



## OpenAI 发布 O3，谷歌推出 Gemini 2.0

2024 年 12 月 23 日

概览【12.6-12.20】

**股指动态：**12月至今，美股科技、港股及中概股普遍上涨，美股大盘下跌。标普 500 指数-0.97%，纳斯达克综合指数+2.84%，费城半导体指数+3.90%；TAMAMA 科技指数+7.20%；纳斯达克中国金龙指数+2.89%；恒生科技指数+3.54%；计算机-1.52%。

**个股表现：**过去两周，热门科技股大部分下跌，苹果、特斯拉及谷歌逆势上涨。据统计，相比 12 月 6 日收盘价，12 月 20 日盘后，苹果合计+4.80%，英伟达 -5.43%，特斯拉 +8.18%，谷歌 +9.68%，亚马逊 -0.93%，META-6.10%，微软 -1.57%，ARM-6.20%，英特尔 -6.69%，高通 -4.15%，AMD-13.98%。

**10 年期国债及汇率：**12 月 20 日，美国 10 年期国债利率为 4.52%，相较 12 月 6 日累计上升 37bps；12 月 20 日，美元兑人民币中间价报 7.19；较 12 月 6 日价累计调贬 53 个基点。

## 核心观点：

OpenAI 发布了新一代 AI 推理模型 O3 及其高性价比版本 O3 Mini，这两款模型在推理能力上实现了显著飞跃，标志着人工智能技术进入了一个新的发展阶段。O3 模型是专为解决编程、数学和通用智能领域的高难度问题而设计的。与前代 O1 模型相比，O3 在多个基准测试中展现了卓越的性能。在编程能力方面，O3 在 Bench Verified 编程基准测试中的准确率达到 71.7%，竞技编程中的 ELO 评分高达 2727，显著超越 O1 的 1891。在数学与科学领域，O3 在 AIME 2024 数学推理测试中的准确率为 96.7%，而在科学基准测试 GPQA Diamond 中，O3 取得了 87.7% 的准确率，远高于 O1 的 78%。

谷歌推出的 Gemini 2.0 系列 AI 模型专为代理时代设计，显著提升了多模态推理、长上下文理解、复杂指令执行和工具调用能力，并支持实时音频、视频流输入及多工具组合的应用开发。Gemini 2.0 引入视频、音频和文本的混合处理、多生成图像和可编译的文本转语音 (TTS) 功能，全面增强低延迟表现和交互能力，通过 Google AI Studio 和 Vertex AI 提供给开发者使用。同时，新的 Multimodal Live API 为开发动态应用提供了强大支持。我们认为，Gemini 2.0 的发布不仅将推动多模态 AI 的技术应用，还将加速 AI 工具链及服务生态的扩展，为 AI 技术的普及和商业落地奠定基础。Gemini 2.0 的推出，预示着 AI 应用将迎来前所未有的创新时机。随着底层模型推理能力的迅速提升，预计明年将会成为 AI 代理的爆发元年。

**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

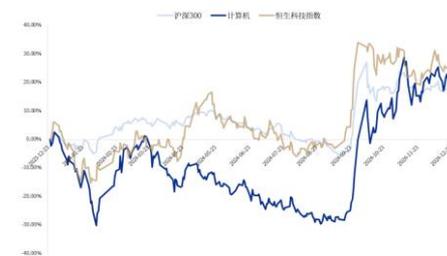
☎：010-66568589

✉：wuyanqing@chinastock.com.cn

分析师登记编码：S0130519070001

## 国内表现

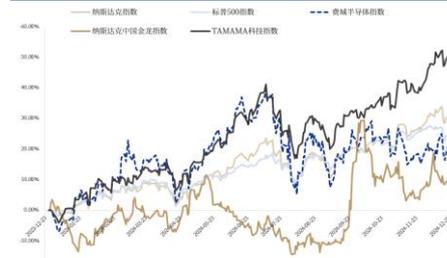
2024-12-23



资料来源：中国银河证券研究院

## 全球行情

2024-12-23



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三季报

【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 .....	5
(二) 大模型及云应用 .....	7
三、风险提示.....	10

## 一、全球市场表现

### (一) 股市动态

12月至今，美股科技、港股及中概股普遍上涨，美股大盘下跌。标普500指数-0.97%，纳斯达克综合指数+2.84%，费城半导体指数+3.90%；TAMAMA科技指数+7.20%；纳斯达克中国金龙指数+2.89%；恒生科技指数+3.54%；计算机-1.52%。

