

OpenAI 公布模型 Voice Engine, AI 安全问题再受关注

计算机行业

推荐

维持评级

摘要 | 03.22-03.29

- 休市情况:** 美股及港股 3 月 29 日全天休市; 沪、深股通(北向)、港股通(南向), 3 月 29 日暂停服务。
- 股指动态:** 美股科技呈下跌走势, 恒生科技指数小幅上涨。标普 500 指数+0.39%, 纳斯达克综合指数-0.30%, 费城半导体指数-0.06%; TAMAMA 科技指数-1.43%; 纳斯达克中国金龙指数-0.30%; 恒生科技指数+0.63%; 计算机-7.03%。
- 个股表现:** 热门科技股表现分化。据统计, 相比 3 月 22 日收盘价, 28 日盘后, 苹果合计-0.46%, 英伟达-4.17%, 特斯拉+2.90%, 谷歌+0.11%, 亚马逊+0.84%, META-4.71%, 微软-1.87%, ARM-6.83%, 英特尔+3.76%, 高通-0.47%, AMD+0.47%。
- 10 年期国债及汇率:** 周内, 中国 10 年期国债利率下降至 2.29%, 累计下跌 1.50bps; 3 月 28 日, 美国 10 年期国债利率下降至 4.20%, 累计下跌 2bps。3 月 29 日, 美元兑人民币中间价报 7.10; 较 3 月 22 日价累计调升 54 个基点。
- 政策与法规:** 美国政府扩大对华芯片限制。当地时间 3 月 29 日, 美国商务部工业和安全局(BIS)公告《实施额外出口管制》, 内容包括对超级计算机和半导体终端用途、及对半导体制造项目的出口管制的更正和澄清。新规则长达 166 页, 将于 4 月 4 日起生效, 明确对向中国出口芯片的限制包含 AI 芯片的笔记本电脑。

核心观点

OpenAI 于 3 月 29 日公布 Voice Engine 模型, 15 秒样本即可复现原声。公司公布的模型介绍和语音样本示例显示, 模型可通过文本输入和一个 15 秒的音频样本生成接近原始说话者的语音。公司表示, 出于安全型考虑现阶段大众还不能使用 Voice Engine。我们认为, 这一技术的实现为语音合成领域带来了革命性突破。但与此同时, 人工智能安全风险加剧, Voice Engine 模型能力可能触及的争议包括: 1) 隐私和数据安全问题; 2) 身份盗用和欺诈问题; 3) 信息真实性失去保障; 4) 知识产权相关问题; 5) 技术滥用问题等。伴随 AI 大模型的迅猛发展, 多模态模型有望加速在娱乐影视及内容创作、广告及营销、实时翻译、AI+教育、AI+医疗、虚拟客服等场景的落地。

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

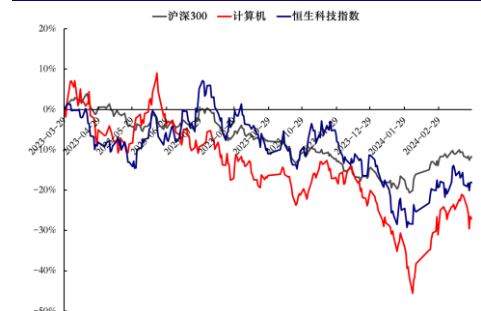
☎: (8610) 66568589

✉: wuyanqing@chinastock.com.cn

分析师登记编码: S0130519070001

国内表现

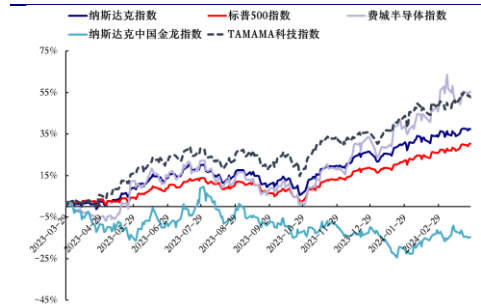
2024-03-29



资料来源: Wind, 中国银河证券研究院

全球行情

2024-03-28



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.....	6
（三）政策与法规.....	10
三、风险提示.....	10

一、全球市场表现

(一) 股市动态

休市情况：美股及港股 3 月 29 日全天休市；沪、深股通（北向）、港股通（南向），3 月 29 日暂停服务。

美股科技呈下跌走势，恒生科技指数小幅上涨。标普 500 指数+0.39%，纳斯达克综合指数-0.30%，费城半导体指数-0.06%；TAMAMA 科技指数-1.43%；纳斯达克中国金龙指数-0.30%；恒生科技指数+0.63%；计算机-7.03%。

