

策略点评

AI 端侧迎来模型催化新逻辑

Gemma 4 的推出与特斯拉 AI5 的流片为端侧 AI 产业提供了新的增量逻辑。

- 谷歌发布 Gemma 4 开源大模型，专为高级推理与智能体工作流打造。本次谷歌推出四种规格的 Gemma 4 通用模型：E2B、E4B、MoE 与 31B，并宣布将 Gemma 4 全系迁移至商业友好的 Apache 2.0 协议，Gemma 4 的发布，不仅仅是一次常规的模型更新，也是端侧 AI 产业的重要拐点事件，Gemma 4 的小模型在手机上实现了离线运行、接近零延迟响应且功耗较低，同时具备处理复杂推理任务的能力。这意味着手机、耳机、智能眼镜、摄像头等终端设备拥有了不依赖云端、可随时调用的高级智能。Apache 2.0 协议之下，硬件厂商无需自研模型，这或将催生大量端侧 AI 原生应用的涌现，可能也会为整个端侧 AI 产业链（芯片、模组、终端）等打开规模化放量的确定性空间。特斯拉 AI5 芯片流片完成，部分场景性能较前代提升 40 倍，预计 2027 年启动量产。AI5 流片的成功是端侧算力的重要进展，表明端侧算力正在成为海外科技巨头们的必争之地，AI 端侧正在从单点突破走向生态竞争。
- Gemma 4 的推出与特斯拉 AI5 的流片，分别从软件和硬件两个层面为端侧 AI 产业提供了新的增量逻辑，尽管目前端侧 AI 仍处于商业化早期阶段，但头部企业在模型轻量化与端侧算力上的持续投入，正在逐步降低终端部署的门槛，并为芯片、模组、终端及应用等环节提供更明确的技术路径与合作基础，AI 端侧产业链有望持续受益。
- **风险提示：**1) 政策落地不及预期，宏观经济波动超预期。2) 市场波动风险。3) AI 端侧景气度不及预期风险。

