华龙证券 CHINA DRAGON SECURITIES

证券研究报告

计算机

报告日期：2024 年 05 月 19 日

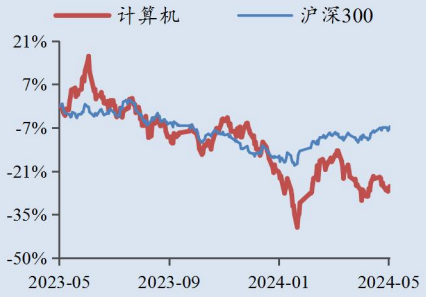
大模型 API 降价趋势下，看好 AI 应用加速落地

——计算机行业周报

华龙证券研究所

投资评级：推荐（维持）

最近一年走势



分析师：孙伯文

执业证书编号：S0230523080004

邮箱：sunbw@foxmail.com

相关阅读

《收入端稳定增长，利润端短期承压—计算机行业 2023 年年报及 2024 年一季报综述》2024.05.17

《Meta 推出开源 Llama 3，关注国内 AI 产业技术进步—计算机行业周报》2024.04.22

摘要：

- **国内大模型持续迭代，规模效应凸显。**字节跳动近期正式推出豆包大模型家族，共包含 9 个大模型。据字节跳动官微，豆包大模型每天处理 1200 亿 tokens 的文本（约 1800 亿汉字），生成 3000 万张图片。大量信息处理任务在提升模型输出效果的同时也能大幅降低模型推理的单位成本，豆包大模型家族中主力模型 API 定价相比行业价格低 99.3%。目前豆包大模型已通过火山引擎开放给企业客户。随着国内外大模型厂商技术角逐进一步激烈，我们认为大模型行业开启价格战信号明显。通过降低价格门槛，大模型厂商有望吸引更广泛的企业用户群体，从而进一步平衡收入和成本。同时，更多 C 端用户有望免费使用基础 AI 应用，庞大的访问量有助于企业进一步提升模型服务能力，完成良性循环。对下游企业来说，推理成本进一步降低，AI 应用行业有望迎来成本拐点。

- **GPT-4o 提效降价，多模态能力再进一步。**2024 年 5 月 13 日，OpenAI 发布新一代大模型 GPT-4o，GPT-4o 能够接受文本、音频、图像和视频的任意组合作为输入，并生成文本、音频和图像输出的任意组合。它可以在最短 232 毫秒内响应音频输入，平均为 320 毫秒，与人类正常对话的响应时间接近。与 GPT-4 Turbo 相比，GPT-4o 的速度快 2 倍，价格减半，速率限制高出 5 倍。我们认为 GPT-4o 进一步验证了大模型厂商 API 降价趋势，其技术路径对国内大模型产业具备积极的映射作用，建议关注国内相关厂商技术突破。

- **投资建议：**我们认为国内开源大模型持续迭代，大模型厂商开启价格战信号明显，AI 应用有望加速普及和创新，AI 应用行业成本拐点将更为明晰。同时算力作为 AI 产业底座，需求预期依旧强劲。另外，各地方低空经济政策陆续出台，产业有望协同发展。个股方面建议关注科大讯飞（002230.SZ）、浪潮信息（000977.SZ）、神州数码（000034.SZ）、海光信息（688041.SH）、中科曙光（603019.SH）、金山办公（688111.SH）、紫光股份（000938.SZ）、创业黑马（300688.SZ）、中科星图（688568.SH）。

- **风险提示：**国产算力建设不及预期；所引用数据资料的误差风险；AI 应用落地速度不及预期；国产大模型迭代速度不及预期；重点关注公司业绩不达预期；政策标准出台速度不及预期。

