



计算机行业研究

买入（维持评级）

行业研究

证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005）

liugaochang@gjzq.com.cn

海外模型与 CSP 更新

本周观点

模型能力升级与成本下降并进，商业收入与科研应用同步落地。Anthropic 第三季度收入 168 亿美元、环比增长 45%，OpenAI 九月年化收入接近 500 亿美元，头部厂商变现持续提速。GPT-6.1 Sol、Sonnet 5.5 与 Haiku 5.5 在专业任务能力、运行速度与调用成本上进一步兼顾，其中 GPT-6.1 Sol 以接近 Astra 的智能水平实现约五分之一的价格，缓存输入成本较 GPT-6 Sol 再降 50%。科研方面，OpenAI 公开一批数学研究手稿及证明材料，Anthropic 披露 Claude 在量子电路设计中的闭环优化成果。专业能力提升叠加使用成本下行，有望扩大模型在企业业务与科研工作中的应用范围。

海外大厂加快推出企业与个人 Agent，向办公协作、专业服务与持续执行延伸。谷歌将 Gemini 接入办公应用，并拓展财务核对、法律研究等专业场景；Meta 以企业平台与 Muse Wonder 强化跨对话记忆（Active Memory）和持续执行能力（Active Focus）；OpenAI Dots 支持跨应用推进任务，AWS Bedrock 托管代理则提供任务状态保存、工具调用与权限管理。模型与办公入口、企业数据及云平台的结合，正在减少企业接入 Agent 的开发和调试环节，推动其进入日常业务流程。

微软扩大自研推理芯片部署，AWS 业务高增长支撑 AI 投入。微软 Maia 200 已用于 Copilot 等产品生产负载，Maia 300 计划于 2027 年推出，并自 2027 年起将 Maia 系列扩展至多吉瓦规模、成为主要推理系统；AWS 管理层看好 AI 投入可持续性，第二季度收入同比增长 36.7%、AI 业务年化收入超过 250 亿美元，Bedrock 客户规模与消费持续扩大。企业应用普及有望带动持续推理需求，自研芯片、算力利用率提升与推理成本优化，则有助于改善云厂商投资回报。

相关标的

国内算力：寒武纪、海光信息、长鑫科技、东阳光、中芯国际、华虹半导体、中科曙光、摩尔线程、华勤技术、利通电子、亿田智能、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、京基智农、浪潮信息、胜蓝股份、盛视科技、神州数码、润泽科技、大位科技、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子等。

海外算力：工业富联、中际旭创、东山精密、胜宏科技、方正科技、麦格米特、先导基电、光智科技、AXTI、新易盛、蓝思科技（元拾科技）、创世纪、海鸥股份、天孚通信、光迅科技、源杰科技、江海股份、生益科技、深南电路、广合科技、奥士康、世运电路、景旺电子、鹏鼎控股、沪电股份、华盛昌、联讯仪器、唯科科技、领益智造（立敏达）、中钨高新、英维克、飞龙股份、冰轮环境、德福科技、南亚新材、火炬电子、三环集团、海星股份、天岳先进、云南锗业、力量钻石、沃尔德、四方达等；英伟达、ALPHABET、台积电、Lumentum、SK 海力士、英特尔、Bloom Energy、美光、闪迪。

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、网宿科技等。

AI 应用：深信服、德才股份、迈富时、汇量科技、卓易信息、鼎捷数致、达梦数据、中控技术、金蝶国际、用友网络、海致科技集团、汉朔科技、金山办公、合合信息、福昕软件、金斯瑞生物科技、晶泰控股-P、英矽智能等

