

产业积极向好，估值安全边际显著提升

--1 月人工智能行业月报

核心观点:

- **板块开年大幅调整，估值泡沫显著消化。**人工智能板块指数（884201.WI）1 月涨跌幅为-24.32%，同期上证综指下跌 6.27%，沪深 300 下跌 6.29%，创业板指数下跌 16.81%。受大盘整体回调、预期转弱以及外界不确定性等因素影响，人工智能板块剧烈回调，行业估值显著回落，安全边际大幅提升，我们认为 2024 年人工智能应用商业化将加速落地，伴随智能终端、多模态迭代升级将推动产业链中上游迎来新一轮爆发，人工智能板块有望迎来反弹。
- **科大讯飞举办星火认知大模型 V3.5 升级发布会，阿里云推出通义千问 Max 版本。**1 月 30 日，科大讯飞举办星火认知大模型 V3.5 升级发布会，展示大模型底层能力到应用场景的新发展，同时也发布了讯飞语音大模型。据介绍，讯飞星火认知大模型 V3.5 文本生成能力提升 7.3%，语言理解提升 7.6%，知识问答提升 4.7%，逻辑推理提升 9.5%，数学能力提升 9.8%，代码能力提升 8.0%，多模态能力提升 6.6%。阿里云 1 月 26 日推出多模态大模型 Qwen-VL-Max 版本。该版本在中文问答和中文文本理解任务上超越了 GPT-4V 和 Gemini，同时在图像相关推理能力和识别、提取和分析图像细节上都有显著提升。
- **海外巨头持续加码 AI 应用，2024 迎来终端 AIPC 元年。**1 月 8 日，苹果官方首次宣布其混合虚拟头显设备 Apple Vision Pro 将于 2 月 2 日开始在美国发售。1 月 9 日，英伟达在国际消费电子展 CES 上宣布，推出三款带有额外组件的新型桌面图形芯片，可以让用户在个人电脑上更好地利用人工智能，并表明在 AIPC 领域处于行业领先地位。1 月 15 日，微软正式发布了 Copilot Pro，这意味着 Copilot Pro 从 B 端走向 C 端，微软在 AI+办公领域业务布局的持续完善。我们认为：1）应用端：苹果、微软登巨头继续加码 AI 应用端，推进人工智能赋能千行百业，降本增效，商业化进程再次提速；2）智能终端：英伟达三款芯片具备强劲游戏和生成式 AI 性能，有望进一步驱动产业加速发展。
- **“数据要素×”三年行动计划正式发布，释放乘数效应。**1 月 4 日，国家数据局等 17 个部门联合印发《“数据要素×”三年行动计划（2024—2026 年）》。《行动计划》选取了工业制造、金融服务、科技创新、医疗健康等 12 个行业和领域，推动发挥数据要素乘数效应，释放数据要素价值，并明确了到 2026 年底的工作目标。我们认为，相比征求意见稿，此次发布的正式稿更强调开辟数据要素增量市场，明确提出推动数据资产“入表”，推动数据资产化、资本化，更好地发挥数据要素的乘数效应。
- **投资建议：**建议关注 1、上游 AI 芯片国产替代相关机会；2、华为产业链及生态伙伴相关机会；3、逐步向国产的自主 AI 算力为底座迁移的 MAAS 及相关应用的机会。4、下游应用端领域。重点推荐：科大讯飞、金山办公、倍轻松、超图软件、萤石网络、大华股份、中科星图、卫宁健康、彩讯股份、万兴科技、中科创达、同花顺、财富趋势、嘉和美康、国能日新、上海钢联、中科曙光、海康威视、柏楚电子、中控技术等公司。
- **风险提示：**技术研发进度不及预期风险；供应链风险；政策推进不及预期风险；消费需求不及预期风险；行业竞争加剧风险。

计算机行业

推荐

维持评级

分析师

吴砚靖

☎：(8610) 66568589

✉：wuyanqing@chinastock.com.cn

分析师证书编码：S0130519070001

鲁佩

☎：(021) 20257809

✉：lupei_yj@chinastock.com.cn

分析师证书编码：S0130521060001

研究助理

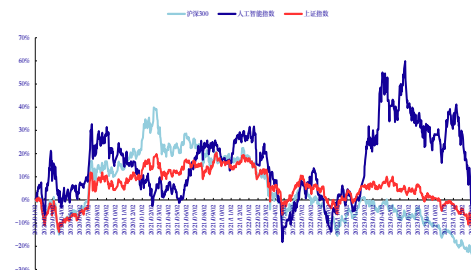
胡天昊

☎：(8610) 80927637

✉：hutianhao_yj@chinastock.com.cn

行业数据

2024



资料来源：中国银河证券研究院

目录

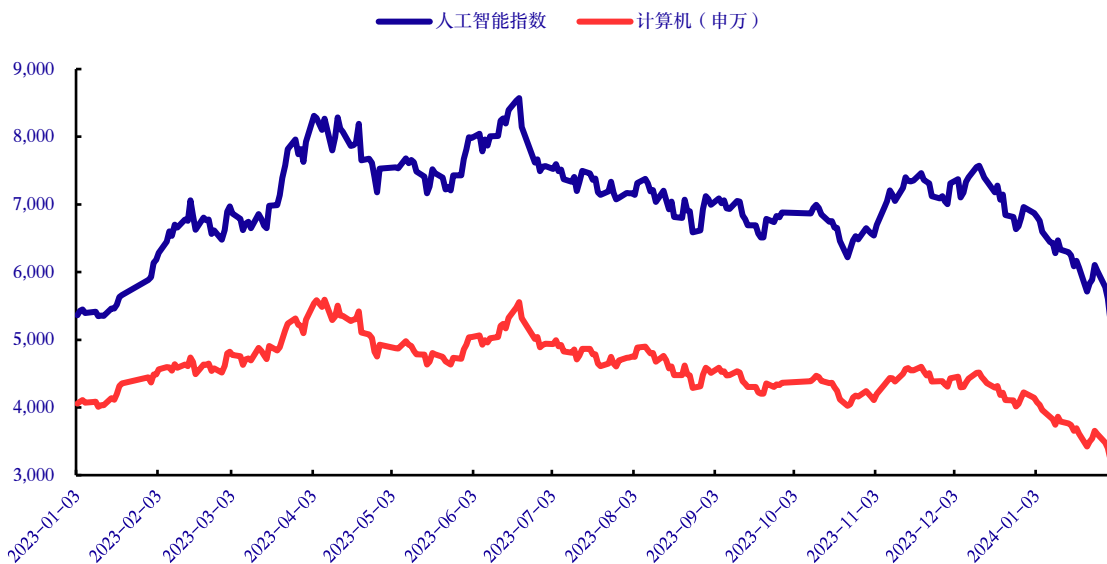
一、市场行情回顾.....	2
(一) 整体行情	2
(二) 代表企业	2
(三) 板块估值	3
二、人工智能产业政策动态.....	5
(一) 数据要素、数据交易所最新新闻及政策	5
(二) 算法端：国内外巨头大模型动态.....	7
(三) 算力端：AI 服务器、AI 芯片最新动态	10
三、人工智能前沿行业动态.....	10
(一) 前沿技术动态	10
1、探索宇宙对人脑演化的奥秘，中国空间站脑电测试完成首个阶段目标	10
2、AI 致命弱点：科学家首次用数学方法证明无法开发出稳定算法	11
3、百度文心大模型落地吉利银河 L6 车型，支持 AI 对话功能.....	12
4、国内首个与 ChatGPT 不同机制的通用大模型发布，算力消耗缩减 50%	12
5、全球首个汽车行业全栈自研全场景 AI 大模型！吉利星睿 AI 大模型正式发布.....	13
6、英特尔(INTC.US)推出车载 AI 芯片挑战英伟达和高通	16
7、AI 初创公司 Rabbit 推出智能终端 R1，挑战智能手机市场	16
(二) 前沿政策动态	17
四、前沿企业动态.....	18
(一) 前沿产品动态	18
1、全球首次！挑战波士顿动力，国产“电驱动”人形机器人实现空翻！	18
2、全球首创，鳍源水下机器人 AI 潜水员追踪功能发布：实现水下视觉锁定和追踪	19
3、“海豚 3 号”水面救生机器人发布	20
4、特斯拉计划明年出货首批人形机器人 Optimus，实用性仍是最大挑战	21
5、科学家制成迄今最轻快全功能机器人：以昆虫为模型，轻至 8 毫克	22
6、世界首个 AI 修路机器人将在英国上路测试：可识别、填平路面裂缝坑洼	23
7、继特斯拉后，宝马将在其美国汽车工厂使用人形机器人	24
(二) 前沿政策动态	25
五、风险提示.....	26

