



德邦证券
Topsperity Securities

计算机

行业投资评级 | 优于大市 (维持)

2024年11月27日

AI Agent：海外创新企业的再次自我革命



证券分析师

姓名：陈涵泊
资格编号：S0120524040004
邮箱：chenhb3@tebon.com.cn

证券分析师

姓名：李佩京
资格编号：S0120524090004
邮箱：lipj@tebon.com.cn

研究助理

姓名：王思
邮箱：wangsi@tebon.com.cn

目录

CONTENTS



- 01 AI Agent：底层技术愈发成熟，算力成本降低开启产业化篇章
- 02 产业进程：国内外科技厂商共振，场景化AI Agent落地加快
- 03 商业成效：Agent用户粘性增强，商业模式完善实现初步兑现
- 04 投资建议和风险提示



01

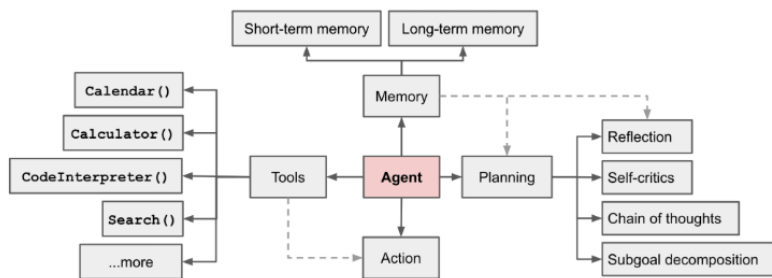
AI Agent：底层技术愈发成熟， 算力成本降低开启产业化篇章

- 1.1、AI Agent：感知环境，自主行动，更广泛的人机融合
- 1.2、大模型：实现复杂任务规划，推理的Scaling Law时刻
- 1.3、算力：国内外算力保障供给，Agent推理成本有望下降

1.1 AI Agent：感知环境，自主行动，更广泛的人机融合

- AI Agent是一种能够感知环境、进行决策和执行动作的智能实体，具备通过独立思考、调用工具去逐步完成给定目标的能力，可拆解为大模型+规划+记忆+工具使用四个组成部分。**例如，与每一步都需要人工输入的传统人工智能工具不同，AutoGPT代理可在后台持续工作。它们会根据所规划的目标开展行动，并可自主运行。
 - 大模型：**充当Agent的大脑,配合规划、记忆、工具高效完成任务。
 - 规划：**子目标和分解，Agent将大任务分解为较小、可管理的子目标，使其能够有效处理复杂任务；反思和完善，Agent可以对过去的行为进行自我批评和自我反思，从错误中学习，并为未来的步骤进行完善，从而提高最终结果的质量。
 - 记忆：**短期记忆，所有的上下文学习都是利用模型的短期记忆进行学习；长期记忆，这为Agent提供了在长时间内保留和回忆（无限）信息的能力，通常通过利用外部向量存储和快速检索来实现。
 - 工具：**学会调用外部API以获取模型权重中缺失的额外信息（经常在预训练后难以更改），包括当前信息、代码执行能力、访问专有信息源等。
- AI Agent带来更广泛的人机融合。**基于大模型的Agent不仅可以让每个人都有增强能力的专属智能助理，还将改变人机协同的模式，带来更为广泛的人机融合。智能体（Agent）模式中，人类设定目标和提供必要的资源（例如计算能力），然后AI独立地承担大部分工作，最后人类监督进程以及评估最终结果。

图表：LLM驱动下的AI Agent System概览



图表：人类与AI协同的三种模式



1.2.1 Agent性能依赖于推理能力，CoT为其重要手段

➤ AI Agent的性能依赖于推理能力，思维链（CoT）提示技术是提升这一能力的重要手段。

■就像人类一样，在从事复杂任务时，每一步之间往往会有一个推理过程。AI Agents也会借助ReAct组件（Reasoning and Acting），将大模型的推理能力和行为决策紧密结合起来，使语言模型可以根据知识进行有逻辑地计划安排。

■根据《Chain of Thought Prompting Elicits Reasoning in Large Language Models》，思维链提示策略可以有效提升大型语言模型的推理能力。通过在输入中加入一系列中间推理步骤作为提示，引导模型按照类似人类的思维方式进行推理，从而提高其在各种复杂任务中的性能，进行更深层次的逻辑推理和问题解决。