风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；特定行业下游资本开支周期性波动的风险。



内容目录

一、模型能力升级与成本下降并进，商业收入与科研应用同步落地.....	3
1.1 头部模型公司收入快速增长.....	3
1.2 模型升级兼顾专业能力与成本.....	3
1.3 数学研究与科学计算取得新进展.....	3
二、海外大厂加快推出企业与个人 Agent，向办公协作、专业服务与持续执行延伸	4
2.1 谷歌拓展财务法律及办公场景.....	4
2.2 Meta 发布企业平台与 Muse Wonder.....	4
2.3 OpenAI 与 AWS 推出持续执行及托管产品	4
三、微软扩大自研推理芯片部署，AWS 业务高增长支撑 AI 投入.....	5
3.1 微软扩大 Maia 推理芯片部署.....	5
3.2 AWS 看好 AI 投入，业务增长提供需求支撑	5
四、相关标的.....	6
五、风险提示.....	6

图表目录

图表 1： Claude 进行闭环量子电路优化.....	4
图表 2： Azure Maia 300 系统	5



一、模型能力升级与成本下降并进，商业收入与科研应用同步落地

1.1 头部模型公司收入快速增长

据路透社 10 月 14 日报道，Anthropic 第三季度收入为 168 亿美元，环比增长 45%，第二季度和第一季度分别为 116 亿美元和 47 亿美元。公司还向参会者表示，第三季度已实现经调整口径盈利。

OpenAI 亦披露收入进展。据路透社 10 月 8 日报道，OpenAI 9 月年化收入接近 500 亿美元。此为年化收入口径，与 Anthropic 披露的单季度收入不同。

我们认为，头部模型公司的收入增长反映出 AI 付费需求的持续释放，专业任务能力提升有望带动更多企业采用模型服务。

1.2 模型升级兼顾专业能力与成本

OpenAI 发布 GPT-6.1 SoI，升级重点包括代码编写与调试、文档理解，以及多步骤业务任务的执行。根据官方介绍，新模型在智能体编程、计算机使用和专业工作方面的能力接近 GPT-6 Astra，标准输入和输出 Token 价格仅为 Astra 的五分之一。

对于需要重复调用同一批上下文的 Agent，GPT-6.1 SoI 提供了更低的缓存输入价格，每百万 Token 为 0.10 美元，较 GPT-6 SoI 下降 50%。

Anthropic 的 Sonnet 5.5 主要面向需求明确的日常工作，例如修复程序错误、制作文档、幻灯片及电子表格。根据公司测试，Sonnet 5.5 完成单项任务的成本较前代最多降低 30%，输出速度较 Sonnet 5 提升 30% 以上。

Haiku 5.5 面向调用量大、对成本较为敏感的任务，可用于生成摘要、查询数据库和信息分类，也可在编程任务中担任子 Agent。Anthropic 表示，相较 Haiku 4.5，新模型平均运行成本降低约 75%。

我们认为，更低的调用成本有利于 AI 进入高频业务场景，企业可以将摘要、分类等重复性工作交给低成本模型，将复杂任务交给能力更强的模型，在控制费用的同时扩大使用范围。我们认为，对于需要多轮调用、反复读取资料的任务，缓存价格下降能够降低重复输入的费用，使企业更容易将模型用于持续运行的工作。

