



国产大模型进入验收期，端侧 AI 迎新纪元

2025 年 1 月 3 日

- **中央经济工作会议召开，12 月板块先扬后抑。** 人工智能板块指数（884201.WI）12 月涨跌幅为-1.66%，同期上证综指上涨 0.76%，沪深 300 上涨 0.47%，创业板指数下跌-3.71%。12 月中央经济工作会议提出要以科技创新引领新质生产力发展，建设现代化产业体系。加强基础研究和关键核心技术攻关，超前布局重大科技项目，开展新技术新产品新场景大规模应用示范行动。开展“人工智能+”行动，培育未来产业。受到相关预期及催化，板块前半月上月迎来上涨，后半月进入下行调整阶段。我们认为，12 月作为政策密集发布窗口期，对 25 年全年具有指导性意义，年底作为业绩空窗期，市场较此前活跃度有所下降，短期内出现一定波动，但不改中长期人工智能产业长期发展趋势。
- **国内头部科技公司持续加码 AI，大模型逐渐兑现商业价值。** 12 月 18 日，火山引擎冬季 FORCE 原动力大会中宣布字节跳动自研大模型豆包全面升级迭代，发布了视觉理解模型，具备更强的内容识别、理解、推理、视觉描述等能力，价格为 0.003 元/千 tokens，比行业平均价格降低 85%，截至 2024 年 11 月底，豆包 APP 累计用户规模超 1.6 亿，MAU 达到 5998 万，排名全球第二（仅次于 ChatGPU），国内第一；DAU 接近 900 万，增速保持两位数并超越 ChatGPT。此外，12 月 26 日小米宣布正在构建万卡 GPU 集群，并计划在大模型领域进行巨额投资，据悉小米的 AI 团队目前规模已超过 3000 人，其 GPU 资源已达到 6500 张。我们认为，豆包 APP 凭借其强大的算法和用户基础，在国内处于领先地位，源自于字节对于人工智能及算力领域庞大投资与积累，同时小米作为国内领先科技公司，在算力领域的投资巨大并且仍保持持续增长态势，验证 AI 应用商业逻辑，AI 应用有望加速落地并进一步推动算力基础设施高景气。
- **国内首款量产 AI 眼镜发布，2025 年迎来端侧 AI 元年。** 闪极科技于 12 月 19 日晚发布国内首款量产 AI 眼镜——闪极 AI 拍拍镜，支持如豆包、文心一言、星火、通义、混元、Kimi、智谱、商汤等数十家大模型接入，也支持接入私有 AI 模型。我们认为，大模型快速迭代升级推动端侧 AI 加速落地，端侧 AI 将成最佳落地场景，大模型赋能 PC、手机、TWS 耳机、可穿戴设备、玩具等终端应用，打开端侧 AI 蓝海市场，根据头豹研究院数据，2023 年中国终端 AI 市场规模为 1939 亿元，预计 2028 年中国端侧 AI 市场规模将达到 19071 亿元，CAGR 为 58%，1 月 7 日-10 日即将召开 2025 CES 展，众多科技企业或将发布创新端侧 AI 产品，2025 将成为端侧 AI 发展新纪元。
- **建议关注：**1、国产算力产业链及生态伙伴；2、算力基础设施产业链；3、逐步向国产的自主 AI 算力为底座迁移的 MAAS 及下游应用端；4、端侧 AI 产业链整机及代工、结构件、开发套件厂商；5、数据要素产业链中供给、流通、应用公司；6、自动驾驶产业链及车路云一体化厂商。建议关注：工业富联、中科曙光、曙光数创、海光信息、龙芯中科、科大讯飞、万兴科技、海康威视、中国软件、达梦数据、润泽科技、软通动力、能科科技、同花顺、财富趋势、中科创达、万集科技、金蝶国际、上海钢联、深桑达 A、云赛智联、易华录、云鼎科技、华勤技术、亿道信息等。

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

鲁佩

☎：(021) 20257809

✉：lupei_yj@chinastock.com.cn

分析师证书编码：S0130521060001

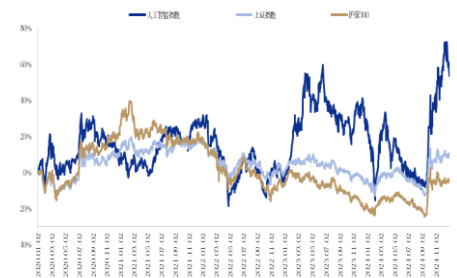
研究助理 胡天昊

☎：(8610) 80927637

✉：hutianhao_yj@chinastock.com.cn

相对沪深 300 表现图

2024-1-3



资料来源：Wind，中国银河证券研究院

- **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。

重点公司盈利预测与估值

股票代码	股票名称	EPS			PE			投资评级
		2023A	2024E	2025E	2023A	2024E	2025E	
002230.SZ	科大讯飞	0.28	0.26	0.42	172.57	185.85	115.05	推荐
688041.SZ	海光信息	0.54	0.76	1.05	277.39	197.09	142.66	推荐
688692.SH	达梦数据	5.19	4.58	5.61	-	70.22	79.57	-
300442.SZ	润泽科技	1.02	1.3	1.88	50.94	39.97	27.64	-
301236.SZ	软通动力	0.56	0.68	0.86	92.79	104.84	86.34	-

资料来源：Wind, 中国银河证券研究院

目录

Catalog

一、 市场行情回顾.....	4
(一) 整体行情	4
(二) 代表企业	4
(三) 板块估值	5
二、 人工智能产业动态.....	8
(一) 数据要素、数据交易所最新新闻及政策	8
(二) 算法端：国内外巨头大模型动态.....	10
(三) 算力端：AI 服务器、AI 芯片最新动态	11
三、 前沿行业动态.....	12
(一) 前沿技术动态	12
(二) 前沿政策动态	24
四、 前沿企业动态.....	25
(一) 前沿产品动态	25
(二) 投融资事件	29
五、 投资建议.....	31
六、 风险提示.....	31

一、市场行情回顾

(一) 整体行情

A 股人工智能指数 (884201.WI) 截至 12 月末收盘价为 8389.80, 月涨跌幅为-1.66%。计算机行业指数 (801750.SI) 截至 12 月末收盘价为 4407.29, 月涨跌幅为-7.07%。

图1: 12月人工智能指数走势图



资料来源: Wind, 中国银河证券研究院

(二) 代表企业

A 股 Wind 人工智能指数 (884201.WI) 截至 12 月 31 日总市值 19410.10 亿, 含成分股 73 支, 权重等分。上市板分布为主板 14 支, 创业板 30 支, 科创板 10 支, 中小板 19 支, 12 月 9 日将合信息 (688615.SH) 新增纳入人工智能指数, 公司是行业领先的人工智能及大数据科技企业, 基于自主研发的领先的智能文字识别及商业大数据核心技术, 为全球 C 端用户和多元行业 B 端客户提供数字化、智能化的产品及服务。

表1: 12月成分股涨幅前十

股票代码	股票简称	12月涨跌幅	12月31日收盘价(元)	相对计算机
300442.SZ	润泽科技	56.60%	51.96	63.67%
688322.SH	奥比中光-UW	46.69%	46.5	53.76%
000681.SZ	视觉中国	38.06%	20.82	45.13%
603893.SH	瑞芯微	30.57%	110.06	37.64%
301316.SZ	慧博云通	28.49%	25.57	35.56%
300383.SZ	光环新网	27.76%	14.59	34.83%
300613.SZ	富瀚微	24.31%	58.5	31.38%

300245.SZ	天玑科技	22.11%	15.63	29.18%
688343.SH	云天励飞-U	20.24%	49.6	27.31%
688256.SH	寒武纪-U	17.29%	658	24.36%

资料来源: Wind, 中国银河证券研究院

表2: 12月成分股跌幅前十

股票代码	股票简称	12月涨跌幅	12月31日收盘价(元)	相对计算机
600410.SH	华胜天成	-35.01%	7.24	-27.94%
300081.SZ	恒信东方	-27.91%	7.18	-20.84%
601360.SH	三六零	-22.76%	10.35	-15.69%
300364.SZ	中文在线	-19.26%	24.53	-12.19%
000409.SZ	云鼎科技	-19.00%	8.91	-11.93%
300078.SZ	思创医惠	-18.93%	3.04	-11.86%
301025.SZ	读客文化	-18.82%	9.53	-11.75%
300222.SZ	科大智能	-16.99%	9.97	-9.92%
002253.SZ	川大智胜	-16.80%	13.82	-9.73%
002235.SZ	安妮股份	-16.56%	5.24	-9.49%

资料来源: Wind, 中国银河证券研究院

(三) 板块估值

人工智能指数(884201.WI)重要成分股过去三年整体营业收入复合增长率1.90%，净利润复合增长率-8.41%，截至12月31日平均估值PE(TTM)96.44倍，PS(TTM)4.32倍。

图2: 12月人工智能指数市场表现



资料来源: Wind, 中国银河证券研究院

表3: 12月人工智能主题基金一览

基金代码	基金简称(官方)	基金类型	基金规模 (亿元)	11月30收盘价 (元)	近1月回报 (%)	近3月回报 (%)	近6月回报 (%)
001986.OF	前海开源人工智能	契约型开放式	7.19	1.49	5.59	6.34	16.84
005729.OF	南方人工智能主题	契约型开放式	4.15	2.19	2.49	4.83	17.98
005844.OF	东方人工智能主题A	契约型开放式	6.15	1.03	-12.53	7.99	27.78
005962.OF	宝盈人工智能A	契约型开放式	5.14	2.60	5.94	15.45	28.40
005963.OF	宝盈人工智能C	契约型开放式	2.06	2.47	5.87	15.22	27.88
006281.OF	万家人工智能A	契约型开放式	16.63	2.46	-4.05	6.07	16.96
008020.OF	华富中证人工智能产业ETF联接A	契约型开放式	2.25	0.85	-1.35	8.24	24.95
008021.OF	华富中证人工智能产业ETF联接C	契约型开放式	1.76	0.84	-1.37	8.15	24.76
008585.OF	华夏中证人工智能主题ETF联接A	契约型开放式	6.20	0.86	-0.38	9.63	25.41
008586.OF	华夏中证人工智能主题ETF联接C	契约型开放式	5.60	0.85	-0.40	9.55	25.22
009239.OF	融通中证人工智能主题C	契约型开放式	1.32	1.43	-0.05	9.57	25.19
011832.OF	西部利得中证人工智能A	契约型开放式	1.43	0.94	0.56	13.01	30.04
011833.OF	西部利得中证人工智能C	契约型开放式	1.01	0.93	0.53	12.89	29.78
011839.OF	天弘中证人工智能主题A	契约型开放式	1.92	0.95	-0.06	9.75	25.51
011840.OF	天弘中证人工智能主题C	契约型开放式	7.17	0.94	-0.09	9.70	25.38
012733.OF	易方达中证人工智能主题ETF联接A	契约型开放式	4.66	1.08	-0.29	10.15	26.10
012734.OF	易方达中证人工智能主题ETF联接C	契约型开放式	6.19	1.08	-0.30	10.13	26.04
014162.OF	万家人工智能C	契约型开放式	19.99	2.40	-4.11	5.86	16.50
014630.OF	汇添富中证人工智能主题联接A	契约型开放式	0.06	1.02	-	-	-
014631.OF	汇添富中证人工智能主题联接C	契约型开放式	0.06	1.02	-	-	-
017811.OF	东方人工智能主题C	契约型开放式	40.18	1.02	-12.56	7.88	27.53
021580.OF	华夏中证人工智能主题ETF联接D	契约型开放式	0.03	0.85	-0.40	9.56	25.24
159363.OF	华宝创业板人工智能ETF	契约型开放式	7.90	0.96	-	-	-
159702.OF	汇添富中证人工智能ETF	契约型开放式	0.13	0.79	-	-	-
159819.OF	易方达中证人工智能ETF	契约型开放式	65.74	0.91	-0.37	9.98	26.89
161631.OF	融通中证人工智能主题A	契约型开放式	6.10	1.46	-0.01	9.68	25.45
512930.OF	平安中证人工智能ETF	契约型开放式	9.51	1.30	-0.31	9.52	26.04
515070.OF	华夏中证人工智能ETF	契约型开放式	25.39	1.16	-0.45	9.89	26.75
515980.OF	华富中证人工智能产业ETF	契约型开放式	18.26	0.97	-1.27	9.02	26.87
517800.OF	方正富邦中证沪港深人工智能50ETF	契约型开放式	1.27	0.72	0.07	8.43	28.84
588733.OF	易方达上证科创板人工智能ETF	契约型开放式	-	-	-	-	-
588763.OF	广发上证科创板人工智能ETF	契约型开放式	-	-	-	-	-

资料来源: Wind, 中国银河证券研究院

表4: 人工智能主要上市公司近况一览 (数据截至 2024 年 12 月 31 日)