一、市场行情回顾

(一) 整体行情

A 股人工智能指数（884201.WI）截至 1 月末收盘价为 5269.63，月涨跌幅为-24.32%。计算机行业指数（801750.SI）截至 1 月末收盘价为 3222.55，月涨跌幅为-23.65%。

图1：1月人工智能指数表现



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 1 月 31 日总市值 11,977.53 亿，含成分股 69 支，权重等分。上市板分布为主板 19 支，创业板 29 支，科创板 6 支，中小板 15 支。

表1：1月份人工智能板块涨幅前十

股票代码	股票简称	1 月 31 日收盘价（元）	月涨跌幅	相对计算机
002226.SZ	江南化工	4.04	-7.76%	15.89%
002415.SZ	海康威视	31.92	-8.06%	15.59%
000977.SZ	浪潮信息	28.07	-15.45%	8.20%
600633.SH	浙数文化	9.26	-16.27%	7.38%
002230.SZ	科大讯飞	38.4	-17.21%	6.44%
300624.SZ	万兴科技	78.1	-17.44%	6.21%
002253.SZ	川大智胜	11.78	-17.62%	6.03%
688256.SH	寒武纪-U	110	-18.49%	5.16%
600560.SH	金自天正	12.03	-18.72%	4.93%
002065.SZ	东华软件	5	-18.96%	4.69%

资料来源：wind,中国银河证券研究院

表2：1月份人工智能板块跌幅前十

股票代码	股票简称	1 月 31 日收盘价	月涨跌幅	相对计算机
------	------	-------------	------	-------

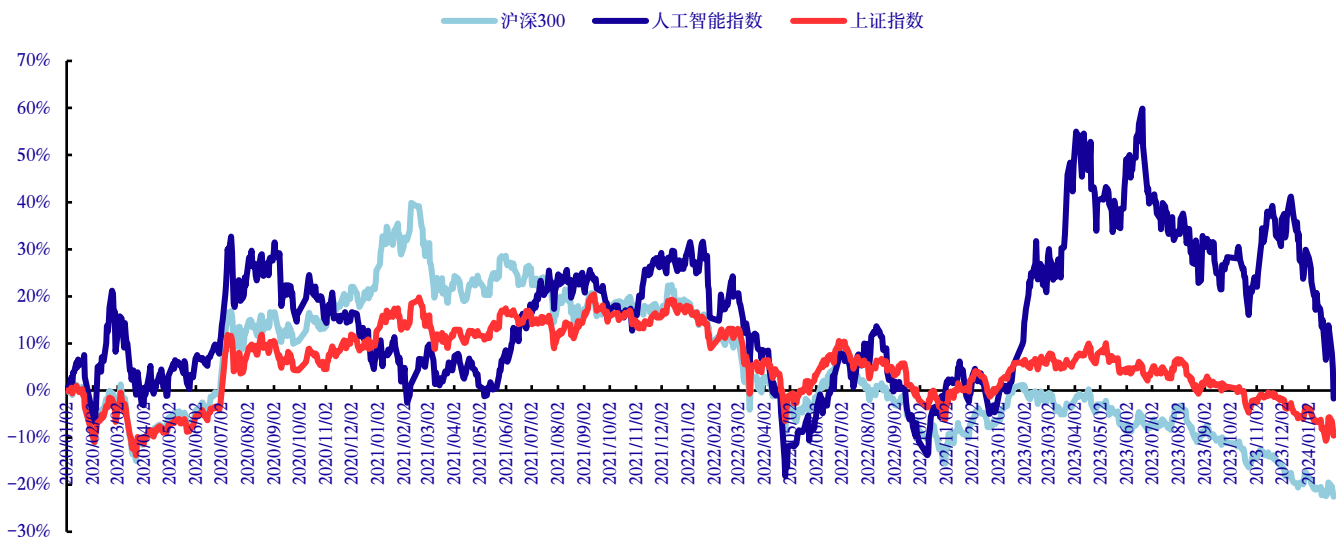
300081.SZ	恒信东方	7.24	-37.32%	-13.67%
688207.SH	格灵深瞳	13.35	-36.31%	-12.66%
688228.SH	开普云	42.06	-35.72%	-12.07%
688088.SH	虹软科技	26.84	-34.60%	-10.95%
300520.SZ	科大国创	14.22	-34.59%	-10.94%
002362.SZ	汉王科技	17.18	-32.94%	-9.29%
300400.SZ	劲拓股份	11.05	-32.17%	-8.52%
300781.SZ	因赛集团	30.65	-31.89%	-8.24%
300078.SZ	思创医惠	3	-31.51%	-7.86%
300571.SZ	平治信息	24.37	-31.02%	-7.37%

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 11.33%, 净利润复合增长率 3.97%, 截至1月31日平均估值 PE(ttm)55.24 倍, PS (ttm) 2.90 倍。

图2: 人工智能板块市场表现



资料来源: wind,中国银河证券研究院

表3: 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)	最新收盘日(元)	近1月回报(%)	近3月回报(%)	近6月回报(%)
001986	前海开源人工智能	契约型开放式	7.17	1.12	-17.79	-19.10	-35.38
005729	南方人工智能主题	契约型开放式	4.21	1.62	-18.16	-18.04	-30.82
005844	东方人工智能主题 A	契约型开放式	6.14	0.73	-25.91	-30.89	-26.66
005962	宝盈人工智能 A	契约型开放式	4.84	1.63	-18.43	-15.86	-28.43
005963	宝盈人工智能 C	契约型开放式	1.95	1.56	-18.48	-16.03	-28.71
006281	万家人工智能 A	契约型开放式	14.13	1.55	-18.24	-10.42	-31.77
008020	华富中证人工智能产业 ETF 联接 A	契约型开放式	1.14	0.57	-20.45	-17.43	-32.10
008021	华富中证人工智能产业 ETF 联接 C	契约型开放式	1.25	0.56	-20.48	-17.50	-32.21

