

# 计算机行业周报

## GPT 更新即将发布，神秘模型现身大模型竞技场

### 投资要点:

#### 本周 AI 新闻速递:

**国内 AI:** (1) 上海 AILab 开源首个可替代 GPT-4V 的多模态大模型。(2) 最新国产开源 MOE 性能表现优异，性价比高。(3) 国产一站式 AI 工作平台正式官宣。(4) 百度文生数字人能力发布，一句话 5 分钟数字人生成。(5) 阿里云发布通义千问 2.5，并开源 1100 亿参数模型。

**国外 AI:** (1) HuggingFace 放出规模最大、质量最高预训练数据集。(2) Google 发布医学大模型 Med-Gemini。(3) 摩根大通发布金融模型 IndexGPT。(4) 微软自有大模型 MAI-1 正在研发中。(5) 全球首个超小型多模态 AI Agent 模型升级版 Octopus V3 正式发布。(6) 美军 AI 战斗机首航，AI 空战不再遥远。(7) OpenAI 推出 DALL·E 3 识别器、媒体管理器。(8) Chatgpt 及 GPT-4 升级版即将在下周一发布。

#### 神秘模型现身大模型竞技场，综合性能直逼 GPT-4 Turbo

近日，神秘模型 GPT2-Chatbot 现身大模型竞技场。目前共有 Im-a-good-gpt2-chatbot 以及 Im-also-a-good-gpt2-chatbot 两个版本，试用这两个模型需要用户在 Arena (battle) 模式下随机抽选。基于团队的自主体验测试，我们得到结论：**新模型综合能力接近 GPT-4 Turbo 水平，在 ASCII (使用电脑字符表示图片) 上表现出色，但在某些测试下，存在简单问题复杂化的情况。**

#### AI 大模型月度数据回顾:

**海外主流 AI 大模型:** 总访问量上 Chatgpt 仍居首位。其余大模型中 Claude、Perplexity、Character.ai 4 月访问量都有一定上涨；Gemini 访问量 4 月出现下滑，环比跌幅为 1.4%。

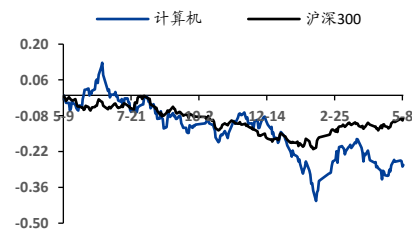
**国内主流 AI 大模型:** 国内 AI 总访问量与海外仍有较大差距。其中 Kimi 于 2024 年 4 月成功超越文心一言位居首位，总访问量达 19.6 百万次，环比提升 60.7%。受益于天工 3.0 发布，天工 AI 4 月访问量达 4 百万次，环比增长 122.2%。文心一言、通义千问、豆包也呈现上升趋势。讯飞星火、百川智能、360 智脑有所下滑。

#### 风险提示

大模型迭代不及预期的风险；大模型商业落地不及预期的风险等。

## 强于大市 (维持评级)

### 一年内行业相对大盘走势



### 团队成员

分析师: 钱劲宇(S0210524040006)  
QJY3773@hfzq.com.cn

### 相关报告



## 正文目录

|     |                                      |   |
|-----|--------------------------------------|---|
| 1   | 本周 AI 新闻速递.....                      | 3 |
| 1.1 | 国内 AI.....                           | 3 |
| 1.2 | 国外 AI.....                           | 3 |
| 2   | 神秘模型现身大模型竞技场，综合性能直逼 GPT-4 Turbo..... | 4 |
| 2.1 | 大模型竞技场——Chatbot Arena.....           | 4 |
| 2.2 | GPT2-Chatbot 现身竞技场.....              | 5 |
| 3   | AI 大模型月度数据回顾.....                    | 7 |
| 4   | 风险提示.....                            | 8 |

## 图表目录

|       |   |   |
|-------|---|---|
| 图表 1: | Chatbot Arena 1V1battle 界面.....               | 5 |
| 图表 2: | Arena 榜单 Top5.....                            | 5 |
| 图表 3: | GPT2-Chatbot 在 ASCII 上表现更佳.....               | 6 |
| 图表 4: | GPT2-Chatbot 能够很好的解决数学问题.....                 | 6 |
| 图表 5: | GPT2-Chatbot 逻辑推理上表现出色.....                   | 6 |
| 图表 6: | GPT2-Chatbot 和 GPT-4 Turbo 都存在简单问题复杂化的情况..... | 7 |
| 图表 7: | 海外主流 AI 大模型月度访问量（单位：百万次）.....                 | 7 |
| 图表 8: | 国内主流 AI 大模型月度访问量（单位：百万次）.....                 | 8 |