中银国际证券股份有限公司
具备证券投资咨询业务资格

策略研究

证券分析师：王君

(8610)66229061

jun.wang@bocichina.com

证券投资咨询业务证书编号：S1300519060003

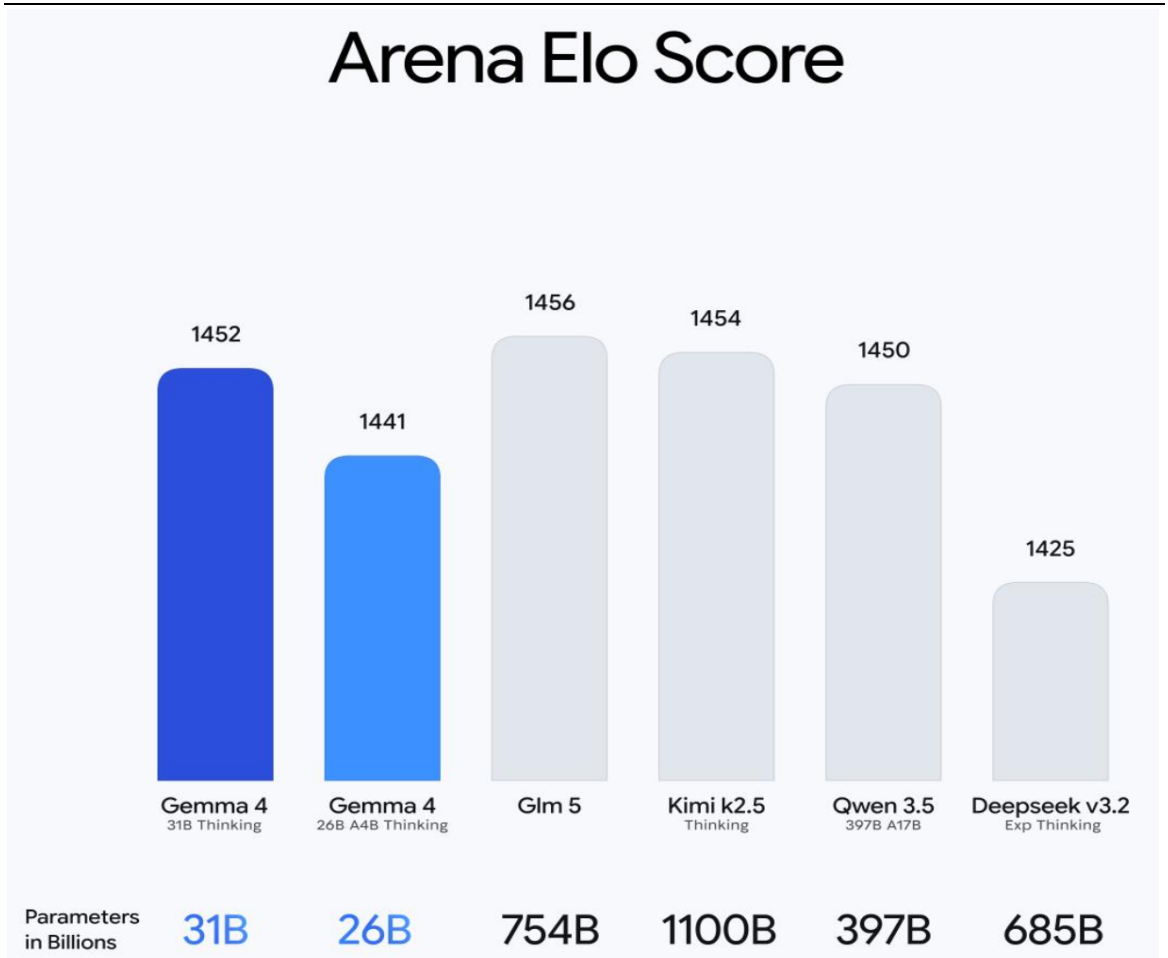
证券分析师：高天然

tianran.gao@bocichina.com

证券投资咨询业务证书编号：S1300522100001

谷歌发布 Gemma 4 开源大模型，或成为端侧 AI 产业拐点性事件。4 月初，谷歌正式推出 Gemma 4 大模型，Gemma 4 专为高级推理与智能体 workflow 打造。本次谷歌推出四种规格的 Gemma 4 通用模型：高效 20 亿参数版 (E2B)、高效 40 亿参数版 (E4B)、260 亿混合专家模型 (MoE) 与 310 亿稠密模型 (31B)。全系产品均超越简单对话场景，可处理复杂逻辑与智能体 workflow。31B 模型目前在行业标准 Arena AI 文本榜单中位列全球开源模型第三，对开发者而言，这种全新的单位参数智能水平意味着，只需更少的硬件开销，即可实现前沿级 AI 能力。E2B、E4B 两个小模型模型则可在手机、树莓派、英伟达 Jetson Orin Nano 等端侧设备上完全离线、近零延迟运行。通过 AICore 开发者预览版，Gemma 4 E2B 和 E4B 被直接集成进 Android 系统级接口。在 Pixel 手机和 Raspberry Pi 上，新模型的推理速度提升了 4 倍，电池消耗则降低了 60%。

图表 1. 31B 模型目前在行业标准 Arena AI 文本榜单中位列全球开源模型第三



资料来源：36kr，爱范儿，中银证券

Google 宣布将 Gemma 4 全系迁移至商业友好的 Apache 2.0 协议，在这套许可证下，开发者可以将模型用于任何目的，包括个人、商业和企业用途，无需支付版权，也无任何用途限制，修改和再分发同样自由。Gemma 4 现在可以合法打包进产品、服务和硬件设备中一并交付。

Gemma 4 的发布，不仅仅是一次常规的模型更新，也是端侧 AI 产业的重要拐点事件，Gemma 4 的小模型在手机上实现了离线运行、接近零延迟响应且功耗较低，同时具备处理复杂推理任务的能力。这意味着手机、耳机、智能眼镜、摄像头等终端设备拥有了不依赖云端、可随时调用的高级智能。Apache 2.0 协议之下，硬件厂商无需自研模型，这或将催生大量端侧 AI 原生应用的涌现，可能也会为整个端侧 AI 产业链（芯片、模组、终端）等打开规模化放量的确定性空间。

特斯拉 AI5 芯片流片完成，部分场景性能较前代提升 40 倍，预计 2027 年启动量产。据界面新闻报道，美国时间 4 月 15 日，特斯拉首席执行官马斯克宣布，旗下芯片设计团队已完成下一代 AI5 自动驾驶芯片的流片工作，该芯片部分场景性能较前代 AI4 提升 40 倍。AI5 单芯片 AI 算力接近 2500TOPS，内存容量达 144GB，其中原始算力较前代提升 8 倍，内存容量提升 9 倍，芯片专门针对 Transformer 引擎优化。AI5 将由三星、台积电分别在美国本土的工厂代工，预计 2027 年启动量产。应用规划方面，AI5 最初立项时计划用于特斯拉无人出租车项目，目前主要应用场景包括擎天柱机器人、特斯拉超级计算机集群及自动驾驶系统。AI5 进展之外，特斯拉 AI6 与 Dojo3 也在同步研发中。AI5 流片的成功是端侧算力的重要进展，表明端侧算力正在成为海外科技巨头们的必争之地，AI 端侧正在从单点突破走向生态竞争。

