

分析师：陈良栋

执业证号：S0100523050005

邮箱：chenliangdong@mszq.com

分析师：李瑶

执业证号：S0100523090002

邮箱：liyao_yj@mszq.com

事件

3月4日晚，美国人工智能初创公司 Anthropic 重磅发布其第三代人工智能(AI)模型——Claude 3 系列模型，包括 Claude 3 Opus、Claude 3 Sonnet 和 Claude 3 Haiku。继 2023 年 11 月推出 Claude 2.1 后，Anthropic 在发布的新一代大语言模型 Claude 3 系列中，按 Haiku、Sonnet、Opus 的次序递增规模，依次提供越来越强大的性能，使模型可以应用于需要高度智能和复杂任务处理的场景，如复杂金融预测、研究和开发等。

► Claude 3 在推理、数学、编码、多语言理解和视觉方面表现出色

1) 快速实时响应，强大视觉能力：以速度和成本效益见长的 Claude 3 Haiku 模型，能在三秒内读取 arXiv 上一篇文章和信息和数据密集的研究论文(约 10k token)，并附带图标和图形。**2) 减少“不必要拒绝”，提高准确度：**Claude 3 系列大模型在面对潜在敏感提示时，大幅降低了错误拒绝的概率，更精准细腻地理解问题的真实意图，减少了“不必要拒绝”，更智能地处理各种请求。**3) 100 万 token 完美展现上下文记忆：**Claude 3 系列模型可提供 200K 上下文窗口，Haiku、Sonnet 和 Opus 三个模型都具备 Gemini 1.5 Pro 处理超过 100 万 token 输入的能力。**4) 用户使用：**Opus 和 Sonnet 模型已经可以通过 Anthropic 的 API 调用，开发者可以立即注册并开始体验；普通用户可以在 claude.ai 上免费体验 Sonnet，最强大的 Opus 仅对 Claude Pro 付费订阅用户开放。

► Claude 与 ChatGPT 训练方法略有不同，Opus 版本表现超越 GPT-4

在模型训练上，Claude 与 ChatGPT 虽然都是靠强化学习(RL)来训练偏好模型，并进行后续微调，但是在训练方法上，Claude 与 ChatGPT 略有不同，ChatGPT 采用人类反馈强化学习(RLHF)，Claude 采用的原发人工智能方法，是基于偏好模型而非人工反馈来进行训练的，这种方法又被称为“AI 反馈强化学习(RLAIF)”。Claude 3 Opus 作为最强版本，其模型在人工智能系统的大多数常见评估基准上都优于同行，包括本科水平专家知识(MMLU)、研究生水平专家推理(GPQA)、基础数学(GSM8K)等，其性能超过 GPT-4。