1.3 数学研究与科学计算取得新进展

10 月 6 日，OpenAI 公开内部前沿模型取得的一批数学研究成果，并通过 GitHub 发布相关手稿与证明材料，覆盖多个研究方向。

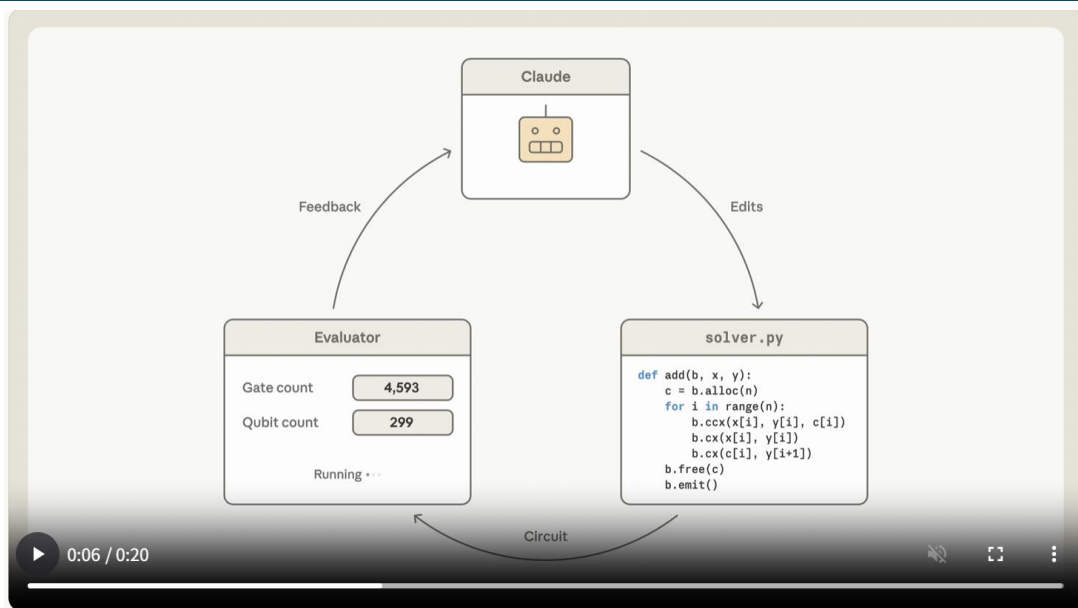
截至 10 月 17 日，OpenAI 公开目录包含 719 篇手稿，按相关研究内容归为 372 个成果组。同一组可包括主要结果、配套论证、推论或不同证明，因此手稿数量与独立问题数量并不相同。约 42% 的主要结果已完成形式化证明，其余成果仍处于不同验证阶段。

10 月 16 日，Anthropic 披露 Claude Mythos 5 在量子电路设计中的研究进展，涉及化学、优化、量子模拟、数论和数值计算等领域的 8 项基础计算任务。模型提出电路方案后，由自动评估程序检查正确性，并根据量子比特和门数量评分；模型再依据反馈改进设计，多轮迭代之间无需人工输入。

我们认为，模型在数学证明和算法设计中的进展，有望扩大 AI 在科研与工程研发中的应用范围，能够反复提出方案、接受验证并改进结果的模型，可承担更多耗时的试验与优化工作，提升专业人员的研发效率。



图表1: Claude 进行闭环量子电路优化



来源: Anthropic, 国金证券研究所

二、海外大厂加快推出企业与个人 Agent，向办公协作、专业服务与持续执行延伸

2.1 谷歌拓展财务法律及办公场景

谷歌在 Google Cloud 上线 Gemini 4 Argon，支持阅读混合文字、表格、图表及图片的长文档，并依据材料生成结果。根据官方介绍，财务方面包括多步骤账簿核对、电子表格公式处理与差异分析；法律方面包括查阅案例和监管文件，提取可核验的引用及关键先例。

谷歌同时将 Gemini 作为通用工作 Agent 接入 Gmail、Drive、Docs、Slides、Sheets、Chat 及 Calendar，用户在不同应用中使用，可沿用相同的记忆、技能及控制设置。

2.2 Meta 发布企业平台与 Muse Wonder

Meta 推出企业平台，初期向企业及开发者提供 Muse agent、Meta Business Agent、Muse API 和 Muse Code 等产品。

Meta 随后发布 Muse Wonder 系列，应用范围进一步延伸至专业服务。官方列举的用途包括保险理赔、租约理解，以及法律、财务和医疗相关应用。持续执行与跨对话记忆是此次更新的两项重点：Active Memory 保留不同对话中的上下文，Active Focus 则支持模型在任务执行过程中处理错误，并在必要时澄清问题、检查关键细节，继续推进工作。产品分发方面，Wonder 已在 Muse API 和 Muse Code 开放；Wonder Full 已向 Muse Code 订阅用户提供，后续还计划接入 Muse。

