

**分析师：吕伟**

执业证号：S0100521110003

邮箱：lvwei_yj@mszq.com

分析师：郭新宇

执业证号：S0100518120001

邮箱：guoxinyu@mszq.com

➤ **事件概述：**2月16日，据中国证券报报道，记者从腾讯集团确认，微信搜一搜在调用混元大模型丰富 AI 搜索的同时，正式灰度测试接入 DeepSeek。

➤ **大模型平权带来大量模型应用需求，有望带动算力需求长期提升。**腾讯方面表示，部分测试用户，可在微信对话框顶部搜索入口，看到“AI 搜索”字样，点击进入后，可免费使用 DeepSeek-R1 满血版模型，获得更多元化的搜索体验。若未显示该入口，说明此次灰度测试暂未覆盖到该用户账号，可耐心等待后续开放。

根据腾讯控股有限公司发布的 2024 年三季报数据显示，在微信及 WECHAT 方面，2024 年第三季度，微信及 WECHAT 的月活跃账户数达到了 13.82 亿。在庞大的用户基础下，大模型的使用频率有望逐步提升，进而带动算力需求长期提升。

根据经济观察报 2 月 15 日报道，近期无论是高性能的 H 系列 GPU，还是 RTX 40 系列的高端显卡，英伟达的 GPU 重新变得供不应求。

➤ **大模型接入、使用量的增加有望带来算力行业景气度进一步提升，Deepseek 或是算力星辰大海的一朵“小浪花”。1) DeepSeek 真实的训练、推理成本不应低估。**据 DeepSeek V3 论文，556 万美元的成本仅包括 DeepSeek-V3 的正式训练，不包括与架构、算法、数据相关的前期研究、消融实验的成本。而基于充足前期准备进行正式训练的成本往往都较低。以同样在 2025 年 1 月发布的加州大学伯克利 Sky-T1-32B-Preview 为例，其正式训练成本仅为 450 美元，但在数学能力等方面跑分超过 OpenAI o1-Preview。

2) DeepSeek 针对算力生态的底层进行优化，有望进一步激发国产算力潜能。在 DeepSeek-V3 的技术博文中，DeepSeek 表示其使用了英伟达的 PTX (Parallel Thread Execution) 语言，可以更精细地控制 GPU 之间传输数据、权重和梯度等，进而实现了算力使用效率的提升。对于国产算力，国内头部大模型（以 DeepSeek 为代表）也有望有较强的能力来进行生态的适配和优化，进而推动国产算力生态持续完善。目前已有多家国产 AI 芯片陆续宣布适配 DeepSeek 大模型。

3) 成本的优化也是促进 AI 推理普及的重要动力。DeepSeek-R1 API 服务定价为每百万输入 tokens 1 元(缓存命中)/4 元(缓存未命中)，每百万输出 tokens 16 元，较 o1 类推理模型的输入输出价格有了大幅度优化。

因此，降低 AI 行业进入门槛与成本，长期看将推动总需求上升而非下降。DeepSeek 所有模型均为开源模型，即所有应用厂商都拥有了可以比肩顶级 AI 的大模型，而且还可自行二次开发、灵活部署，这将加速 AI 应用的发展进程。当模型的成本越低，开源模型发展越好，模型的部署、使用的频率就会越高，使用量就会越大。经济学上著名的“杰文斯悖论”提出：当技术进步提高了资源使用的效率，不仅没有减少这种资源的消耗，反而因为使用成本降低，刺激了更大的需求，最终导致资源使用总量上升。因此从更长的周期来看，DeepSeek 的发展恰恰会加速 AI 的普及和创新，带来算力需求、特别是推理算力需求更大量级