推荐

维持评级

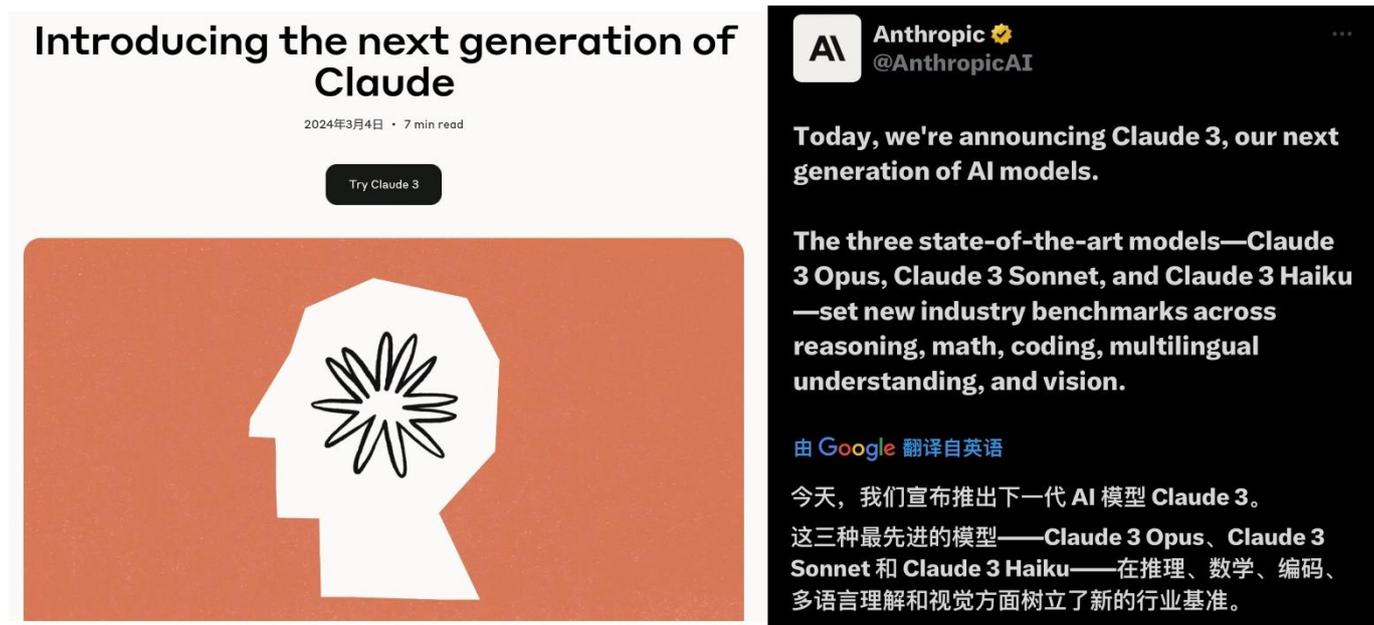
相关研究

1. 传媒互联网行业周报：关注 AI 生视频产业推进，重视 IP 形态升维的价值释放-2024/03/03
2. 传媒互联网行业周报：Sora 展示文生视频能力惊艳，重视内容生产工具层革新-2024/02/25
3. 传媒行业事件点评：OpenAI 发布文生视频模型 Sora，重视内容生产工具层革新-2024/02/21
4. 春节档电影表现回顾及 2024 年展望：春节档票房再创新高，优质影片驱动行业持续复苏-2024/02/18
5. 传媒互联网行业周报：Vision Pro 正式发售，年内首批进口游戏版号发放-2024/02/04

➤ **投资建议：**我们认为 Claude3 系列大模型在推理、数学、编码、多语言理解和视觉等方面表现出色，大模型底层能力持续提升，同时多模态领域也有更多方向性发展，我们认为底层大模型仍处于一个快速发展且竞争的环境中，优质的模型能力有望推动 AI 应用端的使用场景拓展与内容创新，建议重点关注：1) 坚定大模型以及垂直模型布局的公司；2) 核心业务的应用场景有望受益的公司。建议关注昆仑万维、中文在线、上海电影、易点天下、浙文互联、紫天科技、汤姆猫、盛天网络、皖新传媒、南方传媒、世纪天鸿、元隆雅图、视觉中国、奥飞娱乐、蓝色光标、掌趣科技、慈文传媒等。

➤ **风险提示：**AIGC 进展不及预期的风险、AI 应用落地不及预期、政策监管趋严、行业竞争加剧等风险。

图1: Anthropic 发布 Claude 3 系列大模型



资料来源：Anthropic 官网，X 官方账号，民生证券研究院

图2：Claude 3 系列大模型在测试中超越 GPT-4

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge <i>MMLU</i>	86.8% 5-shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	—	—
Grade school math <i>GSM8K</i>	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot	94.4% Maj1@32	86.5% Maj1@32
Math problem-solving <i>MATH</i>	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math <i>MGSM</i>	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	—	79.0% 8-shot	63.5% 8-shot
Code <i>HumanEval</i>	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text <i>DROP, FI score</i>	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed evaluations <i>BIG-Bench-Hard</i>	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Knowledge Q&A <i>ARC-Challenge</i>	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	—	—
Common Knowledge <i>HellaSwag</i>	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84.7% 10-shot