股票代码	股票名称	2023 营收 增速 (%)	2023 净利 润增速 (%)	24Q3 营 收增速 (%)	24Q3 净 利润增速 (%)	总市值 (亿元)	市盈率 PE (TTM)	市销率 PS (TTM)	月涨跌幅 (%)	今年以 来涨跌 幅 (%)
000977.SZ	浪潮信息	-5.41	-12.89	72.26	66.49	763.74	33.17	0.76	8.54	56.27
002230.SZ	科大讯飞	4.41	22.97	17.73	-1039.84	1117.03	521.38	5.10	-5.83	4.18
002236.SZ	大华股份	5.41	230.49	0.77	-3.97	527.19	7.21	1.63	-2.02	-13.28
002362.SZ	汉王科技	3.56	1.72	17.87	29.90	55.39	-	3.41	-5.11	-11.55
002405.SZ	四维图新	-6.72	-171.14	9.06	9.42	228.64	-	6.86	-1.33	8.31
002415.SZ	海康威视	7.42	11.78	6.06	-6.22	2834.59	21.21	3.05	2.06	-11.58
300229.SZ	拓尔思	-13.84	-72.98	2.95	82.98	182.94	315.20	22.88	-9.19	24.20
300474.SZ	景嘉微	-38.19	-79.35	-5.99	53.28	488.60	718.74	71.31	8.22	32.22
601360.SH	三六零	-4.89	77.66	-16.76	-56.39	724.45	-	9.14	-22.76	14.87
603019.SH	中科曙光	10.34	16.12	3.65	2.12	1058.19	57.04	7.23	-14.72	83.14
688088.SH	虹软科技	26.07	54.61	14.09	8.34	154.69	162.63	20.87	7.53	-6.04
688169.SH	石头科技	30.55	73.32	23.17	8.22	405.08	18.73	4.06	2.23	-22.50
688207.SH	格灵深瞳	-25.84	-379.64	-72.99	-684.03	36.88	-	37.49	-11.17	-32.06
688256.SH	寒武纪-U	-2.70	33.72	27.09	12.30	2746.87	-	366.80	17.29	387.55
688787.SH	海天瑞声	-35.33	-203.16	44.90	111.80	59.43	1076.01	27.47	16.55	36.64
688793.SH	倍轻松	42.30	59.50	-11.16	183.41	25.61	-	2.19	-5.64	-11.44
002410.SZ	广联达	-0.42	-88.22	-8.06	-17.59	194.27	287.77	3.14	-8.48	-31.39
688327.SH	云从科技-UW	19.33	28.26	-34.51	-23.69	125.47	-	24.66	-6.78	-28.23
688343.SH	云天励飞-U	-7.36	14.21	112.52	-41.51	176.15	-	23.12	20.24	0.71
688246.SH	嘉和美康	-3.04	-50.21	-11.52	-4748.76	34.12	-	5.39	-18.90	-25.79
603893.SH	瑞芯微	5.17	-54.65	48.47	354.90	460.38	112.49	16.21	30.57	73.60
300033.SZ	同花顺	0.14	-17.07	-1.59	-15.53	1545.60	120.49	43.83	-14.35	83.27
300496.SZ	中科创达	-3.73	-45.54	-4.70	-69.75	273.98	2250.72	5.41	8.19	-25.61
688111.SH	金山办公	17.27	16.23	10.90	17.23	1324.63	90.46	26.96	-2.83	-9.43
688475.SH	萤石网络	12.39	68.80	12.93	-6.69	237.75	44.35	4.49	-5.45	-32.76
300634.SZ	彩讯股份	25.18	40.17	10.72	-37.78	98.36	46.15	6.10	-9.95	6.50
300624.SZ	万兴科技	25.49	68.43	-3.91	-105.42	122.09	769.71	8.49	-0.43	-33.25
301162.SZ	国能日新	26.89	22.93	18.15	7.96	45.57	53.21	8.90	8.05	-12.30
688188.SH	柏楚电子	56.61	53.12	31.19	30.10	399.09	45.22	23.25	7.85	-23.25

资料来源: Wind, 中国银河证券研究院

表5: 境外上市人工智能企业近况一览 (数据截至 2024 年 12 月 31 日)

证券代码	证券简称	3Q24 营 业收入 (亿元)	3Q24 营 业收入同 比增长率 (%)	3Q24 归 母净利润 (亿元)	3Q24 归 母净利润 同比增长 率 (%)	总市值 (原始币 种, 亿 元)	市盈率 PE (TTM)	市销率 PS (TTM)	月涨跌幅 (%)	今年以 来涨跌 幅 (%)
TSLA.O	特斯拉	719.83	0.53	47.74	-32.47	12963.50	102.06	13.34	17.00	62.52

NVDA.O	英伟达	388.19	85.53	174.75	491.57	32887.62	52.14	29.03	-2.86	171.25
GOOGL.O	谷歌	2535.49	14.68	735.82	38.55	23172.21	24.58	6.82	12.17	36.01
META.O	脸书	1161.16	22.50	415.22	65.55	14781.13	26.61	9.46	2.03	66.05
MSFT.O	微软	1803.95	15.84	661.00	26.43	31338.02	34.62	12.33	-0.46	12.93
BIDU.O	百度	990.01	-0.65	185.68	4.81	295.64	9.79	1.55	-0.87	-29.20
AAPL.O	苹果	2961.05	0.79	790.00	6.70	37853.04	40.38	9.68	5.52	30.71
BABA.N	阿里巴巴	7192.94	8.90	766.44	55.97	2019.54	16.49	1.47	-2.95	10.81
0700.HK	腾讯控股	4878.11	7.49	1427.49	61.86	38467.90	20.43	5.39	4.77	43.26
0020.HK	商汤-W	-	-	-	-	551.41	-8.72	13.55	0.00	28.45
0268.HK	金蝶国际	-	-	-	-	305.88	-193.59	4.60	-2.63	-25.04

资料来源: Wind, 中国银河证券研究院

二、人工智能产业动态

(一) 数据要素、数据交易所最新新闻及政策

表6: 数据要素最新新闻及政策

日期	具体内容
12.30	<p>国家发改委、国家数据局、教育部、财政部、金融监管总局、证监会等六部门发布《关于促进数据产业高质量发展的指导意见》</p> <p>《关于促进数据产业高质量发展的指导意见》提出，到2029年，数据产业规模年均复合增长率超过15%，数据产业结构明显优化，数据技术创新能力跻身世界先进行列，数据产品和服务供给能力大幅提升，催生一批数智应用新产品新服务新业态，涌现一批具有国际竞争力的数据企业。在具体措施方面，指导意见提出，做强数据技术企业，培育一批面向数据采集汇聚、计算存储、流通交易、开发利用的技术创新型企业；支持企业面向数据流通交易提供专业化服务，重点围绕业务咨询、交易撮合、合规服务、金融服务等方面，培育一批数据服务企业，发展数据流通交易新模式新业态。壮大数据基础设施企业，重点培育一批具有国际竞争力的数据基础设施企业；支持企业围绕工业制造、现代农业、商贸流通、金融服务、绿色低碳等行业领域，打造一批“数据要素X”典型场景，促进实体经济和数字经济深度融合。投资机会方面，有券商分析师指出，近年来，我国数据战略布局不断演进，相关政策持续出台，有望充分释放企业数据资源价值，构建以数据为关键要素的数字经济，持续高质量推进企业数据资源开发利用。数据产业链拥有万亿广阔空间，在政策、技术、供需等因素共同支撑下有望迎来新一轮高速发展，AIInfra、数据要素、信创三大产业趋势有望成为核心方向。</p>
12.26	<p>企业数据资源开发利用获政策支持，大数据ETF(159739)涨超2%，科华数据涨停</p> <p>A股三大指数午后维持涨势，截至发稿，中证云计算与大数据主题指数涨2.53%。相关ETF方面，大数据ETF(159739)截至发稿上涨2.21%，成交额近1495万元，换手率7.48%。成分股中，科华数据涨停，拓尔思涨超9%，光环新网涨超7%，奥飞数据涨超6%，新易盛、宝信软件、千方科技等多股跟涨。资金流向方面，该ETF昨日获资金净流入435万元，近5个交易日中，有3日获资金净流入。大数据ETF(159739)紧密跟踪中证云计算与大数据主题指数，选取50只业务涉及提供云计算服务、大数据服务以及上述服务相关硬件设备的上市公司证券作为指数样本，以反映云计算与大数据主题上市公司证券的整体表现。</p>
12.20	<p>首家数据科技央企成立明年产业规模将超2000亿元</p> <p>12月19日，中国数联物流信息有限公司在上海正式揭牌成立，这也是首家数据科技中央企业。中国数联物流由国务院国资委直接管理，为股权多元化中央企业，注册资金100亿元。公司引入招商局集团有限公司、中国保利集团有限公司、中国物流集团有限公司、中国民航信息集团有限公司、上海国盛(集团)有限公司、上海数据集团有限公司作为战略投资者。</p>

	<p>中国数联物流将以公路、铁路、水路、航空、口岸等领域数据资源共享和开发利用为核心，整合物流与信息流、资金流，构建国家级物流大数据平台，以数字技术提升产业运营效率，服务实体经济发展，有效降低全社会物流成本。</p> <p>近年来，数据类公司纷纷成立。企查查数据显示，截至最新，我国布局数据要素产业链的企业数目已达到 4.84 万家（剔除产业关联性较弱的企业），其中 6194 家在近 5 年时间成立。按所属地区来看，北京市注册的企业达到 5660 家，超过总数的 10%。</p>
12.11	<p>数据要素产业政策环境：奠定基石，展望未来</p> <p>数据要素产业的政策环境正在不断完善和优化。从“数据二十条”到《关于加快公共数据资源开发利用的意见》，再到国家数据局、地方数据局、大数据中心、数据集团和数据交易所等政府机构的设立与职责明确，中国政府正在为数据要素产业的持续健康发展奠定坚实基础。展望未来，中国政府将继续深化数据要素产业的政策创新与制度建设，推动数据要素产业的规范化、规模化发展。同时，政府还将加强与国际社会的合作与交流，共同推动全球数据要素产业的繁荣发展。</p>
12.5	<p>数据要素深度探索公共数据的价值转化与策略挖掘</p> <p>在数字化浪潮的推动下，数据以其独特的价值，宛如新时代的黄金与石油，迅速成为推动全球经济社会的核心引擎。为了顺应这一波澜壮阔的变革，我国政府不仅通过《关于构建更加完善的要素市场化配置体制机制的意见》（简称“数据二十条”）、《“十四五”数字经济发展规划》等一系列政策措施，以前所未有的力度推进数据要素市场化改革；还在 2023 年成立了国家数据局，展现了国家对数据资源管理的高度重视和长远布局。</p> <p>在一系列政策引领下，我国数据要素行业正朝着健康、有序的方向发展，助力新质生产力充分释放，为数字经济发展注入澎湃动力。未来，数据要素市场将更加成熟、规范，为经济社会发展贡献更大的力量。《2024—2029 年中国数据要素市场前景及投资机会研究报告》显示，我国数据要素市场规模正在迅速扩大，行业整体将进入群体性突破的快速发展阶段。随着国家数据局工作的深入和数据要素市场化改革的持续推进，我国数据要素市场将迎来更加广阔的发展前景，为建设数字中国、实现经济社会高质量发展提供有力支撑。</p>

资料来源：新浪财经、21 财经、Wind、中国投资咨询网、证券时报网、国脉数据资产，中国银河证券研究院

表7：数据交易所新闻及政策

日期	具体内容
12.30	<p>上海数据交易所挂牌的数据产品交易额突破 4000 万元</p> <p>据上海数据交易所消息，《上海市数据产品知识产权登记存证暂行办法》2024 年 12 月 8 日施行以来，上海市知识产权局联合上海数据交易所聚力打造“登记流通链”，实现数据产品知识产权登记即流通，在上海市数据产品知识产权管理平台公示和已登记的 158 件数据产品中，74 件完成在上海数据交易所挂牌，截至 12 月底，已在上海市数据产品知识产权管理平台登记并在上海数据交易所挂牌的数据产品交易额已超过 4000 万元。</p>
12.25	<p>上海数交所今年交易额已突破 40 亿元，数据做支点撬动新质生产力壮大</p> <p>近年来，联和投资公司下属上海数交所以独特的方式不断赋能创新创业，使数据成为新质生产力培育壮大的重要支点。截至目前，上海数交所今年交易额已突破 40 亿元，较去年的 10 亿元和 2022 年的 1 亿元有极为明显的增长，而挂牌上架的数据产品则超过 4500 种，是去年的两倍多。</p> <p>在 AI（人工智能）领域，针对自动驾驶企业对数据的渴求，浦东在上海数交所支持下，将金桥智能网联汽车测试示范区的无人驾驶算法和场景数据挂牌，提供给主机厂、车联网企业、科研机构等，让他们能更高效优化 AI 模型。</p> <p>大模型的发展离不开语料“燃料”。今年，上海数交所率先建成语料专区，引导相关数据产品上线流通，特别是针对高质量学术数据，他们根据语料开放度和质量的不同，为各供方制定差异化策略，以支撑生成式人工智能发展。</p> <p>随着数据资产交易市场的不断发展，上海数交所目前场内数据资产总价值已超过 15 亿元，为“轻资产、重数据”的高科技企业对接金融服务授信金额超过 5 亿元，为写好数字金融大文章作出上海贡献。</p>
12.18	<p>促进数字贸易改革创新安徽数据交易所发放 1000 万元数据券</p> <p>为加速全省数据要素市场化配置改革，推动数据要素交易流通及创新应用，安徽省数据交易所在 2024 年世界制造业大会上重磅发布“1000 万元数据券”助力计划，旨在激活数据要素市场，加快释放数据价值，助力数字经济高质量发展。</p> <p>12 月 12 日，安徽数交所首批数据券集中发放，包括数据产权登记服务券、数据交易券以及算力资源券，可在办理数据产权登记服务业务、数据交易业务和购买算力资源时，享受一定的优惠补贴，降低相关资源和服务的获取成本。下一步，安徽数交所</p>