表1：主要股指周变动

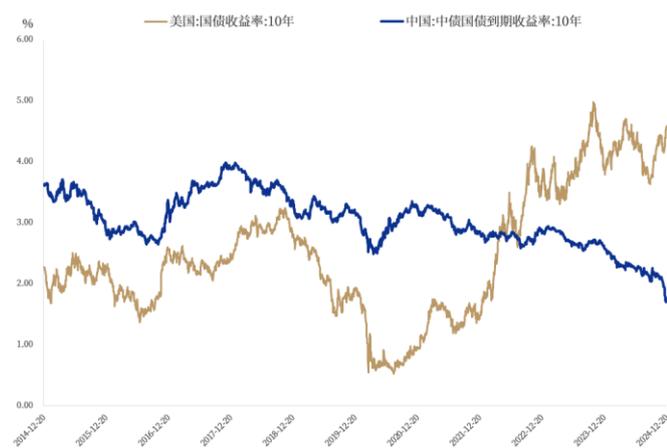
指数代码	指数简称	涨跌幅%					市盈率 PE (TTM)
		[12.13-12.20]周	[12.06-12.13]周	本月	本年度	2023	
SPX.GI	标普500指数	-1.99	-0.64	-0.97	25.25	24.23	28.72
IXIC.GI	纳斯达克指数	-1.78	0.34	2.84	31.67	43.42	46.88
SOX.GI	费城半导体指数	-3.59	1.75	3.90	22.59	64.90	52.61
8884057.WI	TAMAMA科技指数	-0.62	1.32	7.20	51.52	67.81	39.00
HXC.GI	纳斯达克中国金龙指数	-2.21	2.07	2.89	6.40	-3.39	21.17
HSTECH.HI	恒生科技指数	-0.79	0.33	3.54	19.75	-8.83	20.70
CI005027.WI	计算机	0.50	-0.78	-1.52	15.36	8.90	122.42

资料来源：WIND，中国银河证券研究院

### (二) 债市及汇率情况

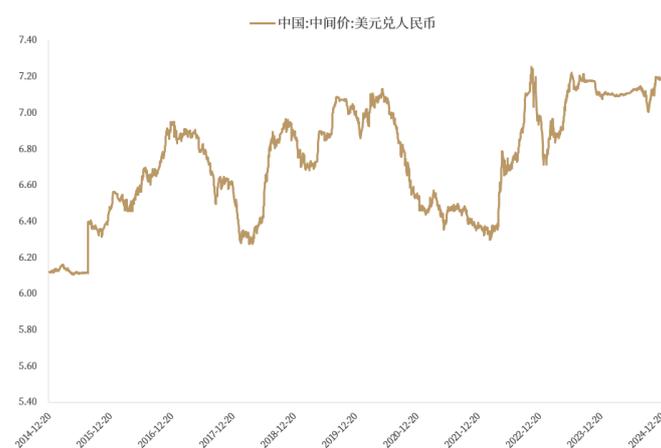
12月20日，美国10年期国债利率为4.52%，相较12月6日累计上升37bps；12月20日，美元兑人民币中间价报7.19；较12月6日价累计调贬53个基点。

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

过去两周，热门科技股大部分下跌，苹果、特斯拉及谷歌逆势上涨。据统计，相比12月6日收盘价，12月20日盘后，苹果合计+4.80%，英伟达-5.43%，特斯拉+8.18%，谷歌+9.68%，亚马逊-0.93%，META-6.10%，微软-1.57%，ARM-6.20%，英特尔-6.69%，高通-4.15%，AMD-13.98%。

