理和运营。

数字体验监控：借助公共云、本地和跨多个数据中心中的人工智能工作负载和数据，ThousandEyes 提供数字体验监控，以提供人工智能驱动的意见，并自动修复从云到本地网络中任何地方发生的问题。

思科可观测性平台使用人工智能功能来跨域实时遥测进行情境化和关联，以便组织能够更好地获得可见性、洞察力和行动，以改善数字体验。

合作伙伴降低风险：当组织计划成功采用人工智能和自动化时，他们将寻求思科的全球合作伙伴生态系统为其提供建议、支持和指导。

### •三星 Neo QLED、OLED 和 Lifestyle TVs 获得 TÜV Rheinland 的“产品碳减排”认证。

三星于 2 月 5 日宣布，其 2024 年的 Neo QLED 4K 和 8K、OLED 和 The Frame 系列中的八个产品系列已获得国际独立第三方测试、检验和认证机构 TÜV Rheinland 的“产品碳减排”认证。该认证展示了这些产品与上一年车型相比的碳足迹减少率，证明了公司逐年减少碳排放的努力。

三星电子视觉显示业务执行副总裁 Yongjae Kim 表示：“作为全球领先的电视制造商，我们有责任打造既重新定义电视功能又推动可持续发展努力的产品。”

TÜV Rheinland 根据国际公认的标准如 ISO 14067 和 ISO 14064-4 以及 TÜV Rheinland 自行开发的标准 2 PfG Q2880/09.23，评估了三星电视产品在其生命周期内的温室气体排放——从生产、运输、使用到处置。它验证并认证所有八个电视产品系列满足 TÜV Rheinland 设定的碳减排标准要求。新获认证的产品包括两款 Neo QLED 8K 系列、三款 Neo QLED 4K 系列、两款 OLED 系列和一款 The Frame 系列。

除了产品认证外，三星视觉显示业务还获得了德国 TÜV Rheinland 在显示和消费电子行业首个“产品碳足迹计算方法”认证。该认证对产品整个生命周期的温室气体排放进行测量和管理，按照管理规定和规定的执行手册收集和分析数据，计算温室气体排放量并确保结果透明和可重复。

随着三星继续努力减少其产品对环境的影响，其目标是在 2024 年为 60 多种型号（包括新推出的条形音箱）获得 TÜV Rheinland 的“产品碳足迹”和“产品碳减排”认证。有了这八个产品系列的减碳结果和分析报告，三星对下一代脱碳电视产品的开发和设计更加有信心和明确的方向。

这一认可建立在之前对三星可持续发展努力的认可之上。2021 年，它的 Neo QLED 4K 和更高

分辨率的电视获得了碳信托的首个“减少二氧化碳排放”认证。2022年，11款电视型号以及5款显示器和标牌产品获得了“减少二氧化碳排放”认证，最近，三星2023年产品线中的24款型号获得了碳信托组织和TÜV Rheinland的认证。

**•AMD 推出嵌入式+架构；将嵌入式处理器与自适应 SoC 相结合，加快边缘 AI 应用的上市时间。**

2月6日，AMD宣布推出AMD Embedded+，这是一种新的架构解决方案，将AMD Ryzen Embedded处理器与Versal自适应SoC集成到单一的集成板上，为原始设计制造商（ODM）合作伙伴提供可扩展且能效高的解决方案，加速产品上市时间。

经AMD验证，Embedded+集成计算平台可帮助ODM客户减少资格认证和构建时间，从而加快上市时间，而无需花费额外的硬件和研发资源。使用Embedded+架构的ODM集成使得能够使用通用软件平台开发具有低功耗、小型化和长生命周期的设计，适用于医疗、工业和汽车应用。

在同类产品中，Embedded+架构首次将领先的AMD x86计算与集成显卡和可编程硬件相结合，适用于关键的AI推理和传感器融合应用。自适应计算在确定性、低延迟处理方面表现出色，而人工智能引擎则提高了每瓦特的高性能推理能力。Ryzen Embedded处理器包含高性能的“Zen”核心和Radeon显卡，还提供世界一流的渲染和显示选项，以提升4K多媒体体验，以及用于4K H.264/H.265编解码的内置视频编解码器。

低延迟处理和高性能每瓦推理相结合，可实现关键任务的高性能，包括将自适应计算与灵活的I/O、用于推理的AI引擎和AMD Radeon显卡实时集成在一个解决方案中，该解决方案利用每种技术的最佳方面。