	将持续跟踪数据券发放及使用情况，及时总结优化实施工作，后续将按照常态化、分批次开展数据券发放活动。同时，加强与各行业的深化合作，推出更多的特色专题数据券，促进数据高效流通，推动产业数字化转型，助力数据赋能行业高质量发展。
12.11	<p>证券时报数据产品登陆深圳数据交易所</p> <p>证券时报数据产品——“证券时报实时财经资讯库(机构版)”在深圳数据交易所正式挂牌上市。证券时报也是首家在深交所挂牌数据产品的财经媒体。这是证券时报顺应数字经济发展要求，积极参与数据要素市场建设的重要尝试，对丰富数据交易场内交易元素、促进数据产业发展都将产生积极的影响。自此以后，证券时报拥有一个新身份——数据产品商。</p> <p>本次挂牌的“证券时报实时财经资讯库(机构版)”为数据聚类数据产品，所含内容由证券时报采编队伍自主原创生产，覆盖宏观政策、产经、上市公司、银行券商、保险、期货、基金等各个经济金融细分领域，极具权威性、准确性、及时性，证券时报对内容享有完整的著作权。产品具备多重应用场景，包括：方便用户及时快速掌握市场动态、政策解读等，可作为企业或机构向用户提供资讯展示服务的优质来源，帮助企业或机构提升用户服务水平；内容可作为底层数据助力数据加工、舆情监测及其他服务类企业进行产品优化和开发。此外，该产品可帮助企业进行大模型训练、学习，为机构和数据厂商做深层次产品开发和应用提供支持。</p>
12.6	<p>全国首位！深圳数据交易所完成跨境数据交易规模超 2.5 亿元</p> <p>2024 全球招商大会金融产业专场在深圳福田会展中心桂花厅盛大启幕。会上，围绕数据跨境流通、科技金融创新等关键领域，河套发展署副署长余杰介绍了河套深港科技创新合作区深圳园区(简称“河套深圳园区”)独特的跨境创新发展优势。河套深圳园区是粤港澳大湾区中唯一以科技创新为主题的重大合作平台。2023 年 8 月，国务院发布其发展规划，明确三大定位：国际科技规则试验区、大湾区中式转化集聚区、深港科技创新合作先导区。余杰介绍，河套深圳园区内的深圳数据交易所，累计完成数据交易规模超过 150 亿元，跨境数据交易规模超过 2.5 亿元，均位居全国首位。</p>

资料来源：新京报、市国资委、人民数据、中新网、证券时报、南方都市报，中国银河证券研究院

(二) 算法端：国内外巨头大模型动态

表8：国内人工智能大模型动态

时间	模型	主要内容
12.30	DeepSeek-V3 大模型	<p>DeepSeekV3 发布，加速 AI 应用落地</p> <p>近日，DeepSeek-V3 的正式发版引起 AI 业内广泛高度关注，其在保证了模型能力的前提下，训练效率和推理速度大幅提升。DeepSeek 新一代模型的发布可能意味着 AI 大模型的应用将逐步走向普惠，助力 AI 应用广泛落地；同时训练效率大幅提升，亦将助力推理算力需求高增。</p>
12.27	“坤安”大模型	<p>有色金属行业新帮手——人工智能大模型“坤安”</p> <p>12 月 26 日，中国有色金属工业协会和中铝集团共同主办的“坤安”人工智能大模型发布会在北京举行。据介绍，“坤安”大模型是为有色金属行业设计的人工智能系统，它包含了大量的有色金属行业数据和专业领域知识，并具备自然语言处理、图像处理、预测以及科学求解器等多项专业能力。</p> <p>该模型将通过人工智能技术来重塑有色金属地质勘探、矿产开采、冶炼加工和再生利用等全场景业务流程，以保障有色金属矿产的安全生产，同时保护网络安全、数据安全和算法安全。中铝集团已经在采矿、冶炼、加工、安全生产、供应链管理、经营管理、科技研发以及智慧服务等 8 个业务领域中应用了“坤安”人工智能大模型，并且已经形成了 7 个品牌。</p>
12.19	光明电力大模型	<p>我国电力行业首个千亿级人工智能大模型发布</p> <p>我国电力行业首个千亿级人工智能大模型——“光明电力大模型”今天对外发布。该模型集成了千亿级数据参数，将在电网规划和运行、电力设备检修、供电服务等 600 多个应用场景发挥智能专家作用，实现电力与算力的协同赋能，助力新型电力系统和新型能源体系建设，更好保障我国能源安全。</p>

12.16	工业 X 射线影像检测的人工智能大模型	<p>自主研发出应用于工业 X 射线影像检测的人工智能大模型</p> <p>日联科技(688531.SH)在投资者互动平台表示, 公司是国内领先的工业 X 射线智能检测装备及核心部件供应商, 在应用场景方面, 公司产品和技术可应用于各类机器人产品上游所涉及到的电子元器件、新能源电池及机械零部件的检测。在产品技术方面, 公司通过针对性的对大量被检测产品及产品缺陷特征的机器学习, 自主研发出应用于工业 X 射线影像检测的人工智能大模型, 实现了工业 X 射线 AI 智能检测, 提高了公司产品的检测效率及准确率, 增强了公司市场竞争力。</p>
12.13	专为媒体场景设计的行业垂类大模型	<p>浙数文化: 旗下两家子公司涉及人工智能大模型的研发</p> <p>浙数文化称公司旗下两家子公司涉及人工智能大模型的研发。其中, 传播大脑公司打造的专为媒体场景设计的行业垂类大模型——传播大模型算法已通过国家网信办生成式人工智能(大语言模型)上线备案, 传播大脑内容生成算法已通过国家网信办境内深度合成服务算法备案, 正在加快推进 AI 垂类应用。以传播大模型为例, 目前已形成了包括文生图文、文生视频在内的智能对话、创作、多模检索、创意设计以及审校 5 大核心功能。杭州城市大脑有限公司打造的社会治理大模型算法、城市大脑参政大模型生成合成类算法已通过国家网信办境内深度合成服务算法备案。</p>

资料来源: 中关村在线、中央广播电视总台、格隆汇、Wind、金融界, 中国银河证券研究院

表9: 海外人工智能大模型动态

时间	模型	主要内容
12.17	Veo2 模型	<p>谷歌宣布推出新一代视频生成模型 Veo2</p> <p>12月17日, 谷歌宣布推出新一代视频生成模型 Veo2, 该模型在理解和呈现现实世界物理现象、人类运动及表情细节方面取得重大突破, 整体视频质量得到显著提升, 具备生成 4K 分辨率、2 分钟时长视频的能力。</p>
12.13	Gemini2.0	<p>谷歌发布了新一代人工智能模型, 希望获得在该技术领域的前沿地位</p> <p>谷歌表示, 其发布的新一代 AI 模型 Gemini2.0 将成为 AI 智能体 (AI Agent) 发展的“跳板”, 可以用于分析智能手机显示的图像、执行各种乏味杂务、记住消费者的对话、帮助视频游戏玩家制定策略, 以及处理在线搜索等任务。在 AI 应用产生过诸如在披萨上涂胶水这样的荒唐建议之后, 谷歌加快了技术改进的步伐, 以最大限度地减少此类失误。谷歌首席执行官孙达尔·皮柴表示, Gemini2.0 将能够进行更类似于人类的推理, 同时解决更高级的数学问题, 甚至编写代码。</p> <p>此次技术升级还将改善仍在实验中的、名为“ProjectAstra”的通用 AI 智能体的多项功能, 使其能够与人们进行更有意义和实用的对话。谷歌还将开始测试名为“ProjectMariner”的功能, 进一步帮助人们进行在线搜索和筛选结果。</p>
12.6	Llama3.3	<p>Meta 发布开源 Llama3.3</p> <p>12月6日, Meta 发布开源 Llama3.3, 拥有 700 亿参数, 其表现可与 Llama3.1 中的 4050 亿参数模型相媲美, 但成本和计算开销显著降低, 例如运行推理所需的 GPU 容量得以大幅缩减, 旨在以更小的规模提供高端性能和高可用性。</p>
12.5	ChatGPTPro	<p>OpenAI 推出 ChatGPTPro</p> <p>12月5日, OpenAI 推出 ChatGPTPro, 月费 200 美元, 可无限访问 OpenAI 最先进的模型和工具, 包括 OpenAIo1、o1-mini、GPT-4o 和高级语音功能等, 还包含 o1 的专业模式, 投入更多计算资源以提供更优质的答案。从 12 月 5 日开始, OpenAI 开启为期 12 天的圣诞直播活动, 每个工作日都会通过直播发布或演示新产品或新功能, 其中包括备受期待的文本生成视频 AI 工具“Sora”以及新的推理模型。</p>

资料来源: 新浪财经、财富中文网、金融界等, 中国银河证券研究院

(三) 算力端: AI 服务器、AI 芯片最新动态

表10: 最新 AI 服务器、AI 芯片动态

时间	主要内容
12.29	<p>小米正搭建 GPU 万卡集群</p> <p>界面新闻报道小米正在着手搭建自己的 GPU 万卡集群，将对 AI 大模型大力投入。小米大模型团队在成立之初便已拥有 6500 张 GPU 资源，此次搭建 GPU 万卡集群计划已实施数月之久，雷军在其中扮演了重要的领导角色。小米自 2016 年起就开始了 AI 的全方位布局，2023 年 4 月成立 AI 实验室大模型团队，2024 年 5 月小米宣布其大语言模型 MiLM 正式通过大模型备案。最近，DeepSeek 开源大模型 DeepSeek-V2 的关键开发者罗福莉将加入小米，担任小米 AI 实验室的领导，并负责大模型团队的建设。</p>
12.25	<p>液冷服务器概念引爆 AI 算力新风口，多家龙头企业布局抢占市场先机</p> <p>近期，液冷服务器概念股在 A 股市场上的表现尤为活跃，如同星科技强势涨停，银轮股份、曙光数创、中科曙光、中兴通讯、英维克等个股纷纷紧随其后，显示出市场对于液冷服务器的热情。值得注意的是，以中兴通讯为代表的通信龙头，盘中两度触及涨停，近三个交易日累计涨幅超过 20%，市值更是达到 1808 亿元。</p> <p>在这一背景下，液冷服务器成为 AI 算力市场的新风口，其背后的推动力不仅来自于技术的进步，也与国家政策的支持息息相关。近年来，国家发改委等相关部门已密集出台政策，以加强算力基础设施建设。同时，各地也在积极布局自己的智算中心项目。比如北京提出到 2027 年实现 100% 自主可控的智算中心建设能力，上海计划到 2025 年新建智算中心时国产芯片使用比例超过 50%，而天津也计划在 2026 年实现其算力中心中国产芯片的使用占比超过 60%。</p>
12.20	<p>ASIC 芯片：未来 AI 计算的新宠，投资风口已开启</p> <p>ASIC 是指专为特定应用设计的集成电路，与通用处理器（如 CPU 和 GPU）相比，ASIC 在特定任务上的计算能力和效率表现更为突出。在 AI 训练和推理的过程中，ASIC 芯片能够提供更高的能效比，最大程度地减少不必要的功耗，这对于大规模运算的 AI 任务尤为重要。例如，谷歌的 TPU 和亚马逊的 Trainium 系列芯片，单位算力成本远低于传统的英伟达 GPU。此外，ASIC 芯片的数据处理速度尤为迅速，能够在处理复杂数据时表现出更高的吞吐量。这些优势使得 ASIC 芯片在 AI 领域尤其受到青睐，预计将成为未来市场竞争的重要一环。</p>
12.12	<p>苹果与博通联手，研发 AI 服务器芯片</p> <p>苹果正在与博通合作开发其首款专为人工智能任务设计的服务器芯片。这款芯片的内部代号为 Baltra，预计将于 2026 年投入量产，这标志着苹果在追求先进人工智能能力以及减少对英伟达等外部芯片供应商的依赖的战略上迈出了大胆的一步。据报道，Baltra 芯片将采用台积电(TSMC)先进的 N3P 制造工艺，体现了苹果致力于在硬件开发中部署最新技术的承诺。台积电长期以来一直是苹果的重要合作伙伴，为其生产先进的处理器提供支持，包括为其 Mac 设备供电的 M 系列芯片。</p>