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	0.39	2.29	3.10	10.16	24.23	26.19
IXIC.GI	纳斯达克指数	-0.30	2.85	1.79	9.11	43.42	42.12
SOX.GI	费城半导体指数	-0.06	3.16	3.77	17.48	64.90	49.94
8884057.WI	TAMAMA科技指数	-1.43	3.77	2.19	13.02	67.81	36.55
HXC.GI	纳斯达克中国金龙指数	-0.30	-2.78	-2.07	-5.55	-3.39	20.10
HSTECH.HI	恒生科技指数	0.63	-2.65	1.35	-7.62	-8.83	21.16
CI005027.WI	计算机	-7.03	3.51	0.98	-8.79	8.90	84.46

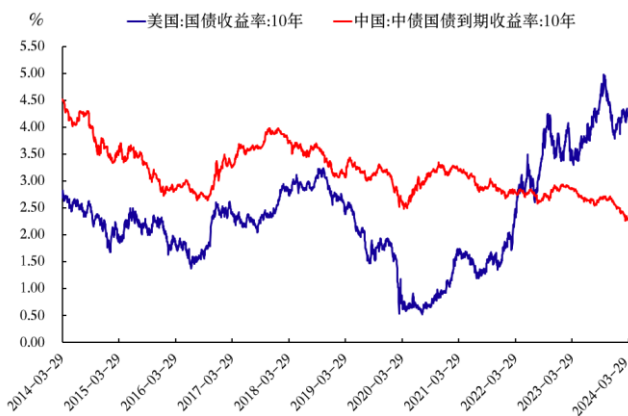
资料来源：Wind，中国银河证券研究院

(二) 债市及汇率情况

债市：周内，中国 10 年期国债利率下降至 2.29%，累计下跌 1.50bps；3 月 28 日，美国 10 年期国债利率下降至 4.20%，累计下跌 2bps。

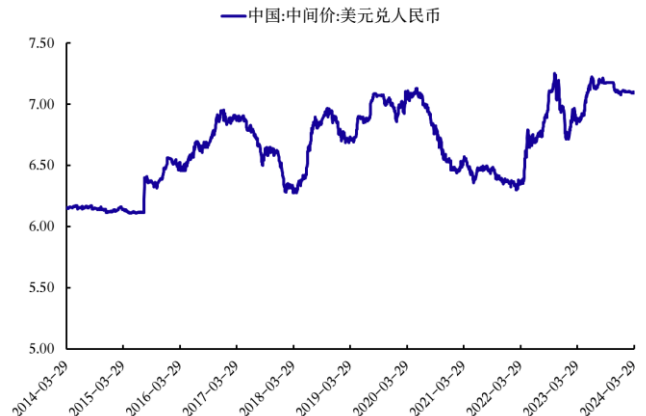
汇率：3 月 29 日，美元兑人民币中间价报 7.10；较 3 月 22 日价累计调升 54 个基点。

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 3 月 22 日收盘价，28 日盘后，苹果合计-0.46%，英伟达-4.17%，特斯拉+2.90%，谷歌+0.11%，亚马逊+0.84%，META-4.71%，微软-1.87%，ARM-6.83%，英特尔+3.76%，高通-0.47%，AMD+0.47%。