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。

008585	华夏中证人工智能主题ETF 联接 A	契约型开放式	4.66	0.58	-19.53	-17.60	-29.95
008586	华夏中证人工智能主题ETF 联接 C	契约型开放式	4.27	0.57	-19.54	-17.65	-30.05
009239	融通中证人工智能主题 C	契约型开放式	1.20	0.96	-19.26	-17.30	-29.70
011832	西部利得中证人工智能 A	契约型开放式	1.38	0.62	-20.83	-18.35	-30.19
011833	西部利得中证人工智能 C	契约型开放式	1.00	0.62	-20.85	-18.43	-30.34
011839	天弘中证人工智能主题 A	契约型开放式	1.22	0.63	-19.73	-17.40	-29.56
011840	天弘中证人工智能主题 C	契约型开放式	4.98	0.63	-19.74	-17.44	-29.63
012733	易方达中证人工智能主题ETF 联接 A	契约型开放式	3.37	0.72	-19.67	-17.63	-30.09
012734	易方达中证人工智能主题ETF 联接 C	契约型开放式	3.80	0.72	-19.67	-17.65	-30.13
014162	万家人工智能 C	契约型开放式	10.66	1.52	-18.30	-10.60	-32.05
014630	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	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	东方人工智能主题 C	契约型开放式	27.98	0.73	-25.94	-30.96	-26.81
159702	汇添富中证人工智能 ETF	契约型开放式	0.13	0.79	-	-	-
159819	易方达中证人工智能 ETF	契约型开放式	44.92	0.60	-20.67	-18.57	-31.53
161631	融通中证人工智能主题 A	契约型开放式	5.05	0.97	-19.23	-17.21	-29.56
512930	平安中证人工智能 ETF	契约型开放式	3.93	0.91	-20.47	-18.46	-31.40
515070	华夏中证人工智能 ETF	契约型开放式	17.00	0.76	-20.62	-18.63	-31.54
515980	华富中证人工智能产业ETF	契约型开放式	10.52	0.63	-21.61	-18.51	-33.88
517800	方正富邦中证沪港深人工智能 50ETF	契约型开放式	0.85	0.45	-21.64	-23.95	-37.37

资料来源: wind,中国银河证券研究院

表4: 国内人工智能上市公司近况一览 (截至 2024.1.31)

证券代码	证券简称	2023Q3 营业收入同比增长率 (%)	2023Q3 营业利润同比增长率 (%)	2023Q3 营收增速 (%)	2023Q3 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	-8.85	-52.46	-8.85	-48.84	413.23	31.29	0.64	-15.45	-15.45
002230.SZ	科大讯飞	-0.37	-99.41	-0.37	-87.81	889.22	369.99	4.74	-17.21	-17.21
002236.SZ	大华股份	5.23	64.89	5.23	58.71	511.96	15.89	1.62	-15.77	-15.77
002362.SZ	汉王科技	5.04	-93.80	5.04	-99.40	42.00	-23.32	2.90	-32.94	-32.94
002405.SZ	四维图新	11.58	-505.16	11.58	-491.03	154.32	-19.98	4.30	-27.08	-27.08
002410.SZ	广联达	7.02	-56.48	7.02	-57.56	203.61	34.51	2.97	-28.65	-8.06
002415.SZ	海康威视	2.60	0.23	2.60	1.33	2978.33	21.10	3.33	-8.06	-29.12
300033.SZ	同花顺	4.25	-17.25	4.25	-12.68	618.46	39.16	16.92	-26.67	-28.27
300229.SZ	拓尔思	-5.57	-78.23	-5.57	-77.63	95.04	203.01	10.90	-29.12	-24.20
300474.SZ	景嘉微	-35.66	-120.07	-35.66	-91.00	232.00	176.53	25.95	-28.27	-20.61
300496.SZ	中科创达	0.57	-15.21	0.57	-14.53	230.32	31.31	4.21	-37.45	-34.60
300624.SZ	万兴科技	30.69	59.29	30.69	83.28	107.55	131.24	7.48	-17.44	1.47
300634.SZ	彩讯股份	24.62	91.22	24.62	88.17	67.37	18.12	4.98	-26.48	-36.31

301162.SZ	国能日新	40.45	51.99	40.45	37.97	36.57	44.40	8.16	-28.97	-18.49
601360.SH	三六零	-2.84	89.69	-2.84	81.05	488.03	-80.30	5.23	-24.20	-26.49
603019.SH	中科曙光	5.03	17.11	5.03	14.27	458.83	27.96	3.43	-20.61	-16.05
603893.SH	瑞芯微	-7.37	-85.09	-7.37	-71.99	193.79	196.23	10.13	-26.89	-28.65
688088.SH	虹软科技	28.67	25.05	28.67	54.78	108.97	126.43	16.93	-34.60	-29.77
688111.SH	金山办公	16.99	7.95	16.99	7.44	989.10	82.59	22.69	-32.25	-28.08
688169.SH	石头科技	29.51	59.07	29.51	59.10	377.50	22.35	4.76	1.47	-32.14
688188.SH	柏楚电子	51.91	46.19	51.91	49.41	371.69	55.89	30.04	0.35	-26.89
688207.SH	格灵深瞳	14.76	8.47	14.76	-8.33	34.57	110.33	9.04	-36.31	-26.67
688246.SH	嘉和美康	18.10	51.01	18.10	105.90	31.42	37.15	3.93	-32.14	-37.45
688256.SH	寒武纪-U	-44.84	15.59	-44.84	15.54	458.25	-40.93	75.06	-18.49	-32.25
688327.SH	云从科技-UW	-24.13	32.16	-24.13	32.20	122.77	-18.04	29.49	-29.77	-29.77
688343.SH	云天励飞-U	-11.37	11.63	-11.37	11.75	125.79	-30.88	24.33	-28.08	-28.08
688475.SH	萤石网络	11.83	92.73	11.83	83.27	219.38	42.56	4.69	-13.14	-13.14
688787.SH	海天瑞声	-38.25	-399.49	-38.25	-272.40	31.97	-150.13	16.07	-26.49	-26.49
688793.SH	倍轻松	45.10	82.64	45.10	72.73	24.28	-29.17	2.04	-16.05	-16.05

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（截至 2024.1.31）

证券代码	证券简称	2023Q3 营业收入（亿元）	2023Q3 营业收入同比增长率（%）	2023Q3 净利润（亿元）	净利润近 1 年增长率（%）	总市值（亿美元）	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	716.06	25.31	70.69	19.44	5964.79	39.77	7.32	-24.63	-24.63
QCOM.O	高通	271.89	-17.12	57.42	-44.09	1657.37	21.35	3.75	2.68	2.68
NVDA.O	英伟达	388.19	85.53	174.75	-55.21	15197.17	80.46	56.34	24.24	24.24
MSFT.O	微软	1557.26	6.37	522.80	18.10	29541.93	35.79	14.90	5.73	5.73
META.O	Meta	947.91	12.25	250.81	68.53	10026.06	33.72	8.53	10.22	10.22
GOOGL.O	谷歌	2210.84	6.91	531.08	23.05	17418.63	23.60	6.16	0.29	0.29
BIDU.O	百度	996.47	9.99	177.16	-26.08	368.16	11.66	2.07	-11.57	-11.57
AAPL.O	苹果	2937.87	-3.42	740.39	-2.81	28511.74	29.40	7.22	-4.22	-4.22
9988.HK	阿里巴巴	6604.87	1.77	491.39	-51.66	1812.76	9.82	1.35	-8.00	-8.00
0700.HK	腾讯控股	4538.19	10.80	881.91	-16.27	3272.14	12.07	4.12	-7.83	-7.83

资料来源：wind,中国银河证券研究院

二、人工智能产业政策动态

（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
1.17	上海医保大数据创新实验室启动医保数据要素流通再进一步 1月17日，上海市大数据中心、上海市医保中心、中保科联共同启动了上海市医保大数据创新实验室（商业保险），旨在推进医保商保数据融通共享，推进商业健康保险在精算研究、产品研发、服务创新等方面的探索应用。国家数据局等部门印发的《“数据要素×”三年行动计划（2024—2026年）》中提到，在医疗健康方面，依法依规探索推进医保与商业健康保险数据融合应用，促进基本医保与商业健康保险协同发展等。实验室的启动，是在该行动计划出台后，上海在医保数据要素融通共享上迈出的第一步。