资料来源：搜狐网、电子发烧友、新浪财经，中国银河证券研究院

三、前沿行业动态

(一) 前沿技术动态

1. 中科院发布多智能体框架 AutoPatent，含 1933 个「草稿-专利」数据对

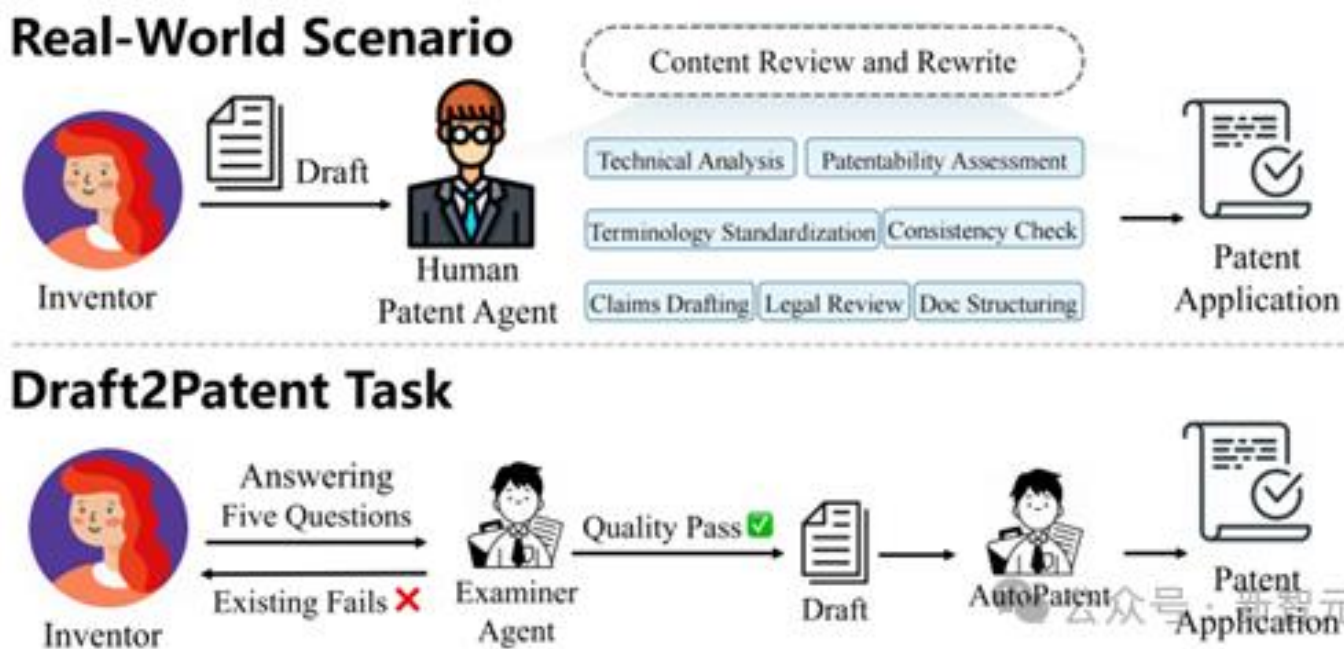
来自中科院深圳先进院和大连理工大学的研究团队联合开发了一个名为 AutoPatent 的自动化的多智能体专利生成框架，能够在十分钟左右的时间依据发明人的专利技术底稿生成一篇完整的专利文档。

以往的专利撰写任务通常关注专利文件内部文本的相互转化，例如专利的摘要与权利要求之间的相互转化等。研究人员通过模拟真实场景中的专利撰写工作，引入了一项新颖的专利起草任务 Draft2Patent，将发明人技术底稿转化为完整的专利文件，提升专利代理人的专利撰写效率。

该任务主要有两个主要困难：

- 底稿和完整专利的平均长度分别为 4,000 Tokens 和 17,000 Tokens。
- 一份高质量专利需要具有专利性（非显而易见性、有用性、新颖性以及符合专利法律规范），同时需要满足专业技术领域和法律的相关规范。

图3: 专利起草任务 Draft2Patent



资料来源: 新智元, 中国银河证券研究院

研究人员构建了 Draft2Patent 任务对应的基准数据集 D2P, 其中包含 1,933 条底稿专利文本和其他元数据, 并将其划分为包含 1,500 条数据的训练集, 133 条数据的验证集和 300 条数据的测试集。

随着大语言模型的发展, 基于大语言模型的智能体展现了在知识密集型领域的意图理解、规划、记忆、重复思考及任务执行的强大能力。

针对 Draft2Patent 任务, 研究人员提出了一个新颖的多智能体自动化专利起草框架 AutoPatent, 其中使用写作、规划及审查三大类八个智能体在短组件生成、专利写作规划树构建 (PGTree)、参考-审查增强生成 (RRAG) 三步骤的驱动下完成完整专利的撰写。

该篇文章的主要贡献有:

- 自动化生成, 提升专利撰写效率

通过自动化生成专利文档, 文章提出的 AutoPatent 框架显著提高了专利撰写的效率, 能够在十分钟左右的时间内生成一篇完整的专利文档, 减轻了专利代理人和发明者的工作负担, 使他们能够更专注于创新和专利内容的优化。

- 多智能体协作, 提升专利质量

AutoPatent 框架通过多智能体协作, 其中规划智能体构建 PGTree (Patent Writing Guideline Tree, 专利写作指南树), 专利详细描述写作智能体与审查智能体基于 RRAG (Reference-Review-Augmented Generation, 参考审查增强生成) 协作撰写占据专利文本 80% 以上的详细描述部分, 确保生成的专利文档符合法律和技术标准。这提高了专利的合规性和保护范围, 增加了专利被授权的可能性。

- 由此及彼, 推动知识产权服务创新

文章提出的 Draft2Patent 任务和 AutoPatent 框架为知识产权服务领域带来了新的思路和方法。这有助于推动知识产权服务的创新和发展，提高整个行业的服务水平和竞争力。

图4: Draft2Patent 三大步骤驱动

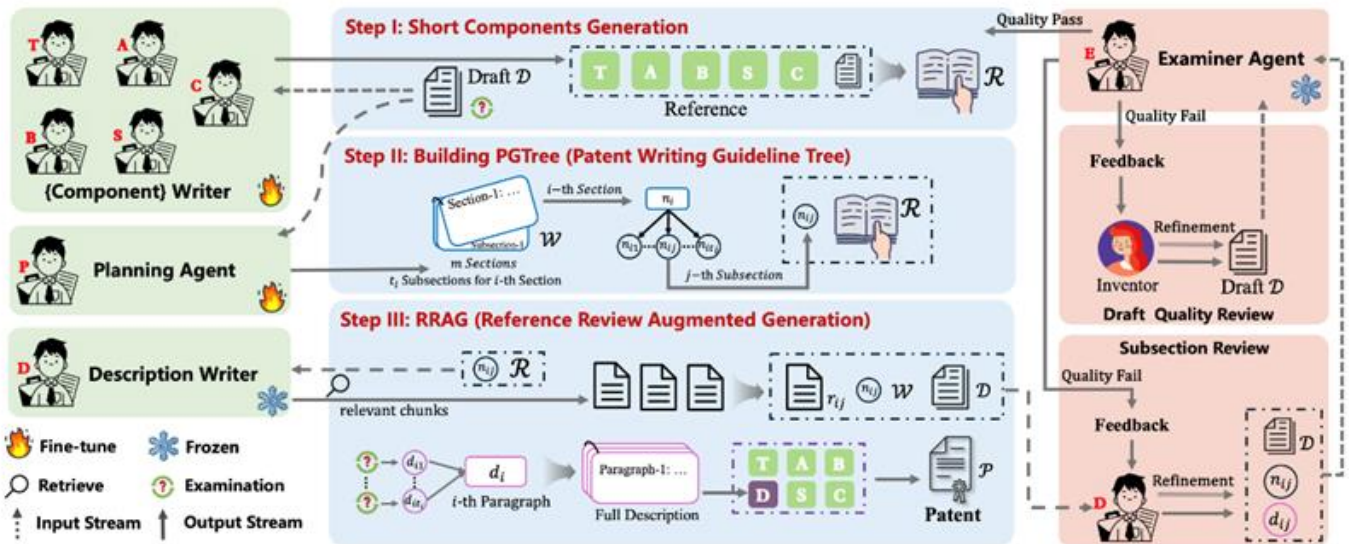


Figure 2: An overview of the AutoPatent framework, which includes eight agents and four steps for automatically generating a patent \mathcal{P} from a draft \mathcal{D} . The fire represents the corresponding agents' parameters are fine-tuned, while the snowflake represents the agents' parameters are frozen. The magnifying glass represents using the description writer for retrieval, while the question mark indicates using the examiner agent for text review. The dashed arrows represent the model's input stream, while the solid arrows represent the model's output stream.

资料来源: 新智元, 中国银河证券研究院

2.ChatGPT 搜索功能全面开放 新增语音搜索服务

12月17日, OpenAI 在其12天发布会的第八天正式宣布, ChatGPT 搜索功能将从即日起向全体用户开放。ChatGPT 搜索功能基于 OpenAI 的 GPT-4o 模型构建, 于今年10月底正式发布, 能让使用聊天机器人的用户像在网上进行常规搜索一样获取即时信息, 但当时仅向 ChatGPT Plus 和 Team 用户开放移动及网络端的付费搜索服务。自当地时间12月5日起, OpenAI 开启了一个密集的新功能发布周期, 计划在接下来的12天内通过12场直播活动陆续推出新产品和功能。在此之前, OpenAI 已经陆续发布了多项创新, 包括 ChatGPT Pro 计划、强化微调技术、Sora、交互界面 Canvas, 高级语音视觉功能, 以及 Projects 功能等。“大约两个月前, 我们为付费用户推出了这项功能,” OpenAI 的首席产品官凯文·韦尔 (Kevin Weil) 在周一的直播中表示。“现在我已经无法想象没有搜索功能的 ChatGPT 会是什么样子。我经常使用它。今天起, 我非常激动地宣布, 我们将免费向所有人提供这项服务。”用户只需拥有一个 OpenAI 账户, 即可开始体验 ChatGPT 搜索服务。登录后, 若用户的查询需要最新信息, ChatGPT 将自动进行网络搜索以提供答案。此外, 用户还可以通过提示栏上的新图标, 随时启动网络搜索功能。OpenAI 还为用户提供了将 ChatGPT 搜索设为浏览器默认搜索引擎的便利选项。

OpenAI 正在将 ChatGPT 搜索功能与高级语音模式相结合。这意味着, ChatGPT 现在能够以语音形式搜索网络, 并以自然、流畅的对话方式回应用户的问题。例如, 如果用户即将前往另一个城市度假, 只需询问 ChatGPT 当地的天气情况, 内置的搜索功能便能提供最新的天气信息。为了

增强这一功能，OpenAI 已经与多家顶级新闻和数据提供商建立了合作关系。因此，用户还将能够直接在 ChatGPT 中查看股票行情、体育赛事得分、天气预报等信息。可以说，ChatGPT 搜索正逐步发展成为谷歌等搜索引擎的有力竞争者。自 ChatGPT 在 2022 年底取得巨大成功后，科技公司纷纷将生成式人工智能融入各类服务之中，在线搜索领域也不例外。OpenAI 的合作伙伴微软与竞争对手谷歌均对各自的搜索产品进行了全面革新，加入了更多的人工智能对话功能。据悉，人工智能搜索初创公司 Perplexity 目前正进行早期融资谈判，估值高达 90 亿美元。

具体来说，UDK-VQA 框架会爬取热门搜索词，结合相关的新闻内容生成问题，并通过视觉搜索引擎匹配问题中涉及的实体，从而为每个问题设定合适的图像。这一过程保证了数据的丰富性和多样性，支持了 SearchLVLMS 的不断更新和优化。

图5: ChatGPT 搜索服务演示



资料来源：腾讯科技，中国银河证券研究院

ChatGPT 搜索关键特性包括以下几点：1. 对话式人工智能。ChatGPT 搜索使用户能够以自然语言提问。2. 跟进问题。允许用户提出跟进问题，使体验更加互动，类似于人与人之间的对话。3. 最新信息回应。ChatGPT 搜索实时捕捉网络资料，提供最新信息。4. 实时数据。让用户能够实时获取体育比分、天气状况、股票报价和突发新闻，以及定位服务和地图。5. 信息摘要。与传统搜索引擎显示一系列带有更多信息的网页链接不同，ChatGPT 搜索提供信息的简洁摘要。6. 来源归属。包含清晰的归属和源链接，使用户能够验证并进一步探索原始网站。7. 视觉结果。ChatGPT 搜索使用视频和图片来回答某些用户查询。

3.OpenAI 发布季第九天：o1 模型开放 API 使用，实时语音 API 大降价

备受期待的 o1 模型终于向开发者开放使用了！这次更新不仅让开发者能够使用更强大的 AI 模型，还大幅降低了使用成本。比如说，过去需要不少费用才能使用的语音功能，现在价格直接降低了 60%。对于预算有限的开发者来说，他们还特别推出了一个“迷你版”的语音服务，费用只有原来的十分之一。具体更新内容包括：