## 1 本周 AI 新闻速递

### 1.1 国内 AI

(1) **上海 AI Lab 开源首个可替代 GPT-4V 的多模态大模型。** InternVL 1.5 是上海 AILab 推出的开源多模态大语言模型 (MLLM)。根据相关论文, InternVL 1.5 在四个特定基准测试中超越了 Grok-1.5V、GPT-4V、Claude-3 Opus 和 Gemini Pro 1.5 等领先的闭源模型。

(2) **最新国产开源 MOE 性能表现优异, 性价比高。** 最新国产开源 MOE 大模型 DeepSeek-V2 综合性能达 GPT-4 级别, 其中中文综合能力超越一众开源模型, 和 GPT-4-Turbo、文心 4.0 等闭源模型同处第一梯队; 英文综合能力和 LLaMA3-70B 同处第一梯队, 超过同是 MoE 的 Mixtral 8x22B。同时 DeepSeek-V2 为开源模型且可免费商用, 输入/输出价格分别为 \$0.14/\$0.28 每百万 tokens, API 价格仅为 GPT-4-Turbo 的百分之一。

(3) **国产一站式 AI 工作平台正式官宣。** 零一万物正式官宣一站式 AI 工作平台万知。万知平台根据中国人的使用习惯, 量身打造一系列生产力工具, 能够实现会议纪要、周报、写作助手, 文档速度, PPT 制作等功能, 帮助打工人实现工作效率的快速提升。

(4) **百度文生数字人能力发布, 一句话 5 分钟数字人生成。** 百度智能云曦灵平台基于文心一言 4.0 推出数字人定制能力。相较于过去高精度 3D 数字人按天及月算的定制周期, 此次数字人定制周期缩短至分钟级, 且在效率大幅提升的前提下, 细节质量仍保持高水准。

(5) **阿里云发布通义千问 2.5, 并开源 1100 亿参数模型。** 相较通义千问 2.1, 2.5 版本的理解能力、逻辑推理、指令遵循、代码能力分别提升 9%、16%、19%、10%, 中文能力更是持续领先业界, 综合能力全面赶超 GPT-4。

### 1.2 国外 AI

(1) **HuggingFace 放出规模最大、质量最高预训练数据集。** HuggingFace 机器学习团队宣布开源迄今为止规模最大、质量最高的预训练数据集 FineWeb。包含 15T+ 个 tokens 的 Web 数据集, 主要用作英语领域的公共数据研究。

(2) **Google 发布医学大模型 Med-Gemini。** Med-Gemini 融合了高级推理、多模态理解和长文本处理能力。在 14 个医疗基准测试中, Med-Gemini 在 10 个基准上取得最佳表现, 远超 GPT-4。

(3) **摩根大通发布金融模型 IndexGPT。** 其基座模型为 GPT-4, 能通过特定主题关键字创建投资篮子主题索引, 再通过大模型进行深度分析扫描新闻等内容生成

投资指数，以展示该公司的整体投资价值。IndexGPT 将在“主题投资篮子”里获得使用。

(4) **微软自有大模型 MAI-1 正在研发中。**根据 Information 报道及微软首席技术官 Kevin Scott 确认，微软大模型 MAI-1 正在开发过程中，拥有 5000 亿参数。为实现新模型训练，微软已预留大批英伟达 GPU 服务器，并一直编制训练数据以优化模型。

(5) **全球首个超小型多模态 AI Agent 模型升级版 Octopus V3 正式发布。**相较于 V2 版本，V3 具有图像处理和多语言文本处理能力，为端侧应用打下良好基础。此外，在函数调用性能上，V3 远超同类模型。相较于传统大语言模型，V3 体积更小，能耗更低，能够轻松在各种小型终端运行，未来在手机、AR/VR、机器人、汽车等端侧得到广泛应用。

(6) **美军 AI 战斗机首航，AI 空战不再遥远。**近期，美国空军部长 Kendall 亲自试驾国防部正在研制的 X-62AAI 飞机。1 小时的飞行过程中，Kendall 全程未干预，所有动作都由 AI 自主完成。X-62A 能够实现全自动空战，通过 AI 软件控制完成复杂动作，代码量达 10 万行，已试飞 21 次。