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)	市销率 PS (TTM)	市盈率 PE (TTM)	PE(2022)	PE(2023)	PE(2024E)	
				[2024-12-06]	[2024-12-20]								
算力及终端	美股	AAPL.O	苹果(Apple)	242.84	254.49	4.80	38,468.26	9.84	41.04	20.71	30.87	37.91	
		NVDA.O	英伟达(NVIDIA)	142.44	134.70	-5.43	32,988.03	29.12	52.30	36.86	280.04	0.00	
		TSLA.O	特斯拉(TESLA)	389.22	421.06	8.18	13,516.28	13.91	106.41	30.98	52.67	184.17	
		HPQ.N	惠普(HP)	36.20	32.94	-8.25	308.91	0.58	11.13	8.24	9.14	0.00	
		CSCO.O	思科(CISCO)	59.89	58.52	-2.29	2,330.71	4.40	24.81	16.57	16.28	0.00	
		ASML.O	阿斯麦	708.98	705.68	-0.47	2,784.54	9.62	36.44	36.70	34.34	36.63	
		AMD.O	超威半导体(AMD)	138.59	119.21	-13.98	1,934.55	7.96	105.94	79.12	278.85	91.73	
		INTC.O	英特尔(INTEL)	20.92	19.52	-6.69	841.90	1.55	-5.27	13.61	125.43	-2556.96	
		QCOM.O	高通(QUALCOMM)	159.51	152.89	-4.15	1,698.61	4.36	16.75	9.51	23.36	18.21	
		ARM.O	ARM	140.89	132.15	-6.20	1,388.90	39.29	216.68	0.00	147.03	0.00	
	ON.O	安森美半导体(ON SEMICONDUCTOR)	66.52	65.50	-1.53	278.90	3.78	15.89	14.18	16.47	19.15		
	港股	0909.HK	明源云	2.92	2.77	-5.14	53.85	3.05	-13.01	-10.65	-8.66	-22.05	
9698.HK		万国数据-SW	19.74	18.82	-4.66	286.90	2.36	-6.49	-21.99	-2.92	-22.70		
1686.HK		新意网集团	4.38	4.14	-5.48	96.84	3.61	10.67	11.66	7.98	18.80		
台股	2330.TW	台积电	1,065.00	1,035.00	-2.35	329,043.66	10.12	25.55	11.71	18.05	24.66		
	2454.TW	联发科	1,305.00	1,390.00	6.51	22,263.46	4.27	20.56	8.46	21.09	0.00		
互联网	美股	GOOGL.O	谷歌(ALPHABET)-A	174.71	191.41	9.68	23,516.28	6.89	24.85	19.04	23.69	24.60	
		AMZN.O	亚马逊(AMAZON)	227.03	224.92	-0.93	23,650.36	3.81	47.43	-314.82	51.61	45.73	
			脸书										
		META.O	(META PLATFORMS)	623.77	585.25	-6.10	14,774.57	9.46	26.60	13.75	23.27	27.17	
		NFLX.O	奈飞(NETFLIX)	934.74	909.05	-2.75	3,885.81	10.34	49.94	29.21	39.40	46.17	
		PDD.O	拼多多	99.89	99.51	-0.38	1,381.97	2.60	8.94	22.77	22.94	9.07	
		NTES.O	网易	91.29	92.09	0.88	590.56	3.92	15.06	16.34	14.46	14.76	
		BIDU.O	百度	86.43	85.80	-0.73	295.89	1.57	9.96	36.42	14.52	10.38	
		TCOM.O	携程网	69.07	72.44	4.88	488.28	6.82	21.41	116.72	17.58	21.60	
		BABA.N	阿里巴巴	85.93	82.28	-4.25	1,965.42	1.43	16.00	23.78	18.64	0.00	
	港股	9988.HK	阿里巴巴-SW	83.95	80.10	-4.59	15,269.97	1.43	16.05	23.80	18.53	0.00	
		0700.HK	腾讯控股	412.60	426.40	3.34	39,517.16	5.54	20.99	15.17	21.90	19.63	
		80700.HK	腾讯控股-R	385.60	401.20	4.05	36,571.16	5.78	21.90	0.00	21.86	19.93	