OpenAI o1 模型：API 现已支持功能调用、开发者消息、Structured Outputs 以及视觉识别

等功能。

实时 API 升级：集成简单的 WebRTC 功能，GPT-4o 音频价格下调 60%，同时推出 GPT-4o mini，音频费用仅为原来的十分之一。

偏好微调 (Preference Fine-Tuning)：新一代模型定制技术，帮助开发者更轻松地根据用户需求和偏好优化模型。

Go 和 Java SDK：现已开放测试版，开发者可快速上手集成。

这些进步意味着什么呢？简单来说，就是 AI 应用的开发变得更简单、更便宜，也更容易根据具体需求进行定制了。不管是想开发智能客服、自动翻译，还是其他创新应用，开发者们都能以更低的成本实现自己的想法。

·API 中 OpenAI o1 的实际应用

全新的 o1 模型，这是一次重要的技术升级。相比之前的版本，新模型不仅变得更聪明，还能更快地回应用户需求。目前，该模型已经向高级开发者开放使用，并计划逐步扩大开放范围。这次升级带来了哪些实际改进？

更快的响应速度：处理速度提升显著，相比旧版本快了整整 60%，开发者可以根据需要调节模型的"思考时间"，在速度和准确度之间找到平衡。

更强的实用功能：视觉识别，现在可以"看懂"图片，这对工业生产、科研等领域特别有帮助。精准输出，能够按照开发者的要求，生成格式严格统一的内容。个性化对话：支持调整语气和风格，让 AI 的回应更符合具体场景需求。系统集成，可以轻松与其他软件系统对接，实现更复杂的功能。

实际应用案例：已经有开发者使用新模型开发出了多个实用的应用，比如智能客服，能更好地理解 and 解决客户问题。供应链优化，帮助企业更好地管理库存和物流。金融分析，协助预测市场趋势。

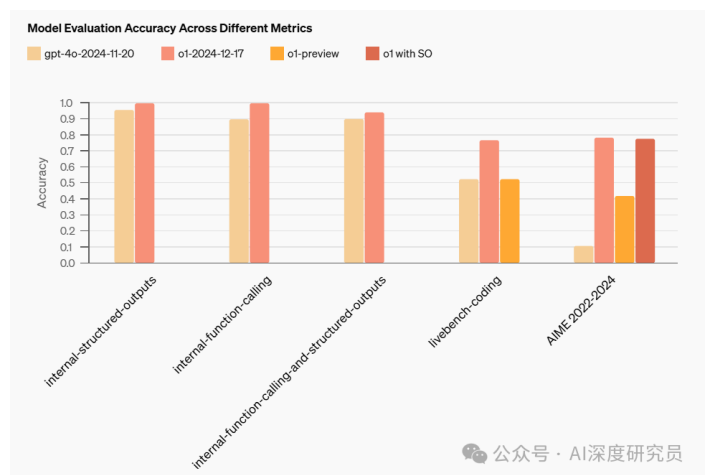
性能测试成绩：在最新的测试中，o1 模型在多个领域都展现出了优秀的表现，基础能力测试准确率达到 75.7%，编程能力在代码开发测试中达到 76.6% 的高分，数学计算在某些测试中甚至达到了 96.4% 的惊人准确率。

图6: OpenAI o1 基准测试结果

类别	评估项目	o1-2024-12-17	o1-preview
通用能力	GPOA diamond	75.7	73.3
	MMLU (pass @1)	91.8	90.8
编程能力	SWE-bench Verified	48.9	41.3
	LiveCodeBench	76.6	52.3
数学能力	MATH (pass @1)	96.4	85.5
	AIME 2024 (pass @1)	79.2	42.0
	MGSM (pass @1)	89.3	90.8
视觉能力	MMMU (pass @1)	77.3	—
	MathVista (pass @1)	71.0	—
事实性	SimpleQA	42.6	42.4
智能体	TAU-bench (零售)	73.5	—
	TAU-bench (航空)	54.2	—

资料来源: Founder Park, 中国银河证券研究院

图7: o1 模型数据应用情况



资料来源: Founder Park, 中国银河证券研究院

·让 AI 更懂你的需求

OpenAI 最新推出的偏好微调 (Preference Fine-Tuning) 功能，让 AI 模型的个性化定制变得更加简单和高效。这项技术采用了一种叫做"直接偏好优化" (DPO) 的创新方法，通过对比不同

的 AI 回答，让模型学会识别什么样的回答更符合用户的偏好。

这与传统的模型训练方法有着本质的区别。传统的监督式微调就像是让 AI 按照标准答案来学习，而新推出的偏好微调则更像是让 AI 理解"用户更喜欢哪种回答方式"。这种方法特别适合那些没有标准答案、需要理解用户个人喜好的场景，比如写作风格的调整、对话语气的把握，以及创意内容的生成等。

下方表格展示了偏好微调与监督式微调的核心区别：

图8：偏好微调与监督式微调对比

监督微调 (SFT) 与偏好微调 (PFT) 对比

对比项目	监督微调 (SFT)	偏好微调 (PFT)
目标	通过复制标记的输出来鼓励模型生成正确的输出	通过强化偏好响应并降低非偏好响应的可能性来优化模型的行为
训练数据	精确的输入输出对	偏好与非偏好模型输出对，通过人工标注、A/B测试或合成数据生成获得
使用场景	适用于理想输出易于准备且需要严格正确性的任务，如自定义代码格式	适用于"更好"的响应具有主观性的任务，如创意写作或文本摘要

资料来源: Founder Park, 中国银河证券研究院

偏好微调成功案例：Rogo AI 正在开发一款面向金融分析师的 AI 助手，该助手能将复杂的查询任务拆分为子查询。在其专业基准 Rogo-Golden 的测试中，Rogo AI 发现，监督式微调在处理非分布式查询扩展时表现不佳，例如在“公司 X 的增长速度如何”这类查询中，可能遗漏关键指标如 ARR。而通过偏好微调，这一问题得到有效解决，模型性能从基础模型的 75% 准确率提升至 80% 以上。

偏好微调将正式上线，支持 gpt-4o-2024-08-06，并即将拓展至 gpt-4o-mini-2024-07-18。定价将与监督式微调保持一致，同时，明年初将支持 OpenAI 最新发布的模型版本。

·WebRTC 支持，简化流程

OpenAI 在其实时 API 中集成了 WebRTC，这项专为互联网设计的实时通信技术，极大地增强了 AI 应用在会议和视频流传输等领域的性能。通过自动调整比特率和执行回声消除，WebRTC 帮助 AI 应用适应网络变化，确保了语音应用的稳定性和流畅度。


WebRTC 在简化代码方面相较于旧的 Websockets 集成有了显著进步。过去，开发者使用 Websockets 时，常常需要编写 200-250 行代码来处理诸如反压等复杂问题。现在，利用 WebRTC，同样的功能仅需 12 行代码即可轻松实现。这种简化在实时语音聊天应用的开发中尤为突出，

WebRTC 让开发者能够快速构建出功能完备的应用程序，大幅提升了开发效率。

在最新的演示中，我们见识了 WebRTC 如何简化实时语音应用的开发。

开发者仅用几行简单的 HTML 代码就创建了对等连接，轻松实现音频流的发送与接收。这样的技术支持让开发者可以专注于应用逻辑，而不必深陷复杂的网络通信细节，显著降低了开发难度。

图9: WebRTC 简化代码演示

```
JavaScript 
```

```
1  async function createRealtimeSession(localStream, remoteAudioEl, token) {
2    const pc = new RTCPeerConnection();
3    pc.ontrack = e => remoteAudioEl.srcObject = e.streams[0];
4    pc.addTrack(localStream.getTracks()[0]);
5    const offer = await pc.createOffer();
6    await pc.setLocalDescription(offer);
7    const headers = { Authorization: `Bearer ${token}`, 'Content-Type': 'applicat
8    const opts = { method: 'POST', body: offer.sdp, headers };
9    const resp = await fetch('https://api.openai.com/v1/realtime', opts);
10   await pc.setRemoteDescription({ type: 'answer', sdp: await resp.text() });
11   return pc;
12 }
```

f

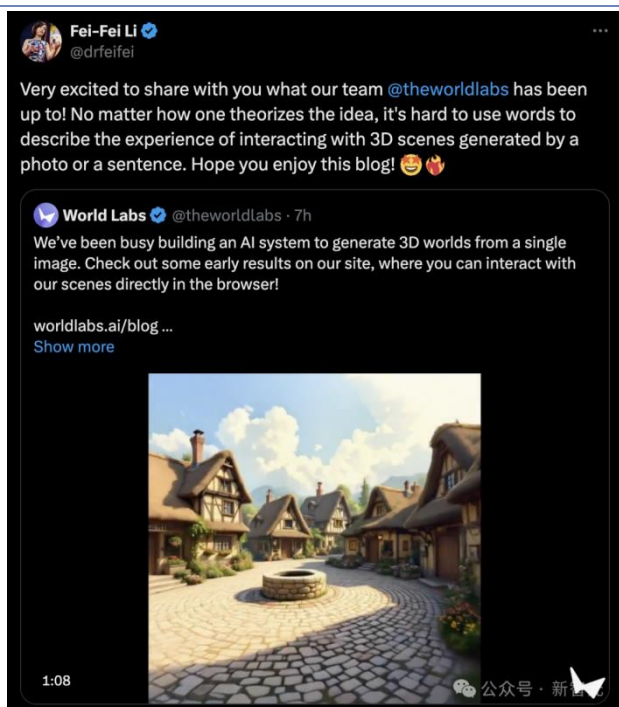
资料来源: 雪球易创科技, 中国银河证券研究院

OpenAI 为了简化开发者集成实时 API 的过程, 不仅推出了 Python SDK, 还大幅下调了价格, 让实时 API 的集成变得更加便捷和经济。

4. 李飞飞空间智能首个模型诞生! 一张图生成 3D 世界, 视频游戏要变天

AI 教母李飞飞创立的 World Labs 首次官宣「空间智能」模型, 一张图, 即可生成一个 3D 世界。用李飞飞的话说, 「无论怎样理论化这个想法, 都很难用语言描述一张照片或一句话生成 3D 场景的互动体验。」

图10: 李飞飞 World Labs 官宣



资料来源: 新智元, 中国银河证券研究院

图11: 李飞飞 TED 演讲



资料来源: 新智元, 中国银河证券研究院

这是迈向空间智能的第一步。所有场景都能在浏览器中实时渲染，还能实现可控的相机效果、可调节的模拟景深。未来，游戏 NPC 的虚拟世界可以随意切换，都是分分钟生成的事情。英伟达高级研究科学家、李飞飞高徒 Jim Fan 总结道，「GenAI 正在创造越来越高维度的人类体验快照。Stable Diffusion 是 2D 快照；Sora 是 2D+时间维度的快照；而 World Labs 是 3D、完全沉浸式的快照」。

·探索一个新世界

不论是 Midjourney、FLUX，还是 Runway、DreamMachine，我们熟知的大多数 GenAI 工具仅能制作图像/视频 2D 内容。若是实现了在 3D 中生成，视频的控制性、一致性能得到极大的改善。这也就意味着，制作电影、游戏、模拟器等其他物理世界的数字表现形式，将会发生翻天覆地的变化。World Labs 成立开始的初衷便是，空间智能的 AI 对世界进行建模，还能 3D 时空中物体/地点/交互进行推理。这次，他们首次展示了这个 3D 生成的世界。

·3D 特效

大多数生成式模型都是预测像素的。而预测 3D 场景有很多好处：- 场景持久性：一旦生成了一个世界，它就会稳定存在。即使你转开视线后再次观看，场景也不会在你看不见时发生变化。- 实时控制：生成场景后，你可以在其中实时移动。你可以仔细观察花朵的细节，或是探头查看角落后面有什么。- 几何精确性：这个生成的世界遵循基本的 3D 几何物理规则。它们具有真实的立体感和空间深度，与某些 AI 生成视频的虚幻效果形成鲜明对比。可视化 3D 场景最简单的方法是，就是使用深度图（depth map）。在深度图中，每个像素都会根据其到相机的距离来着色：

·空间智能，计算机视觉下一个前沿

此前，李飞飞在一次活动中，首次详细揭秘了何谓「空间智能」：视觉化为洞察，看见成为理解，理解导致行动。她将人类智能归结为两大智能，一是语言智能，另一个便是空间智能。虽然语言智能备受关注，但空间智能将对 AI 产生重大的影响。

正如 World Labs 的官博所阐述的那样，人类智能包含了诸多方面。语言智能，可以通过语言与他们进行交流和联系。而其中最为基础的便是——空间智能，能够让我们理解，并与周围世

界进行互动。此外，空间智能具备了极强的创造力，可以将我们脑海中的画面，在现实中呈现。正是有了空间智能，人类能够推理、行动和发明。从简单的沙堡到高耸的城市可视化设计，都离不开它。

5.谷歌新旗舰模型 Gemini 2.0 鲨疯了，免费不限量，网友：我读论文能力提高 10 倍

谷歌 Gemini 2.0 一大波网友试玩已新鲜出炉，目前看来，众人一致认为这是谷歌对 OpenAI 的一次成功反击。

根据官方跑分结果，Gemini 2.0 在各项基准测试上的得分和 Claude 3.5 Sonnet 差不多，但相比 Gemini 1.5 Pro，其生成速度提升了 2 倍。同时，根据 Deepmind CEO 哈萨比斯的说法，Gemini 2.0 主打「多模态+Agent」，是一款面向智能体时代的 AI 模型。