表 1：重点关注公司及盈利预测

股票代码	股票简称	2024/05/17	EPS（元）				PE				投资评级
		股价（元）	2023A	2024E	2025E	2026E	2023A	2024E	2025E	2026E	
000034.SZ	神州数码	30.4	1.79	2.10	2.82	2.99	16.95	13.54	10.11	9.52	增持
000938.SZ	紫光股份	21.96	0.74	0.90	1.08	1.22	29.88	24.51	20.42	18.05	未评级
000977.SZ	浪潮信息	39.12	1.18	1.47	1.81	2.14	33.14	26.64	21.67	18.25	未评级
002230.SZ	科大讯飞	43.19	0.28	0.38	0.52	0.65	154.25	113.60	83.65	65.98	未评级
300688.SZ	创业黑马	25.29	0.06	0.39	0.60	0.76	421.5	64.14	42.33	33.20	未评级
603019.SH	中科曙光	44.22	1.25	1.49	1.71	2.07	35.38	29.68	25.86	21.36	增持
688041.SH	海光信息	73.87	0.54	0.74	1.02	1.38	136.8	99.65	72.51	53.54	未评级
688111.SH	金山办公	281.08	2.86	3.61	4.76	6.18	98.28	77.76	59.08	45.49	未评级
688568.SH	中科星图	53.68	0.94	1.33	1.83	2.51	57.11	40.44	29.32	21.37	未评级

数据来源：Wind，华龙证券研究所

内容目录

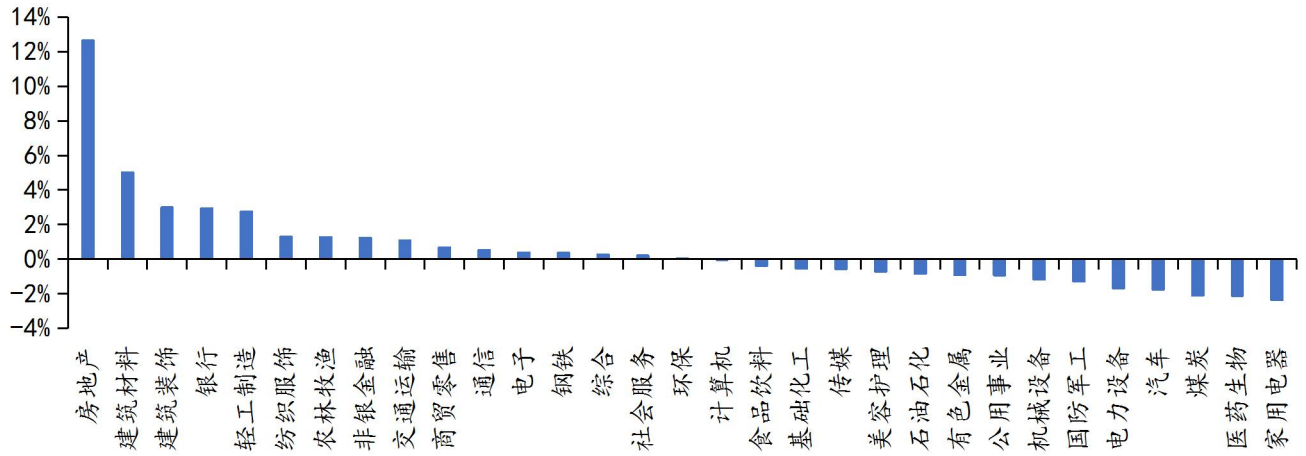
图目录

表目录

1 一周市场表现

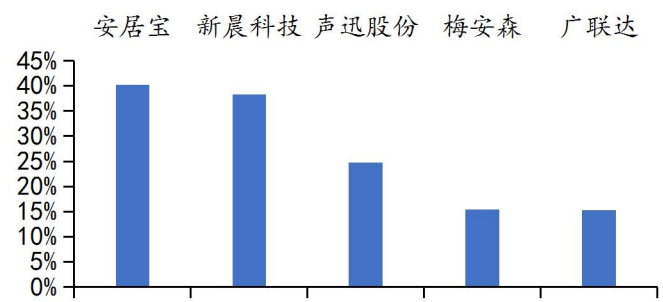
2024 年 5 月 13 日-5 月 17 日，申万计算机指数下跌 0.01%。板块个股涨幅前五名分别为安居宝（300155.SZ）、新晨科技（300542.SZ）、声迅股份（003004.SZ）、梅安森（300275.SZ）、广联达（002410.SZ）。板块个股跌幅前五名分别为 ST 汇金（300368.SZ）、ST 峡创（300300.SZ）、*ST 龙宇（603003.SH）、ST 证通（002197.SZ）、ST 英飞拓（002528.SZ）。