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。

1.15	朝阳区启动“数据要素生态合作伙伴计划”，打造国际数据要素产业集聚区 朝阳区启动“数据要素生态合作伙伴计划”，打造国际数据要素产业集聚区，该计划旨在促进数据流通交易，激发数据价值释放，服务产业发展。数据要素生态合作伙伴面向数据供给、加工、治理、流通、交易、应用等全生命周期的各类企事业单位。伙伴应具备数据要素技术、数据价值驱动、数据要素服务等能力，且有意愿有能力参与朝阳数据要素产业建设。伙伴遴选范围包括数据要素技术型企业、数据要素价值驱动企业、数据要素价值驱动企业三类。
1.10	浙江：探索建立数据产权交易相关制度发展数据要素市场 1月10日，浙江发布《浙江省人民政府办公厅关于加快人工智能产业发展的指导意见》，在培育数据要素市场方面，浙江将加快一体化智能化公共数据平台建设，有序开发利用公共数据，探索建立数据产权交易相关制度，发展数据要素市场，培育优质数据服务商和第三方服务机构。在确保数据安全的前提下，推动医疗、交通等公共数据与企业数据融合创新应用，探索面向具体场景的数据衍生产品和衍生服务交易。
1.4	国家数据局等17个部门联合印发《“数据要素×”三年行动计划（2024—2026年）》 1月4日，国家数据局等17个部门联合印发《“数据要素×”三年行动计划（2024—2026年）》。这也是国家数据局挂牌成立后，主导发布的首个文件。《行动计划》选取了工业制造、金融服务、科技创新、医疗健康等12个行业和领域，推动发挥数据要素乘数效应，释放数据要素价值，并明确了到2026年底的工作目标。
1.2	四川发布数据要素市场化配置综合改革“路径图” 1月2日，四川省大数据中心、四川省发改委、四川省经济和信息化厅、中共四川省委网信办印发《关于推进数据要素市场化配置综合改革的实施方案》。《实施方案》明确，到2027年底，与高质量发展要求相适应的数据要素市场化配置体制机制基本建立、配套制度逐步健全、安全治理能力持续提升。数据资源体系将基本完善，数据大体量汇聚、高质量治理、高效化流动能力不断增强，数据作为关键生产要素作用进一步凸显。数据要素价值体系将基本形成，数据要素市场主体规模持续壮大、质量稳步提升，数据要素服务经济社会发展能力显著增强。

资料来源：wind，中国银河证券研究院整理

表7：数据交易所最新新闻及政策

日期	具体内容
1.26	北京石景山签约大数据交易所共建数据资产流通创新中心 1月26日，北京市石景山区与北京国际大数据交易所宣布达成战略合作，共同建设“数据资产流通创新中心”，充分挖掘数据资产价值，推动数字资产的流通模式不断演进，为各行各业发展提供支持。此次合作是在由北京市科委、中关村管委会、北京市经济和信息化局等主办的数据资产高质量发展大会上宣布的。
1.23	深圳数据交易所与福田区政数局联合发布“福田区公共数据授权运营专区” 近日，深圳数据交易所与福田区政务服务数据管理局联合发布“福田区公共数据授权运营专区”，以规范、促进福田区公共数据开发利用，加快释放公共数据价值。福田区在全市率先推动公共数据授权运营改革，在全国率先出台了区县级公共数据授权运营管理办法，在鼓励公共数据和社会数据融合应用的同时，深化数字政府改革建设，提升政府治理能力和公共服务水平，推动数字经济、数字社会发展。
1.8	郑州：联合共建国家级区域性数据交易所、数据要素化综合改革研究院（郑州分院） 郑州市人民政府印发郑州市数字政府建设行动方案。其中提出，积极开展创新试点示范建设。立足郑州国家中心城市建设大局，围绕重点领域、关键环节、共性需求等有序开展试点示范。加快推进数据要素市场化配置改革创新，谋划建设数据要素产业载体，联合共建国家级区域性数据交易所、数据要素化综合改革研究院（郑州分院），努力打造全国数据交易改革创新策源地和数字产业发展高地。加快未来产业抢滩布局，前瞻布局量子信息、区块链、网络安全等创新试点。

资料来源: wind, 中国银河证券研究院整理

(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
2024.1.30	讯飞星火大模型 V3.5	1 月 30 日, 科大讯飞将发布星火大模型 V3.5。据悉, 该平台基于科大讯飞与华为联合发布中国首个全国产支持万亿参数大模型训练的“飞星一号”平台完成训练。星火大模型 V3.5 在逻辑推理、语言理解、文本生成、数学答题、代码、多模态各个能力方面均实现大幅提升, 进一步逼近 GPT-4Turbo 的最新水平; 同时星火语音大模型将首次发布, 在首批 37 个主流语种上已整体超越 OpenAI 公司推出的 Whisper-large-v3, 保持科大讯飞智能语音技术的国际领先水平; 此外, 公司还将同时发布首个深度适配国产算力的讯飞星火开源大模型。
2024.1.29	百川智能 Baichuan3	1 月 29 日, 百川智能发布了超千亿参数的大语言模型 Baichuan3。该模型采用“动态数据选择”、“重要度保持”及“异步 CheckPoint 存储”等技术, 显著提升了性能, 相较业界主流框架提升超过 30%。其稳定训练时间长达一个月以上, 且故障恢复迅速, 不超过 10 分钟。在多项能力测评中, 如 CMMLU、GAOKAO、AGI-Eval 以及数学和代码专项评测 MATH、HumanEval、MBPP 等, Baichuan3 均表现出色。
2024.1.26	值得买 AI 问答机器人	1 月 26 日, 值得买科技公开宣布, 什么值得买 APP 的大模型产品“AI 问答机器人”, 正式通过国家《生成式人工智能服务管理暂行办法》备案。据介绍, 什么值得买“AI 问答机器人”是一款消费领域垂类产品, 将通过对话形式与用户进行交互, 旨在为用户提供智能的消费辅助决策服务, 根据用户需求指令, 帮助用户快速甄选全网优质好物链接。
2024.1.25	街远科技品商大模型、多模态大模型“摹小仙”	1 月 25 日, 街远科技发布文章, 宣布公司自主研发的品商大模型和多模态大模型“摹小仙”成功通过备案。值得注意的是, 这批通过备案的企业中, 街远科技是唯一有两个大模型产品通过备案的企业。据公司有关负责人表示, 品商大模型为企业提供从市场洞察到营销创新的全链路支持, 已在消费品(食品饮料、3C、家电、鞋服)、金融、政务等多个行业得到广泛应用。“摹小仙”则是街远科技在多模态领域的最新突破, 其参数规模超过 35 亿, 专为服装、美妆、食品饮料等行业设计。据介绍, 这也是行业内首个实现此类功能的模特 AI 商拍平台。
2024.1.24	岩芯数智 Yan 模型	1 月 24 日, 上海岩芯数智人工智能科技有限公司, 正式发布了国内首个非 Attention 机制的通用自然语言大模型——Yan 模型。作为行业内少有的非 Transformer 大模型, Yan 模型用全新自研的“Yan 架构”代替 Transformer 架构, 用百亿级参数达成千亿参数大模型的性能效果——记忆能力提升 3 倍、速度提升 7 倍的同时, 实现推理吞吐量的 5 倍提升。它去除了 Transformer 中高成本的注意力机制, 代之以计算量更小、难度更低的线性计算, 大大提高了建模效率和训练速度, 效率翻倍的同时实现了成本的骤降。
2024.1.23	第四范式式说大模型	1 月 23 日, 第四范式公开宣布, 公司的式说大模型正式通过国家《生成式人工智能服务管理暂行办法》备案。据介绍, 式说大模型于去年 2 月发布并经过不断迭代, 致力于用生成式 AI 改造企业软件(AIGS), 并先后与金融、零售、制造、物流、教育、医疗等行业客户开展大模型合作。后续, 式说将继续深化与企业客户在大模型侧的联合探索, 提供更加可靠的服务。
2024.1.23	阿里通义千问 GPT-BI	1 月 23 日消息, 由中国一汽联合阿里云通义千问打造的大模型应用 GPT-BI 日前已率先落地, 为中国一汽的数字化转型升级增添新活力。该应用可接收自然语言查询, 结合企业数据自动生成分析图表, 目前可达到近 90% 的准确率。另外, 相比传统 BI (BusinessIntelligence) 的“固定问答”, 它能实现问答任意组合, 数据随时穿透, 做到