➤ Altman提出了AGI五级理论，推理能力提升是向L3智能体过度的重要环节。其中，AI Agent处于通过AGI的L3等级，智能体不仅能思考，还可以采取行动的AI系统，能够执行全自动化业务。L2等级所展现的强大推理和问题解决能力是达到Agent等级的重要环节，OpenAI o1模型为L2等级的代表。

图表：AGI五级理论

| 等级 | 内容 |
|----|---|
| L1 | 聊天机器人具有对话能力的AI，能够与用户进行流畅的对话，提供信息、解答问题、辅助创作等，比如聊天机器人。 |
| L2 | 推理者像人类一样能够解决问题的AI，能够解决类似于人类博士水平的复杂问题，展现出强大的推理和问题解决能力，比如OpenAI o1。 |
| L3 | 智能体不仅能思考，还可以采取行动的AI系统，能够执行全自动化业务。 |
| L4 | 创新者能够协助发明创造的AI，具有创新的能力，可以辅助人类在科学发现、艺术创作或工程设计等领域产生新想法和解决方案。 |
| L5 | 组织者可以完成组织工作的AI，能够自动掌控整个组织跨业务流程的规划、执行、反馈、迭代、资源分配、管理等，基本上已经与人类差不多。 |

图表：大模型引入CoT提高复杂任务准确性

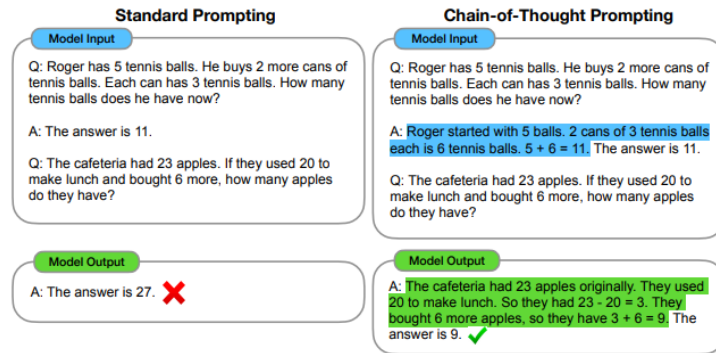
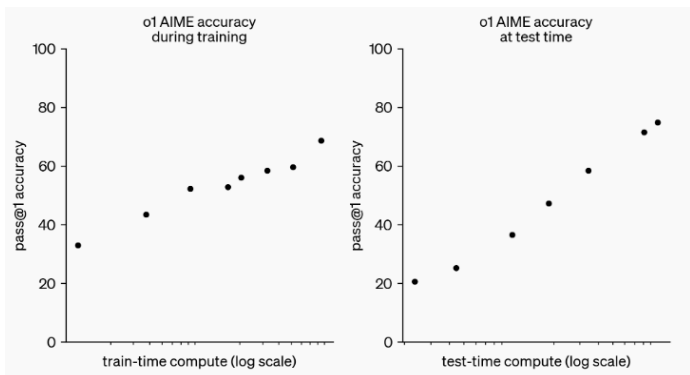