Gemma 4 的开源与特斯拉 AI5 的流片，分别从软件和硬件两个层面为端侧 AI 产业提供了新的增量逻辑，尽管目前端侧 AI 仍处于商业化早期阶段，但头部企业在模型轻量化与端侧算力上的持续投入，正在逐步降低终端部署的门槛，并为芯片、模组、终端及应用等环节提供更明确的技术路径与合作基础，AI 端侧产业链有望持续受益。

风险提示

1) 政策落地不及预期，宏观经济波动超预期。2) 市场波动风险。3) AI 端侧景气度不及预期风险。

披露声明

本报告准确表述了证券分析师的个人观点。该证券分析师声明，本人未在公司内、外部机构兼任有损本人独立性与客观性的其他职务，没有担任本报告评论的上市公司的董事、监事或高级管理人员；也不拥有与该上市公司有关的任何财务权益；本报告评论的上市公司或其它第三方都没有或没有承诺向本人提供与本报告有关的任何补偿或其它利益。

中银国际证券股份有限公司同时声明，将通过公司网站披露本公司授权公众媒体及其他机构刊载或者转发证券研究报告有关情况。如有投资者于未经授权的公众媒体看到或从其他机构获得本研究报告的，请慎重使用所获得的研究报告，以防止被误导，中银国际证券股份有限公司不对其报告理解和使用承担任何责任。

评级体系说明

以报告发布日后公司股价/行业指数涨跌幅相对同期相关市场指数的涨跌幅的表现为基准：

公司投资评级：

- 买入：预计该公司股价在未来 6-12 个月内超越基准指数 20% 以上；
- 增持：预计该公司股价在未来 6-12 个月内超越基准指数 10%-20%；
- 中性：预计该公司股价在未来 6-12 个月内相对基准指数变动幅度在-10%-10% 之间；
- 减持：预计该公司股价在未来 6-12 个月内相对基准指数跌幅在 10% 以上；
- 未有评级：因无法获取必要的资料或者其他原因，未能给出明确的投资评级。

行业投资评级：

- 强于大市：预计该行业指数在未来 6-12 个月内表现强于基准指数；
- 中性：预计该行业指数在未来 6-12 个月内表现基本与基准指数持平；
- 弱于大市：预计该行业指数在未来 6-12 个月内表现弱于基准指数；
- 未有评级：因无法获取必要的资料或者其他原因，未能给出明确的投资评级。

沪深市场基准指数为沪深 300 指数；新三板市场基准指数为三板成指或三板做市指数；香港市场基准指数为恒生指数或恒生中国企业指数；美股市场基准指数为纳斯达克综合指数或标普 500 指数。

风险提示及免责声明

本报告由中银国际证券股份有限公司证券分析师撰写并向特定客户发布。

本报告发布的特定客户包括：1) 基金、保险、QFII、QDII 等能够充分理解证券研究报告，具备专业信息处理能力的中银国际证券股份有限公司的机构客户；2) 中银国际证券股份有限公司的证券投资顾问服务团队，其可参考使用本报告。中银国际证券股份有限公司的证券投资顾问服务团队可能以本报告为基础，整合形成证券投资顾问服务建议或产品，提供给接受其证券投资顾问服务的客户。

中银国际证券股份有限公司不以任何方式或渠道向除上述特定客户外的公司个人客户提供本报告。中银国际证券股份有限公司的个人客户从任何外部渠道获得本报告的，亦不应直接依据所获得的研究报告作出投资决策；需充分咨询证券投资顾问意见，独立作出投资决策。中银国际证券股份有限公司不承担由此产生的任何责任及损失等。

本报告期内含保密信息，仅供收件人使用。阁下作为收件人，不得出于任何目的直接或间接复制、派发或转发此报告全部或部分内容予任何其他人，或将此报告全部或部分内容发表。如发现本研究报告被私自转载或转发的，中银国际证券股份有限公司将及时采取维权措施，追究有关媒体或者机构的责任。所有本报告期内使用的商标、服务标记及标记均为中银国际证券股份有限公司或其附属及关联公司（统称“中银国际集团”）的商标、服务标记、注册商标或注册服务标记。