		9999.HK	网易-S	142.30	142.90	0.42	4,588.21	3.91	15.04	16.52	13.96	14.64	
		9888.HK	百度集团-SW	84.95	81.95	-3.53	2,298.88	1.55	9.79	36.49	14.48	10.35	
		89888.HK	百度集团-SWR	79.50	76.95	-3.21	2,127.50	1.61	10.20	0.00	14.50	10.52	
		1024.HK	快手-W	47.60	42.75	-10.19	1,843.93	1.35	11.10	-19.96	32.60	11.40	
		81024.HK	快手-WR	44.35	40.25	-9.24	1,706.46	1.41	11.59	0.00	32.58	11.56	
		9626.HK	哔哩哔哩-W	150.70	150.00	-0.46	631.89	2.24	-20.85	-8.90	-7.41	-42.23	
		2518.HK	汽车之家-S	55.35	49.10	-11.29	250.11	3.15	12.91	14.79	12.86	13.36	
软件及应用	美股	MSFT.O	微软(MICROSOFT)	443.57	436.60	-1.57	32,460.69	12.77	35.86	24.58	38.62	0.00	
		SNOW.N	SNOWFLAKE	183.64	166.29	-9.45	548.92	16.08	-48.68	-67.89	-82.25	0.00	
		ORCL.N	甲骨文(ORACLE)	191.69	169.66	-11.49	4,745.32	8.64	40.82	32.81	34.08	0.00	
		CRM.N	赛富时(SALESFORCE)	361.99	343.65	-4.96	3,288.73	8.84	55.41	91.82	1224.61	0.00	
		ADBE.O	奥多比(ADOBE)	552.96	447.17	-19.13	1,968.44	9.15	35.40	32.69	50.01	34.90	
		INTU.O	财捷(INTUIT)	646.47	643.39	-0.48	1,801.72	10.86	61.72	52.92	73.39	0.00	
		SNPS.O	新思科技(SYNOPSYS)	517.28	492.18	-4.85	760.80	12.42	33.61	49.43	63.63	0.00	
		CDNS.O	铿腾电子(CADENCE)	307.51	301.47	-1.96	826.82	18.99	79.57	51.91	71.17	0.00	
		ADSK.O	欧特克(AUTODESK)	307.71	298.12	-3.12	640.96	10.75	58.75	81.13	63.29	0.00	
		U.N	Unity	27.72	22.51	-18.80	90.68	4.61	-11.42	-9.34	-18.90	-12.05	
	港股	0020.HK	商汤-W	1.71	1.50	-12.28	555.11	13.64	-8.77	-10.98	-5.46	-12.55	
		80020.HK	商汤-WR	1.61	1.42	-11.80	513.73	14.15	-9.10	0.00	-5.46	-12.83	
		3888.HK	金山软件	33.00	34.00	3.03	454.27	4.19	31.59	-5.25	61.69	27.77	
		0268.HK	金蝶国际	9.99	9.03	-9.61	323.81	4.87	-204.94	-133.49	-178.41	-504.88	
		9878.HK	汇通达网络	18.34	17.18	-6.32	96.65	0.12	26.87	70.68	34.57	18.66	
		3650.HK	KEEP	6.31	6.05	-4.12	31.80	1.33	-11.49	0.00	6.00	-9.86	
		0354.HK	中国软件国际	5.43	5.22	-3.87	142.61	0.78	20.08	24.18	22.39	17.59	
		1357.HK	美图公司	3.09	3.09	0.00	140.14	4.21	28.17	57.75	38.62	23.11	
		3896.HK	金山云	4.44	4.69	5.63	178.47	2.21	-7.84	-2.69	-3.18	-11.86	
		2013.HK	微盟集团	1.63	2.57	57.67	86.30	4.17	-9.19	-8.29	-9.62	-32.89	
1675.HK	亚信科技	5.96	5.96	0.00	55.74	0.67	19.73	13.73	13.69	8.89			
2121.HK	创新奇智	6.56	6.70	2.13	37.86	2.47	-5.94	-31.65	-8.15	-8.35			
2400.HK	心动公司	25.70	26.25	2.14	128.90	3.05	369.07	-16.75	-53.48	18.37			
0777.HK	网龙	10.30	9.93	-3.59	52.75	0.71	10.70	9.71	10.84	6.30			