推荐**维持评级****相关研究**

1. 计算机行业动态研究：AI 搜索深度：微信接入 DeepSeek-2025/02/17
2. 计算机周报 20250216：还有哪些云待重估？-2025/02/16
3. DeepSeek 系列报告：AI 编程或为 B 端最先崛起的 AI 应用-2025/02/13
4. 计算机行业动态报告：DeepSeek 系列报告之 AI+教育-2025/02/12
5. 计算机行业深度报告：为什么 DeepSeek 最先激活方向是 SaaS-2025/02/11

提升。

➤ **国民级应用模型使用持续放大算力需求。**根据前文提到，2024 年第三季度，微信及 WECHAT 的月活跃账户数达到了 13.82 亿，同比和环比均有所增长。在如此庞大的客户使用需求下，其背后对算力的需求将迎来新的增长空间。

搜一搜：根据在 2023 微信公开课 PRO 上微信搜一搜团队披露，2022 年微信搜一搜月活跃用户就已达 8 亿。据新榜日常监测的百万级公众号样本库，2024 年公众号累计产出超 4.44 亿篇文章，有 30.78 万篇阅读量达到 10 万+，相比 2023 年高于 5.33 万篇。

视频号：根据视频号互选平台公开显示的账号数据（不包括未入驻账号），截至 2025 年 1 月 10 日，视频号至少有 1288 个账号粉丝量超过 100 万，相比 2023 年增长 2.6 倍；有 44 个账号粉丝量超过 500 万，相比 2023 年增长近 3 倍。

DeepSeek 接入知名头部应用，带来的推理算力需求测算：我们仅根据微信目前的日活量，以及日均 token 调用量为基础，做出保守、中性、乐观 3 种假设，结合大模型推理算力需求计算公式，对 DeepSeek 接入带来的推理算力需求进行测算。

在 3 种假设下，预计微信搜索接入 DeepSeek 且全面普及使用后，所带来的推理需求提升，就将带来 294、529、882 亿元的 AI 服务器资本开支需求（具体测算表格在下一页）。

同时应考虑到，国内有多个主流 APP 也具有较高的使用和搜索频率，根据 QuestMobile 的数据，2024Q3 微信的平均 MAU 超过 10 亿，但国内有多个 APP 的 MAU 接近于这个水平，如淘宝（9 亿）、抖音（8 亿）、支付宝（9 亿）、美团（5 亿）、微博（5 亿）等，因此我们可以合理推断，未来国内 APP 接入 DeepSeek 等大模型且全面普及使用后，推理算力的需求有望持续快速提升。

➤ **投资建议：**DeepSeek 等头部模型的发展让 AI 应用、AI Agent 等加速落地，使用量快速提升推动 AI 算力需求进入从训练推动到推理推动的拐点，Deepseek 或是算力星辰大海的一朵“小浪花”。从微信等头部 APP 接入 DeepSeek 后可以看到，国民级应用模型使用持续放大算力需求，进而带来推理等算力需求的大量增长，算力行业迎来重要发展机遇。

建议重点关注国产 AI 算力产业链：1) 芯片设计：重点关注寒武纪、海光信息等 AI 芯片龙头。2) 国产 AI 算力液冷：高澜股份、英维克等。3) AI 服务器领军企业浪潮信息、中兴通讯、紫光股份、中科曙光、工业富联等。4) 昇腾一体机与服务器产业链：科大讯飞、软通动力、高新发展、中国长城、神州数码、拓维信息、烽火通信等。