(7) **OpenAI 推出 DALL·E 3 识别器、媒体管理器。**该媒体管理器主要受众是媒体行业和内容创作者，帮助控制自己的知识产权内容是否会被 OpenAI 搜集用于训练 AI 模型。目前，OpenAI 已经开放了 DALL·E 3 识别器的 API，开发人员从今天开始至 7 月 31 日可申请试用。OpenAI 会在 8 月 31 日之前公布获取使用权限的名单。

(8) **Chatgpt 及 GPT-4 升级版即将在下周一发布。**根据 OpenAI 官推，OpenAI 将于太平洋时间 10 点开启直播，发布 Chatgpt 及 GPT-4 更新的相关 demo。根据 Sam Altman 推文，此次发布并不是之前提及的 GPT-5 或搜索引擎。

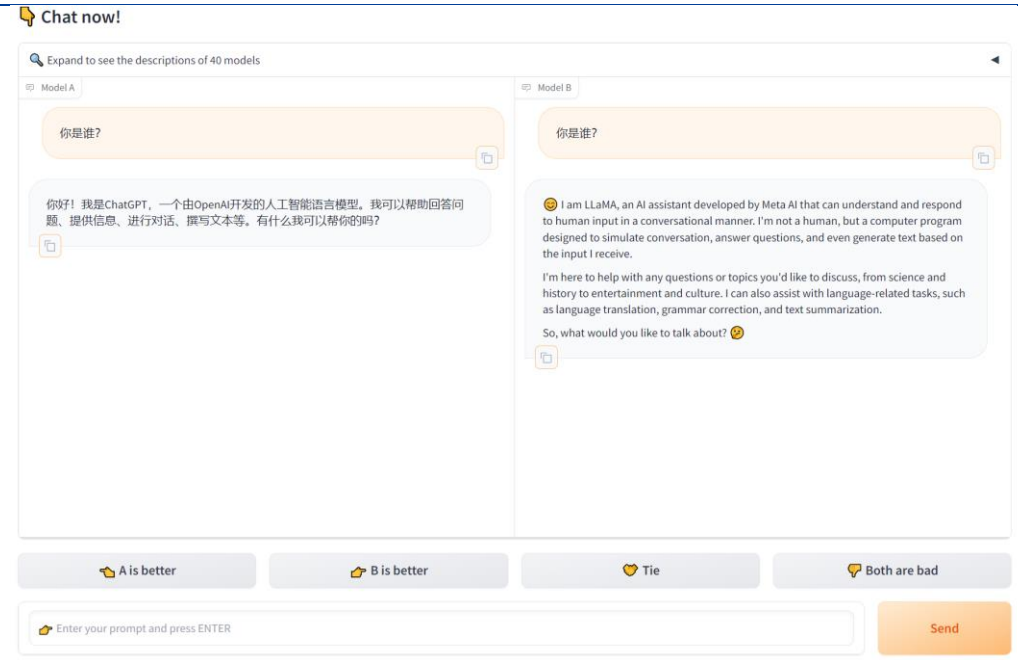
## 2 神秘模型现身大模型竞技场，综合性能直逼 GPT-4 Turbo

### 2.1 大模型竞技场——Chatbot Arena

Chatbot Arena 是一个由 LMSYS 和加州大学伯克利分校 Skylab 共同开发的开源项目，是目前大模型权威榜单之一。该榜单基于匿名 1V1 battle 的投票规则，基于 ELO 评级对于模型排名。



图表1: Chatbot Arena 1V1battle 界面



来源: Chatbot Arena, 华福证券研究所

截至 2024 年 5 月 8 日, 网站共有 95 个模型参与排名位居前五名的模型分别为 GPT-4-Turbo-2024-04-09、GPT-4-1106-preview、Gemini 1.5 Pro API-0409-Preview、Claude 3 Opus、GPT-4-0125-preview, 得分分别为 1258、1252、1249、1248、1246。