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

### 【NVIDIA 推出其最经济实惠的生成式 AI 超级计算机】

NVIDIA 推出了新型 Jetson Orin Nano Super 开发者套件，这是一款紧凑型生成式 AI 超级计算机，通过软件升级提供更高性能，价格从 499 美元降至 249 美元。与上一代相比，其生成式 AI 推理性能提升 1.7 倍，总性能提升 70%，达到 67 INT8 TOPS，内存带宽提升 50%，达到 102GB/s。该套件配备 NVIDIA Ampere 架构 GPU 和 6 核 Arm CPU，支持多个并发 AI 应用管道、四个摄像头以及更高分辨率和帧率，适用于开发生成式人工智能、机器人和计算机视觉应用。Jetson Orin Nano Super 运行 NVIDIA AI 软件生态系统，包括用于机器人技术的 NVIDIA Isaac、用于视觉 AI 的 Metropolis 和用于传感器处理的 Holoscan，并支持开发者通过 NVIDIA Omniverse Replicator 和 TAO Toolkit 快速构建 AI 应用。其增强性能对生成式 AI 模型和基于变换器的计算机视觉尤为有利，现有的 Jetson Orin Nano 开发套件用户可通过升级 JetPack SDK 解锁性能提升。

图3: NVIDIA 推出的 Jetson Orin Nano Super 开发者套件



资料来源：彭博社消息，中国银河证券研究院

### 【谷歌利用 Compute Engine X4 机器实现 SAP S/4HANA 的巅峰性能】

随着企业数据量的不断增加和对实时数据分析的需求日益增强，迁移到公共云面临着巨大的挑战，尤其是像 SAP S/4HANA 这样的复杂企业工作负载。为了应对这一挑战，Google Cloud 推出了专为 SAP HANA 设计的 X4 机器系列。X4 机器提供强大的内存和计算能力，支持 SAP HANA 在线事务处理（OLTP）和在线分析处理（OLAP）工作负载，能够处理最苛刻的 SAP 环境需求。X4 系列具有高达 32TB 内存配置和数千个 vCPU 核心，并通过与 Google Cloud 的 Hyperdisk 集成提供高性能和扩展性，帮助企业在不受性能瓶颈限制的情况下扩展其 SAP HANA 部署。

X4 机器的设计不仅关注性能，还重视高可用性和安全性。其内存优化的计算引擎提供 99.95% 的 SLA，最大限度地减少了任务关键型 SAP 应用程序的停机时间。结合 Titanium 技术，它能够

显著提高工作负载的性能和效率，通过专用硬件加速存储和网络功能。X4 还支持增强的安全性和合规性，包括对数据加密和密钥管理的硬件级保护。对于企业来说，X4 提供了一种简化的云原生体验，使得 SAP 实例能够无缝迁移并与 Google Cloud 的生态系统紧密集成，从而推动数字化转型和业务的持续增长。

#### 【英特尔携生态伙伴夯实智能底座，激活 AI 应用落地新潜能】

在火山引擎 2024 FORCE 原动力大会上，英特尔与火山引擎联合发布了搭载英特尔至强 6 性能核处理器的第四代通用计算型实例 g4il。这一实例以卓越的性能和技术创新，为传统企业应用、AI 场景及未来复杂需求提供了强大支持。通过架构升级和技术优化，g4il 显著提升了计算密度，采用双单路架构降低爆炸半径，保障系统高稳定性。其在视频转码、Web 应用和数据库应用中的性能分别提升了 17%、19% 和 20%，支持分钟级镜像部署，极大降低了开发门槛，加速 AI 应用的快速落地。

英特尔至强处理器还通过数据保护与压缩加速技术 (QAT) 提升了存储效率和数据可靠性。QAT 技术优化了数据压缩与解压缩效率，硬件加速生成校验码，并结合可信域拓展 (TDX) 和 GPU 机密计算技术，为云上 AI 场景提供安全保障。

在智能体开发领域，英特尔与 Coze 合作推出 Coze-AIPC 平台，通过端插件机制无缝集成 PC 功能与大模型能力，降低技术门槛，为开发者提供高效端云协同工具。此外，英特尔在大语言模型推理、生命科学研究、视频云服务和企业知识库等领域也提供了优化方案，通过软硬件协同提升效率，满足行业需求。

生成式 AI 等技术正推动产业与社会快速变革，英特尔将持续与火山引擎深化合作，通过高性能算力与创新解决方案助力 AI 应用落地，为未来产业发展注入新动力。

#### 【马斯克 xAI 扩展孟菲斯超算集群，与 OpenAI 竞争】

马斯克的人工智能初创公司 xAI 计划将其位于孟菲斯的超级计算机集群 Colossus 扩展至至少 100 万个图形处理单元 (GPU)，这一举措标志着该超级计算机集群的大规模扩张。目前，Colossus 拥有 10 万个 GPU，用于训练 xAI 的聊天机器人 Grok。此次扩展计划涉及英伟达、戴尔和超微等公司在孟菲斯建立业务以协助扩张，预示着马斯克与 OpenAI 及其首席执行官萨姆·奥特曼之间的竞争进一步加剧。马斯克通过扩大 xAI 的超级计算机集群，提升其在人工智能领域的竞争力，以对抗 OpenAI 和其他行业领导者。