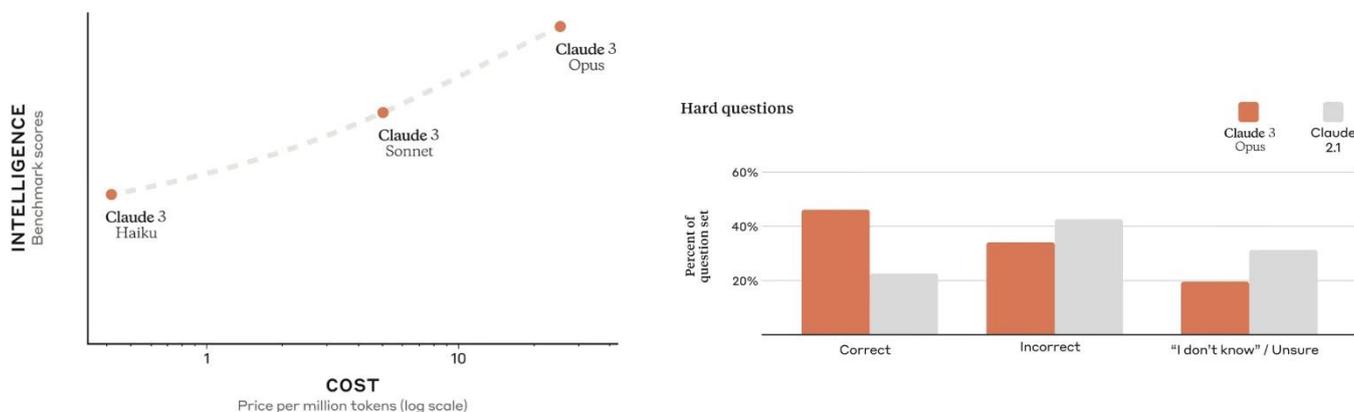
资料来源：Anthropic 官网，民生证券研究院

图3：Claude 3 系列大模型具备与其他领先模型相媲美的高级视觉识别能力

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4V	Gemini 1.0 Ultra	Gemini 1.0 Pro
Math & reasoning <i>MMMU (val)</i>	59.4%	53.1%	50.2%	56.8%	59.4%	47.9%
Document visual Q&A <i>ANLS score, test</i>	89.3%	89.5%	88.8%	88.4%	90.9%	88.1%
Math <i>MathVista (testmini)</i>	50.5% CoT	47.9% CoT	46.4% CoT	49.9%	53.0%	45.2%
Science diagrams <i>AI2D, test</i>	88.1%	88.7%	86.7%	78.2%	79.5%	73.9%
Chart Q&A <i>Relaxed accuracy (test)</i>	80.8% 0-shot CoT	81.1% 0-shot CoT	81.7% 0-shot CoT	78.5% 4-shot CoT	80.8%	74.1%

资料来源：Anthropic 官网，民生证券研究院

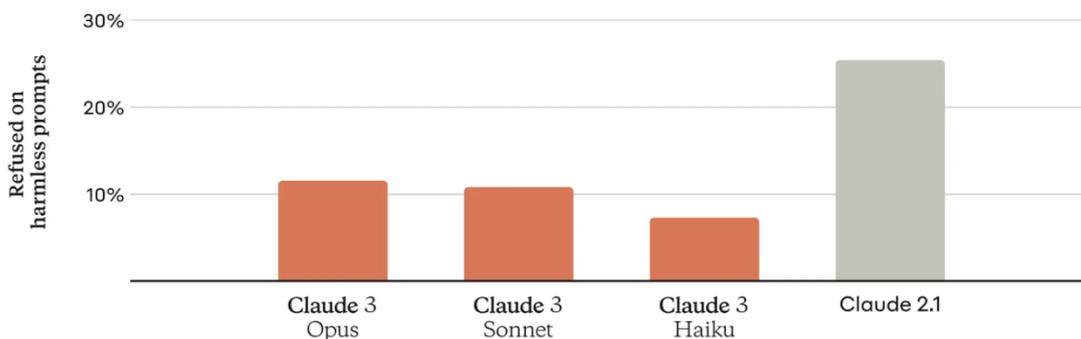
图4: Claude 3 Opus 模型为 Claude 3 系列最强版本——精准度大幅提高



资料来源: Anthropic 官网, 民生证券研究院

图5: Claude 3 系列大模型减少“不必要拒绝”

Incorrect refusals



资料来源: Anthropic 官网, 民生证券研究院

图6: Claude 3 Opus、Claude 3 Sonnet、Claude 3 Haiku 模型横向比较

Model	Key Features & Differentiator
Claude 3 Opus	<p>Claude 3 Opus is our most intelligent model, with best-in-market performance on highly complex tasks. It can navigate open-ended prompts and sight unseen scenarios with remarkable fluency and human-like understanding. Opus shows us the outer limits of what's possible with generative AI.</p> <p>Cost: \$15 \$75 [Input \$/million tokens Output \$/million tokens]</p> <p>Context window: 200K*</p> <p>Potential uses:</p> <ul style="list-style-type: none"> Task automation: plan and execute complex actions across APIs and databases, interactive coding R&D: research review, brainstorming and hypothesis generation, drug discovery Strategy: advanced analysis of charts & graphs, financials and market trends, forecasting <p>Differentiator: Higher intelligence than any other model available.</p>
Claude 3 Sonnet	<p>Claude 3 Sonnet strikes the ideal balance between intelligence and speed—particularly for enterprise workloads. It delivers strong performance at a lower cost compared to its peers, and is engineered for high endurance in large-scale AI deployments.</p> <p>Cost: \$8 \$15 [Input \$/million tokens Output \$/million tokens]</p> <p>Context window: 200K</p> <p>Potential uses:</p> <ul style="list-style-type: none"> Data processing: RAG or search & retrieval over vast amounts of knowledge Sales: product recommendations, forecasting, targeted marketing Time-saving tasks: code generation, quality control, parse text from images <p>Differentiator: More affordable than other models with similar intelligence; better for scale.</p>
Claude 3 Haiku	<p>Claude 3 Haiku is our fastest, most compact model for near-instant responsiveness. It answers simple queries and requests with unmatched speed. Users will be able to build seamless AI experiences that mimic human interactions.</p> <p>Cost: \$0.25 \$1.25 [Input \$/million tokens Output \$/million tokens]</p> <p>Context window: 200K</p> <p>Potential uses:</p> <ul style="list-style-type: none"> Customer interactions: quick and accurate support in live interactions, translations Content moderation: catch risky behavior or customer requests Cost-saving tasks: optimized logistics, inventory management, extract knowledge from unstructured data <p>Differentiator: Smarter, faster, and more affordable than other models in its intelligence category.</p>

资料来源: Anthropic 官网, 民生证券研究院

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个别客户的特殊状况、目标或需要，客户应当充分考虑自身特定状况，不应单纯依靠本报告所载的内容而取代个人的独立判断。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：广东省深圳市福田区益田路 6001 号太平金融大厦 32 层 05 单元； 518026