➤ **风险提示：**政策落地不及预期；AI 算力技术路线变化存在不确定性；行业竞争加剧。

表1: 微信搜索接入 DeepSeek 且全面普及使用后, 带来的推理算力需求测算

| 推理算力需求测算 | 保守 | 中性 | 乐观 |
|--|---|-------------|-------------|
| DAU (万人) | 10000 | 12000 | 15000 |
| DeepSeek V3 大模型参数量 (亿) | 6710 | 6710 | 6710 |
| 每个日活每天使用 AI 搜索的次数 (次) | 10 | 15 | 20 |
| 每次 AI 搜索调用的 token 量 | 500 | 500 | 500 |
| 日均 token 调用量 (亿) | 5000 | 9000 | 15000 |
| 推理计算时间 (s) | 2 | 2 | 2 |
| 日均每秒 token 计算数量 (亿) | 0.04 | 0.07 | 0.12 |
| 峰值倍数 | 4 | 4 | 4 |
| 算力需求公式 | 云端 AI 推理算力需求 $\approx 2 \times$ 模型参数量 \times 数据规模 \times 峰值倍数 | | |
| 算力需求结果 (FLOPS) | 2.07099E+19 | 3.72778E+19 | 6.21296E+19 |
| H20 单卡算力 (TFLOPS) | 148 | 148 | 148 |
| MFU | 50% | 50% | 50% |
| 需要 H20 卡数量 (万张) | 28 | 50 | 84 |
| H20 单价 (万元) | 8.4 | 8.4 | 8.4 |
| H20 服务器均价 (万元) | 84 | 84 | 84 |
| 微信搜索接入 DeepSeek 且全面普及使用后, 长期内对 AI 服务器需求 (亿元) | 293.86 | 528.94 | 881.57 |

资料来源: 36 氪、新浪财经、CDSN、腾讯网、IT 时报, 民生证券研究院测算

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并登记为注册分析师，基于认真审慎的工作态度、专业严谨的研究方法与分析逻辑得出研究结论，独立、客观地出具本报告，并对本报告的内容和观点负责。本报告清晰地反映了研究人员的研究观点，结论不受任何第三方的授意、影响，研究人员不曾因、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。

评级说明

| 投资建议评级标准 | 评级 | 说明 |
|---|------|---------------------|
| 以报告发布日后的 12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。 | 推荐 | 相对基准指数涨幅 15%以上 |
| | 谨慎推荐 | 相对基准指数涨幅 5% ~ 15%之间 |
| | 中性 | 相对基准指数涨幅-5% ~ 5%之间 |
| | 回避 | 相对基准指数跌幅 5%以上 |
| 行业评级 | 推荐 | 相对基准指数涨幅 5%以上 |
| | 中性 | 相对基准指数涨幅-5% ~ 5%之间 |
| | 回避 | 相对基准指数跌幅 5%以上 |

免责声明

民生证券股份有限公司（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。

本报告仅供本公司境内客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅为参考之用，并不构成对客户的投资建议，不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告所包含的观点及建议并未考虑获取本报告的机构及个人的具体投资目的、财务状况、特殊状况、目标或需要，客户应当充分考虑自身特定状况，进行独立评估，并应同时考量自身的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见，不应单纯依靠本报告所载的内容而取代自身的独立判断。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而导致的任何可能的损失负任何责任。

本报告是基于已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及预测仅反映本公司于发布本报告当日的判断，且预测方法及结果存在一定程度局限性。在不同时期，本公司可发出与本报告所刊载的意见、预测不一致的报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，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、咨询服务等相关服务，本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

若本公司以外的金融机构发送本报告，则由该金融机构独自为此发送行为负责。该机构的客户应联系该机构以交易本报告提及的证券或要求获悉更详细的信息。本报告不构成本公司向发送本报告金融机构之客户提供的投资建议。本公司不会因任何机构或个人从其他机构获得本报告而将其视为本公司客户。

本报告的版权仅归本公司所有，未经书面许可，任何机构或个人不得以任何形式、任何目的进行翻版、转载、发表、篡改或引用。所有在本报告中使用的商标、服务标识及标记，除非另有说明，均为本公司的商标、服务标识及标记。本公司版权所有并保留一切权利。

民生证券研究院：

上海：上海市浦东新区浦明路 8 号财富金融广场 1 幢 5F； 200120

北京：北京市东城区建国门内大街 28 号民生金融中心 A 座 18 层； 100005

深圳：深圳市福田区中心四路 1 号嘉里建设广场 1 座 10 层 01 室； 518048