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) [2024-03-28]	市销率 PS (TTM)	市盈率 PE (TTM)	PE(2022)	PE(2023)	PE(2024E)
				[2024-03-22]	[2024-03-28]							
算力及终端	美股	AAPL.O	苹果(Apple)	172.28	171.48	-0.46	26,479.74	6.87	26.24	20.71	30.87	25.92
		NVDA.O	英伟达(NVIDIA)	942.89	903.56	-4.17	22,589.00	37.08	75.90	36.86	280.04	79.85
		TSLA.O	特斯拉(TESLA)	170.83	175.79	2.90	5,598.54	5.79	37.33	30.98	52.67	48.14
		HPQ.N	惠普(HP)	30.05	30.22	0.57	295.70	0.56	8.66	8.24	9.14	0.00
		CSCO.O	思科(CISCO)	49.78	49.91	0.26	2,020.95	3.53	15.03	16.57	16.28	0.00
		ASML.O	阿斯麦	979.96	970.47	-0.97	3,829.37	12.52	44.02	36.70	34.34	47.11
		AMD.O	超威半导体(AMD)	179.65	180.49	0.47	2,916.97	12.86	341.57	79.12	278.85	69.14
		INTC.O	英特尔(INTEL)	42.57	44.17	3.76	1,880.26	3.47	111.32	13.61	125.43	45.00
		QCOM.O	高通(QUALCOMM)	170.10	169.30	-0.47	1,889.39	5.21	24.34	9.51	22.36	19.62
		ARM.O	ARM	134.15	124.99	-6.83	1,284.99	43.74	1,511.75	0.00	147.03	0.00
		ON.O	安森美半导体 (ON SEMICONDUCTOR)	74.68	73.55	-1.51	314.30	3.81	14.39	14.18	16.47	16.86
		港股	0909.HK	明源云	2.52	2.44	-3.17	47.39	2.59	-7.33	-10.65	-8.66
9698.HK	万国数据-SW		8.01	6.37	-20.47	97.11	0.88	-2.03	-21.99	-2.92	-5.24	
1686.HK	新意网集团		2.55	2.55	0.00	59.65	2.35	6.57	11.66	7.98	0.00	
台股	2330.TW	台积电	785.00	769.00	-2.04	221,693.85	9.22	23.78	11.71	18.34	20.00	
	2454.TW	联发科	1,125.00	1,160.00	3.11	18,555.63	4.28	24.10	8.46	21.09	0.00	
互联网	美股	GOOGL.O	谷歌(ALPHABET)-C	150.77	150.93	0.11	18,842.06	6.10	25.43	19.04	23.69	22.19
		AMZN.O	亚马逊(AMAZON)	178.87	180.38	0.84	18,736.76	3.26	61.58	-314.82	51.61	44.13
		META.O	脸书 (META PLATFORMS)	509.58	485.58	-4.71	12,379.40	9.18	31.66	13.75	23.27	25.50
		NFLX.O	奈飞(NETFLIX)	628.01	607.33	-3.29	2,628.28	7.79	48.60	29.21	39.40	36.59
		PDD.O	拼多多	122.99	116.25	-5.48	1,544.48	4.42	18.22	22.77	22.94	12.80
		NTES.O	网易	105.52	103.47	-1.94	670.80	4.57	16.06	16.34	14.46	15.73
		BIDU.O	百度	102.18	105.28	3.03	368.24	1.94	12.87	36.42	14.52	11.70
		TCOM.O	携程网	45.02	43.89	-2.51	301.59	4.77	21.42	116.72	17.58	23.66
		BABA.N	阿里巴巴	72.13	72.36	0.32	1,829.46	1.41	13.05	23.78	18.64	12.19
	港股	9988.HK	阿里巴巴-SW	71.00	70.25	-1.06	14,312.90	1.40	12.97	23.80	18.53	12.10
		0700.HK	腾讯控股	288.80	303.80	5.19	28,745.35	4.27	22.61	15.17	21.90	16.54
		80700.HK	腾讯控股-R	268.60	282.60	5.21	26,067.72	4.39	23.21	0.00	21.86	16.96
		9999.HK	网易-S	166.80	162.80	-2.40	5,248.06	4.60	16.17	16.52	13.96	15.82
		9888.HK	百度集团-SW	98.20	102.70	4.58	2,880.96	1.94	12.85	36.49	14.48	11.67
		89888.HK	百度集团-SWR	91.25	95.50	4.66	2,612.60	1.99	13.19	0.00	14.50	11.97
		1024.HK	快手-W	49.00	49.05	0.10	2,130.21	1.69	30.18	-19.96	32.60	16.06
		81024.HK	快手-WR	45.55	45.70	0.33	1,931.79	1.73	31.03	0.00	32.58	16.50
		9626.HK	哔哩哔哩-W	86.75	88.30	1.79	371.97	1.50	-6.99	-8.90	-7.41	-18.19
软件及应用	美股	2518.HK	汽车之家-S	50.85	51.80	1.87	263.86	3.33	12.72	14.79	12.86	12.06
		9898.HK	微博-SW	70.15	73.95	5.42	173.98	1.26	6.50	52.13	7.52	6.56
		MSFT.O	微软(MICROSOFT)	428.74	420.72	-1.87	31,261.33	13.74	37.87	24.58	38.62	35.57
		SNOW.N	SNOWFLAKE	159.03	161.60	1.62	540.07	19.24	-64.59	-67.89	-82.25	0.00
		ORCL.N	甲骨文(ORACLE)	127.79	125.61	-1.71	3,452.41	6.57	32.44	32.81	34.08	26.09
		CRM.N	赛富时(SALESFORCE)	307.77	301.18	-2.14	2,921.45	8.38	70.63	91.82	1224.61	0.00
		ADBE.O	奥多比(ADOBE)	499.52	504.60	1.02	2,260.61	11.34	47.09	32.69	50.01	39.84
		INTU.O	财捷(INTUIT)	643.74	650.00	0.97	1,819.86	12.06	65.70	52.92	73.39	65.52
		SNPS.O	新思科技(SYNOPSYS)	594.20	571.50	-3.82	871.79	14.22	61.94	49.43	63.63	56.14
	港股	CDNS.O	铿腾电子(CADENCE)	322.74	311.28	-3.55	848.55	20.75	81.50	51.91	71.17	0.00
		ADSK.O	欧特克(AUTODESK)	262.86	260.42	-0.93	557.08	10.13	61.49	81.13	63.29	0.00
		U.N	Unity	26.99	26.70	-1.07	103.05	4.71	-12.47	-9.35	-18.80	0.00
		0020.HK	商汤-W	0.80	0.71	-11.25	237.63	5.82	-3.34	-10.98	-5.46	-6.11
		80020.HK	商汤-WR	0.75	0.66	-12.00	215.49	5.97	-3.43	0.00	-5.46	-6.27
		3888.HK	金山软件	25.00	24.10	-3.60	322.74	3.28	60.50	-5.25	61.69	26.63
		0268.HK	金蝶国际	9.40	8.83	-6.06	317.33	4.97	-137.01	-133.49	-178.41	-879.94
		9878.HK	汇通达网络	30.40	30.65	0.82	172.43	0.19	34.86	70.68	34.57	27.15
		3650.HK	KEEP	3.99	4.06	1.75	21.34	0.90	1.75	0.00	6.00	-8.26
港股	0354.HK	中国软件国际	4.96	4.72	-4.84	137.21	0.73	17.43	24.18	22.39	11.01	
	1357.HK	美图公司	3.50	3.28	-6.29	148.74	4.88	35.63	57.75	38.62	26.18	
	3896.HK	金山云	1.83	1.66	-9.29	63.17	0.81	-2.63	-2.69	-3.18	4.39	
	2013.HK	微盟集团	2.05	1.92	-6.34	53.66	2.18	-6.41	-8.29	-9.62	-41.96	
	1675.HK	亚信科技	7.50	7.17	-4.40	67.06	0.77	11.40	13.73	13.69	6.67	
	2121.HK	创新奇智	6.73	7.34	9.06	41.47	2.15	-6.45	-31.65	-8.15	0.00	
	2400.HK	心动公司	16.82	15.90	-5.47	76.39	2.04	-83.37	-16.75	-53.48	23.48	
	0777.HK	网龙	12.36	10.96	-11.33	58.23	0.72	9.59	9.71	10.84	5.31	