我们认为，跨对话记忆有助于减少重复提供背景资料的工作，持续执行能力则有望减少任务中途对人工接管的依赖，两者结合能够提升 Agent 处理复杂事务的便利性。

2.3 OpenAI 与 AWS 推出持续执行及托管产品

OpenAI 的 Dots 可在自己的电脑环境中跨应用执行任务。用户指定目标和允许自主操作的范围后，Dots 可在两次对话之间继续工作，完成后将结果交给用户审阅，遇到需要用户判断的问题时主动联系。目前，Dots 正逐步向符合条件的账户开放。

AWS 发布 Bedrock 托管代理预览版，由 AWS 与 OpenAI 联合开发，基于适配 AWS 环境的 OpenAI Agents API 定制版本构建。该服务负责保存任务状态、调用工具、执行代码及协调多个步骤，会话中保留消息、工具调用和中间结果。每个代理使用独立 IAM 角色，重要操作前可设置人工审批，受支持的 API 活动则通过 CloudTrail 留存记录。

我们认为，办公应用中的直接调用、跨对话的任务延续，以及云平台提供的权限和运行管理，有望减少企业接入 Agent 所需的开发工作，使模型更容易用于连续的业务流程。

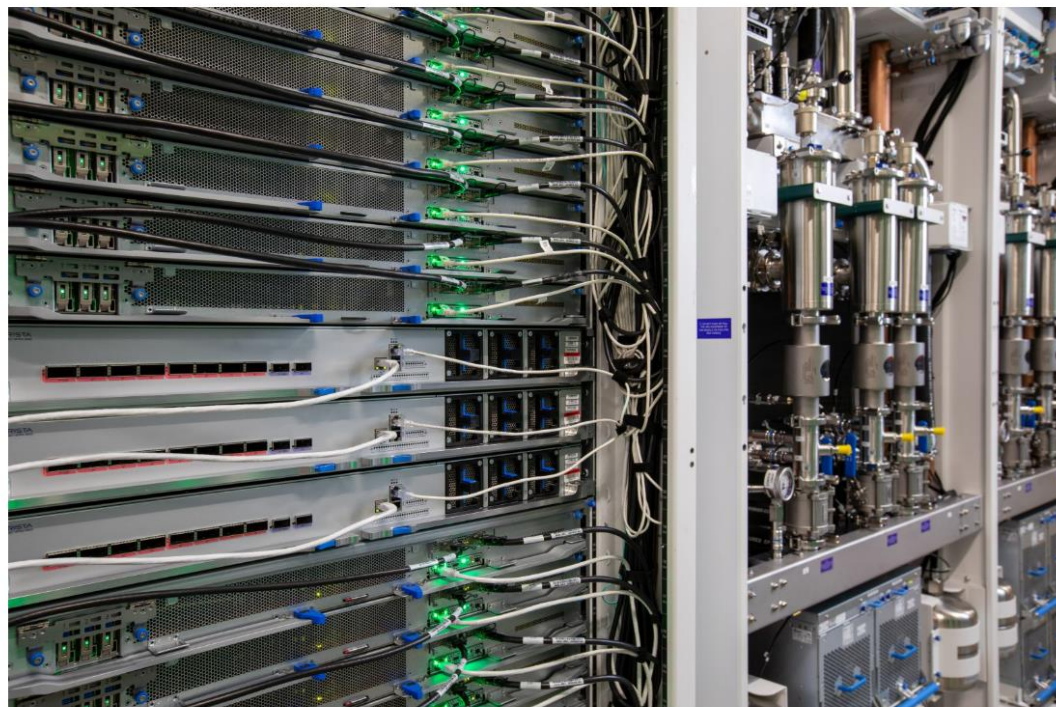


三、微软扩大自研推理芯片部署，AWS 业务高增长支撑 AI 投入

3.1 微软扩大 Maia 推理芯片部署

微软宣布，Maia 200 已用于 Microsoft Copilot 等产品的生产工作负载，Maia 300 计划于 2027 年推出。微软计划自 2027 年起将 Maia 系列扩展至多吉瓦规模，并使其成为微软产品与服务的主要推理系统。Maia 300 的架构以推理为重点，针对长上下文、内存密集和解码密集型任务进行优化。