·Gemini 2.0 主打「多模态+Agent」，网友已玩疯

首先，随着 Gemini 2.0 Flash 实验版一同上线的，还有一项名为深度研究（Deep Research）的智能体新功能。这个功能在 Gemini Advanced 中推出，仅支持英文版本，有点像最近比较火的科研 AI 搜索。按照谷歌 CEO 皮猜的说法，用户只需给一个科研主题，它就能帮你整理出一份带参考来源的完整报告。

·浏览器智能体 Project Mariner

此外，基于 Gemini 2.0，谷歌这次还发布了一个类似 Anthropic「Computer Use」的智能体——Project Mariner。简单来说，用户只需一句简单指令，它就能自动在浏览器中完成操作，包括处理像素和文本、代码、图像和表单等。而且使用时会在右边展开侧边栏，实时显示 AI 的思考及操作过程。（用户也能继续输入指令）

不过目前该智能体尚未全面推出，仅限受信任的测试人员使用。

·编程智能体 Jules

最后，基于 Gemini 2.0 Flash，谷歌这次为开发者专门构建了一个代码智能体 Jules。具体而言，Jules 以异步方式工作，并与用户的 GitHub 工作流集成，在专注于实际想要构建的内容时，它负责处理 Bug 修复和其他耗时的任务。

据官方介绍，Jules 在 SWE-bench Verified 测试中达到了 51.8% 的通过率，而 Claude 3.5 是 49%。不过，目前 Jules 仅开放给一小部分可信赖的测试者，并计划在 2025 年初向其他有兴趣的开发者提供。

总之，目前发布的上述 3 个智能体均未开放给大众，要想体验 Gemini 2.0 的能力，目前更多还是在多模态上。

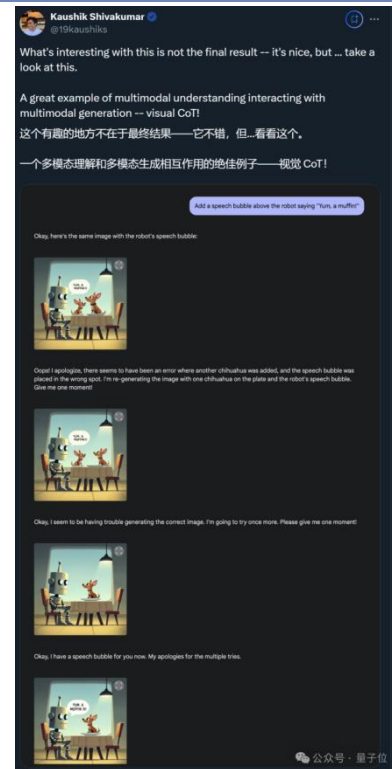
进一步深扒发现，Gemini 2.0 在图像生成时竟然具备 COT 思维能力。在第一次生成后，它很快就发现了自己的错误，并一步步自动调整，最终生成了符合指令的正确图像。

图12: 能全程语音交流论文的科研助手



资料来源: 量子位, 中国银河证券研究院

图13: 自动调整思维能力



资料来源: 量子位, 中国银河证券研究院

另外, 谷歌 AI Studio 负责人还出来预告了一把, 好的都在后头。

目前 2.0 版本已提供给一些开发者内测, 谷歌正在迅速将其集成在 Gemini 和搜索等产品线中。正式产品将于 1 月份全面上市, 届时将推出更多型号。

6. 通信巨头入局视频生成, 直接霸榜权威评测: 人物跨越多场景依然一致

视频生成模型评测权威榜单 VBench, 突遭“屠榜”。

这个突然杀出来的模型, 就是中国电信人工智能研究院 (TeleAI) 刚刚发布的视频大模型 (VBench 上的代号为 TeleAI-VAST)。

而且 16 项子指标有 9 项都是第一, 大幅领先第二名。

其中更是有 5 项得分超过 99%, 并在物体分类和人体动作两项拿了满分 100%。

图14: 视频生成模型评测权威榜单 VBench

Model Name (clickable)	Source	Date	Total Score	Quality Score	Semantic Score
TeleAI-VAST	TeleAI	2024-12-02	89.71%	88.98%	92.63%
StratoLens	StratoLensStudio	2024-11-22	83.78%	84.39%	81.34%
JT-CV-9B	JiuTianCV	2024-11-18	83.69%	84.62%	79.97%
T2V-Turbo-v2	T2V-Turbo Team	2024-10-02	83.52%	85.13%	77.12%
MiniMax-Video-01	VBench Team	2024-10-01	83.41%	84.85%	77.65%
Data-Juicer (2024-09-23, T2V-Turbo)	Data-Juicer Team	2024-09-23	82.53%	83.38%	79.13%

资料来源: 量子位, 中国银河证券研究院

在样片中，四位女主角为寻找宝石，分别踏上了各自的冒险之旅。具体细节不展开太多，但关键之处在于，每个人物在各自的分镜中，形象都和开头的合影保持了一致。要知道对于视频生成来说，保持单个人物的前后一致就已经不是一件容易的事，TeleAI 视频生成大模型却一下处理了四个。

·人物动作精准控制，还有音画同步

除了开头三分钟的宣传片，TeleAI 展示了另外几段作品。

第一段视频中，女主角先后出现在公交车、马路、酒吧吧台和座位四个不同的场景。在这四个场景中，人物的外貌、发型、衣着全都保持了一致，画面整体的色调风格也很统一。这说明，TeleAI 视频生成大模型已经实现了不同的场景变化下主体的时空一致性。

接下来的画面里，两只猴王展开了一场近距离对战，期间的人物动作、镜头跟随，都已经呈现出了电影级的水准。如果细节再继续打磨优化，拿来拍电影指日可待。

人物特征、时空一致性和动作精准性都有了，还能再做些什么呢？目前的视频生成模型大多生成的都是默片，这次 TeleAI 把声音也加进来了。

TeleAI 视频生成大模型生成的声音，不是单纯地来上一段音乐那么简单。仔细听会发现，舰船的鸣笛声、飞机起飞的轰鸣声、潜水艇的水声，以及最后火箭冲出水面的声音，都与画面中看到的内容做到了同步。也就是说，TeleAI 视频生成大模型作品中体现的一致性，已经跨越模态了。

·独创两阶段生成架构

不仅效果和成绩优秀，TeleAI 视频生成大模型的技术架构也十分独特。它没有采用传统的路径，而是全自研了一个“二阶段视频生成技术”——VAST (Video As Storyboard from Text)。TeleAI 团队没有选择一步到位，而是将视频的生成分解成了两个过程。

在第一阶段，采用多模态大型模型根据文本输入生成中间素材，包括姿势、分割图和深度信息。TeleAI 团队把这些中间表示称作“故事板”，是模型能够体现场景语义和结构本质的关键。第二阶段，才是真正的视频生成。利用基于 DiT 架构的扩散模型，TeleAI 以这些表示为条件，结合目标对象的文本描述和外观信息，生成最终的视频。

这种分段式的方式，使得生成视频时能够精确控制主体的位置、运动和视觉外观。而且，TeleAI 视频生成大模型不仅能在模型上分“阶段”，还可以把视频按场景分成“片段”。当创作者想要生成一段长视频时，视频生成工具可以先设计具体的分镜头，开头 3 分钟的视频就是用这种方法创作出来的。针对每一个分镜，生成具有一致性的生成中间素材，这些中间素材不仅作用于模型内部，对创作者也是可见的，甚至还能进行调整修改。因为对片段进行了划分，所以，只要在每个片段中都能保持人物一致性，可以生成的视频长度将是无限长。

中国电信介绍，本次发布的视频生成大模型将于明年开启公测，可以期待一波新鲜的 AI 大片了。

·各种模态全都有，还要做智能体

此次亮相的视频生成大模型，是 TeleAI 整个大模型布局中的一个环节。此前，TeleAI 已经自主了研发覆盖语义、语音、视觉、多模态的“星辰”大模型能力体系。星辰大模型在央企中唯一开源，还完成首个全国产化万卡万参大模型训练，并打造业界首个支持 40 种方言自由混说的语音识别大模型。基础模型之外，TeleAI 还面向工业、教育等领域推出 50 多个场景大模型，并构建了“星海”数据智能中台，形成了 9 万亿 Tokens 高质量数据集。在使用方式上也另辟蹊径，甚至推出了发短信和大模型对话的功能。

除了模态、场景、数据和使用方式，TeleAI 也全新推出了大模型智能体平台，与视频生成大模型一同登上 TeleAI 开发者大会。

图15: 大模型智能体平台



资料来源: 量子位, 中国银河证券研究院

7.清华、智谱团队: 「6000 亿合成交错语音文本」预训练, 问答性能提升近 3 倍

与基于文本的大语言模型 (LLM) 相比, 语音语言模型 (SpeechLM) 接受语音输入并生成语音输出, 能够实现更自然的人机交互。然而, 传统的 SpeechLM 因缺乏无监督语音数据和并行语音-文本数据, 很难像 LLM 一样进行大规模扩展。

为解决这一问题, 来自清华大学和智谱的研究团队提出了一种新方法, 利用从文本语料库中提取的大规模合成交错数据来扩展语音-文本预训练, 从而消除了对并行语音-文本数据集的需求。这一方法从现有文本库中抽取文本片段, 并使用 text-to-token 模型合成相应的语音片段, 从而高效地构建语音-文本交错数据, 而无需生成实际语音。此外, 通过在编码器中加入向量量化的瓶颈层, 他们还采用了从自动语音识别 (ASR) 模型中提取的监督语音 tokenizer。即使在较低的采样率 (如 12.5Hz) 下, 这种有监督的训练方法也能生成语义保存较好的离散语音 token, 同时保持语音重构质量。

研究团队表示, 从预训练的语言模型开始, 并将预训练扩展到 1 万亿个 token (其中包括 6000 亿合成交错语音文本数据), 他们在语音语言建模和口语问题解答方面取得了 SOTA, 将语音问答任务方面的性能从之前的 13% (Moshi) 提高到 31%。

他们进一步证明, 通过使用语音对话数据对预训练的模型进行微调, 他们开发出一种端到端语音聊天机器人, 在对话能力和语音质量方面都达到了与现有基线相当的性能, 展现了在纯语音领域中应用的巨大潜力。

·研究方法

研究团队提出了一种通过从文本语料库合成高质量交错语音-文本数据来扩展语音-文本预训练的新方法, 使得大规模预训练无需依赖大量的语音数据集成为可能。完整方法框框架结合了语音 token 化、语音-文本交替数据生成以及两阶段训练。

在语音 token 化的设计上, 他们采用了一种基于自动语音识别 (ASR) 的监督语音 tokenizer, 通过在编码器中加入向量量化的瓶颈层和平均池化层来有效生成离散语音 token。同时, 为提高实时语音处理能力, tokenizer 还优化为支持流式推理从而降低在线应用中的延迟。

语音-文本交错数据是 SpeechLM 预训练的关键。他们先是基于现有的文本到语音 (TTS) 数据集训练了一个 text-to-token 模型, 能直接将文本片段转换为对应的语音 token, 较传统多步生

成方法效率更高且潜在错误的累积更少。

接着，他们使用训练好的 text-to-token 模型从大规模文本语料库中采样文本片段，并将其转换为语音 token。这些语音 token 与原始文本片段交错组合，从而生成高质量语音-文本交错数据。最终，他们基于 FineWeb 语料库为模型预训练生成了 6000 亿 token 的语音-文本交错数据。

之后，他们开展了一个两阶段训练，在第一阶段，使用合成语音-文本交错数据对模型进行预训练，在第二阶段，使用语音对话数据集微调预训练模型。在语音-文本预训练阶段，他们通过使用不同用途的 4 类数据：语音-文本交错数据、无监督文本数据、无监督语音数据和有监督语音-文本数据进行预训练，使模型学习语音和文本之间的对齐关系。到了监督微调阶段，他们利用语音对话数据进行微调，使模型能够支持端到端语音输入与输出。

图16: 研究方法概述

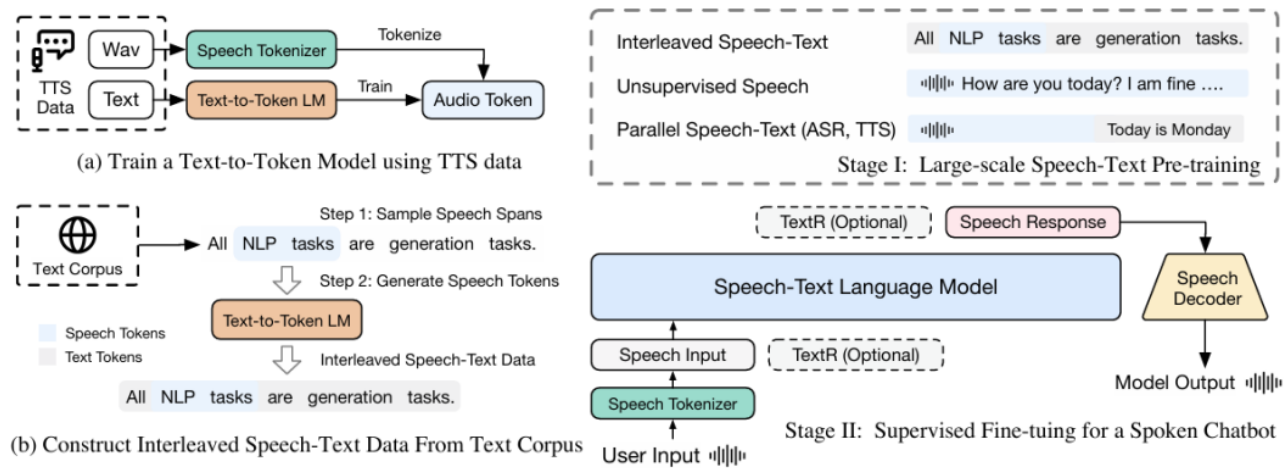


Figure 2: **Overview of our method.** First we train a text-to-token model to construct interleaved speech-text data. The speech language model’s training contains two stages. In the stage 1 the model is pre-trained with synthetic speech-text interleaved data. In the stage 2 the the model is fine-tuned with a speech dialogue dataset.