图表2: Arena 榜单 Top5

| Rank* (UB) ▲ | Model   | ★ Arena Elo ▲ | 📊 95% CI ▲ | 🗳️ Votes ▲ | Organization ▲ | License     | ▲ Knowledge Cutoff |
|--------------|---|---------------|------------|------------|----------------|-------------|--------------------|
| 1            | <a href="#">GPT-4-Turbo-2024-04-09</a>          | 1258          | +3/-3      | 44592      | OpenAI         | Proprietary | 2023/12            |
| 2            | <a href="#">GPT-4-1106-preview</a>              | 1252          | +2/-3      | 76173      | OpenAI         | Proprietary | 2023/4             |
| 2            | <a href="#">Gemini 1.5 Pro API-0409-Preview</a> | 1249          | +3/-3      | 61011      | Google         | Proprietary | 2023/11            |
| 2            | <a href="#">Claude 3 Opus</a>                   | 1248          | +2/-2      | 101063     | Anthropic      | Proprietary | 2023/8             |
| 3            | <a href="#">GPT-4-0125-preview</a>              | 1246          | +3/-2      | 70239      | OpenAI         | Proprietary | 2023/12            |

来源: Chatbot Arena, 华福证券研究所

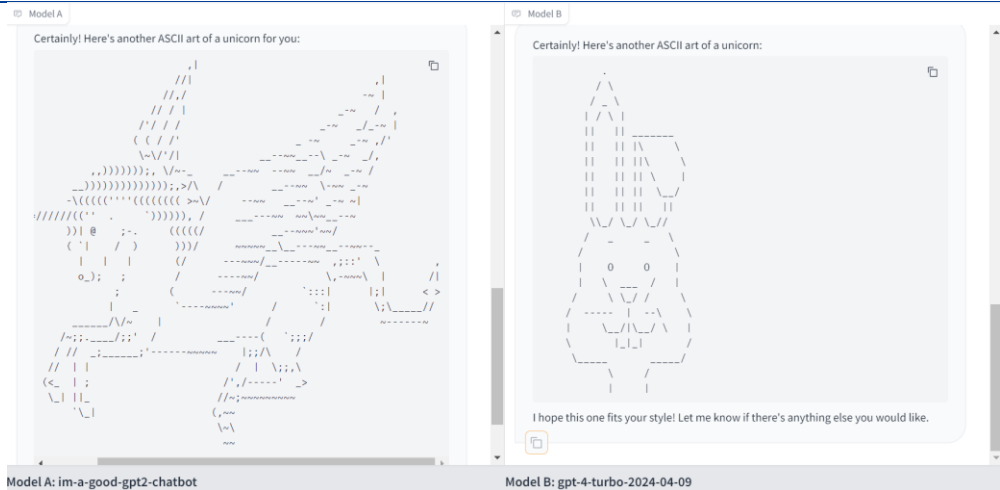
## 2.2 GPT2-Chatbot 现身竞技场

近日, 神秘模型 GPT2-Chatbot 现身大模型竞技场。目前共有 Im-a-good-gpt2-chatbot 以及 Im-also-a-good-gpt2-chatbot 两个版本, 试用这两个模型需要用户在 Arena (battle) 模式下随机抽选。基于团队的自主体验测试, 我们得到结论: **新模型综合能力接近 GPT-4 Turbo 水平, 在 ASCII (使用电脑字符表示图片) 上表现出色, 但在某些测试下, 存在简单问题复杂化的情况。**

**基于 prompt: create a ascii art unicorn**, GPT2-Chatbot 和 GPT-4 Turbo 均能够生成 ASCII 图片, 但 GPT2-Chatbot 能够使用多种电脑字符, 生成的图片更加复杂, 处理效果更好。



图表3: GPT2-Chatbot 在 ASCII 上表现更佳



来源: Chatbot Arena, 华福证券研究所

基于 prompt: 已知哥哥 4 岁时, 弟弟比哥哥小 2 岁。求哥哥 100 岁时, 弟弟多少岁?, GPT2-Chatbot 和 GPT-4 Turbo 均能够找到哥哥与弟弟的固定年龄差, 并根据年龄差计算哥哥 100 岁时, 弟弟的年纪。

图表4: GPT2-Chatbot 能够很好的解决数学问题



来源: Chatbot Arena, 华福证券研究所

基于 prompt: 我今天有 3 个苹果, 昨天吃了一个苹果。现在有几个苹果? GPT2-Chatbot 和 GPT-4 Turbo 均能给出正确答案, 并能给出昨天吃掉的苹果并不影响今天拥有的苹果数量的理由。