图 1：申万一级行业周涨跌幅一览



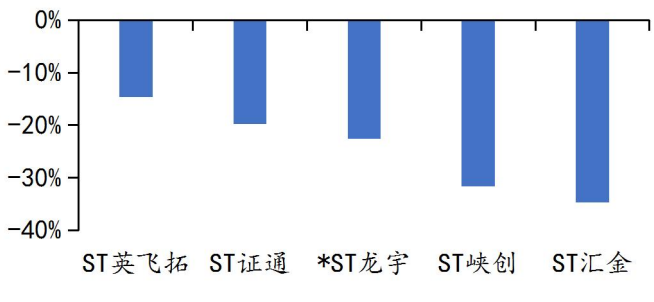
数据来源：Wind，华龙证券研究所

图 2：计算机股票周涨幅前五



数据来源：Wind，华龙证券研究所

图 3：计算机股票周跌幅前五



数据来源：Wind，华龙证券研究所

2 行业要闻

2024 年 5 月 17 日消息，北京市经信局近日就《北京市促进低空经济产业高质量发展行动方案（2024—2027 年）（征求意见稿）》向社会公开征求意见。征求意见稿明确，力争通过三年时间，北京低空经济相关企业数量突破 5000 家，低空技术服务覆盖全国，带动北京市经济增长超 1000 亿元。到 2027 年，北京将围绕应急救援、物流配送、空中摆渡、城际通勤、特色文旅等，新增 10 个以上应用场景，开通 3 条以上面向周边地区的低空航线，基本建成网络化的基础设施体系及低空应用生态。

3 重点公司公告

【宝信软件】2024 年 5 月 13 日消息，宝信软件将以自有土地（宝信张江软件园）建设智慧制造研发中心，项目总投资 53,987 万元，项目建设资金为公司自有资金。

【新致软件】2024 年 5 月 13 日消息，新致软件拟定增募资不超 3 亿元。本次发行股票的种类为境内上市人民币普通股（A 股），每股面值人民币 1.00 元。

4 本周观点

4.1 国内外大模型持续迭代，豆包大模型 API 大幅降价

4.1.1 商汤、阿里发布对标 GPT-4 Turbo 大模型，国产大模型技术突破可期

2024 年 5 月 9 日，阿里云正式发布通义千问 2.5 版本。同时推出 1100 亿参数开源模型 Qwen1.5-110B，在多个基准测评中，Qwen1.5-110B 超过 GPT-4 和 Llma3-70B，中文性能追平 GPT-4 Turbo。自 2023 年 4 月以来，通义千问从初代模型升级至 2.5 版本。相比此前的通义千问 2.1 版，通义千问 2.5 的理解能力、逻辑

推理、指令遵循、代码能力分别提升 9%、16%、19%、10%。

从 2023 年至今，通义已经推出全尺寸不同应用场景的八款大语言模型：1）可在手机、PC 等端侧设备部署的 20B 以下小尺寸模型；2）可供企业和科研场景使用的 72B、110B 大尺寸模型；3）注重性价比的 32B 中等尺寸模型。

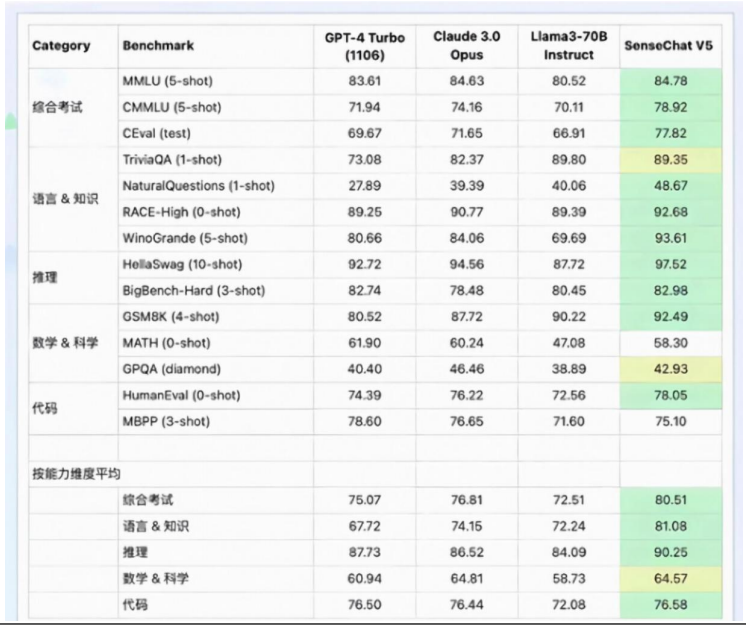
目前，通义千问有 9 万企业用户数量，700 万开源模型下载量，B 端客户有新浪微博、中国一汽、完美世界、蓝凌科技等。

