



2024 年深度行业分析研究报告



目 录

1. 原理解析：思维链铸就智能体，多体交互拓展应用.....	6
1.1 Agent 模式架构解析.....	6
1.2 单智能体 vs 多智能体.....	7
1.3 Prompt 提示词的艺术.....	8
2. 应用场景：赋能两类实体领域，成本与效益的博弈.....	8
2.1 自动化类：微软智能体 AutoGen	9
2.2 情感需求类：陪伴型智能体.....	11
2.3 效能讨论：成本与价值间的博弈	12
3. 未来展望：人工智能发展迅猛，智能体商业化落地	13
3.1 海内外政策推动人工智能发展加速	13
3.2 算力规模高增，企业对 AIGC 的投入意愿强.....	15
3.3 多模态智能体有望实现大规模商业化	17
3.4 全球智能体市场增速加快	18
4. 潜在标的：Agent 企业千帆竞发，垂类应用引人期待.....	19
4.1 LeewayHertz: 提供多领域解决方案	19
4.2 Markovate: 提供多模态智能体解决方案	21
4.3 Replika: 情感类智能体热度不减.....	2
4.4 CharacterAI : 自研大模型，打造智能体社区.....	23
4.5 星火智能体：布局智能体平台，发布垂类智能体	24
4.6 SkyAgents: 自研模型降低成本	25
4.7 商汤：联合清华、上海人工智能实验室打造GITM	26
4.8 字节跳动：豆包聊天助手.....	27
4.9 腾讯：AppAgent.....	28
5. 投资建议	31
6. 风险提示	32

图目录

图 1	人类与 AI 的交互模式	6
图 2	AI Agent 智能体工作原理	7
图 3	单智能体强化学习原理图	7
图 4	多智能体强化学习原理图	7
图 5	CrewAI 整体框架	8
图 6	AutoGen 运作模式图解	9
图 7	Agent 交流模式展示	10
图 8	AutoGen 实验数据对比	11
图 9	林间聊愈室产品介绍	12
图 10	OpenAI 不同模型 tokens 消耗价目表	12
图 1 1	单次运行 Agent 对 Tokens 的消耗	13
图 1 2	单次运行 Agent 的实际成本	13
图 1 3	中国人工智能市场规模预测	14
图 1 4	全球人工智能市场规模预测	14
图 1 5	美国增强型人工智能研发投入2015-2030	15
图 1 6	美国及其 AI 盟友	15
图 1 7	中国智能算力规模及预测	16
图 1 8	企业 AI Agent 使用情况	16
图 1 9	专注于横向应用的 AI 代理初创企业率先获得融资	17
图 2 0	多模态大模型发展	18
图 2 1	全球自主人工智能和智能体市场规模	18
图 2 2	中国前五大人工智能市场应用	19
图 2 3	LeewayHertz 公司发展历程与合作公司	19
图 2 4	LeewayHertz 的智能体产品架构	20
图 2 5	LeewayHertz 为私募股权提供的 AI Agent 解决方案	20
图 2 6	Character.ai 智能体架构	21
图 2 7	ShopSpot 用户界面展示	21
图 2 8	ShopSpot 库存管理系统	21
图 2 9	ShopSpot 多模态智能体图片识别功能展示	22
图 3 0	Replika 付费模式	22

图 3 1	截止2024年8月 Replika 各年龄段用户占比	23
图 3 2	截止2024年8月 Replika 用户全球分布	23
图 3 3	Character .AI 用户界面	23
图 3 4	Character.ai 2024年7月每日访问量	24
图 3 5	Character. AI 各年龄段用户占比.....	24
图 3 6	Character . AI用户全球分布	24
图 3 7	科大讯飞星火智能体介绍.....	25
图 3 8	部分星火AI 智能体展示	25
图 3 9	天工大模型能力展示	26
图 4 0	天工Agent 计费模式.....	26
图 4 1	RL架构对比GITM 创新架构	27
图 4 2	解锁科技树的成功率(蓝色是GITM).....	27
图 4 3	各模型完成任务成功率对比	27
图 4 4	豆包智能体的对话界面	28
图 4 5	腾讯AppAgent展示软件操作.....	29
图 4 6	AppAgent 的 App 探索学习阶段	30
图 4 7	AppAgent 的实际部署阶段	31