资料来源：学术头条，中国银河证券研究院

(二) 前沿政策动态

表11: 相关政策法规

时间	部门	文件	内容
12月24日	上海市教育委员会	《上海市推进实施人工智能赋能基础教育高质量发展的行动方案(2024—2026年)》	鼓励并支持全市中小学（含中职校）、幼儿园广泛开展人工智能教学实践与体验，积极探索人工智能赋能拔尖创新人才早期发现与培养机制，全面提升师生数字素养与技能，推进人工智能赋能教育变革和样态重塑，打造人工智能与上海基础教育融合发展的高地、新业态和新模式。
12月12日	工业和信息化部 财政部 中国人民 银行 金融监管总 局	《中小企业数字化赋能专项行动方案(2025—2027年)》	该行动方案旨在通过数字化赋能，推动中小企业实现专精特新发展，其中人工智能作为重要手段被广泛应用和深度赋能。方案提出，到2027年，中小企业数字化转型将取得显著成效，形成一批数字化水平较高的转型标杆，其中人工智能技术的应用将发挥关键作用。

12月6日	国家知识产权局	《人工智能相关发明专利申请指引（征求意见稿）》	人工智能技术创新不断取得新突破，已成为新一轮科技革命和产业变革的重要驱动力量，人工智能相关专利申请量快速增长。专利制度作为 激励和保护创新成果 的有效手段，在促进人工智能技术的发展、规范引导人工智能技术的应用等方面都发挥着重要作用。为全面、深入地诠释我国现行专利法律制度框架下的人工智能领域专利审查政策，回应创新主体普遍关注的热点法律问题，提升专利申请质量，发布人工智能相关发明专利申请指引
12月4日	重庆市经济和信息化委员会	《重庆市支持具身智能机器人产业创新发展若干政策措施》	旨在推动 具身智能机器人领域的技术突破与产业升级 ，提升重庆在这一前沿科技领域的竞争力。政策涵盖了对 关键技术研发 的大力支持、 产业培育 方面的税收优惠和场地补贴、 高端专业人才的激励措施 以及加强产业链协同等内容。
12月18日	市工业和信息化局	《深圳市打造人工智能先锋城市的若干措施》	旨在通过加大普惠服务资源供给、发放“训力券”“语料券”“模型券”、支持行业应用、开放政务场景、推动产业集聚等措施，加速深圳人工智能产业发展，打造 人工智能先锋城市 。
12月13日	中国信息通信研究院	《人工智能发展报告（2024年）》	总结了人工智能在全球范围内的技术创新、产业升级、行业应用和安全治理等方面的最新进展。报告指出，人工智能正迈向 通用智能初始阶段 ，工程化加速推进，新产品新模式层出不穷。同时，生成式人工智能在技术创新、用户培养、产业发展等方面取得突破性进展，推动各行各业转型升级。此外，报告还关注了人工智能的 可靠性评估、伦理规范 等问题，并展望了未来的发展机遇。

资料来源：各政府网站，中国银河证券研究院

四、前沿企业动态

（一）前沿产品动态

1. 图森未来转型 6 个月：推出 AI 大模型，布局游戏生态，思考和底气是什么？

宣布转型 AIGC 的图森未来正式宣布启用全新品牌 CreateAI，并公布多个业务进展：将基于武侠 IP《金庸群侠传》开发一款大型开放世界 RPG 游戏。而此前公布的《三体》项目将与“超时空要塞之父”、著名动画导演河森正治和顶级动画制作公司株式会社白组合作。

更为重要的是，图森未来还正式发布首个大模型产品 Ruyi，该模型定位于开源图生视频领域。作为其自研 AI 基础设施的底层核心，图森未来将基于 Ruyi 大模型，打造多个针对游戏和动画制作的 AI 生成式工具。

这距离图森未来宣布转型仅半年时间，并且由单一的 AI 游戏开发上升至生态战略：

一是将大模型聚焦游戏动漫制作领域。

二是通过自研游戏，让大模型从流程上开始真正参与游戏及动画制作，从而让大模型具备更多商业价值，同时缩短其自研游戏 IP 的开发周期。

“我们希望通过生成式 AI 技术重新定义数字化娱乐内容的制作。”图森未来董事长陈默表示，“我们的目标是 5 年内将制作成本和开发周期降 70%。”

·为什么是大模型+游戏

2024 年国产游戏大作《黑神话·悟空》验证了优质游戏的吸金能力。

根据贝恩咨询公司数据，在全球大部分产业增长乏力情况下，游戏和动漫行业依然蓬勃发展。其中 2023 年全球视频游戏市场规模超过 2000 亿美元，而全球动漫市场年复合增长率达到 10%，预计 2030 年将达到 600 亿美元。

然而高制作的成本和漫长的开发周期，限制了高质量内容的供应。根据行业数据和经验，动画电影/电视剧制作成本约占总承包的 60-70%，平均制作周期 4-6 年。视频游戏开发成本则占总预算的 50-70%，开发周期约 5-7 年。生成式 AI 技术将彻底改变这一现状，这意味着巨大的商机。就连

埃隆马斯克旗下的人工智能公司 xAI，也在不久前宣布进军游戏领域，成立一家 AI 驱动的游戏工作室。

图森未来显然也看到了这个蛋糕，经过公司内部和董事会的多次评估，最终踏上数字化内容娱乐转型之路。据陈默介绍，在图森未来之前他曾创办过游戏公司，多年来一直保持对游戏和动漫行业的关注。而图森未来作为一家技术基因公司，能够以边际成本最小的方式快速进入大模型领域。

当然最重要的是，新的方向给了图森未来“翻身”的机会：此前因无人驾驶产生的高额成本已经降低，现有业务架构下有至少 3 年的资金储备，游戏具有成熟的商业化模式，叠加 AI 能力将加快游戏推出的进度，能够更快的形成规模收入和利润。根据陈默的规划，新业务乐观能在 2026 年实现收支平衡，他也希望通过最终的成绩来给市场和投资人“一个交代”。

技术复用和经验思考

图森未来于 12 月 17 日上线图生视频模型 Ruyi，用户提供一张图片作为输入，选择输出时长、分辨率、运动幅度等参数，就会得到不超过 5 秒的视频。虽然相较于可灵、海螺等国内第一梯队的图生视频大模型，Ruyi 还具有一定的差距，且存在手部畸形、多人时面部细节崩坏、不可控转场等问题，但能够在不到半年时间内从训练数据到发布模型，图森未来展示了其作为一家技术公司的实力和水平。

据图森未来 COO 郝佳男介绍，虽然无人驾驶和图生视频模型是两个不同的领域，但依然有非常多的技术可以复用。

首先是技术人才方面，虽然因业务调整自动驾驶算法的技术人员已经离开，但感知算法的技术骨干依然留任，而大模型与此前无人驾驶的感知算法较为接近，技术人员能够在很短时间内上手。

在数据积累和处理层面，复用了此前无人驾驶非常多的数据处理基础设施工具和流程，基本没有边际成本：比如高质量的数据清洗和打标签，如何大规模处理视频和数据，如何把视频保存下来，如何调动上千块的 GPU，如何快速发现问题并进行迭代，如何快速部署代码和更新。这些是非常复杂和琐碎的工作，但非常依赖于经验，决定着开发的效率。

事实上，无人驾驶的数据模型虽然体量小，但数据处理训练要比图生视频大模型复杂太多，因为无人驾驶涉及安全问题，必须保证 100% 不能出错。而外界对于图生视频模型的交付效果容忍度较高，对于图森未来而言，这甚至是“有点轻松”的工作。郝佳男表示，在 Ruyi 推出的过程中也踩了很多算力算法的坑，目前已发布的也是 7B 的小体量模型，还需要做很多调整的实验，去解决一致性、流畅性以及审美等问题。图森未来的目标是尽快追赶第一梯队，根据内部评估，以现有发展速度，2025 年下半年有望达到第一梯队水平。

既然图森对大模型如此有信心，为何还要大费力气自研游戏和动画电影？郝佳男表示，目前已有导演或博主通过市面上的大模型工具完成短片，但在落实到工业电影的实际生产中，依然存在不小的困难。因此做大模型最大的挑战，其实是如何让大模型真正参与或融入现有的制作流程，这不是一个技术问题，而是商业和流程的问题。

这也是图森未来做无人驾驶业务获得最深刻的教训：与生态方密切合作共建生态，一方面配合度达不到最佳状态，另一方面太受制于人风险过高。“因此最好的解决方案就是自己研发游戏，内部游戏团队和大模型团队去磨合，双方不断进行交互数据反馈，才能让大模型真正成为制作的一环。”郝佳男说。

郝佳男预计，内部的游戏与大模型团队的磨合至少还需要 1 年时间，不过一旦解决了这个难题，不仅有利于图森缩短开发游戏的时间，也能够赋予大模型真正的商业价值，并实现图森转型后的愿景——用生成式 AI 技术重新定义数字化娱乐内容的制作。

2. 华为：全球首个海拔 5000 米高原露天矿无人驾驶成果发布

12 月 25 日消息，华为官方今晚发文，宣布该公司今日与西部矿业、中铁十九局在西藏昌都玉

龙铜矿共同举办了一场以“智绘高原露天矿开启智驾新征程”为主题的发布会。

三方在现场共同宣布全球首个 5000 米高原露天矿无人驾驶项目成果交付，并展示了无人驾驶技术在极端环境中的适应性与可行性。

中铁十九局党委书记、董事长李华伟介绍称，玉龙铜矿的多编组无人驾驶车队目前已经在数千米长、最小宽度为 20 米、最大坡度为 8% 的矿道上运行作业。

据介绍，在项目建设过程中，华为以无人驾驶技术赋能。项目组预计，在 2025 年第一季度实现无人驾驶常态化作业，并且整体效率与有人驾驶模式的持平。

官方介绍称，三方团队密切合作、克服高原环境的各种极限挑战，玉龙铜矿项目从今年 9 月份启动建设，仅用了 3 个月时间就实现了成果展示。

华为表示，无人驾驶技术的智能化调度系统优化了作业流程，减少道路维护次数和车辆维修频率，同时降低了燃料和轮胎的消耗。IT之家从官方获悉，这种模式相比传统模式优势显著，两个编组 10 辆无人驾驶矿卡每年可节省约 600 万元的成本。

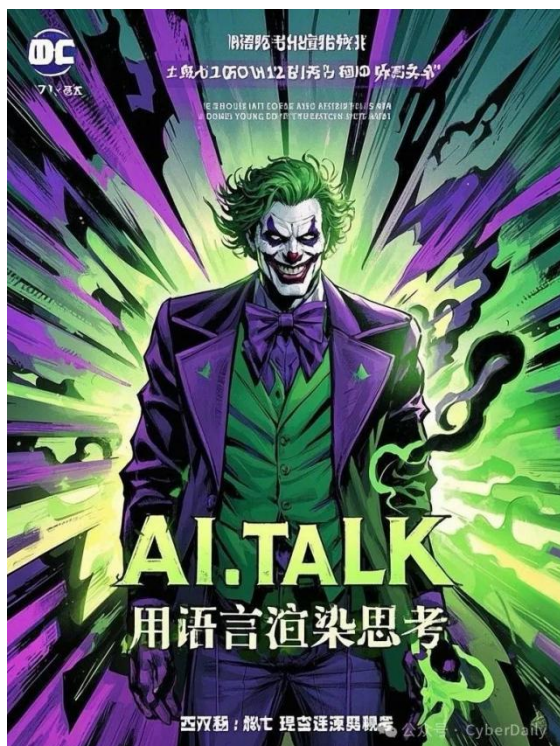
3.字节新图片模型搞定了中文，能生成各种中英文海报|AI 新大陆

作为 AIGC 的鼻祖之一，文生图一直撑起生成式 AI 的门面，但过去两年，文生图一直偏爱英文世界，流窜在社交网络的图片，都是各种英文字体风格的海报、封面、插图.....就是不能生成中文。最稳定的则是 Ideogram，花式输出各种设计，它也是 CyberDaily 主力设计师。