此前，商汤科技于 4 月 23 日发布国内首个对标 GPT-4Turbo 的大模型——日日新 SenseNova 5.0，该版本具备更强的知识、数学、推理及代码能力，综合性能全面对标 GPT-4 Turbo，并在主流客观评测上达到或超越 GPT-4 Turbo。大模型指标方面，日日新 SenseNova 5.0 具有 6000 亿参数，基于超过 10TB tokens 训练。

我们认为商汤日日新 5.0 大模型率先开启了国内大模型行业对标 GPT-Turbo 的新局面，也是国内大模型厂商技术迭代加速的又一信号，未来国内大模型整体性能水平提升可期。

同时，开源大模型在推动产业进步方面有重要意义。首先，我国有大量中小企业，自研大模型带来的高额研发成本对中小企业来说难以消化，开源模式能够显著降低成本，企业和个人可免费使用和改进模型，减少研发投入，尤其利好资源有限的中小企业和初创公司。其次，开源模式有利于促进技术交流与共享，加快模型的更新迭代，提升整体研发效率。开源社区的活跃也能够推动 AI 人才培养，更多开发者参与其中，共同成长，形成良性循环。最后，开源大模型有助于推动产业 AI 化，随着大模型在各行业的深入应用，各类行业 AI 应用全面落地，产业智能化转型加速可期。

图 4：日日新 5.0 在大部分核心测试集指标上对标甚至超过 GPT-4 Turbo



Category	Benchmark	GPT-4 Turbo (1106)	Claude 3.0 Opus	Llama3-70B Instruct	SenseChat V5
综合考试	MMLU (5-shot)	83.61	84.63	80.52	84.78
	CMMLU (5-shot)	71.94	74.16	70.11	78.92
	CEval (test)	69.67	71.65	66.91	77.82
语言 & 知识	TriviaQA (1-shot)	73.08	82.37	89.80	89.35
	NaturalQuestions (1-shot)	27.89	39.39	40.06	48.67
	RACE-High (0-shot)	89.25	90.77	89.39	92.68
	WinoGrande (5-shot)	80.66	84.06	69.69	93.61
推理	HellaSwag (10-shot)	92.72	94.56	87.72	97.52
	BigBench-Hard (3-shot)	82.74	78.48	80.45	82.98
数学 & 科学	GSM8K (4-shot)	80.52	87.72	90.22	92.49
	MATH (0-shot)	61.90	60.24	47.08	58.30
	GPQA (diamond)	40.40	46.46	38.89	42.93
代码	HumanEval (0-shot)	74.39	76.22	72.56	78.05
	MBPP (3-shot)	78.60	76.65	71.60	75.10
按能力维度平均					
	综合考试	75.07	76.81	72.51	80.51
	语言 & 知识	67.72	74.15	72.24	81.08
	推理	87.73	86.52	84.09	90.25
	数学 & 科学	60.94	64.81	58.73	64.57
	代码	76.50	76.44	72.08	76.58

资料来源：商汤科技官方公众号，华龙证券研究所

4.1.2 字节跳动大模型 API 大幅降价，AI 应用有望迎来成本拐点

2024 年 5 月 15 日，字节跳动正式推出豆包大模型家族，共包含 9 个大模型。据字节跳动官微，豆包大模型每天处理 1200 亿 tokens 的文本（约 1800 亿汉字），生成 3000 万张图片。大量信息处理任务在提升模型输出效果的同时也能大幅降低模型推理的单位成本，豆包大模型家族中主力模型 API 定价相比行业价格低 99.3%。目前豆包大模型已通过火山引擎开放给企业客户。

本次豆包大模型的降价策略反映出，通过持续的技术优化和规模经济效应，大模型厂商能够有效地控制成本，并将这些优势转化为更具吸引力的价格。此外，这一策略也顺应了国内外大模型行业通过提效降价来争夺客源、保持用户日活量的大趋势。

从行业角度分析，随着国内外大模型厂商技术角逐进一步激烈，我们认为大模型行业开启价格战信号明显，更多大模型厂商有望加入价格比拼。通过降低价格门槛，大模型厂商有望吸引更广泛的企业用户群