		“问答即洞察”。目前，中国一汽正在积极探索“GPT+”的大模型创新范式，GPT-BI 不仅是中国一汽的首个大模型落地案例，也是汽车行业的首个大模型 BI 应用。
2024.1.22	零一万物 Yi-VL 多模态语言模型	1 月 22 日，零一万物 Yi 系列模型家族迎来新成员，YiVisionLanguage（Yi-VL）多模态语言大模型正式面向全球开源。凭借卓越的图文理解和对话生成能力，Yi-VL 模型在英文数据集 MMMU 和中文数据集 CMMMU 上取得了领先成绩，展示了在复杂跨学科任务上的强大实力。据悉，Yi-VL 模型基于 Yi 语言模型开发，包括 Yi-VL-34B 和 Yi-VL-6B 两个版本。在全新多模态基准测试 MMMU 中，两个版本均有不俗表现。
2024.1.21	猎户星空大模型	1 月 21 日，猎户星空大模型发布。这款大模型专为企业应用而生，旨在为企业提供更高效、智能的 AI 解决方案。据傅盛介绍，猎户星空大模型通过 140 亿参数实现了千亿参数大模型才能实现的效果，这意味着它在性能和效果上具有很高的竞争力。同时，该模型还可以面向素质教育等七大应用领域进行微调，以满足不同行业的实际需求。此外，猎户星空大模型还具有在千元显卡算力上运行的优点，这意味着企业可以更轻松地部署和运行该模型。
2024.1.18	新壹视频大模型	1 月 18 日，新壹视频大模型，通过了《生成式人工智能服务管理暂行办法》备案，全面向社会开放服务，新壹科技也进一步夯实了在 AI 生成视频领域领头羊的地位。新壹视频大模型发布于 2023 年 7 月 27 日，现已具备从视频脚本生成、素材匹配原生 AI 视频生成、声音克隆、智能剪辑到数字人播报的全流程 AI 视频生产能力，是目前国内首款聚焦视频领域的人工智能大模型。
2024.1.16	智谱 AIGLM-4	1 月 16 日智谱 AIDevDay 正式发布新一代基座大模型 GLM-4，宣布性能全面逼近 GPT-4。同时上线 GLM-4AllTools 平台、GLMs 智能体创建能力和 GLMs 商店，后续将推出开发者收入计划。全栈自主创新 GLM-4 全家桶，在多模态理解、复杂长文本、和自动解决复杂任务上都有较大提升，全面对标 GPT-4
2024.1.16	比亚迪璇玑 AI 大模型	比亚迪发布智能化领域最新技术，其中整车智能化架构被命名为“璇玑”，由“中央大脑”、车端 AI 和云端 AI，车联网、5G 网、卫星网，及传感链、控制链、数据链、机械链组成。其中，“中央大脑”可实现对多种 Soc 芯片的兼容，还能通过芯片解耦带来算力的扩充。主控芯片 AI 模块为通用的 GPU 架构，搭配模块化设计，能够实现存算的任意分配，按照功能需求调整算力方式，做到无缝切换和适配未来的算法模型，让汽车的功能可以实现快速迭代和进化。据悉，该“中央大脑”由比亚迪自研、自产。
2024.1.11	深度求索 DeepSeekMoE	幻方量化旗下组织深度求索发布了国内首个开源 MoE 大模型——DeepSeekMoE，全新架构，免费商用。混合专家模型(MixedExpertModels, 简称 MoEs)是用于提高大语言模型效率和准确度的技术。这种方法的核心是将复杂任务划分为更小、更易管理的子任务，每个子任务由专门的小型模型或“专家”负责，然后根据输入数据的特性选择性地激活这些“专家”。
2024.1.8	字节跳动 Vista-LLaMA	1 月 8 日，字节跳动与浙江大学合作推出了多模态大语言模型 Vista-LLaMA，该模型专为视频内容理解而设计，能够输出高质量视频描述。Vista-LLaMA 在多个开放式视频问答基准测试中表现卓越，尤其在 NExT-QA 和 MSRVT-TQA 测试中取得了突破性成绩。其在零样本 NExT-QA 测试中实现了 60.7%的准确率，在 MSRVT-TQA 测试中达到了 60.5%的准确率，超过了目前所有的 SOTA 方法。
2024.1.4	阿里巴巴 AnyText	阿里达摩院宣布推出名为“AnyText”的算法，旨在解决 AI 绘图在处理文字创作时面临的挑战。AnyText 算法的核心优势在于其强大的自然语言理解和图像生成能力。这项新技术可以准确理解给定文本的语境并有效地将文字与视觉元素结合，无论是古典书法的模仿还是现代字体的创意设计，都能游刃有余，大大提升 AI 绘画的多样性和创造力。 目前，AnyText 支持中英日韩四种语言。该项目在魔搭社区放出了 Demo，可以通过部署到本地使用，也可以直接在魔搭社区试用。

2024.1.4	网易有道子曰教育大模型 2.0	网易有道在北京举行子曰教育大模型创新成果发布会，正式推出子曰教育大模型 2.0，上下文窗口扩大到 16K，并宣布将基于子曰教育大模型自研的 RAG 引擎“QAnything”正式开源。此外，网易有道还发布了有道 AI 学习机 X20。
2024.1.2	美图 AI 视觉大模型 MiracleVision	1 月 2 日，美图公司自研 AI 视觉大模型 MiracleVision(奇想智能)通过《生成式人工智能服务管理暂行办法》备案，将面向公众开放。截至目前，在通过备案的国内大模型当中，美图公司 MiracleVision(奇想智能)是唯一专注视觉领域的大模型，也是福建省唯一通过备案的大模型。