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

【美国起诉苹果非法垄断智能手机】

美国司法部对苹果在智能手机市场非法垄断提起诉讼, 拜登政府和 16 个州指控苹果非法滥用 iPhone 在智能手机市场的垄断权力。

政府指出涉嫌非法维持其垄断地位的几种不同方式包括:

1) **iMessage**: 有意降低向 Android 手机发送 iMessage 的质量, 包括加载速度延迟, 错过表情符号响应、编辑功能和端到端加密等;

2) **Apple Pay**: 苹果以安全为由, 不允许第三方应用程序访问 iPhone 的移动支付芯片;

3) **Apple Watch**: 与 Android 手机不兼容, 司法部声称这是故意将客户锁定在苹果的硬件和软件生态系统中, 迫使 Apple Watch 用户购买 iPhone;

4) **应用商店**: iPhone 上获取应用程序的唯一方法是通过苹果专有的应用商店, 美国司法部认为其迫使应用程序受到繁重限制和 30% 佣金。

5) **超级应用**: 苹果防止应用程序成为“超级应用程序”(可在 iOS、Android 操作系统上同时运行的应用程序), 并限制“小程序”(应用程序中的应用程序)。美国司法部表示, 该要求旨在将开发人员锁定在苹果的系统。

【英特尔推动 AI PC 加速计划, 助力亿级设备智能化】

英特尔于 2024 年 3 月 6 日宣布, 作为 AI PC 加速计划的一部分, 将推出两项新的人工智能计划: AI PC 开发者计划和将独立硬件供应商纳入 AI PC 加速计划。这两项计划是英特尔为了在 2025 年前在 1 亿台应用英特尔技术的 AI PC 上优化 AI 应用。

新计划旨在: AI PC 开发者计划专为软件开发人员和独立软件供应商 (ISV) 设计, 能够为开发人员提供顺畅的体验, 并使开发人员能够轻松大规模采用新的 AI 技术。AI PC 开发人员计划允许访问一系列工具、工作流、AI 部署框架以及搭载英特尔酷睿至尊处理器 (i9-10980XE) 的英特尔硬件的开发套件。更新后的开发人员资源页面可以为开发人员提供了便捷的一站式访问服务, 其中涵盖 AI PC 相关以及以客户端为中心的工具包、文档和培训材料。这些资源能帮助开发者充分发挥英特尔酷睿至尊处理器 (i9-10980XE) 的技术潜力, 提升 AI 及机器学习应用的性能, 并快速实现新的应用场景。

