国联证券 GUOLIAN SECURITIES

证券研究报告

行业研究|行业点评研究|计算机（2171）

Meta 发布开源大模型 Llama 3.1，开源模型能力进一步提升

2024年07月28日

报告要点

Meta 发布开源大模型 Llama3.1，此模型包含 405B、70B、8B 三个版本。该模型在多项基准测试中超越 GPT-4o 和 Claude 3.5 Sonnet，开源模型的能力或已追赶上闭源 SOTA 模型。Meta 外发的开源大模型 Llama3.1 最高版本参数量达到 4050 亿，该版本性能与最好的闭源模型性能接近。Llama3.1 开源/免费使用权重和代码，并允许进行模型微调、蒸馏到其他模型以及在任何地点部署。模型提供 128k 上下文窗口，在多语言处理、优秀的代码生成、复杂问题理解推理能力上大幅提升，并包含模型工具使用。

分析师及联系人


黄楷　　陈安宇

SAC：S0590522090001　SAC：S0590523080004

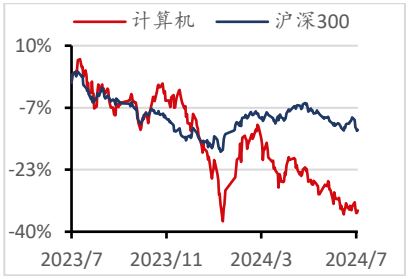
计算机

Meta 发布开源大模型 Llama 3.1，开源模型能力进一步提升

投资建议： 强于大市（维持）

上次建议： 强于大市

相对大盘走势



相关报告

1、《计算机：重仓持股比例环比继续下降，信创持股市值增加》2024.07.23

2、《计算机：车路云一体化系列之三—V2X 车载终端篇》2024.07.19

扫码查看更多

行业事件

Meta 发布开源大模型 Llama3.1，此模型包含 405B、70B、8B 三个版本。该模型在多项基准测试中超越 GPT-4o 和 Claude 3.5 Sonnet，开源模型的能力或已追赶上闭源 SOTA 模型。

模型能力进一步提升，开源使用权重与代码

Meta 外发的开源大模型 Llama3.1 最高版本参数量达到 4050 亿，该版本性能与最好的闭源模型性能接近。Llama3.1 开源/免费使用权重和代码，并允许进行模型微调、蒸馏到其他模型以及在任何地点部署。模型提供 128k 上下文窗口，在多语言处理、优秀的代码生成、复杂问题理解推理能力上大幅提升，并包含模型工具使用。Llama StackAPI 可以轻松集成。整个生态系统包含 25 个合作伙伴，其中包括亚马逊、英伟达、Databricks、Groq、微软云和谷歌云。

强算力叠加 Transformer 架构，开发高质量开源模型

Meta 在 Llama3.1 的报告中指出：数据、规模和复杂性管理是开发高质量模型的关键因素。数据上，改进了用于前训练和后训练的数据的数量和质量；规模上，模型在预训练时使用浮点运算规模几乎为最大版本 Llama2 的 50 倍，在 15.6T 文本上预训练了 4050 亿参数的模型；复杂性管理上，采用了 Transformer 架构并稍作调整，而不是 MoE 架构，在后训练中采用了监督微调（SFT）、拒绝采样（RS）和直接偏好优化（DPO）。模型 405B 版本使用 16K 个 H100 GPU 训练，对应的服务器配备 8 个 GPU 和两个 CPU，强算力平台叠加 Transformer 架构进一步提升模型质量。

投资建议

模型开源可以保护用户的数据，帮助用户微调/蒸馏适用于自己的模型，促使更多用户使用 AI 模型，从而长期推动 AI 生态体系进步。AI 大模型产业的发展有望带动四方面投资机遇。（1）算力基础设施建设机遇：大模型能力提升或推动算力需求改变，国产 GPU 生态体系加速发展，建议关注中科曙光、紫光股份、浪潮信息等领军企业；**（2）端侧 AI 软件开发机遇：**端侧设备将是人机交互的重要中介，或将带动端侧 AI 应用的软件开发机遇，建议关注中科创达等相关公司；**（3）生产力工具革新机遇：**大模型有望为生产力工具带来降低专业门槛、减少重复劳动等变化，建议关注金山办公、用友网络、泛微网络等；**（4）行业信息化创新机遇：**受益于人机交互能力提升，大模型有望率先在政务、金融等领域窗口服务或培训场景落地，行业信息化厂商将是连接基础大模型厂商和行业客户的重要环节，建议关注恒生电子、宇信科技、中控技术、卫宁健康等行业信息化头部企业。

风险提示：客户转化程度不及预期；商业化进程不及预期风险；行业竞争加剧风险等。

1. 风险提示

客户转化程度不及预期：由于传统搜索引擎存在多年，已在使用者中日常生活中形成固定行为模式，新型的 AI 搜索模式需要使用者一定时间适应。

商业化进程不及预期风险：AI 搜索模式逐步推出，成熟的盈利商业模式仍未形成，商业化进程或有不及预期风险。

行业竞争加剧风险：AI 搜索市场规模大，有取代传统搜索市场可能性，行业竞争或出现加剧。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即：以报告发布日后的 6 到 12 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场以沪深 300 指数为基准，北交所市场以北证 50 指数为基准；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准；韩国市场以柯斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于 10%
		增持	相对同期相关证券市场代表性指数涨幅在 5%～10%之间
		持有	相对同期相关证券市场代表性指数涨幅在-5%～5%之间
		卖出	相对同期相关证券市场代表性指数涨幅小于-5%
	行业评级	强于大市	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		弱于大市	相对表现弱于同期相关证券市场代表性指数

一般声明

特别声明

版权声明

联系我们

北京：北京市东城区安外大街 208 号玖安广场 A 座 4 层
上海：上海市虹口区杨树浦路 188 号星立方大厦 8 层
无锡：江苏省无锡市金融一街 8 号国联金融大厦 16 楼
深圳：广东省深圳市福田区益田路 4068 号卓越时代广场 1 期 13 楼