

传媒

行业快报

头部模型官宣降价，应用端或加速落地

投资要点

- ◆ **热点事件:**5月21日上午,阿里云宣布,通义千问 GPT-4 级主力模型 Qwen-Long 的 API 输入价格降至 0.0005 元/千 Tokens, 直降 97%。随后, 百度通过官方微信宣布, 旗下文心大模型两款主力模型 ENIRE Speed 和 ENIRE Lite 全面免费。头部大模型厂商宣布大幅降价, 有望扩充市场, 降低 B 端和 C 端用户使用成本, 加速 AI 应用端爆发。
- ◆ **头部大模型官宣降价, 或开启应用端“大航海时代”。**5月15日, 字节跳动旗下云服务平台火山引擎官宣豆包大模型。豆包通用模型 pro-32k 的推理输入价格为 0.0008 元/千 Tokens, 彼时价格比行业便宜 99.3%。5月21日, 阿里云和百度先后宣布旗下大模型大幅降价。通义千问 GPT-4 级主力模型 Qwen-Long, 最高支持 1000 万 tokens (约 1500 万字或 1.5 万页文档) 长文本输入。其 API (应用程序编程接口) 输入价格从 0.02 元/千 tokens (文本中最小的语义单元) 降至 0.0005 元/千 tokens, 直降 97%, 降价后约为 GPT-4 价格的 1/400。除了输入价格降至 0.0005 元/千 tokens, Qwen-Long 输出价格也直降 90% 至 0.002 元/千 tokens。不久前发布的通义千问旗舰款大模型 Qwen-Max, API 输入价格降至 0.04 元/千 tokens, 降幅达 67%。阿里云方面预计, 公共云+API 将成为企业使用大模型的主流方式, 未来大模型 API 的调用量会有成千上万倍的增长。头部大模型公司先后宣布重要产品降价, 有望推动应用端爆发, 内容端成本进一步下降, B 端智慧产业和 C 端创作或迎不同程度赋能。
- ◆ **轻量模型门槛降低, 端侧 AI 或迎深度赋能。**大模型如何有效服务好已有的场景, 如移动设备、学习机、扫地、语音交互等, 相关问题备受关注。在未来的 2024 年乃至 2025 年, 能够在适配、性能与价格上率先拔得头筹的算法团队, 将更有望赢得端侧 AI 领域的最终话语权。国内大模型团队在端侧上发力迅猛, 以多模态能力为例, 海外头部公司 OpenAI、Google, 国内阿里、面壁智能等均在 20B 以内参数规模的小模型有布局。面壁智能最新发布的多模态模型 MiniCPM-Llama3-V 2.5 性能飞跃: 参数规模仅 8B, 综合性能超越 Google Gemini Pro 与 OpenAI 的 GPT-4V; OCR 能力质的飞跃; 对手机端多模态大模型进行系系统级加速, 速度提升 150 倍; 多语种能力提升, 支持语种 30+。面壁在短短 3 个月内, 旗下端侧模型不断迭代, 成绩亮眼。端侧 AI 迭代加速, 叠加头部厂商降价, 利好端侧应用端加速落地。
- ◆ **投资建议:**头部大模型厂商降价, 端侧 AI 快速迭代, 有望加速形成规模经济, AI 应用端有望迎来爆发。建议关注: 腾讯控股、美图公司、汤姆猫、昆仑万维、恺英网络、巨人网络、世纪天鸿、蓝色光标、南方传媒、天娱数科、引力传媒、捷成股份、视觉中国、华策影视、值得买、中广天择、万兴科技等。
- ◆ **风险提示:**政策不确定性、AI 技术迭代不及预期、AI 商用进度不及预期等。

 投资评级 **领先大市-A维持**
首选股票 **评级**

一年行业表现



资料来源: 聚源

升幅%	1M	3M	12M
相对收益	-1.73	-6.42	-16.5
绝对收益	2.3	0.16	-23.09

分析师

倪爽

 SAC 执业证书编号: S0910523020003
 nishuang@huajinsec.com

相关报告

- 传媒: 11 部影片定档端午, 关注暑期档爆款安排-华金证券-传媒-行业快报 2024.5.19
- 恺英网络: 情怀+创新筑 IP 护城河, 重点游戏赋能稳健增长-华金证券-传媒-恺英网络-公司点评 2024.5.16
- 姚记科技: 创新营销赋能, 游戏+卡牌主业稳健增长-华金证券-传媒-姚记科技-公司点评 2024.5.12
- 传媒: 通义千问 2.5 发布, 中文大模型纵深突破-华金证券-传媒-行业快报 2024.5.10
- 传媒: 多片合力提振观影热情, 五一档延续影视复苏-华金证券-传媒-行业快报 2024.5.6
- 传媒: Vidu 发布, 国内视频生成持续突破-华金证券-传媒-行业快报 2024.5.6
- 南方传媒: 数智赋能教育出版, “AI+教育”催生新业态-华金证券-传媒-南方传媒-公司点评 2024.4.30



行业评级体系

收益评级：

领先大市—未来 6 个月的投资收益率领先沪深 300 指数 10%以上；

同步大市—未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-10%至 10%；

落后大市—未来 6 个月的投资收益率落后沪深 300 指数 10%以上；

风险评级：

A —正常风险，未来 6 个月投资收益率的波动小于等于沪深 300 指数波动；

B —较高风险，未来 6 个月投资收益率的波动大于沪深 300 指数波动；

分析师声明

倪爽声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

本公司具备证券投资咨询业务资格的说明

华金证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。

免责声明：

本报告仅供华金证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议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，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息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。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准。

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到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦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。

本报告版权仅为本公司所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发、篡改或引用本报告的任何部分。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华金证券股份有限公司研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。

华金证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

风险提示：

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出价或询价。投资者对其投资行为负完全责任，我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。

华金证券股份有限公司

办公地址：

上海市浦东新区杨高南路 759 号陆家嘴世纪金融广场 30 层

北京市朝阳区建国路 108 号横琴人寿大厦 17 层

深圳市福田区益田路 6001 号太平金融大厦 10 楼 05 单元

电话：021-20655588

网址：www.huajinsec.cn