通过将独立硬件供应商 (IHV) 纳入 AI PC 加速计划, 英特尔为这些供应商提供了准备、优化及使其硬件适配英特尔 AI PC 的条件。合格的合作伙伴能够使用英特尔的开放实验室, 在硬件解决方案和平台的开发初期阶段获得技术支持和协作工程服务。此外, 计划还包括提供参考硬件, 使合作伙伴能够测试和完善自己的技术, 确保在推向市场时能达到最高效率。

新计划重要性: 1) 借助由英特尔平台 (中央处理器、神经处理单元和图形处理单元) 提供支持的优化软件和硬件, 让在 AI PC 上使用 AI 普及化; 2) 改进兼容性: 通过使用最新的英特尔酷睿至尊处理器 (i9-10980XE) 套件以及优化和软件工具, 开发人员可以确保他们的软件 and 应用程序在最新的英特尔处理器上平稳运行, 从而增强兼容性和整体最终用户体验; 3) 优化性能: 在开发周期的早期针对特定硬件架构优化软件可以提高性能和效率; 4) 扩大市场机会和全球规模: 与英特尔及其人工智能合作伙伴生态系统合作, 可以提供在生态系统中扩展以及进入新市场和跨行业领先的机会。

【微软和 OpenAI 筹划“星际之门”超算计划】

据 The Information 在 2024 年 3 月 29 日的报道, OpenAI 正在制定一个数据中心项目的计划, 该项目可能耗资达 1000 亿美元, 其中包括一台名为“星际之门 (Stargate)”的人工智能超级计算机, 该项目预计将于 2028 年推出。报道称, 微软可能会为该项

目提供资金，预计该项目的成本将比一些最大的现有数据中心高出 100 倍。The Information 将 1000 亿美元的暂定成本来源于一位与 Altman（Open AI CEO）交谈过的人，以及一位查看过微软的一些初始成本估算的人。报告称，Microsoft 和 OpenAI 正处于五阶段计划的第三阶段，未来两个阶段的很大一部分成本涉及购买所需的人工智能芯片。并且新项目将设计成可以使用来自不同供应商的芯片。

【微星推出新款 SSD：Spatium M580 Frozr】

微星于 3 月 28 日推出新款 SSD：Spatium M580 Frozr。这款固态硬盘搭载 E26 PCIe Gen 5 控制器，配备 232 层 3D NAND 闪存模块，可实现 14.6 GB/s 的读取速度和 12.7 GB/s 的写入速度，与当前的 PCIe 4.0 SSD 相比，传输速度快 2 倍。归功于这款 SSD 实现了在 DRAM 缓存缓冲区和 SLC 缓存的性能提升。此外，M580 FROZR 还配备了 LPDC ECC 和 E2E 数据保护功能，确保数据完整性和可靠性。

MSI 还提供内置于 MSI Center 中的软件解决方案，具有系统状态监控和数据备份功能。使用 MSI Center 时，用户可以看到有关 SSD 状态的全面信息，包括驱动器运行状况、已用容量和当前工作温度。

（二）大模型及云应用

【xAI 发布 Grok-1.5，编码和数学能力提升】

xAI 于 3 月 28 日发布 Grok-1.5。Grok-1.5 最显著的改进是它在编码和数学相关任务中的表现。Grok-1.5 在 MATH 基准测试中取得了 50.6% 的得分，在 GSM8K 基准测试中取得了 90% 的得分，这两个数学基准测试涵盖了从小学到高中的各种竞赛问题。此外，在评估代码生成和解决问题能力的 HumanEval 基准测试中，Grok-1.5 获得了 74.1% 的得分。在与竞品比较上表现突出。

图3: Grok-1.5 在各种测试的得分以及竞品对比

Benchmark	Grok-1	Grok-1.5	Mistral Large	Claude 2	Claude 3 Sonnet	Gemini Pro 1.5	GPT-4	Claude 3 Opus
MMLU	73% 5-shot	81.3% 5-shot	81.2% 5-shot	75% 5-shot	79% 5-shot	83.7% 5-shot	86.4% 5-shot	86.8 5-shot
MATH	23.9% 4-shot	50.6% 4-shot	—	—	40.5% 4-shot	58.5% 4-shot	52.9% 4-shot	61% 4-shot
GSM8K	62.9 8-shot	90% 8-shot	81% 5-shot	88% 0-shot CoT	92.3% 0-shot CoT	91.7% 11-shot	92% 5-shot	95% 0-shot CoT
HumanEval	63.2% 0-shot	74.1% 0-shot	45.1% 0-shot	70% 0-shot	73% 0-shot	71.9% 0-shot	67% 0-shot	84.9% 0-shot

资料来源: xAI 官网, 中国银河证券研究院

在长文理解上，Grok-1.5 的上下文窗口中处理达到了 128K。这使得 Grok-1.5 的文本处理容量增加到之前版本上下文长度的 16 倍，从而能够处理更长文档中的信息。此外，该模型还能处理更长、更复杂的指令，同时在扩展上下文窗口时仍能保持指令跟踪能力。在“大海捞针”（Needle In A Haystack, NIAH）评估中，Grok-1.5 可检索长度达到 128K 字节中的嵌入文本。