现在，中文世界有自己的 Ideogram 了，字节旗下的即梦 AI 正在灰度测试文生图 2.1 模型：可生成各种风格的中文、英文字体混合的海报。以后，要做日常文章插图、封面、节假日海报、庆典等之类的设计，写个 Prompt 就可以让它来完成。

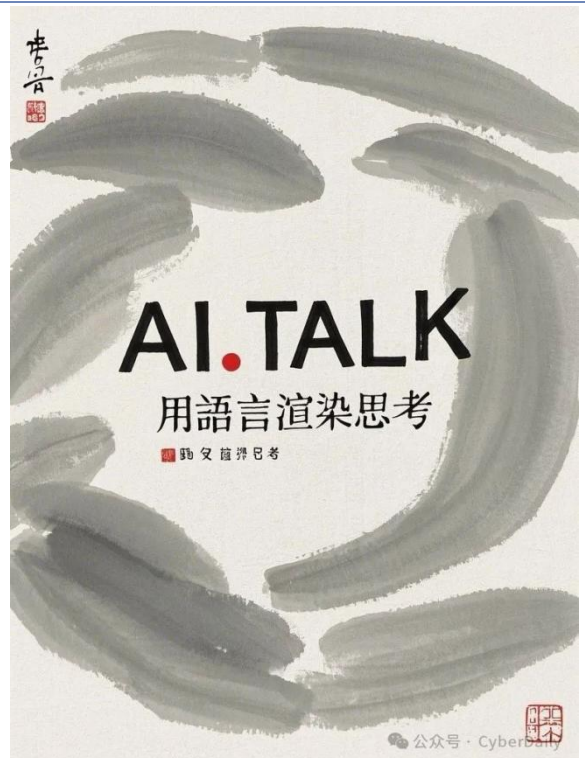
分享来自 AI Talk 创始人汗青的制作，可以看到很完整地生成中英文排版的海报，尽管有些字不能完全显现，但这只是初代版本的模型。

图17: AI Talk 创始人汗青制作作品



资料来源: CyberDaily, 中国银河证券研究院

图18: AI Talk 创始人汗青制作作品



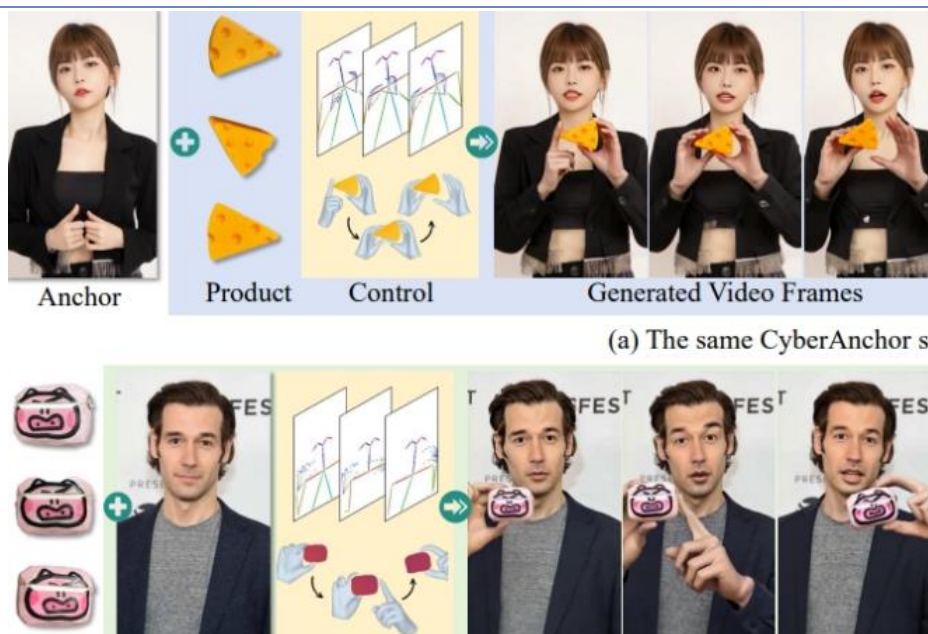
资料来源: CyberDaily, 中国银河证券研究院

4. 中科院联手腾讯打造 AI 带货王 AnchorCrafter，动作表情自然到位

中国科学院和腾讯联手，推出了一款名为 AnchorCrafter 的 AI 虚拟主播带货神器!它能像一位经验丰富的带货主播一样，手持商品，配合流畅自然的动作和表情，为产品疯狂打 call!

它可以让你从众多网络主播中挑选一位你最中意的，然后用 AI 技术将她的形象“复制”到虚拟世界中，让她成为你的专属带货主播!你只需要提供主播的一张照片，以及你想让她展示的商品，AnchorCrafter 就能自动生成一段精彩的带货视频!AnchorCrafter 还可以让你精确控制主播的动作和表情，就像导演一样，你可以根据自己的需求，设计各种各样的互动动作，让主播与商品完美融合，展现出最吸引人的一面。

图19: AnchorCrafter 效果展示



资料来源: 腾讯研究院, 中国银河证券研究院

AnchorCrafter 是如何做到这一切的呢?

首先，它会从个角度拍摄商品的照片，并利用强大的 AI 算法提取商品的特征信息，确保生成的虚拟商品与真实商品一模一样，连细节都完美还原!

其次，它会捕捉真人主播的动作和表情，并将其转化为数据，然后将这些数据“注入”到虚拟主播的模型中，让虚拟主播的动作和表情更加自然流畅!

最后，它还采用了一种特殊的训练方法，让 AI 模型更加注重手部和商品之间的互动细节，避免出现手穿模、商品漂浮等尴尬的穿帮镜头!

实验结果表明，AnchorCrafter 生成的虚拟主播带货视频，无论是画面质量还是动作自然度，都达到了业界领先水平，甚至可以和真人主播相媲美!有了 AnchorCrafter，再也不用担心虚拟主播“不接地气”了!它可以帮助轻松制作出高质量的带货视频，吸引更多消费者，提升产品销量!

相信在不久的将来，AnchorCrafter 将会成为电商平台的标配，为消费者带来更加便捷、高效的购物体验!

5. 谷歌推出视频生成模型 Veo 与 Imagen3，将颠覆创意 workflow

近日，谷歌云部门宣布，其最新的视频和图像生成模型——Veo 和 Imagen3，现已在 Vertex AI 平台上线，进一步推动了 AI 在营销与广告领域的应用。

Veo 模型首次亮相于 Google I/O 开发者大会，是谷歌 DeepMind 对竞争对手如 Runway 的 Gen-3 和 OpenAI 的 Sora 等产品的回应。该模型能够将文本或图像提示转化为长达 60 秒的高清电影风格视频，具备出色的帧级一致性，确保主体在镜头内流畅移动。

Imagen3 则专注于图像生成，能够根据文本生成照片级逼真的视觉效果，并在细节、照明和伪影减少方面超越前代产品。此外，Imagen3 还引入了编辑功能，用户可以通过文本提示优化生成的图像，包括图像升级、修复、外绘和背景替换等，满足个性化的创意需求。

通过集成 Veo 和 Imagen3，Google Cloud 为营销、销售等领域的团队提供了更强大的创作工具。Imagen3 的应用简化了高质量产品图片和社交媒体内容的制作，而 Veo 则将这些视觉效果转化为精美的视频，帮助团队快速迭代创意并加速内容生产。

谷歌产品管理高级总监 Warren Barkley 表示，客户如 Agoda 正借助 Veo 和 Imagen 等 AI 模型大幅缩短视频广告的制作周期，提升创意效率。此外，Veo 和 Imagen3 还内置了数字水印和内容审核等安全功能，以应对生成式 AI 带来的风险。

这些创新为全球企业带来了重新构想视觉内容创作和交付的机会。早期采用者包括亿滋国际（奥利奥、吉百利等品牌）和全球营销巨头 WPP。随着谷歌基础模型的持续扩展，各行业企业将能更高效地利用 AI 推动创意变革。

谷歌在视频生成领域的布局意味着竞争愈发激烈。与亚马逊 AWS 在 re:Invent 大会上发布的 Nova Reel 相呼应，谷歌的 Veo 和 Nova Reel 都在 AI 视频生成领域设立了新的标杆。Nova Reel 能够基于文本和图像提示生成六秒长的高质量视频，且同样通过 Amazon Bedrock 平台提供。

微软目前尚未推出视频生成模型，尽管其 AI Foundry 平台提供了图像生成工具。业内预计，随着 OpenAI 的 Sora 视频生成模型的发布，微软将在这一领域迎头赶上。

（二）投融资事件

表12: 12月 AI 相关投融资事件

融资方	赛道	公司简介	融资日期	融资轮次	融资金额	本轮投资方
云鲸智能	新兴技术及应用	家用机器人研发商	2024-12-16	战略轮	数亿元人民币	深圳和无锡两大国资
矽光科技	硬件	微型显示器研发商	2024-12-16	战略轮	数千万元	浙江大学教育基金 卓戴资本 能达资本
维享时空	智能大空间运营平台	城市元宇宙商业化解决方案提供商，致力于将空间计算智能平台及空间 AIGC 的技术应用于数实共生的城市元宇宙空间	2024-12-20	战略轮	数千万元	芯原股份、镭目科技战略投资， 蓝驰创投和优山资本继续跟投
丰通伟业	通用 AI 开放平台	通用 AI 开放平台运营商	2024-12-16	战略轮	五百万人民币	山行资本
万勋科技	特种机器人研发提供商	万勋科技是一家特种机器人研发提供商，通过减少现场人员投入和安全风险，提升作业效率，产品已成功应用于反恐安防、电力检修、应急救援、核电能源、工业协作、科研教育等多种领域。	2024-12-16	A+轮	数亿元人民币	德虎资本 元钛基金
闪极 SHARGE	智能硬件	闪极 SHARGE 是一家充电器制造生产商，致力于为用户提供氮化镓充电器，充电效率高、能量损失小、发热幅度低、尺寸小易于携带。	2024-12-18	A+轮	数千万元人民币	绿洲资本

莱斯能特	传感器芯片制造商	莱斯能特是一家传感器芯片制造商，公司主要聚焦于 MEMS 传感技术，从事工业汽车类 MEMS 传感器芯片的研发、生产和销售。	2024-12-18	A+轮		云岫资本
------	----------	--	------------	-----	--	------

资料来源：百度，投资界，网思科技官网，36 氪，Wind 万得，搜狐，新浪财经，中国银河证券研究院

五、投资建议

建议关注：1、国产算力产业链及生态伙伴；2、算力基础设施产业链；3、逐步向国产的自主 AI 算力为底座迁移的 MAAS 及下游应用端；4、端侧 AI 产业链整机及代工、结构件、开发套件厂商；5、数据要素产业链中供给、流通、应用公司；6、自动驾驶产业链及车路云一体化厂商。
建议关注：工业富联、中科曙光、曙光数创、海光信息、龙芯中科、科大讯飞、万兴科技、海康威视、中国软件、达梦数据、润泽科技、软通动力、能科科技、同花顺、财富趋势、中科创达、万集科技、金蝶国际、上海钢联、深桑达 A、云赛智联、易华录、云鼎科技、华勤技术、亿道信息等。

六、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: 12 月人工智能指数走势图	4
图 2: 12 月人工智能指数市场表现	5
图 3: 专利起草任务 Draft2Patent	13
图 4: Draft2Patent 三大步骤驱动	14
图 5: ChatGPT 搜索服务演示	15
图 6: OpenAI o1 基准测试结果	16
图 7: o1 模型数据应用情况	16
图 8: 偏好微调与监督式微调对比	17
图 9: WebRTC 简化代码演示	18
图 10: 李飞飞 World Labs 官宣	19
图 11: 李飞飞 TED 演讲	19
图 12: 能全程语音交流论文的科研助手	21
图 13: 自动调整思维能力	21
图 14: 视频生成模型评测权威榜单 VBench	21
图 15: 大模型智能体平台	23
图 16: 研究方法概述	24
图 17: AI Talk 创始人汗青制作作品	27
图 18: AI Talk 创始人汗青制作作品	27
图 19: AnchorCrafter 效果展示	28
表 1: 12 月成分股涨幅前十	4
表 2: 12 月成分股跌幅前十	5
表 3: 12 月人工智能主题基金一览	6
表 4: 人工智能主要上市公司近况一览 (数据截至 2024 年 12 月 31 日)	7
表 5: 境外上市人工智能企业近况一览 (数据截至 2024 年 12 月 31 日)	7
表 6: 数据要素最新新闻及政策	8
表 7: 数据交易所新闻及政策	9
表 8: 国内人工智能大模型动态	10
表 9: 海外人工智能大模型动态	11
表 10: 最新 AI 服务器、AI 芯片动态	12
表 11: 相关政策法规	24
表 12: 12 月 AI 相关投融资事件	29

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。

鲁佩机械行业首席分析师，伦敦政治经济学院经济学硕士，证券从业9年，2021年加入中国银河证券研究院。曾获新财富最佳分析师、IAMAC最受欢迎卖方分析师、万得金牌分析师、中证报最佳分析师、Choice最佳分析师、金翼奖等。

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