图表5: GPT2-Chatbot 逻辑推理上表现出色



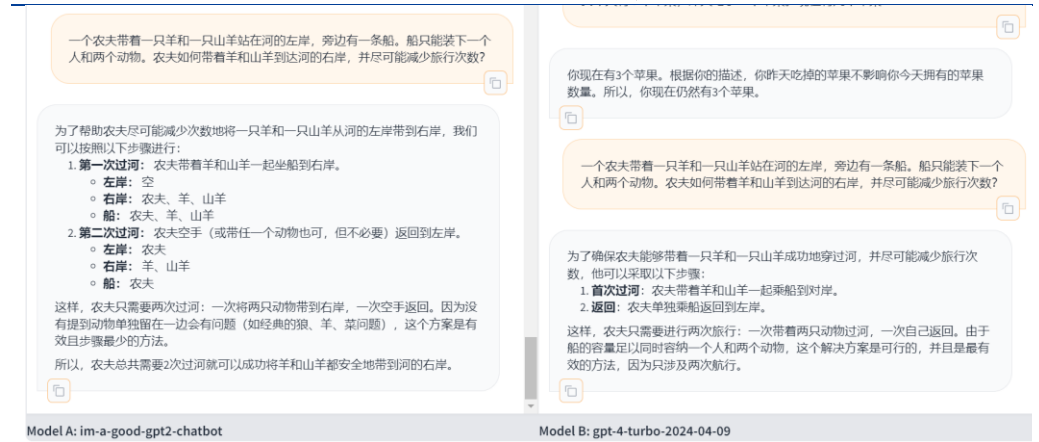
来源: Chatbot Arena, 华福证券研究所

基于 prompt: 一个农夫带着一只羊和一只山羊站在河的左岸, 旁边有一条船。船只能装下一个人和两个动物。农夫如何带着羊和山羊到达河的右岸, 并尽可能减少旅行次数? GPT2-Chatbot 和 GPT-4 Turbo 均假设了农夫把羊和山羊带到右岸后自



己需要乘船返回，但基于人类的理解来看，并不需要让农夫返回到左岸。所以两者均给出了错误答案。

图表6: GPT2-Chatbot 和 GPT-4 Turbo 都存在简单问题复杂化的情况

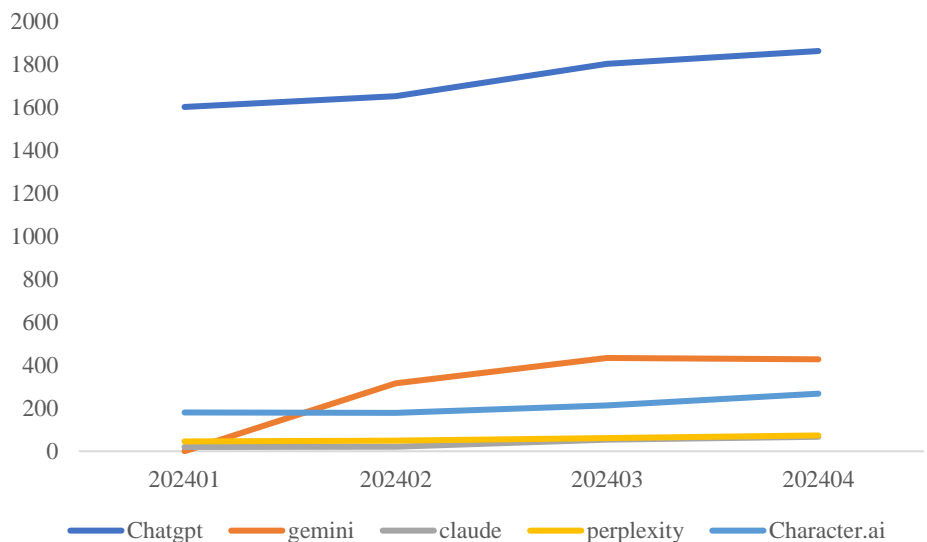


来源: Chatbot Arena, 华福证券研究所

### 3 AI 大模型月度数据回顾

**海外主流 AI 大模型:** 总访问量来看, Chatgpt 仍遥遥领先, 2024 年 4 月访问量达 1860 百万次, 环比上升 3.3%。Claude、Perplexity、Character.ai 都有一定的上涨, 4 月访问量环比涨幅分别为 23.7%、19.5%、25.7%。Gemini 出现下滑, 环比跌幅为 1.4%。

图表7: 海外主流 AI 大模型月度访问量 (单位: 百万次)