Grok-1.5 的基础架构建立在 JAX、Rust 和 Kubernetes 的定制分布式训练框架之上。这种训练堆栈够以最小的投入，进行大规模地新架构训练。在大型计算集群上训练 LL

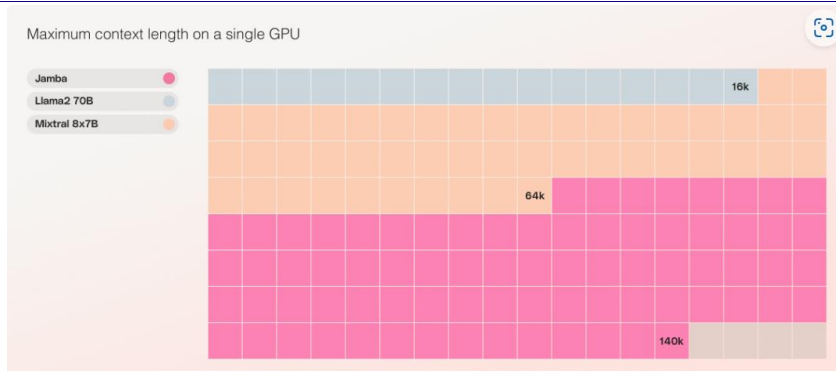
的主要挑战是最大限度提高训练工作的可靠性和正常运行时间。XAI 定制的训练协调器可确保自动检测有问题的节点，并将其从训练任务中剔除。此外，xAI 团队优化了检查点、数据加载和训练任务的重启，以尽量减少故障停机时间。

【 AI21 Labs 推出全球首款 Mamba 架构模型 Jamba 】

AI21 Labs 于 3 月 28 日推出全球首款基于 Mamba 架构的生产级模型 Jamba。Jamba 用传统的 Transformer 架构元素增强了 Mamba 结构化状态空间模型(SSM)技术，弥补了纯 SSM 模型的固有局限性。

Jamba 的主要特点：1) 首个基于新型 SSM 变压器混合架构的 Mamba 生产级机型；2) 长上下文吞吐量是 Mixtral 8x7B 的 3 倍；3) 256K 超大上下文窗口；4) 在其规模级别中，唯一可在单个 GPU 上容纳多达 140K 上下文的模型 5) 模型开源，可在 "Hugging Face "网站上下载。

图4: Jamba 能够在单个 GPU 上处理 140K 的上下文



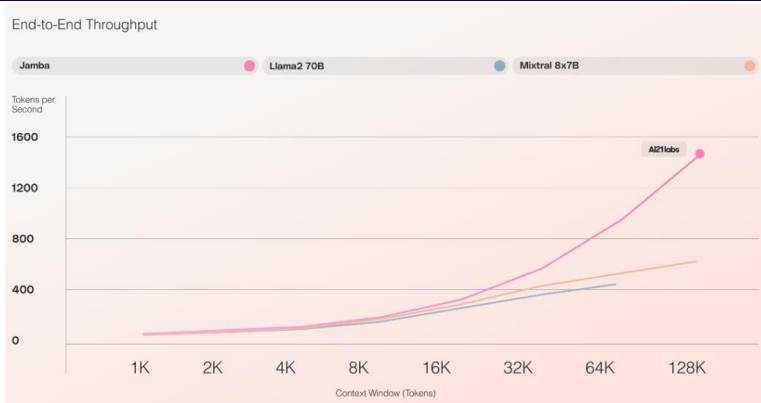
资料来源: AI21 Labs 官网, 中国银河证券研究院

到目前为止，LLM 主要建立在传统的 Transformer 架构上，这种架构存在两个主要缺点：1) 内存占用量大：Transformer 的内存占用量随上下文长度而扩展，这使得在没有大量硬件资源的情况下运行长上下文窗口或大量并行批处理变得困难，从而限制了广泛的试验和部署；2) 由于 Transformer 的注意力机制与序列长度的平方成正比，随着上下文的增长，推理速度变慢，原因在于每个标记都依赖于它之前的整个序列，这使得长上下文的使用案例超出了高效生产的范围。

Mamba 优化了传统 Transformer 架构缺点。Jamba 由 Transformer、Mamba 和 MoE 层组成，可同时优化内存、吞吐量和性能。Jamba 的 MoE 层允许它在推理时仅利用其可用的 52B 参数中的 12B，这种混合结构使这些 12B 有源参数比同等大小的仅使用 Transformer 模型更有效。在此之前从没有公司或机构将 Mamba 缩放到 3B 参数以上，Jamba 是同类产品中第一个达到生产级规模的混合架构。