#### 【iGenius 基于英伟达 Grace Blackwell 超级芯片建设大型超算 Colosseum】

意大利人工智能创新企业 iGenius 宣布正在部署一台名为 Colosseum 的 AI 超级计算机，该超算基于英伟达先进的 Grace Blackwell 超级芯片构建，将成为全球范围内采用该芯片的最大规模英伟达 DGX SuperPOD 之一。Colosseum 超算将配备数千颗 Grace Blackwell 超级芯片，采用前沿的液体冷却技术，提供高达 115 exaflops 的惊人计算性能。

#### 【英特尔 Xe3 “Celestial” GPU 设计接近完成，预计 2025 年末亮相】

英特尔公司在推进其独立显卡业务方面取得显著进展，特别是下一代 Xe3 “Celestial” GPU 的研发工作。根据英特尔院士 Tom Petersen 在 PCWorld 的 The Full Nerd 播客中透露的信息，Xe3 “Celestial” GPU 的 IP 设计已基本完成，硬件设计接近成熟。目前，英特尔的 GPU 软件团队仍在对 Xe3 进行深入开发，而硬件团队已经开始着手下一代产品 Xe4 “Druid” 的研发工作，以保持英特尔 GPU 部门的发展节奏。

Xe3 “Celestial” GPU 预计将在 2025 年末与英特尔的 Panther Lake 移动处理器核显一同亮相，该处理器将提供配备 4 个或 12 个 Xe Core 的变体。这一消息为期待英特尔显卡性能提升的消费者带来积极信号。尽管英特尔在 2022 年 2 月发布的“视觉计算路线图”中计划在 2023 至 2024 年间推出锐炫 “Battlemage” 产品线，但实际情况似乎并未完全达到预期，发布时间有所推迟，性能方面也

未完全达到预期。

## （二）大模型及云应用

### 【OpenAI 推出全新推理模型 O3】

OpenAI 在其“12 天 Shipmas”活动的最后一天发布了新一代 AI 模型 O3 及其高性价比版本 O3 Mini，这两款模型在推理能力上实现了显著飞跃，标志着人工智能技术进入了一个新的发展阶段。

O3 模型是专为解决编程、数学和通用智能领域的高难度问题而设计的。与前代 O1 模型相比，O3 在多个基准测试中展现了卓越的性能。在编程能力方面，O3 在 Bench Verified 编程基准测试中的准确率达到 71.7%，竞技编程中的 ELO 评分高达 2727，显著超越 O1 的 1891。在数学与科学领域，O3 在 AIME 2024 数学推理测试中的准确率为 96.7%，而在科学基准测试 GPQA Diamond 中，O3 取得了 87.7% 的准确率，远高于 O1 的 78%。

在被誉为 AI 通用智能金标准的 ARC AGI 测试中，O3 的表现刷新了行业纪录。在低计算设置下，O3 的准确率达到 76%，而在高计算设置下，O3 的准确率达到 88%，首次超过人类表现水平（85%）。这一成就不仅展示了 O3 在解决全新、未见过问题时的强大适应性和泛化能力，也显示了其在通用推理能力上的重大突破。

O3 Mini 作为 O3 的轻量化版本，旨在在保持高推理性能的同时降低使用成本。它通过自适应思维时间根据任务复杂度调整推理深度，实现成本效益平衡。在实时演示中，O3 Mini 成功生成了一个交互式 UI 的 Python 脚本，展现了其在复杂编程任务中的实用性和高效性。

安全创新方面，O3 和 O3 Mini 引入了深度对齐（Deliberative Alignment）方法，这一方法超越了传统的 RLHF 和 RLAIIF，使模型在推理过程中实时评估提示词，提供思维链解释，动态识别潜在风险或歧义，显著提升了安全性和可靠性。

O3 Mini 预计将于 2025 年 1 月底发布，主要面向高效成本场景，而 O3 的全面发布将在安全测试完成后进行，具体时间取决于测试反馈和调整进度。OpenAI 采取谨慎的发布策略，优先确保模型的安全性与可靠性，同时保持与用户社区的透明沟通。这些新一代模型的发布，预示着未来 AI 应用场景将变得更加多元、精细和高效，同时也为实现通用人工智能（AGI）的目标迈出了重要一步。