**•微软被 DARPA 选择，继续开发公用事业规模的量子计算机。**

2月8日，美国国防高级研究计划局(DARPA)运行的一项计划旨在确定建造和运行公用事业规模量子计算机的可行性，该计算机能够可靠、准确地解决传统计算机无法解决的有意义的问题。该计划被称为“未充分开发的实用规模量子计算系统”(US2QC)，投资于那些使用有前景的创新方法来设计和构建功能性、可扩展量子计算机的公司。最近，DARPA选择Microsoft Azure Quantum继续支持基于拓扑量子位的公用事业规模量子计算机的开发。

DARPA选择支持微软Azure Quantum项目，突显了该机构对量子计算领域潜在革命性技术的认可和投资。通过研发基于拓扑量子比特的量子计算机，微软不仅在理论上寻求超越当前量子计算的局限性，而且还致力于实现一种更加稳定和可扩展的量子计算平台。拓扑量子比特因其潜在的高错误容忍率和稳定性而被认为是量子计算领域的一个重要突破。

随着DARPA和微软之间的合作进入下一阶段，微软Azure量子团队将为基于拓扑量子位的量子计算机的容错原型(FTP)开发详细设计。FTP的设计将确定该小型量子计算机所有组件和子系统的最低性能要求，从而证明构建和运行公用事业规模量子计算机的可行性。

微软的FTP旨在处理存储在多个逻辑量子位中的信息，代表数千个物理量子位，这将为可扩展的量子计算机建立基线。DARPA将获得FTP的架构设计，该设计将由DARPA测试和评估团队的技术专家进行评估。此外，微软在此过程中取得的科学发现将由DARPA在微软的设施中进行独立验证。

微软认为，DARPA决定扩大对Microsoft Azure Quantum的投资，这为具有挑战性但前景光明的拓扑方法以及微软设计和构建大规模量子超级计算机的能力注入了信心。这一目标的实现将使当前棘手的问题得到迅速解决，为世界带来巨大的价值。

## (二) 软件及应用

### •谷歌推出人工智能模型 Gemini 1.5 Pro.

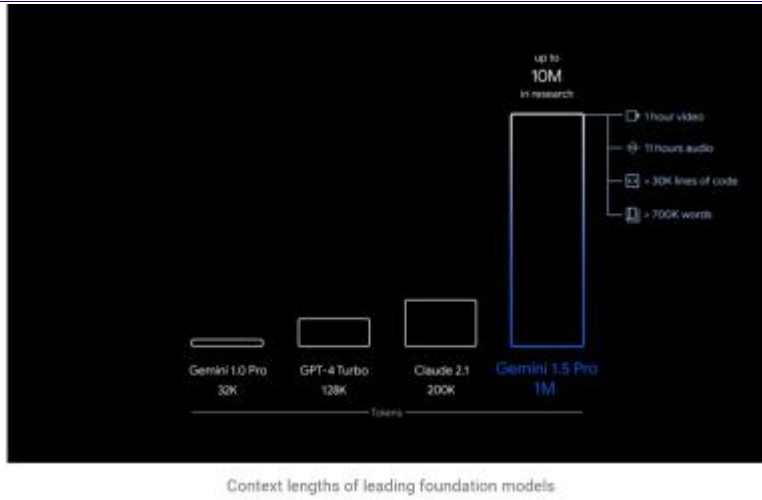
自推出 Gemini 1.0 以来，谷歌一直在测试、完善和增强其功能。2月15日，谷歌宣布推出新一代型号：Gemini 1.5 Pro。Gemini 1.5 Pro 显著增强了性能。它代表了谷歌方法的重大改变，建立在几乎谷歌所有基础模型开发和基础设施的研究和工程创新之上。Gemini 1.5 Pro 建立在对 Transformer 和 MoE 架构的领先研究之上：传统 Transformer 是一个大型神经网络，而 MoE 模型则被划分为更小的“专家”神经网络，可使 Gemini 1.5 的训练和服务更加高效。

谷歌发布用于早期测试的第一个 Gemini 1.5 型号是 Gemini 1.5 Pro：中等大小的多模态模型，针对跨广泛任务的扩展进行了优化，并且与谷歌迄今为止最大的模型 1.0 Ultra 的性能相当。它还引入了一个在长上下文理解方面的突破性实验功能。