资料来源：wind，中国银河证券研究院整理

表9：国外人工智能大模型动态

时间	企业名称	主要内容
2024.1.29	Meta	Meta 宣布开源 CodeLlama70B 参数模型 当地时间 1 月 29 日，Meta 首席执行官马克·扎克伯格宣布免费开源 CodeLlama70B 参数模型。据介绍，CodeLlama70B 可以处理比先前版本更多的查询，这意味着开发人员可以在编程时提供更多提示，并且它可以更准确。CodeLlama70B 在 HumanEval 基准测试中的准确率为 53%，优于 GPT-3.5(48.1%)，接近 OpenAI 论文中报告的 GPT-4(67%)。
2024.1.26	OpenAI	OpenAI 连更 5 款模型：修复 GPT-4 “变懒” BUG，GPT-3.5Turbo 模型输出降价 50% OpenAI 今天一口气宣布了 5 个新模型，包括两个文本嵌入模型、升级的 GPT-4Turbo 预览版和 GPT-3.5Turbo、一个审核模型。更新版的 GPT-4Turbo——gpt-4-0125-preview 不仅大幅改善了模型「偷懒」的情况，而且还极大地提升了代码生成的能力。GPT-3.5Turbo 更新——gpt-3.5-turbo-0125 输入价格大降 50%，输出价格下降 25%。性能上也有升级。迄今为止最强的审核模型 text-moderation-007，帮助开发者有效地识别有害文本。两个新一代 embedding 模型：text-embedding-3-small，体积更小效率更高，价格直接砍到了原模型 20%！支持 512 维和 1536 维两种大小；text-embedding-3-large，性能相比之前模型大幅提升，高达 3072 维的嵌入向量，价格是小模型的 6.5 倍。最后，OpenAI 还强调称，用户通过 API 传送的数据，都不会被用于训练新的模型。
2024.1.24	谷歌	Google 团队发布文本生成视频模型 Lumiere 1 月 24 日，Google 研究员 HilaChefer 在社交媒体 X 上宣布，文生视频模型 Lumiere 正式上线。研究团队为 Lumiere 引入了时空 U-Net 架构，使得生成视频在运动幅度和一致性表现良好，同时提供多种风格选择，支持生成内容的局部修改。
2024.1.17	StabilityAI	StabilityAI 开年首个大模型：媲美 70 亿 CodeLlama，没 GPU 也能跑 StabilityAI 发布了自家在 2024 年的第一个模型——StableCode3B。StableCode3B 专注于代码能力，在仅仅 3B 参数的规模之下，达到了比肩 CodeLlama7B 的效果。相较于 CodeLlama7B，StableCode3B 的体积减少了 60%，但在多种编程语言上保持了相当的水准，在 Python 和 C++的代码补全中甚至反超了 CodeLlama7B。另外，由于模型仅有 3B 大小，StableCode3B 可以在 MacBookAir 等普通笔记本电脑上实时运行，甚至没有独立 GPU 也可以！
2024.1.2	微软	微软研究人员推出 WaveCoder：增强指令调优与精炼数据生成 近期的研究表明，通过在高质量指令数据集上进行微调，生成的模型可以在广泛的任务上展现出出色的能力。然而，现有的指令数据生成方法通常会产生重复数据，并且在数据质量上不够可控。微软研究人员最新研究通过将指令数据分类为 4 个与代码相关的任务，并提出了基于 LLM 的生成器-鉴别器数据处理框架，从开源代码中生成多样、高质量的指令数据，从而扩展了指令调优的泛化能力。

资料来源：wind，中国银河证券研究院整理

(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

日期	具体内容
1.31	苏姿丰：2024 年 AMD 的 AI 芯片营收将达 35 亿美元 当地时间 1 月 30 日，AMD 公布了 2023 年第四季度及 2023 年全年财报。2023 年全年，AMD 营收为 227 亿美元，毛利率为 46%，营业收入为 4.01 亿美元，净利润为 8.54 亿美元，稀释每股收益为 0.53 美元。AMDCEO 苏姿丰在 1 月 30 日的电话会议上预测，2024 年 AMD 的 AI 芯片营收将达 35 亿美元，高于先前预测的 20 亿美元。另外，去年四季度的 AI 芯片营收也超越此前预测的 4 亿美元，但她并未披露确切的数字。
1.15	英特尔推出汽车版 AI 芯片，与高通、英伟达展开竞争 在 2024 年的国际消费电子展 (CES) 上，英特尔正式发布了一款专为汽车领域设计的人工智能 (AI) 芯片。这一创新产品标志着英特尔正式进军车载 AI 市场，与高通和英伟达等强劲对手展开正面竞争。这款 AI 芯片旨在为汽车提供强大的 AI 功能，满足车载信息娱乐系统、自动驾驶和高端仪表盘等领域的多样化需求。英特尔表示，这款芯片具有高性能和低功耗的特点，能够为汽车带来出色的 AI 处理能力，提升驾驶体验。
1.9	英伟达在 RTX40SUPER 系列上“加量降价” 北京时间 1 月 9 日凌晨，英伟达发布 RTX40SUPER 系列，其主要产品 RTX4080SUPER 效能是 RTX3080Ti 的两倍，而价格却比后者便宜 200 美元。英伟达公布的数据显示，创作者使用生成式人工智能的应用进行内容生产时效率有 1.5 倍左右的提升。不仅如此，英伟达还展示了针对设计者的人工智能赋能案例：在英伟达 ACE 平台中，游戏开发者仅需进行语音输入，便可以通过大模型自动生成 NPC 的回复，并还原至游戏场景，提升开发效率；同时，英伟达与 iStock 合作，通过生成式 AI 实现背景绘制，甚至是单独图案元素的替换和重新生成。
1.9	技嘉科技(GIGABYTE)推出人工智能/高性能计算(AI/HPC)服务器 在 CES 期间，技嘉科技及其子公司 GigaComputing 展示了四款顶级的 AI/HPC 服务器。其中 G383-R80 服务器是最新推出的一款产品，采用 AMDInstinct™MI300AAPU，并集成了 CPU、GPU，以及高达 128GB 的统一 HBM3 内存，可满足实时大数据分析的需要。另一款是 G593 系列人工智能服务器，搭载 NVIDIAHGXH1008-GPU，以无可匹敌的高密度 5U 配置为人工智能模型训练和推理提供强力支持。这款服务器在 MLPerfTraining 基准测试中表现突出，可轻松处理各种人工智能工作负载。针对可扩展和灵活的计算需求，技嘉科技推出了 XH23-VG0，这款服务器搭载 NVIDIAGraceHopper 超级芯片，遵循 MGX 模块化设计，并提供了适用于大型 AI/HPC 应用的 FHFL 扩展槽。G493-SB0 服务器支持 8 个双槽 Gen5GPU，适用于广泛的工作负载，包括生成式人工智能、虚拟化、渲染和 3D 图形等。
1.8	AMD 正式发布桌面锐龙 8000GAPU 时隔长达 2 年 9 个月，AMD 终于升级了桌面 APU 产品线，带来了全新的锐龙 8000G 系列，这也是第五代桌面锐龙 APU。AMD 推出锐龙 8000G 系列，通过核显强调 AI 能力。不同于专注在图形处理技术上的英伟达，AMD 此次的最高规格产品锐龙 78700G 则展示了在脱离独显的环境下，凭借核显 Radeon780M 仍能畅玩《赛博朋克 2077》的强大能力。在 AI 能力方面，AMD 计算和图形事业部高级副总裁兼总经理 JackHuynh 介绍，8000G 采用了 Zen4CPU 内核，并集成了具有 16TOPS 算力的 NPU，加上 CPU 和 GPU，最高算力可达 39TOPS。
1.1	英伟达 H20AIGPU 芯片有望于 2024 年第二季度开始量产 1 月 1 日，据外媒报道，英伟达 H20AIGPU 芯片有望于 2024 年第二季度开始量产，纬创为独家供应商。报道中提到，这款芯片为英伟达专为中国大陆设计，计算性能进行了限制，以符合最新出口要求。此前，英伟达方面表示，正在为中国大陆开发新的“合规芯片”。据悉，英伟达 H20 降低了 AI 算力，但对比起来，有着更低的售价、支持 NVLink 高速互联技术以及 CUDA 等。