### 【NVIDIA 推出 NeMo Retriever 微服务，用于由数据驱动的多语言生成 AI】

NVIDIA 的 NeMo Retriever 提供了一种强大的多语言信息检索解决方案，帮助企业克服语言和语境障碍，满足全球用户的需求。通过集成 NeMo Retriever 和重新排序的 NIM 微服务，企业能够从多样化的大型数据集中提取知识，提供准确的上下文感知响应，同时支持全球主要语言。该技术不仅显著提高了检索效率，还通过长上下文支持和动态嵌入大小调整等功能，将数据存储需求减少 35 倍，使 AI 系统更高效、更具成本效益。与合作伙伴 DataStax、Cohesity、Cloudera 和 SAP 等的协作进一步推动了生成式 AI 应用的开发和扩展。

NeMo Retriever 在实际应用中展现出强大的性能，例如与 DataStax 合作，Wikipedia 将 1000 万条 Wikidata 条目矢量化为 AI 可处理的格式，仅需不到三天的时间，相较以往的 30 天大幅提升了效率。这种技术支持更实时的知识更新和多语言的全球访问，帮助企业在复杂文档处理（如合同或医疗记录）中实现准确、高效的信息提取和决策优化。

通过与 NVIDIA 合作的生态系统，企业能够轻松部署生成式 AI 解决方案，结合自身的独特需求和资源，构建高效的多语言信息检索系统。开发者可以通过 NVIDIA API 目录或免费提供的 NVIDIA AI Enterprise 开发者许可证，访问包括 NeMo Retriever 在内的微服务，快速开发和优化用于搜索、问答、摘要和推荐等领域的 AI 应用，为全球化业务提供更精准和可扩展的解决方案。