在处理长上下文时，Jamba 的吞吐量是 Transformer 基础模型（如 Mixtral 8x7B 这类相似大小的模型）的 3 倍，比 Transformer 基础模型更高效。

图5: 在处理长上下文时, Jamba 是 Transformer 基础模型的 3 倍

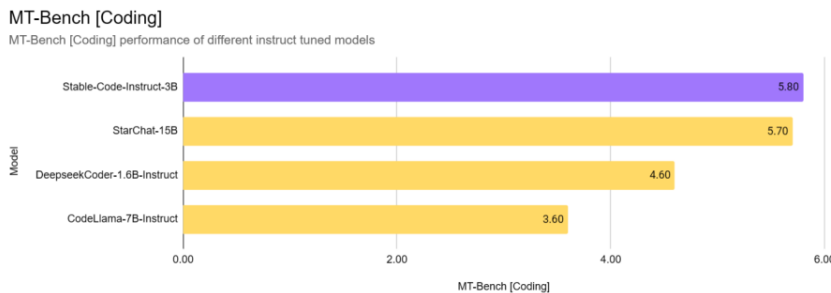


资料来源: AI21 Labs 官网, 中国银河证券研究院

【Stability AI 发布 Stable Code Instruct 3B 模型, 表现优于同类等参数模型】

Stability AI 于 3 月 25 日发布了新的代码模型 Stable Code Instruct 3B。Stable Code Instruct 3B 是 Stability AI 在 Stable Code 3B 基础上建立的新的指令调优大型语言模型。该模型增强了代码完成并支持自然语言交互, 能够提高编程和软件开发相关任务的效率和直观性。分析表明, Stable Code Instruct 3B 在各种与编码相关的任务中优于 Codellama 7B Struct 和 DeepSeek-Coder Instruct 1.3B 等同类模型。

图6: Stable Code Instruct 3B 与 Codellama 7B Instruct 和 DeepSeek-Coder Instruct 1.3B 等模型比较



资料来源: Stability AI 官网, 中国银河证券研究院

【OpenAI 发布 Voice Engine 模型, 15 秒样本即可复现原声】

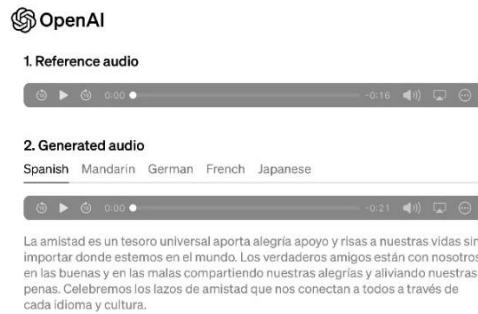
OpenAI 于 3 月 29 日在其官方网站公布了名为 Voice Engine 模型的介绍和语音样本示例。该模型可通过文本输入和一个 15 秒的音频样本生成接近原始说话者的语音。公司表示出于安全考虑, 现阶段 Voice Engine 未对大众开放使用。

OpenAI 于 2023 年底开始与合作伙伴进行测试。一些早期的应用包括: 1) 通过自然、富有感情的语音, 为尚未掌握阅读技能的儿童提供阅读帮助; 2) 翻译视频和播客等内容, 让创作者能够用自己的声音流利地向全世界更多的人交流; 3) 改善偏远地区的基本服务, 如通过 voice engine 用当地小众语言与当地当地人交流; 4) 为失语者提供支持, 例如为患有影响语言的疾病的人提供治疗应用, 以及为有学习需求的人提供教育支持; 5) 为患有突发性或退行性言语疾病的患者恢复声音。

展望未来, 为了面对 Voice Engine 可能带来的风险, OpenAI 提出了以下意见: 1) 逐步淘汰基于语音的身份验证作为访问银行账户和其他敏感信息的安全措施; 2) 教育公众了解人工智能技术的能力和局限性, 包括欺骗性人工智能内容的可能性; 3) 加速开发跟踪视听内容来源的技术, 以便分辨真人或人工智能; 4) 制定保护人工智能中个人声音使用的政策。

我们认为，这一技术的实现为语音合成领域带来了革命性突破。但与此同时，人工智能安全风险加剧，Voice Engine 模型能力可能触及的争议包括：1) 隐私和数据安全问题；2) 身份盗用和欺诈问题；3) 信息真实性失去保障；4) 知识产权相关问题；5) 技术滥用问题等。伴随 AI 大模型的迅猛发展，多模态模型有望加速在娱乐影视及内容创作、广告及营销、实时翻译、AI+教育、AI+医疗、虚拟客服等场景的落地。