图表2: Azure Maia 300 系统



来源：微软，国金证券研究所

我们认为，将自研芯片用于自身产品，有利于微软围绕具体模型和应用优化软硬件配置；若部署规模扩大，芯片研发及系统适配投入也有望通过更多推理任务分摊。

3.2 AWS 看好 AI 投入，业务增长提供需求支撑

据 Axios 报道，AWS 首席财务官认为，AI 基础设施及服务支出带来了实质性创新，相关投入具有可持续性。

此前披露的经营数据为此次表态提供了背景。亚马逊 2026 年第二季度业绩公告显示，AWS 收入同比增长 36.7%，为 18 个季度以来最快增速；AWS 的 AI 业务年化收入超过 250 亿美元，同比增长达到三位数。

模型服务的客户使用规模也在扩大。同一公告披露，已有数十万客户使用 Amazon Bedrock，过去六个月新增客户数超过产品推出后前两年的总和，第二季度客户支出超过此前所有季度合计。

投入端，亚马逊截至 2026 年 6 月末的过去十二个月自由现金流为净流出 76 亿美元，公司将其主要归因于物业及设备净购置支出同比增加 661 亿美元，并表示这一增量主要反映 AI 投资。

我们认为，AI 业务收入与模型服务消费的增长，有利于云厂商将新增算力转化为付费使用，为持续扩容提供需求支撑。我们认为，随着企业将 Agent 用于日常业务，多步骤执行、工具调用及任务重试有望增加推理需求，带动计算、存储和配套云服务的使用。我们认为，新增需求能否持续覆盖设备投入、折旧和运行成本，将影响扩容的回报；提高算力利用率、优化推理成本，有助于缓解投入增长对现金流的压力。



四、相关标的

国内算力：寒武纪、海光信息、长鑫科技、东阳光、中芯国际、华虹半导体、中科曙光、摩尔线程、华勤技术、利通电子、亿田智能、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、京基智农、浪潮信息、胜蓝股份、盛视科技、神州数码、润泽科技、大位科技、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子等。

海外算力：工业富联、中际旭创、东山精密、胜宏科技、方正科技、麦格米特、先导基电、光智科技、AXTI、新易盛、蓝思科技（元拾科技）、创世纪、海鸥股份、天孚通信、光迅科技、源杰科技、江海股份、生益科技、深南电路、广合科技、奥士康、世运电路、景旺电子、鹏鼎控股、沪电股份、华盛昌、联讯仪器、唯科科技、领益智造（立敏达）、中钨高新、英维克、飞龙股份、冰轮环境、德福科技、南亚新材、火炬电子、三环集团、海星股份、天岳先进、云南锗业、力量钻石、沃尔德、四方达等；英伟达、ALPHABET、台积电、Lumentum、SK 海力士、英特尔、Bloom Energy、美光、闪迪。

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、网宿科技等。

AI 应用：深信服、德才股份、迈富时、汇量科技、卓易信息、鼎捷数致、达梦数据、中控技术、金蝶国际、用友网络、海致科技集团、汉朔科技、金山办公、合合信息、福昕软件、金斯瑞生物科技、晶泰控股-P、英矽智能等。

五、风险提示

行业竞争加剧的风险：模型厂商若竞争加剧可能影响利润率。

技术研发进度不及预期的风险：模型能力迭代若不及预期可能影响大模型厂商商业化。

特定行业下游资本开支周期性波动的风险：云厂商、运营商等客户对 AI 服务器与算力基础设施投资力度波动风险。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；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，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国门内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究