资料来源：wind，中国银河证券研究院整理

三、人工智能前沿行业动态

(一) 前沿技术动态

1、探索宇宙对人脑演化的奥秘，中国空间站脑电测试完成首个阶段目标

IT 之家 1 月 25 日消息，据央视新闻今日报道，1 月 24 日是中国空间站天和核心舱发射入轨 1000 天，截至目前，中国空间站已经安排在轨实施了 110 个空间科学研究与应用项目，涉及空间生

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。

命科学与人体研究、微重力物理和空间新技术领域。其中，在轨脑电测试已完成首个阶段性目标任务。

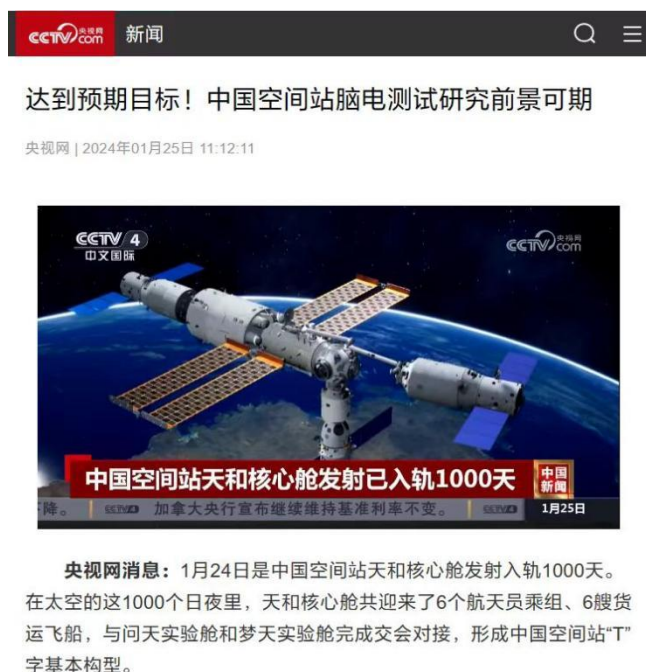
据中国航天员科研训练中心科技人员王波介绍，首次脑电实验可追溯到神舟十一号，通过脑控机器人验证了脑机交互技术的在轨适用性。从空间站任务开始，研究人员陆续开展了数十次在轨脑电测试，达到了空间站建造期的预期目标。

在轨脑电测试研究的目的有三个：观察空间环境对脑造成了何种影响；观察人脑在空间如何适应环境，重塑自己的神经；发展或验证脑能力提升的技术。

从长远来看，空间站开展在轨脑电测试研究的目的是探索宇宙对人脑演化的奥秘，揭示生命体演化过程中的重要机制，从而为开发类脑智能体提供新思路。

中国航天员科研训练中心科技人员曹勇谈到了脑电测试实验的操作方式：整个实验开始时，航天员会被引导进入一个静息状态，采集一段相对平稳的状态的数据，作为后续任务数据的基线。在静息完之后，航天员就会根据任务进程操作操纵杆、按键，对任务进行回应。研究人员会结合脑电采集的数据情况，综合判断航天员的脑力负荷状态。

图3：中国空间站脑电测试完成首个阶段目标



资料来源：IT之家，中国银河证券研究院

2、AI 致命弱点：科学家首次用数学方法证明无法开发出稳定算法

IT之家 1月25日消息，哥本哈根大学的研究人员在世界上首次用数学方法证明，在各种基础问题之上，不可能开发出始终稳定的人工智能算法。

哥本哈根大学的研究人员已经证明，处理复杂问题的完全稳定机器学习算法近乎无法实现，这凸显了对人工智能局限性。

以 ChatGPT 为代表的 AI 浪潮下，机器能比医生更准确解读医学扫描图像、能比人类更安全地驾驶汽车，但再优秀的算法也存在弱点。

以自动驾驶汽车读取路标为例，如果有人在路标上贴有标签，人类驾驶员的注意力通常不会被分散，而机器由于和接受训练的路标存在差异，很容易出现分心。

该小组负责人阿米尔·耶胡达约夫（Amir Yehudayoff）教授说：“现实生活中存在着各种各样的噪音，人类习惯于忽略这些噪音，而机器却会感到困惑”。

3、百度文心大模型落地吉利银河 L6 车型，支持 AI 对话功能

IT 之家 1 月 22 日消息，据 Apollo 智能驾驶消息，近日，百度 IDG 与吉利汽车合作的 AI 对话产品，在银河 L6 车型上成功量产。这是汽车行业首个基于大语言模型底座能力落地的 AI 车载对话产品。

据介绍，从去年 3 月份文心一言大模型发布以来，IDG 智舱团队就开始了座舱大模型产品重构方向的探索，经过了近 10 个月的产品设计和体验打磨，百度打造的车载 AI 原生应用“小度车载智慧助手”成功的完成 0 到 1 阶段，进入到规模化商业落地阶段。

吉利银河 L6 的 AI 对话主打四个核心功能点，IT 之家总结如下：

知识问答：应对用户的提出的各类问题，聊新闻、娱乐、健康等知识均可应答，能够准确理解用户的意图，提供个性化的回应和建议。

高情商回复：比如在带娃出行的场景下，可以借助 AI 对话能力让语音助手扮演成一个小小朋友陪小孩聊天；在长途自驾场景下，可以让语音助手扮演成某个喜爱的电影角色。

行程规划：用户在自驾旅行时，可以通过 AI 对话，快速制定自驾游线路和行程。如自驾到一个新城市，车主可以通过 AI 对话快速了解当地的天气状况、景点建议以及风土人情。根据上下文理解，更准确的给用户生成回复内容。

美食推荐：可以根据用户语音输入的口味喜好和位置信息推荐优质的美食和餐厅。比如，驾车过程中想寻找附近合适的餐厅，只需要把你的就餐地点、口味偏好、人数等信息告诉语音助手，就能获得推荐餐厅清单。

Apollo 智能驾驶表示，银河 L6 量产的 AI 对话产品是百度 IDG 与吉利智能化合作项目的一个重要里程碑，展望未来，双方将继续深化合作关系，共同推动大模型技术在汽车领域的广泛应用。

吉利银河 L6 于去年 9 月上市，该车长宽高分别为 4782/1875/1489mm，轴距 2752mm，拥有银河炫彩白、夜魅紫、山岚青、晨曦白、拂晓灰等配色，售价为 11.58 万-14.98 万元。该车标配银河 NOS 系统，内置高通骁龙 8155 芯片，出厂预装了 42 款常用 App，后续可提供 OTA 升级。

银河 L6 搭载新一代雷神电混 8848 动力系统，匹配 1.5T 混动专用发动机+3 挡变频电驱 DHTPro，可实现 6.3 秒百公里加速，车速最快达到 247km/h；能耗方面，该车拥有 60km/125km 纯电续航，最大 1370km 综合续航、百公里亏电油耗 4.55L。

4、国内首个与 ChatGPT 不同机制的通用大模型发布，算力消耗缩减 50%

1 月 24 日下午，国内 AIGC 企业上海岩芯数智人工智能科技有限公司（岩芯数智，RockAI）在上海发布了国内首个自研的非 TransformerAttention 机制的低算力通用自然语言大模型——Yan 模型，记忆能力提升 3 倍、速度提升 7 倍，推理吞吐量提升 5 倍。