本报告及其所载的任何信息、材料或内容只提供给阁下作参考之用，并未考虑到任何特别的投资目的、财务状况或特殊需要，不能成为或被视为出售或购买或认购证券或其它金融票据的要约或邀请，亦不构成任何合约或承诺的基础。中银国际证券股份有限公司不能确保本报告中提及的投资产品适合任何特定投资者。本报告的内容不构成对任何人的投资建议，阁下不会因为收到本报告而成为中银国际集团的客户。阁下收到或阅读本报告须在承诺购买任何报告中所指之投资产品之前，就该投资产品的适合性，包括阁下的特殊投资目的、财务状况及其特别需要寻求阁下相关投资顾问的意见。

尽管本报告所载资料的来源及观点都是中银国际证券股份有限公司及其证券分析师从相信可靠的来源取得或达到，但撰写本报告的证券分析师或中银国际集团的任何成员及其董事、高管、员工或其他任何个人（包括其关联方）都不能保证它们的准确性或完整性。除非法律或规则规定必须承担的责任外，中银国际集团任何成员不对使用本报告的材料而引致的损失负任何责任。本报告对其中所包含的或讨论的信息或意见的准确性、完整性或公平性不作任何明示或暗示的声明或保证。阁下不应单纯依靠本报告而取代个人的独立判断。本报告仅反映证券分析师在撰写本报告时的设想、见解及分析方法。中银国际集团成员可发布其它与本报告所载资料不一致及有不同结论的报告，亦有可能采取与本报告观点不同的投资策略。为免生疑问，本报告所载的观点并不代表中银国际集团成员的立场。

本报告可能附载其它网站的地址或超级链接。对于本报告可能涉及到中银国际集团本身网站以外的资料，中银国际集团未有参阅有关网站，也不对它们的内容负责。提供这些地址或超级链接（包括连接到中银国际集团网站的地址及超级链接）的目的，纯粹为了阁下的方便及参考，连结网站的内容不构成本报告的任何部份。阁下须承担浏览这些网站的风险。

本报告所载的资料、意见及推测仅基于现状，不构成任何保证，可随时更改，毋须提前通知。本报告不构成投资、法律、会计或税务建议或保证任何投资或策略适用于阁下个别情况。本报告不能作为阁下私人投资的建议。

过往的表现不能被视作将来表现的指示或保证，也不能代表或对将来表现做出任何明示或暗示的保障。本报告所载的资料、意见及预测只是反映证券分析师在本报告所载日期的判断，可随时更改。本报告中涉及证券或金融工具的价格、价值及收入可能出现上升或下跌。

部分投资可能不会轻易变现，可能在出售或变现投资时存在难度。同样，阁下获得有关投资的价值或风险的可靠信息也存在困难。本报告中包含或涉及的投资及服务可能未必适合阁下。如上所述，阁下须在做出任何投资决策之前，包括买卖本报告涉及的任何证券，寻求阁下相关投资顾问的意见。

中银国际证券股份有限公司及其附属及关联公司版权所有。保留一切权利。

中银国际证券股份有限公司

中国上海浦东
银城中路 200 号
中银大厦 39 楼
邮编 200121
电话: (8621) 6860 4866
传真: (8621) 5888 3554

相关关联机构:

中银国际研究有限公司

香港花园道一号
中银大厦二十楼
电话: (852) 3988 6333
致电香港免费电话:
中国网通 10 省市客户请拨打: 10800 8521065
中国电信 21 省市客户请拨打: 10800 1521065
新加坡客户请拨打: 800 852 3392
传真: (852) 2147 9513

中银国际证券有限公司

香港花园道一号
中银大厦二十楼
电话: (852) 3988 6333
传真: (852) 2147 9513

中银国际控股有限公司北京代表处

中国北京市西城区
西单北大街 110 号 8 层
邮编: 100032
电话: (8610) 8326 2000
传真: (8610) 8326 2291

中银国际(英国)有限公司

2/F, 1 Lothbury
London EC2R 7DB
United Kingdom
电话: (4420) 3651 8888
传真: (4420) 3651 8877

中银国际(美国)有限公司

美国纽约市美国大道 1045 号
7 Bryant Park 15 楼
NY 10018
电话: (1) 212 259 0888
传真: (1) 212 259 0889

中银国际(新加坡)有限公司

注册编号 199303046Z
新加坡百得利路四号
中国银行大厦四楼(049908)
电话: (65) 6692 6829 / 6534 5587
传真: (65) 6534 3996 / 6532 3371