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

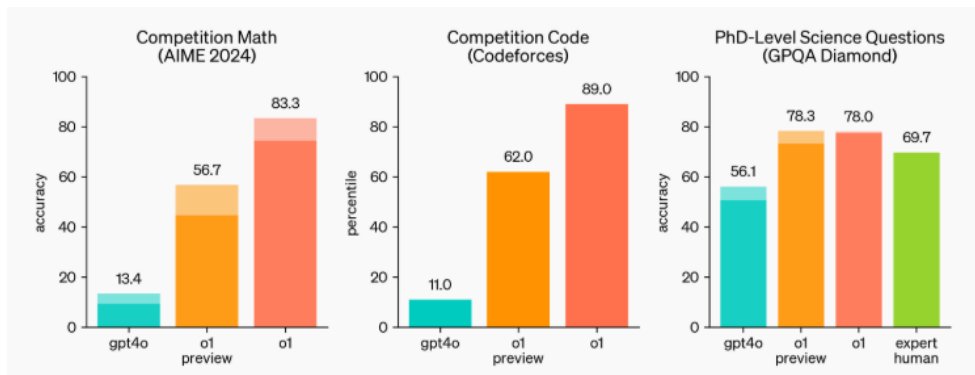
1.2.2 传统Scaling Law已放缓，o1掀起推理的划时代革命

- 受制于高质量文本数据越来越少等原因，传统的大模型Scaling Law已放缓。根据The Information，传统的大模型Scaling Law已经撞墙，OpenAI下一代旗舰Orion遭遇瓶颈，仅仅增加模型规模已经不能保证性能的线性提升，且伴随着高昂成本和显著的边际效益递减。路透社发文表示，由于当前方法受到限制，OpenAI和其他公司正在寻求通向更智能AI的新途径。
- OpenAI发布了全新的多模态Self-play RL模型o1，通过多步推理的方式，其在语言推理方面取得了傲人的成绩，开启了推理的Scaling Law时刻。
 - o1开创模型scaling的新维度，模型性能不再仅仅受限于预训练阶段，现在可以通过增加推理计算资源来提升模型表现。OpenAI发现，随着强化学习的增加（训练时间计算）和思考时间的增加（测试时间计算），o1的性能会持续提高。
 - o1模型实现了结构化推理，展现思维链（COT）能力，将单个任务拆解成多个简单任务，提升回答的逻辑性。它通过让模型在回答复杂问题时，逐步解释每一步的推理过程，而不是直接给出答案。因此模型在回答问题时就像是人类在解题时那样，先思考每一步的逻辑，再逐步推导出最终的结果。**具体来看，o1至少已经能够实现：提出假设，验证思路，反思过程这三种主要的逻辑推理能力。**
- 我们认为，大模型已从GPT 3.5的简单对话能力发展至o1的复杂任务规划解决能力，为AI Agent发展建立了技术根基，通过强化推理的方式迈开通过AGI的重要一步。

图表：o1 性能随着训练时间和测试时间计算而平稳提升



图表：o1 在具有挑战性的推理基准上大大优于 GPT-4o

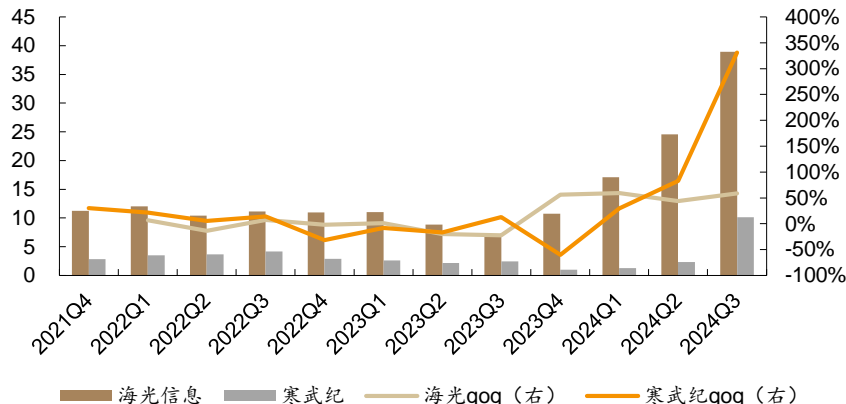


1.3 算力：国内外算力保障供给，Agent推理成本有望下降

国内外算力供需错配现象改善，算力部署成本有望下降，有利于Agent产业化推广。

- 1) **国外英伟达B卡明年有望加速出货。**根据英伟达FY25Q3最新季报，CEO黄仁勋表示，在经历过多次修改后，Blackwell芯片目前已全面投产；首席财务官Colette Kress表示，Blackwell芯片计划于本季度开始出货，并将在明年加速出货。
- 2) **国产芯片供应逐步向好。**以海光与寒武纪为代表的国产算力2024年存货、预付款表现均出现好转，或预示着国内AI芯片出货迎来好转。我们认为，华为昇腾、寒武纪思元和海光深算是国产算力的重要供给方，多轮迭代后可与英伟达主流芯片对标，基本能够满足小模型训练和下游应用的推理需求，国内算力供给好转有望满足Agent规模化推广推理算力需求。
- 从AI算力租赁成本来看，**租赁价格今年较去年缩水70%。**根据PANews，去年由于AI算力供应紧张，H100的租赁价格高达每小时8美元，但现在市场上算力供应过剩，价格降到了每小时2美元以下。目前，市场上H100的供应量远超需求，因此租用H100比购买更划算，投资购买新的H100已不再具有收益。

