

# Claude新版本发布，大模型能力持续高歌猛进

2024 年 06 月 24 日

## 【事项】

- ◆ 北京时间 6 月 21 日，Anthropic 正式发布 Claude 3.5 Sonnet。从 Anthropic 透露的信息来看，在阅读、编码、数学和视觉等几个主要 AI 基准测试中，新模型能力不仅超过了 Claude 3 Sonnet，还超越了 Anthropic 此前最强版本的旗舰模型 Claude 3 Opus。另外，在其释出的一系列与行业竞争对手的测试结果榜单中，Claude 3.5 Sonnet 的大部分成绩优于 OpenAI、Google、Meta 的旗舰模型，包括刚推出不久的 GPT-4o。细致来看，Claude 3.5 Sonnet 在研究生水平推理 (GPQA)、编码能力 (HumanEval)、文本推理 (DROP) 等方面的成绩均好于 GPT-4o，在数学问题解决能力上稍逊于后者。

## 【评论】

- ◆ **性能优异，多模态能力优秀。**Claude 3.5 具有 20 万个 token 的上下文窗口，约合 15 万个单词。其图像理解能力也尤为突出，能轻易解读图表图形、转录低画质图像中的文字。应用场景上，可以做到从商品的包装、标签或收据上读取信息，或者识别大量的货物标签和运输文件，在零售、物流和金融服务领域发挥关键能力。
- ◆ **价格降低，运行速度提高。**从价格上来看，Claude 3.5 Sonnet 每百万输入 3 美元，每百万输出 15 美元，是上一代 Opus 的 1/5，运行速度却得到了一倍的提升。
- ◆ **实际应用场景不局限于聊天机器人。**Anthropic 推出名为「Artifacts」的新功能。用户需要 Claude 生成代码片段、文本、或网站设计等内容时，Claude 会以「Artifacts」窗口呈现。用户可以在独立窗口中边聊边看边改，将 AI 生成的内容无缝地融入到项目的工作中。团队协作功能也在研发中，未来各团队将能在一个共享空间中集中管理工作，企业更多的应用场景得以满足。

强于大市 (维持)

东方财富证券研究所

证券分析师：方科

证书编号：S1160522040001

电话：021-23586361

相对指数表现



相关研究

- 《华为开发者大会召开，关注华为产业链》  
2024.06.20
- 《多地车路云项目落地进行中》  
2024.06.20
- 《陆家嘴论坛召开，聚焦科技创新与资本市场高质量发展》  
2024.06.19
- 《《上海低空经济发展白皮书 2024》发布，产业发展态势良好》  
2024.06.18
- 《数据要素政策密集出台催化产业落地》  
2024.06.17

图表 1: Claude 3.5 Sonnet 与其他大模型效果对比

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro	Llama-400b (early snapshot)
Graduate level reasoning <i>GPQA, Diamond</i>	<b>59.4%*</b> 0-shot CoT	50.4% 0-shot CoT	53.6% 0-shot CoT	—	—
Undergraduate level knowledge <i>MMLU</i>	<b>88.7%**</b> 5-shot	86.8% 5-shot	—	85.9% 5-shot	86.1% 5-shot
	<b>88.3%</b> 0-shot CoT	85.7% 0-shot CoT	<b>88.7%</b> 0-shot CoT	—	—
Code <i>HumanEval</i>	<b>92.0%</b> 0-shot	84.9% 0-shot	90.2% 0-shot	84.1% 0-shot	84.1% 0-shot
Multilingual math <i>MGSM</i>	<b>91.6%</b> 0-shot CoT	90.7% 0-shot CoT	90.5% 0-shot CoT	87.5% 8-shot	—
Reasoning over text <i>DROP, F1 score</i>	<b>87.1</b> 3-shot	83.1 3-shot	83.4 3-shot	74.9 Variable shots	83.5 3-shot Pre-trained model
Mixed evaluations <i>BIG-Bench-Hard</i>	<b>93.1%</b> 3-shot CoT	86.8% 3-shot CoT	—	89.2% 3-shot CoT	85.3% 3-shot CoT Pre-trained model
Math problem-solving <i>MATH</i>	<b>71.1%</b> 0-shot CoT	60.1% 0-shot CoT	76.6% 0-shot CoT	67.7% 4-shot	57.8% 4-shot CoT
Grade school math <i>GSMSK</i>	<b>96.4%</b> 0-shot CoT	95.0% 0-shot CoT	—	90.8% 11-shot	94.1% 8-shot CoT

\* Claude 3.5 Sonnet scores 67.2% on 5-shot CoT GPQA with maj@32

\*\* Claude 3.5 Sonnet scores 90.4% on MMLU with 5-shot CoT prompting

资料来源：机器之心公众号，东方财富证券研究所

我们认为，海外大模型厂商之间的技术竞赛持续演绎，利于AI行业整体的发展，建议关注国内AI产业链大模型侧的新变化。

建议关注：科大讯飞，三六零，商汤等

## 【风险提示】

技术发展不及预期；  
市场竞争加剧。

东方财富证券股份有限公司（以下简称“本公司”）具有中国证监会核准的证券投资咨询业务资格  
**分析师申明：**

作者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，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

#### **投资建议的评级标准：**

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3到12个月内的相对市场表现，也即：以报告发布日后的3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；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以标普500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15%以上；  
增持：相对同期相关证券市场代表性指数涨幅介于5%~15%之间；  
中性：相对同期相关证券市场代表性指数涨幅介于-5%~5%之间；  
减持：相对同期相关证券市场代表性指数涨幅介于-15%~-5%之间；  
卖出：相对同期相关证券市场代表性指数跌幅15%以上。

#### **行业评级**

强于大市：相对同期相关证券市场代表性指数涨幅10%以上；  
中性：相对同期相关证券市场代表性指数涨幅介于-10%~10%之间；  
弱于大市：相对同期相关证券市场代表性指数跌幅10%以上。

#### **免责声明：**

本研究报告由东方财富证券股份有限公司制作及在中华人民共和国（香港和澳门特别行政区、台湾省除外）发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人作出邀请。

本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。

那些涉及期货、期权及其它衍生工具的交易，因其包括重大的市场风险，因此并不适合所有投资者。

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所引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东方财富证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。