体，帮助大模型厂商平衡收入和成本。同时，更多 C 端用户有望免费使用基础 AI 应用，庞大的访问量有助于企业进一步提升模型服务能力，完成良性循环。

从应用角度分析，大模型厂商开启价格战有望助力 AI 应用加速普及和创新。AI 应用行业成本拐点将更为明晰，AI 技术在各行各业的渗透和应用有望提速，整体产业的数字化转型和创新预期加快。

图 5：豆包大模型家族

图 6：豆包智能体

资料来源：字节跳动官方微信公众号，华龙证券研究所 资料来源：字节跳动官方微信公众号，华龙证券研究所

4.2 OpenAI 发布 GPT-4o，多模态能力再提升

2024 年 5 月 13 日，OpenAI 发布新一代大模型 GPT-4o，GPT-4o 能够接受文本、音频、图像和视频的任意组合作为输入，并生成文本、音频和图像输出的任意组合。它可以在最短 232 毫秒内响应音频输入，平均为 320 毫秒，与人类正常对话的响应时间接近。

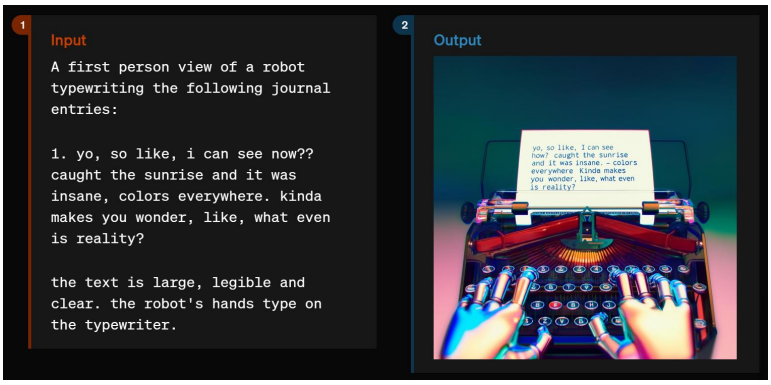
提效降价，多模态能力再进一步。GPT-4o 在英语文本和代码上的 GPT-4 Turbo 性能大致相当，在非英语语言的文本上也有显着改进，与现有模型相比，GPT-4o 在视觉和音频理解方面尤其出色。响应速度方面，在 GPT-4o 之前，使用语音模式与 ChatGPT 交谈的平均延迟为 2.8 秒（GPT-3.5）和 5.4 秒（GPT-4），而 GPT-4o 支持的 ChatGPT 可以实现平均响应时间 320 毫秒，且可随时打断。功能实现方面，此前语音模式是一个由三个独立模型组成的通路：一个简单的模型

将音频转录为文本，GPT-3.5 或 GPT-4 接收文本并输出文本，第三个简单模型将该文本转换回音频。这意味着智能的主要来源 GPT-4 会丢失大量信息，包括音调、多个扬声器或背景噪音，也无法输出笑声、歌声或表达情感。借助 GPT-4o，OpenAI 在文本、视觉和音频上端到端地训练了一个新模型，即所有输入和输出都由同一个神经网络处理。开发人员可以通过调用 API 来使用 GPT-4o 的文本和视觉能力。与 GPT-4 Turbo 相比，GPT-4o 的速度快 2 倍，价格减半，速率限制高出 5 倍。GPT-4o 的文本和图像功能也开始在 ChatGPT 中推出并在免费套餐中提供。

新音频及视频功能将迭代推出。OpenAI 将在未来几周内在 ChatGPT Plus 中推出带有 GPT-4o 的新版本语音模式。另外，音频和视频功能的 API 也将在几周内向向部分合作伙伴推出。