图表：海光信息和寒武纪存货（亿元）情况



图表：海光信息和寒武纪预付款项情况

| 预付款项水平 (亿元) | 2021Q4 | 2022Q1 | 2022Q2 | 2022Q3 | 2022Q4 | 2023Q1 | 2023Q2 | 2023Q3 | 2023Q4 | 2024Q1 | 2024Q2 | 2024Q3 | |
|-------------|--------|---------|--------|---------|---------|---------|--------|--------|----------|--------|--------|---------|--------|
| 海光信息 | 2.37 | 6.84 | 8.56 | 12.10 | 9.37 | 14.49 | 18.69 | 18.51 | 23.88 | 26.14 | 30.90 | 22.13 | |
| qoq | | 188.09% | 25.19% | 41.28% | -22.51% | 54.62% | 28.93% | -0.94% | 29.02% | 9.44% | 18.21% | -28.38% | |
| 寒武纪 | 0.85 | 1.06 | 0.90 | 1.78 | 0.08 | 0.08 | 0.08 | 1.33 | 1.48 | 2.05 | 5.50 | 8.54 | |
| qoq | | 91.94% | 24.34% | -14.74% | 96.67% | -95.34% | -0.53% | -7.99% | 1651.68% | 11.47% | 38.50% | 168.71% | 55.22% |

02

产业进程：国内外科技厂商共振， 场景化AI Agent落地加快

- 2.1、大模型能力增长遇瓶颈，国内外科技大厂转向AI Agent
- 2.2、软件开发门槛大幅降低，千行百业爆款Agent/应用将至
- 2.3、垂类Agent：行业know-how显现，下一代AI应用重要场景

2.1 大模型能力增长遇瓶颈，国内外科技大厂转向AI Agent

- **OpenAI、Google、Anthropic正在将注意力从模型转移至一系列AI Agent工具上，“贾维斯”未来将至。**
- **Anthropic**：10月23日，Anthropic发布了Claude 3.5 Sonnet升级版，推出革命性功能——**Computer use**，可通过API即可实现AI与PC交互，替代人力进行使用计算机控制鼠标、键盘，包括查看屏幕、移动光标、点击按钮和输入文本等操作。
- **Google**：根据The Information，Google正在开发一个专门针对Chrome浏览器进行了优化的AI Agent，其代号为「Project Jarvis」，将为用户执行任务，包括“收集研究资料、购买产品或预订航班”。Project Jarvis或在今年12月公布，将由 Gemini 2.0提供AI支持，只能在Web浏览器上运行。Project Jarvis在执行不同操作时，中间会有几秒钟的思考时间。此外，根据至顶网，谷歌云计划推出AI Agent合作伙伴计划以及新的AI Agent Space，其目的是让客户在谷歌的在线市场上轻松查找、购买和部署合作伙伴的人工智能代理，推动生成式人工智能销售和客户增长。
- **OpenAI**：据彭博社，OpenAI 正准备推出代号为“**Operator**”的新型AI Agent工具，计划于2025年1月发布。该工具能够在用户的计算机上自动执行任务，如编写代码或预订行程。OpenAI计划在明年1月发布该工具的研究预览版，并通过公司的应用程序接口（API）向开发者开放。

图表：Claude实现数据下载及分析的展示demo



图表：海外大模型龙头厂商AI Agent布局情况

| 公司 | 产品名称/发布/计划发布时间 | 功能表现 |
|-----------|---------------------------|---|
| Anthropic | Computer use 2024.10 | 可通过API即可实现AI与PC交互，替代人力进行使用计算机控制鼠标、键盘，包括查看屏幕、移动光标、点击按钮和输入文本等操作 |
| Google | Project Jarvis 2024.12 | 将为用户执行任务，包括“收集研究资料、购买产品或预订航班” |
| OpenAI | Operator 2025.01 | 能够在用户的计算机上自动执行任务，如编写代码或预订行程 |