图7: Voice Engine 可以翻译视频和播客等内容



资料来源: OpenAI 官网, 中国银河证券研究院

【Databricks 推出 DBRX 通用开源大型语言模型】

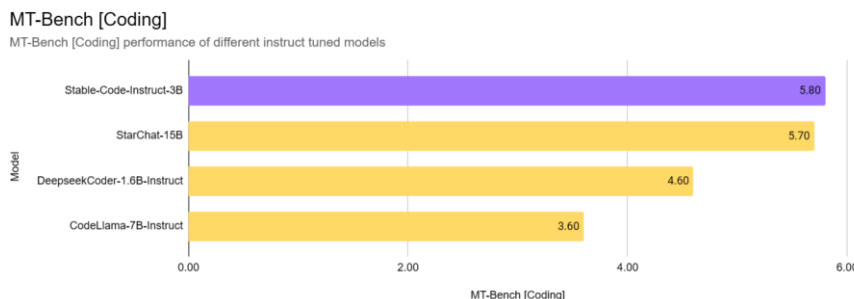
Databricks 于 3 月 27 日推出了 DBRX，一个通用的大型语言模型。DBRX 在标准测试中优于所有现有的开源模型以及一些专有模型。

DBRX 使用“专家混合 (mixture-of-experts)”架构 (MoE 架构)，这种架构将学习过程划分为多个称为“专家”的专用子网络，每个“专家”都精通指定任务的特定方面，由“门控网络 (gating network)”决定了如何在专家之间以最佳方式分配输入数据。

MoE 架构建立在 MegaBlocks 开源项目之上。Databricks 表示，与其他 MoE 架构相比，DBRX 项目的训练速度可以提高一倍以上，计算效率最高可达两倍。它目前仅适用于文本，不适用于视频和图像 (非多模态模型)。

Databricks 表示，虽然 DBRX 的参数是 Llama2 的两倍，有 1320 亿个参数，但它的推理速度也是 Llama 2 的两倍。Databricks 声称 DBRX 在语言理解、编程、数学和逻辑基准测试方面优于现有的开源 LLMs Llama 2 70B 和 Mixtral-8x7B (另一种采用 MoE 方法的模型) 和专有模型 GPT-3.5。据 Databricks 表示，该模型耗资 1000 万美元，使用 3,000 个 H100 GPU 在两个月的时间内对公共和许可数据源进行了训练。

图8: DBRX 与其他开源模型的对比



资料来源: Databricks 官网, 中国银河证券研究院

【AI 初创公司 Suno AI 发布可制作广播级音乐的生成模型】

AI 初创公司 Suno AI 于 3 月 22 日公布其可制作广播级的音乐生成模型: Suno V3。V3 可以使用户在几秒钟内以任何主要语言制作一首歌，只需几个简短的词描述就可

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。

以制作出完整的两分钟歌曲。Suno AI 称其的核心用户群是第一次制作音乐的普通人，熟练的音乐家也可以操作。

V3 在 V3 Alpha 的基础上做出了一些改进，包括：1) 更好的音频质量；2) 更多风格和流派；3) 提高了指示要求的遵循性，减少幻觉现象，能够以更优雅的方式结束音乐制作。

在安全和音乐版权保护上。Suno AI 表示 V3 专为创作原创音乐而设计，模型不会引用其他音乐家的原创作品。此外，为了防止滥用，Suno AI 开发了专有的、听不见的水印技术，可以检测歌曲是否是使用 Suno 创作的。

Suno AI 表示 Suno V3 现在向所有用户免费开放，Suno V4 也正在研发当中。

(三) 政策与法规

【美国政府扩大对华芯片限制】

去年十月，拜登政府宣布停止向包括中国、伊朗和俄罗斯在内的国家出口英伟达和其他公司设计的人工智能芯片。**当地时间 3 月 29 日，美国商务部工业和安全局(BIS) 公告《实施额外出口管制》，内容包括对超级计算机和半导体终端用途、及对半导体制造项目的出口管制的更正和澄清。**新规则长达 166 页，将于 4 月 4 日起生效，明确对向中国出口芯片的限制包含 AI 芯片的笔记本电脑。

图9: 关于《实施额外出口管制》的公告



资料来源：联邦公报，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: Grok-1.5 在各种测试的得分以及竞品对比	6
图 4: Jamba 能够在单个 GPU 上处理 140K 的上下文.....	7
图 5: 在处理长上下文时, Jamba 是 Transformer 基础模型的 3 倍	8
图 6: Stable Code Instruct 3B 与 Codellama 7B Instruct 和 DeepSeek-Coder Instruct 1.3B 等模型的比较.....	8
图 7: Voice Engine 可以翻译视频和播客等内容.....	9
图 8: DBRX 与其他开源模型的对比.....	9
图 9: 关于《实施额外出口管制》的公告.....	10

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