

传媒

行业快报

Gemma 2 发布，降本提效延展 AI 生态

投资要点

- ◆ **热点事件:** 6月27日, Google 面向全球研究人员和开发人员发布 Gemma 2 大语言模型, 目前有 90 亿参数 (9B) 和 270 亿参数 (27B) 两种大小, 并将在未来几个月发布参数为 26 亿的 Gemma 2 模型, 更适合智能手机的人工智能应用场景。大语言模型持续迭代, 降本增效有望加速 AI 赋能 B 端和 C 端应用和生产。
- ◆ **Gemma 2 带来效率和性能的提升。** 目前, Gemma 2 共推出 9B 和 27B 两种参数, 与上一代相比性能大幅度提升, 但部署要求大幅度下降, 只需一块 NVIDIA H100 Tensor Core GPU 或 TPU 主机就能使用。**架构方面:** 基于上一代进行了全方位改良, 使用局部滑动窗口注意力和全局注意力, 其中局部注意力层的滑动窗口大小设置为 4096 个 tokens, 全局注意力层的跨度设置为 8192 个 tokens。**训练数据方面:** Gemma 2 的 27B 模型使用了 13 万亿 tokens 的英文数据进行训练, 9B 模型使用了 8 万亿 tokens 数据, 这些数据包括网页文档、代码、论文和科学文章等。**训练策略方面:** Gemma 2 使用了知识蒸馏方法, 通过学习大型教师模型给出的输出概率进行训练, 以较小规模复制卓越性能。通过 MBPP、MMLU、ARC-C、GSM8K、BBQ Disambig 等知名基准测试平台的综合评测, Gemma 2 在多项基准测试中表现出色, 例如在 MMLU 5-shot (Gemma-2 得分 75.2%, 显著高于 Gemma-1 的 42.3%)、ARC-C 25-shot、GSM8K 5-shot 等测试中, 27B 模型相比之前版本和其他标准模型 (如 Mistral 和 LLaMA-3 等) 均有显著提升。在其他数学推理、逻辑问题解决和常识问答等任务上, Gemma 2 同样展示了在处理复杂认知任务方面的强大效能。Gemma 2 在降低成本同时显著提升能效, 在硬件端降低使用门槛, 利好铺展 AI 生态。
- ◆ **Gemma 2 (2.6B) 有望赋能端侧 AI 生态。** 预计将在未来几个月内发布的 26 亿 Gemma 2 模型将更适合智能手机的 AI 应用场景。相较 Gemma-1 2.5B, Gemma-2 2.6B 几乎在所有指标测评中实现超越, 尤其在 MMLU 5-shot (提升 9 个百分点) 和 GSM8K 5-shot (提升 8.8 个百分点) 提升相对显著。随着 Apple、Samsung 等移动终端和联想、惠普等 PC 对端侧 AI 需求的逐渐提升, 优质轻量模型的迭代和发展符合需求端的增长。我们看好轻量模型迭代催化对端侧 AI 的持续赋能。
- ◆ **投资建议:** 优质模型持续迭代, 降本增效, 有望拓展 AI 生态, 多元赋能千行百业发展。建议关注: 腾讯控股、网易-S、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、因赛集团、天下秀、上海电影、华扬联众、掌阅科技、顺网科技、中广天择、万兴科技等。
- ◆ **风险提示:** 政策不确定性、AI 落地不及预期、宏观环境波动风险等。

 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	-4.21	-16.44	-24.01
绝对收益	-9.2	-17.82	-34.18

分析师

倪爽

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

相关报告

- 传媒: 钉钉举办重磅会议, 构建开放 AI 生态-华金证券-传媒-行业快报 2024.6.27
- 传媒: OpenAI 禁令发布, 国产大模型或迎发展机遇-华金证券-传媒-行业快报 2024.6.27
- 传媒: Claude 3.5 Sonnet 发布, AI 应用场景持续拓展-华金证券-传媒-行业快报 2024.6.23
- 传媒: 视频生成赛道火热, 下游产业有望受益-华金证券-传媒-行业快报 2024.6.16
- 传媒: Apple 携手 OpenAI, 端侧 AI 迎重磅催化-华金证券-传媒-行业快报 2024.6.11
- 传媒: 微短剧新规生效, 助力优质内容稳健发展-华金证券-传媒-行业快报 2024.6.4
- 传媒: 影视产业持续回暖, 关注档期潜在爆款-华金证券-传媒-行业快报 2024.6.2



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