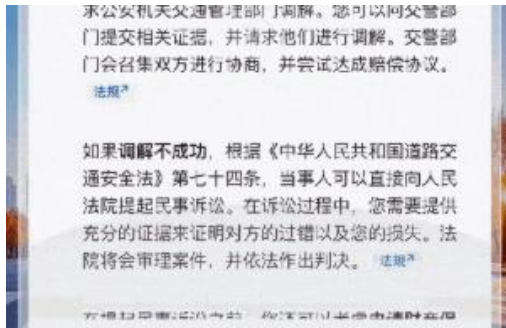
2.1 大模型能力增长遇瓶颈，国内外科技大厂转向AI Agent

- 国内智谱抢跑发布AutoGLM，百度、腾讯、钉钉等互联网大厂纷纷跟进。
- 智谱**：10月24日，智谱发布“Web Browser Use” AutoGLM-Web，一个能模拟用户访问网页、点击网页的浏览器助手。具体来看，1) 可以根据用户指令在私域网站上完成高级检索并总结信息；2) 模拟用户看网页的过程，进行批量、快速的浏览并总结多个网页内容；3) 结合历史邮件信息回复邮件。10月25日，智谱发布了AI手机端的“phone use” AutoGLM，只需接收简单的文字/语音指令，它就可以模拟人类操作手机，通过一句话的方式手机便可自动点外卖、发微信、写点评、给朋友圈评论。**根据科创板日报，目前，智谱的AutoGLM可适配微信、淘宝、美团、小红书、大众点评、12306、携程、高德地图共8款应用。**
- 百度**：11月12日，百度世界2024现场，李彦宏分享了百度在过去一年的技术创新最新成果，并发布了两大AI应用方向：智能体和产业应用。李彦宏现场展示了4种不同类型的智能体，包括公司类智能体，角色类智能体，工具类智能体和行业类智能体。例如行业类智能体中，大会现场展示的法律行业智能体—法行宝，它可以帮助用户解答各类法律相关的问题。上线半年多以来，法行宝为940万人提供了高效、可信赖的法律服务，已经回答了1660万个问题，成为每个人免费的AI律师。
- 钉钉**：11月13日，钉钉迎来重磅升级：AI深入业务场景，钉钉面向企业关键场景推出“精选AI助理”，首批已上线工单助理、Excel助理、法务助理等6大AI助理。同时，钉钉联合生态伙伴发布行业 AI 解决方案，覆盖制造、医疗、零售、教育等行业。
- 腾讯**：11月15日，腾讯出品的ima.copilot（以下简称ima）正式上线了Windows版本。ima由腾讯混元大模型提供技术支持，是面向学习、办公场景，以知识库为核心的AI智能工作台，是搜、读、写一体的效率工具。

图表：智谱AutoGLM通过语音与美团APP互动



图表：百度法行宝帮写一封起诉状



图表：腾讯ima首页：搜、读、写三大功能板块



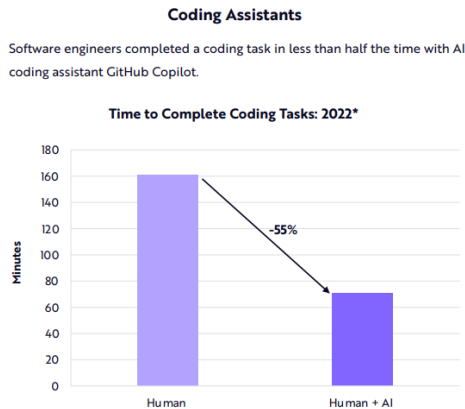
2.2 软件开发门槛大幅降低，千行百业爆款Agent/应用将至

➤ 人人编写软件时代或即将来临，软件开发门槛大幅降低。

- 根据ARK，在代码编程领域，AI作为代码助手能够帮助人类减少55%的编程时间。此外，AI编程不仅提高了开发效率，还将降低对专业开发者技能的需求，未来人人都是开发者成为可能，使得软件开发更加民主化，更多的人有机会和能力参与到软件开发中。
- 根据Gartner预测，到2025年，70%的应用程序将由低代码和无代码开发平台构建，其中80%的应用程序由非IT工作者开发而成。