这是国内首个发布的与 ChatGPT 不同机制的通用大模型，参数规模达百亿。该公司称，这是用百亿级参数达成千亿参数大模型的性能效果。

岩芯数智 CEO 刘凡平表示，“在全球范围内，一直以来都有不少优秀的研究者试图从根本上解决对 Transformer 架构的过度依赖，寻求更优的办法替代 Transformer。历经超过几百次的设计、修改、优化、对比、重来后，岩芯数智自主研发出了不再依赖 Transformer 的全新架构——‘Yan 架构’，同时，基于 Yan 架构的低算力通用大模型应运而生。我们期望 Yan 架构可作为 AI 领域的基础设施，并以此建立 AI 领域的开发者生态，最终让任何人在任何设备上都能使用通用大模型，获取更加经济、便捷、安全的 AI 服务。”

据悉，成立于 2023 年 6 月的岩芯数智，是 A 股上市公司岩山科技（002195.SZ）旗下宣布拥抱 AIGC 新浪潮、向数智化转型并更名后成立的新公司，专注于打造为百业赋能的一站式 AIGC 数智化服务平台。而岩芯数智的团队则于 2022 年在内部组建。

Yan 模型是岩芯数智最新自研的百亿级参数模型，主要针对 ToB 行业领域，目前公司已提供多种场景的解决方案。例如 RockAI 模型大脑可以让企业拥有私域的知识大模型，实现跨文档知识整合、实时信息记录，可用于银行金融、政务办公等。

Yan 模型的最大特点,是采用非 TransformerAttention(注意力)机制。据悉,包括火热的 ChatGPT、BERT 等大模型,普遍采用 Transformer 基础架构,而 Attention 机制区别于 CNN 技术,是通过参数和大数据利用算力训练形成的技术体系,则根据应用场景的不同而输出不同信息,通俗来说是把注意力集中放在重要的点上而忽略其他不重要的因素。Attention 主要分为空间注意力和时间注意力,前者用于图像处理,后者用于自然语言处理。

刘凡平此前对钛媒体 App 表示,他们在对 Transformer 模型不断的调研和改进过程中,意识到了重新设计大模型的必要性:一方面,在 Attention 机制下,现有架构的调整几乎已经达到瓶颈;另一方面,岩芯数智更期望降低企业对大模型的使用门槛,让大模型在更少的数据、更低的算力下具备更强的性能,以应用于更广泛的业务。因此,岩芯数智希望研发非 Transformer 架构模型。

据介绍,相比 Transformer 架构的模型,Yan 模型算力消耗缩减 50%、推理效率提升 7 倍、准确率提高 3 倍、推理吞吐量提升 5 倍。针对长序列环境,在单张 409024G 显卡上,当模型输出 token 的长度超出 2600 时,Yan 模型显存使用始终稳定在 14G 左右,理论上能够实现无限长度的推理,而 Transformer 模型会出现显存不足。

此外,Yan 模型也同时具备私密化、经济化、精准化和实时性、专业性、通用性等六个商业化落地能力。而且,该团队首次研发了一种合理的关联特征函数和记忆算子,可降低模型内部结构的复杂度,发掘决策过程的透明度和可解释性,从而助力全新架构下的 Yan 模型在医疗、金融、法律等领域的广泛运用。

刘凡平透露,接下来岩芯数智将打造全模态实时人机交互系统,全面打通感知、认知、决策与行动,构建通用人工智能 (AGI) 的智能循环体 (类似于 Agents),为通用机器人等具身智能方向的研究提供大模型基础底座的“更多选择”。他强调,通过提供基于 Yan 架构的专业生产力工具,实现端侧训练、训推一体,推动企业数智化发展。

岩芯数智董事长陈代千表示,随着 Yan 模型的进一步落地和应用,未来公司将为机器人、嵌入式设备、物联网设备等领域提供所需的智能化能力,为企业和用户创造更多的价值。

图4: 岩芯数智 CEO 刘凡平



资料来源:钛媒体,中国银河证券研究院

5、全球首个汽车行业全栈自研全场景 AI 大模型! 吉利星睿 AI 大模型正式发布

继 1 月 5 日,吉利汽车集团 CEO 淦家阅在吉利银河 E8 上市发布会上,发布了 AI 大模型的相关能力及多款应用后,1 月 11 日,吉利正式发布全球首个汽车行业全栈自研全场景 AI 大模型——吉利星睿 AI 大模型。吉利星睿 AI 大模型以强大的星睿智算中心为算力底座,将自研基础大模型与

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。

吉利全球领先的 NPDS 研发体系、巨量造车全链路场景数据库深度融合，将成为汽车行业应用场景最丰富、算力最强大、汽车专业知识体系最完备，数据与模型最安全可靠的大模型。

全面践行吉利“造每个人的智能精品车”品牌新使命，吉利星睿 AI 大模型开创四大先河、首发十余项行业首创技术，不仅率先实现 AI 原生应用规模化量产上车，重新定义智能座舱新体验，成为每个用户的超级 AI 智能伙伴，更以大模型技术全链路赋能造车，成为汽车行业 AI 造车引领者，全面引领中国汽车全场景 AI 时代。

以生成式 AI 为代表的大模型技术是当下最前沿的人工智能技术，国内外科技巨头正积极布局各自大模型。2023 年 5 月，国家标准委也正式启动了大模型标准化工作。而吉利早在 2021 年就已在 AIGC 相关技术领域进行研究和布局。从去年 2 月上线星睿智算中心，到如今发布星睿 AI 大模型，吉利已形成从数据、算力到大模型应用的智能生态闭环。

星睿 AI 大模型包括语言大模型、多模态大模型、数字孪生大模型 3 大基础模型，并由此衍生出 NLP 语言大模型、NPDS 研发大模型、多模态感知大模型、多模态生成大模型、AIDRIVE 大模型、数字生命大模型 6 大能力模型，打破了汽车行业“重应用、轻技术”的现状，开创了车企自研基础模型的先河。在中国信通院大模型通用能力官方评测中，吉利星睿 AI 大模型荣获中国权威最高 4+ 评级，吉利也成为中国人工智能产业发展联盟汽车行业推进组的副组长单位，牵头汽车行业大模型行业标准制定工作。

图5：吉利星睿 AI 大模型



资料来源：搜狐网，中国银河证券研究院

同时，星睿 AI 大模型拥有超千亿参数量，海量常识和情感模块，能够更聪明、精准的和用户对话。它可以是一个对景点如数家珍的金牌导游、一个对汽车知识答疑解惑的技术专家、一个对新闻热搜了如指掌的时尚达人、甚至是一个能讲故事、画画的创作大师，为用户带来不止于车的全场景陪伴 AI 智能体验。汽车将加速进化成为可以像真人一样自然交互的“智慧生命体”，为用户打造颠覆式的智能交互体验。

值得一提的是，不同于科技巨头开发的通用大模型，星睿 AI 大模型深度聚焦汽车垂直领域，对车辆功能使用、常见车辆问题、交通法规、售后服务等海量知识库进行了大规模学习，拥有汽车行业最完备的专业知识储备，仅星睿语言大模型训练过程中就特别补充了汽车领域 39 类知识库。未来，星睿 AI 大模型知识结构还将持续更新迭代，成为用户最趁手的“汽车百科全书”。

在吉利丰富的大模型集群中，星睿 NLP 语言大模型基于星睿智算中心纯自研，总训练数据量超过 3 万亿 Tokens，全球领先，并具备首创的情感模块，实现出色的逻辑推理能力、上下文记忆能力，综合能力达到了中文前十水平，可以实现类人化情感交互。星睿 NPDS 大模型是吉利星睿语言大模型在研发端的创新应用，可以支持研发人员进行造型设计、机械设计、质量管控和自动驾驶虚拟训练。

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/948071001103006034>