图4: NVIDIA 推出的 NeMo Retriever 微服务



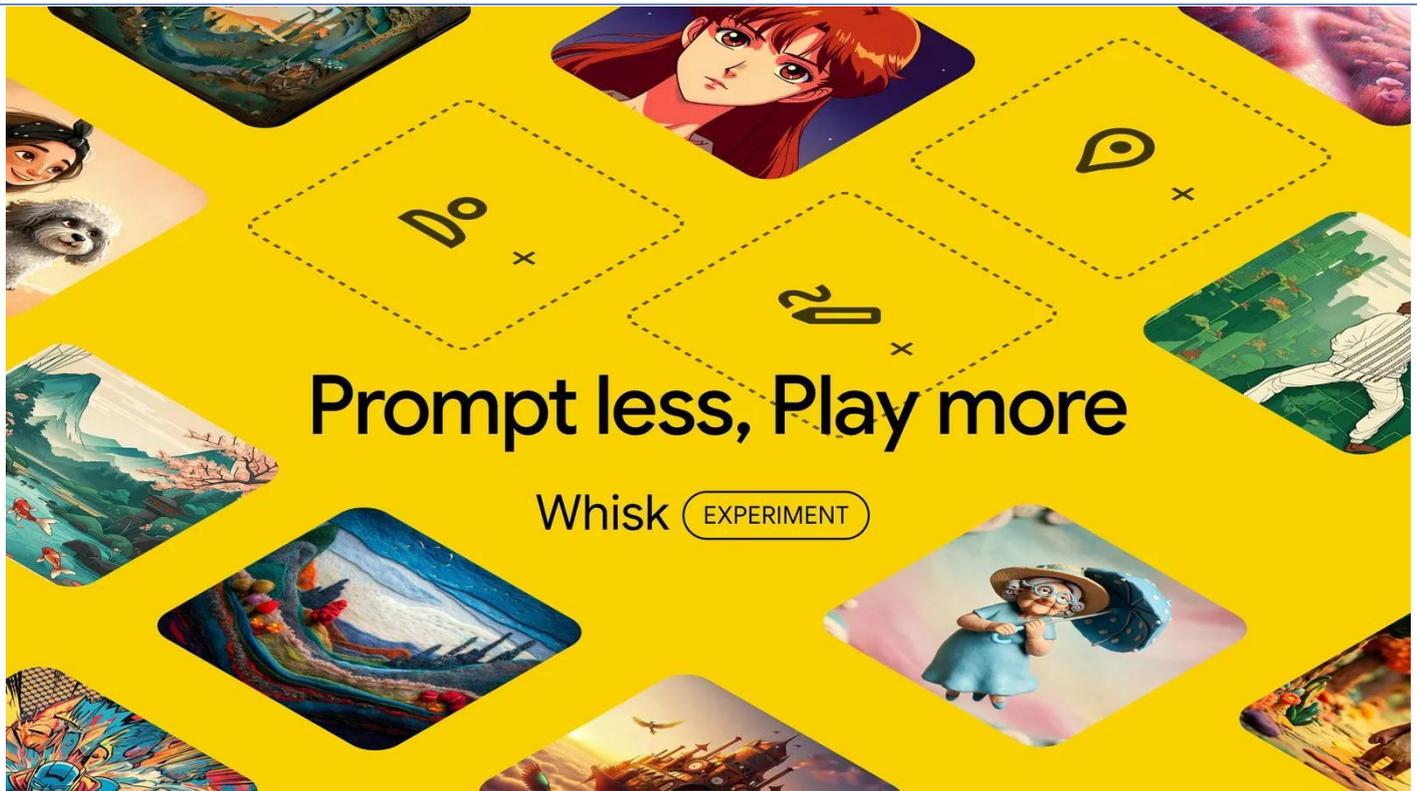
资料来源: NVIDIA 官网, 中国银河证券研究院

#### 【Whisk: 使用图像和人工智能将想法可视化并重新混合】

NVIDIA 推出了新的生成式人工智能工具 Whisk, 它通过图像提示而不是传统的文字描述, 帮助用户创作独特的视觉作品。用户只需拖入主题、场景和风格图片, Whisk 会自动将这些元素混合并生成全新的图像。该工具利用了 Google 的 Gemini 模型和 Imagen 3 图片生成模型, 能够捕捉图像的核心特征, 并以创意和新颖的方式重新组合它们。无论是创作数字毛绒玩具、珐琅别针, 还是奇幻场景, Whisk 都能提供轻松的图像生成体验, 适用于各种创意项目。

在实际使用中, Whisk 允许用户通过简单的拖放操作, 将不同类型的图像输入并调整主题、场景和风格, 生成多样化的视觉效果。虽然生成的图像可能与用户预期有所不同, 尤其在细节如人物的身高、体型或肤色等方面, 但用户可以通过查看和编辑底层提示来调整结果。Whisk 被描述为一种新型的创意工具, 旨在快速探索创意, 而非进行像素级的编辑, 允许用户尝试多种选项并下载最终作品。它提供了一种更加灵活和创意驱动的方式来进行视觉创作, 适合艺术家和创意人员进行视觉探索和灵感激发。

图5: NVIDIA 推出的生成式人工智能工具 Whisk



资料来源：谷歌官网，中国银河证券研究院

#### 【OpenAI 首日连更：发布“满血版”o1 模型和高端 ChatGPT Pro 订阅服务】

在 OpenAI 为期 12 天的“12 天 12 场直播”活动首日，公司推出“满血版”推理模型 o1 和全新的 ChatGPT Pro 订阅服务。“满血版”o1 模型相较于之前的 o1-preview 版本，在智能水平、处理速度和多模态能力上都有显著提升，特别是在数学、代码和博士级别科学问题上，智能水平提升 40%-50%，处理速度加快，重大错误率降低 34%。同时，新模型支持多模态输入，能够处理图像并生成更详细的响应。ChatGPT Pro 订阅服务则以 200 美元/月的价格提供无限制访问 o1 模型、o1-mini 和高级语音模式，以及 o1 pro 模式，后者使用更多计算资源进行深入思考，解答更难的问题。

#### 【Sora: OpenAI 的革命性 AI 视频工具即将正式发布】

OpenAI 的 Sora AI 视频工具预计将在 2024 年 12 月 5 日开始的为期 12 天的活动中正式发布。这一系列直播活动将从太平洋时间 12 月 5 日上午 10 点（北京时间 12 月 6 日凌晨 2 点）开始，每个工作日都会进行一次直播，展示一个新产品或者新样品。

Sora 是 OpenAI 即将发布的一款备受瞩目的 AI 视频工具，其核心功能是通过简单的文字描述生成高质量的视频内容。例如，用户输入“日落时的海滩”，Sora 便能自动创作出符合该描述的视频。这一技术对于创作者和内容生产行业来说具有重大意义，它不仅能显著提升生产效率，还能释放更多的创意空间。预计 Sora 将在短视频、广告和影视等领域产生深远的影响。

尽管在 Alpha 测试阶段，部分艺术家因版权争议和未获补偿等问题对 OpenAI 提出抗议，但这并未影响 Sora 技术能力的期待值。

### 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