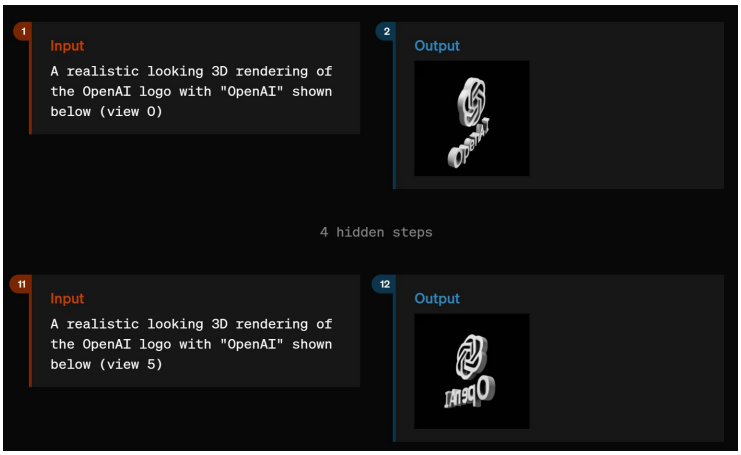
我们认为 GPT-4o 进一步提振了国内外 AI 行业的信心，对国内大模型产业具备积极的映射作用。未来世界范围内 AI 技术竞争有望加剧，建议关注国内科技企业的技术追赶和突破。

图 7：GPT-4o 文生图示例



资料来源：OpenAI 官网，华龙证券研究所

图 8：GPT-4o 3D 图像生成示例



资料来源：OpenAI 官网，华龙证券研究所

5 风险提示

（1）国产算力建设不及预期。算力是 AI 应用基石，国产算力建设不达预期将会延缓 AI 应用的落地速度。

（2）所引用数据资料的误差风险。本报告数据资料来源于公开数据，将可能对分析结果造成影响。

（3）AI 应用落地速度不及预期。当前市场上 AI 应用的定价、商业模式以及市场监管等方面仍处于探索阶段。

（4）国产大模型迭代速度不及预期。国内大模型厂商技术起步较晚，国产大模型受算力、算法等因素影响较大。

（5）重点关注公司业绩不达预期。重点关注公司业绩会受到各种因素影响，如果业绩不达预期，会使得公司股价受到影响。

（6）政策标准出台速度不及预期。当前 AI 相关技术发展速度较快，数据需求量大，往往伴随数据安全、数据所有权等问题，因此需要政策提供支持和引导。

免责及评级说明部分

分析师声明：

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉尽责的职业态度，独立、客观、公正地出具本报告。不受本公司相关业务部门、证券发行人士、上市公司、基金管理公司、资产管理公司等利益相关者的干涉和影响。本报告清晰准确地反映了本人的研究观点。本人不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。据此入市，风险自担。

投资评级说明：

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的 6-12 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：A 股市场以沪深 300 指数为基准。	股票评级	买入	股票价格变动相对沪深 300 指数涨幅在 10%以上
		增持	股票价格变动相对沪深 300 指数涨幅在 5%至 10%之间
		中性	股票价格变动相对沪深 300 指数涨跌幅在-5%至 5%之间
		减持	股票价格变动相对沪深 300 指数跌幅在-10%至-5%之间
		卖出	股票价格变动相对沪深 300 指数跌幅在-10%以上
	行业评级	推荐	基本面向好，行业指数领先沪深 300 指数
		中性	基本面稳定，行业指数跟随沪深 300 指数
		回避	基本面向淡，行业指数落后沪深 300 指数

免责声明：

本报告的风险等级评定为 R4，仅供符合华龙证券股份有限公司（以下简称“本公司”）投资者适当性管理要求的客户（C4 及以上风险等级）参考使用。本公司不会因为任何机构或个人接收到报告而视其为当然客户。

本报告信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证。编制及撰写本报告的所有分析师或研究人员在此保证，本研究报告中任何关于宏观经济、产业行业、上市公司投资价值等研究对象的观点均如实反映研究分析人员的个人观点。报告中的内容和意见仅供参考，并不构成对所述证券买卖的价格的建议或询价。本公司及分析研究人员对使用本报告及其内容所引发的任何直接或间接损失及其他影响概不负责。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提及的公司所发行的证券并进行证券交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务，投资者应充分考虑本公司及所属关联机构就报告内容可能存在的利益冲突。

华龙证券研究所

北京	兰州	上海	深圳
地址：北京市东城区安定门外大街 189 号天鸿宝景大厦西配楼 F4 层	地址：兰州市城关区东岗西路 638 号文化大厦 21 楼	地址：上海市浦东新区浦东大道 720 号 11 楼	地址：深圳市福田区民田路 178 号华融大厦辅楼 2 层
邮编：100033	邮编：730030	邮编：200000	邮编：518046
	电话：0931-4635761		