Gemini 1.5 Pro 上下文窗口容量增至 100 万个 tokens，意味着可以一次性处理大量信息，包括 1 小时的视频、11 小时的音频、超过 30,000 行代码或超过 700,000 个单词的代码库。研究中谷歌成功测试了多达 1000 万个 tokens。从今天开始，少数开发人员和企业客户可以通过 AI Studio 和 Vertex AI 的私人预览版在最多 100 万个 tokens 上下文窗口中进行尝试。

当谷歌推出完整的 100 万个 tokens 上下文窗口时，谷歌正在积极进行优化，以改善延迟、减少计算要求并增强用户体验。谷歌下一代模型的这些持续进步将为人们、开发人员和企业使用人工智能进行创造、发现和构建开辟新的可能性。

图 3 : Gemini 1.5 Pro 与其他领先基础模型 tokens 对比



资料来源：Google，中国银河证券研究院

**• OpenAI 发布新的生成式 AI 模型 Sora，只需输入描述性句子即可创建逼真的视频。**

2 月 15 日，OpenAI 公司周四推出了新的生成式 AI 模型 Sora。Sora 的工作原理与 OpenAI 的图像生成人工智能工具 DALL-E 类似。用户输入所需的场景文字，Sora 将生成高清视频。Sora 还可以生成受静止图像启发的视频，并扩展现有视频或填充缺失的帧。

OpenAI 希望通过 Sora 与谷歌、Meta 等公司的视频生成人工智能工具展开竞争，谷歌公司在 1 月份宣布了 Lumiere。其他初创公司也提供类似的人工智能工具，例如 Stability AI，它有一款名为 Stable Video Diffusion 的产品。亚马逊还发布了 Create with Alexa，这是一个专门生成基于提示的短片动画儿童内容的模型。

Sora 目前仅限于生成一分钟或更短的视频。OpenAI，由微软支持，将多模态（文本、图像和视频生成的结合）作为提供更广泛的人工智能模型套件的目标。目前，Sora 只对一小群安全测试人员或“red teamers”开放，他们测试模型在错误信息和偏见等领域的脆弱性。该公司尚未发布除其网站上的 10 个样本剪辑之外的任何公开演示，并表示其随附的技术论文将在 2 月 15 日晚些时候发布。

与 GPT 模型类似，Sora 使用 Transformer 架构，解锁卓越的缩放性能。视频和图像使用 patch（视觉编码块）的较小数据单元的集合，每个 patch 都类似于 GPT 中的令牌。通过统一表示数据的方式，可在比以前更广泛的视觉数据上训练扩散 Transformer，跨越不同的持续时间、分辨率和长宽比。Sora 建立在过去对 DALL-E 和 GPT 模型的研究基础上。它使用来自 DALL-E3 的重捕获技术，该技术涉及为视觉训练数据生成高度描述性的标题。因此，该模型能够更忠实地遵循用户在生成的视频中的文字指令。除了能够从文本指令生成视频外，该模型还能够获取现有的静止图像并从中生成视频，并精确地将图像内容动画化，并添加小细节。该模型还可以获取现有的视频并对其进行扩展或填充缺失的帧。Sora 是能够理解和模拟现实世界的模型的基础，这一能力将是实现 AGI 的重要里程碑。

**图 4：文生视频模型 Sora 可生长长达 60 秒的视频**



资料来源：OpenAI，中国银河证券研究院

请务必阅读正文最后的中国银河证券股份有限公司免责声明。

• Meta 发布 V-JEPA，迈向 AMI 愿景的下一步。

2月15日，Meta公开发布关于联合嵌入预测架构(V-JEPA)模型的视频，迈向Yann LeCun先进人工智能(AMI)愿景的下一步。此物理世界模型的早期示例擅长检测和理解对象之间高度详细的交互。本着负责的开放科学的精神，Meta在知识共享非商业许可下发布此模型，供研究人员进一步探索。

V-JEPA是一种非生成模型，通过预测抽象表示空间中视频的缺失或屏蔽部分来进行学习。这与Meta的图像联合嵌入预测架构(I-JEPA)比较图像的抽象表示(而不是比较像素本身)的方式相似。与尝试填充每个缺失像素的生成方法不同，V-JEPA可以灵活地丢弃不可预测的信息，从而将训练和样本效率提高1.5至6倍。

由于V-JEPA采用自我监督的学习方法，因此完全使用未标记的数据进行预训练。标签仅用于在预训练后使模型适应特定任务。事实证明，这种类型的架构比以前的模型更有效，无论是在所需

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/026125221053010104>