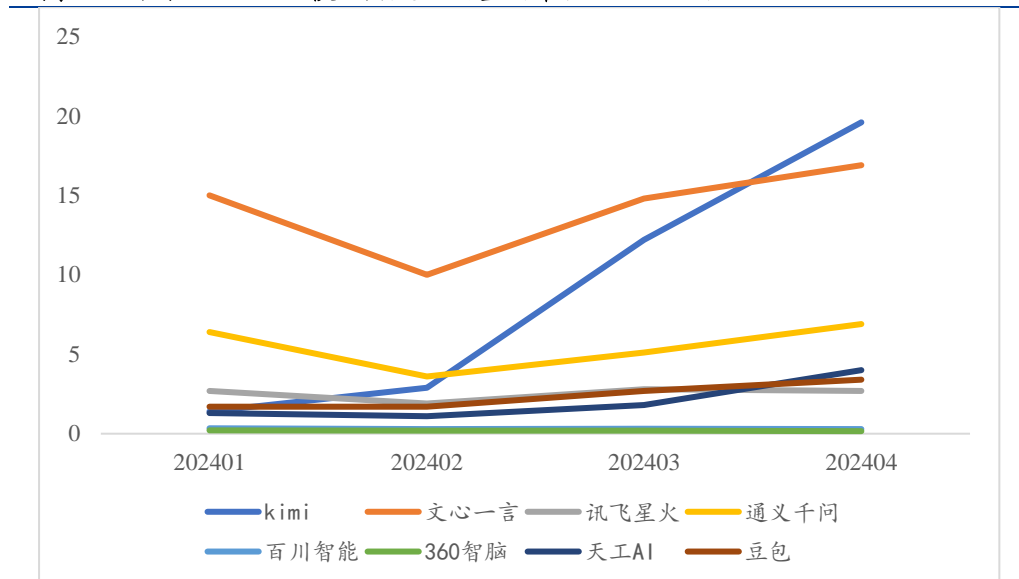
来源: Similarweb, 华福证券研究所

**国内主流 AI 大模型:** 总访问量来看, 国内大模型与海外仍有较大差距。Kimi 于 2024 年 4 月成功超越文心一言位居首位, 总访问量达 19.6 百万次, 环比提升 60.7%。受益于天工 3.0 发布, 天工 AI 4 月访问量达 4 百万次, 环比增长 122.2%。文心一言、通义千问、豆包也呈现上升趋势。讯飞星火、百川智能、360 智脑有所下滑。其中,



由于 360AI 搜索的更新，360 智脑访问量一定程度上被分流。而 360AI 搜索 4 月访问量达 1.2 百万次，环比增幅约 400%。

图表8: 国内主流 AI 大模型月度访问量 (单位: 百万次)



来源: Similarweb, 华福证券研究所

#### 4 风险提示

**大模型迭代不及预期的风险:** 目前大多数大模型的研发仍处于早期投入阶段, 存在研发失败以及迭代进度缓慢的风险。

**大模型商业落地不及预期的风险:** 当前大模型与各个场景结合处于早期磨合阶段, 仍需要进一步的训练、微调及试用, 存在落地时间点不及预期的风险。





## 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，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将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。

## 一般声明

华福证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。

本报告版权归“华福证券有限责任公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

## 特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

## 投资评级声明

| 类别   | 评级   | 评级说明                               |
|------|------|------------------------------------|
| 公司评级 | 买入   | 未来 6 个月内，个股相对市场基准指数涨幅在 20%以上       |
|      | 持有   | 未来 6 个月内，个股相对市场基准指数涨幅介于 10%与 20%之间 |
|      | 中性   | 未来 6 个月内，个股相对市场基准指数涨幅介于-10%与 10%之间 |
|      | 回避   | 未来 6 个月内，个股相对市场基准指数涨幅介于-20%与-10%之间 |
|      | 卖出   | 未来 6 个月内，个股相对市场基准指数涨幅在-20%以下       |
| 行业评级 | 强于大市 | 未来 6 个月内，行业整体回报高于市场基准指数 5%以上       |
|      | 跟随大市 | 未来 6 个月内，行业整体回报介于市场基准指数-5%与 5%之间   |
|      | 弱于大市 | 未来 6 个月内，行业整体回报低于市场基准指数-5%以下       |

备注：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）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）

## 联系方式

华福证券研究所 上海

公司地址：上海市浦东新区浦明路 1436 号陆家嘴滨江中心 MT 座 20 层

邮编：200120

邮箱：hfjys@hfzq.com.cn