➤ 千行百业爆款智能体应用有望陆续涌现。我们认为，软件开发门槛的降低有助于涌现大量智能体应用，而具备客户真实场景，致力于解决客户实际需求的厂商有望率先重塑人与技术互动方式，打造出现象级的AI应用，从而占据价值链当中的核心位置。随着AI Agent变得越来越易用和高效，“Agent+”的产品越来越多，未来AI Agent有望成为AI应用层的基本架构，涵盖toC和toB产品等不同领域。目前，零售与电子商务、教育、房地产、人力资源、金融服务等多领域AI智能体应用已落地发布。

图表：AI作为代码助手，缩减55%编码时间



图表：千行百业AI Agent陆续涌现

| 行业 | 项目/产品 | 说明 |
|--------|-----------------------|---|
| 零售与电商 | Alexa | 亚马逊的Alexa是一个AI智能体，可以推荐产品，下订单和跟踪发货。可以为用户提供个性化推荐，改善供应链管理，增强客户体验。 |
| 教育 | Aida | 为学生提供反馈，并建议个性化的学习路径。 |
| | Jagoda.ai | Jagoda.ai是一款人工智能驱动的在线辅导工具，可为广泛的学科提供个性化帮助，包括数学、生物学、化学等。有了Jagoda.ai，你就可以轻松解决问题。 |
| 房地产 | Epique | epiquAI工具使用人工智能生成房地产博客文章、房产描述和电子邮件活动。它还可以生成Instagram 报价并提供经纪人建议和法律服务。 |
| | PropertyPen | 人工智能驱动的房地产列表作家，可以生成属性描述，撰写房地产清单，以及自动化房地产营销。 |
| 旅游和酒店 | Trava | Trava 是一家 AI旅行社，可以帮助用户计划旅行、搜索酒店和预订餐厅等。它提供了入门提示，并提供了有关旅行的建议。 |
| 客户支持 | Ebi.Ai | 人工智能聊天机器人是基本且有限的。EBI的 AI助手。从第一天起，AI就可以处理您 85% 的日常查询。只需几分钟即可完成设置。 |
| 金融 | Intelligent Portfolio | 根据客户的投资目标创建和管理投资组合。 |
| 制造和供应链 | Predix | AI实时监控机器，以预测和防止设备故障 |
| | Autopilot | 帮助自动驾驶车辆，以及停车、变道和安全驾驶。 |

场景

- **随着企业数字化进程不断发展，部分企业已积累了大量的数字文档、资料等，AI Agent是实现了行业know-how的具象化。** 公司运营积累的丰富经验、专业知识及行业know-how对于企业来说非常重要，竞争对手难以复制。
- **AI Agent通过挖掘企业流程、文档及相关信息化知识完成知识库的具象化。** 专家知识的数字化是企业AI Agent成功的关键，通过不断标注和反馈，专家的个人知识可以丰富完善知识库，形成知识闭环，使系统能够持续学习并提升性能。

图表：企业数字化转型过程当中积累的know-how

| 行业know-how | 详细内容 |
|------------|---|
| 技术窍门 | 指的是企业在生产过程中掌握的特定技术，这些技术可能未被广泛知晓，是企业保持竞争力的关键。 |
| 操作技巧 | 涉及日常操作中的技巧和方法，这些技巧可能来自于员工的个人经验或企业多年的实践。 |
| 工艺流程 | 企业在生产过程中形成的特定工艺流程，这些流程可能经过多次优化，以达到提高效率和质量的目的。 |
| 商业策略 | 企业在市场竞争中形成的特定商业策略，包括市场定位、产品定价、营销推广等。 |
| 管理方法 | 企业在管理过程中形成的有效管理方法，如人力资源管理、财务管理、供应链管理等。 |
| 市场洞察 | 企业对市场的深入理解和预测，包括消费者行为、市场趋势、竞争对手分析等。 |
| 知识产权 | 虽然知识产权通常被视为正式的知识，但它们也是know-how的一部分，因为它们保护了企业的创新成果，如专利、商标、版权等。 |
| 企业文化 | 企业的价值观，行为准则和工作氛围，这些文化因素可以影响员工的行为和企业的决策。 |
| 客户关系 | 企业与客户建立的长期关系，包括客户信任、忠诚度和口碑等。 |

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/328026066130007004>