表目录

表 1	AI Agent 类型盘点以及具体应用.....	9
表 2	国内政策推动人工智能发展.....	13
表 3	海外政策推动人工智能发展.....	15

1. 原理解析：思维链铸就智能体，多体交互拓展应用

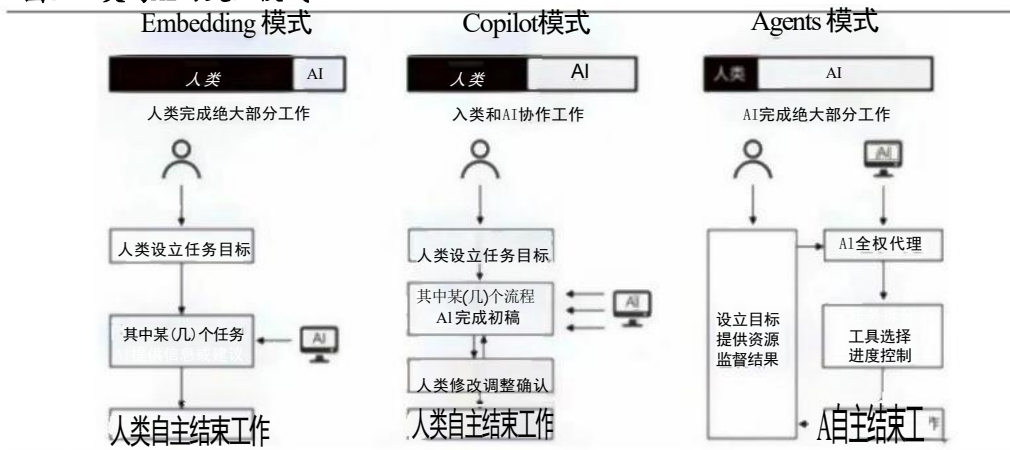
早在上世纪50年代，阿兰图灵把“高度智能有机体”扩展到了人工智能。如今随着大模型的快速发展，这个概念又被重新拾起。大模型成为了智能体目前最完美的载体，有望完成从概念到实际应用的蜕变。

用户在Agent (智能体)模式中给AI 设置目标和身份，并提供 Prompt (提示词)。AI自主拆分任务、使用工具、完成工作，用户仅负责设立目标、提供工具资源和监督结果。OpenAI 定义的智能体具有长期和短期记忆、自主规划、工具使用和自动执行任务的能力，能提高工作效率和用户体验。另外，智能体也分为单智能体和多智能体。单智能体通过试错学习在单一环境中行动，追求最大奖励，多用于简易任务。多智能体在博弈环境中行动，追求长期累积奖励，多用于复杂测试。

1.1 Agent模式架构解析

Agent有效减少人类工作总量，人与AI协作才是最终形态。人类与AI交互可大致分为三种模式。Embedding 模式中大模型可以填补一些信息缺失，完成少量子任务，例如总结信息等等。用户最终会整合挑选AI提供的信息，并自主完成任务。Copilot 模式则更加智能化，AI 可根据用户设定的流程去执行任务。例如让AI根据写一段稿件或者根据需求编程，但其对Prompt 的要求也更高。在AI完成流程后，用户需要对内容结果进行调整并自主结束工作。Agent 智能体模式的AI参与度更高，但也不是完全由AI代理。用户需要给AI设计一个目标和身份，以及需要使用的工具。配上更为复杂的Prompt, AI能自主进行任务拆分，使用工具并结束任务。用户只负责设立目标、提供资源、监督结果。

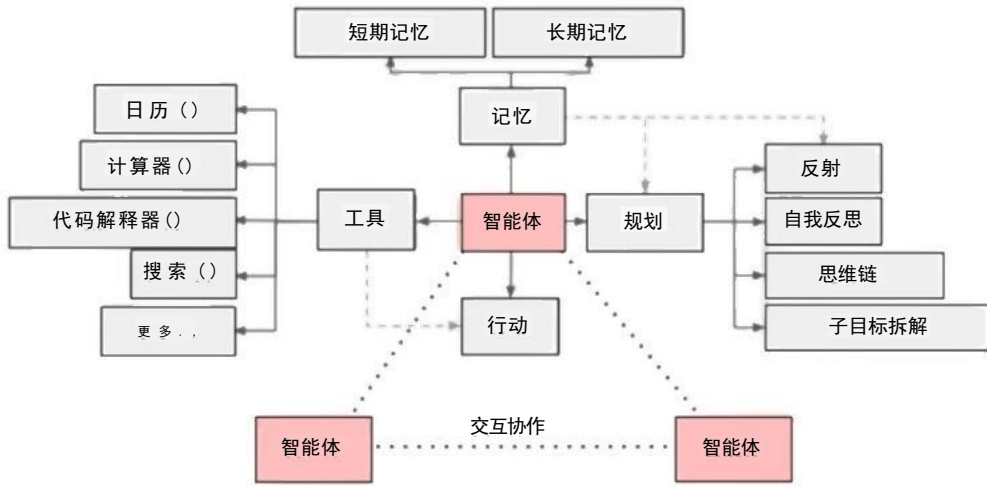
图1 人类与AI的交互模式



资料来源：腾讯研究院，海通证券研究所

以 LLM 为核心，四模块铸造AI Agent。从 OpenAI 的定义来看，智能体以大语言模型为核心，其拥有长期和短期记忆、自主规划能力、能自动化执行复杂任务、能够使用工具等四个特点。1) 记忆模块：智能体像人类一样，能留存学到的知识以及交互习惯等，这样的机制能让智能体在处理重复工作时调用以前的经验，从而避免用户进行大量重复交互。短期记忆适用于所有上下文的学习，类似平常我们与ChatGPT沟通的模式；长期记忆则保留知识和交互回忆，例如智能体在特定行业积累的大量数据和经验，则能提供更专业、更具深度和个性化的回答，提升用户体验。2) 规划模块：将复杂任务分解成子目标并逐一解决，完成任务后进行反思总结。例如反思自己大量输出重复内容或在单一子目标耗时过长等问题，将经验存入长期记忆以规避类似错误。3) 工具模块：智能体可利用工具来弥补自身短板，通过调用外部API来实现功能拓展。例如调用连接互联网的API去搜索实时信息。4) 行动模块：智能体会形成完整的计划流程。例如先读取以前工作的经验和记忆，之后规划子目标并使用相应工具去处理问题，最后输出给用户并完成反思。

图2 AI Agent智能体工作原理



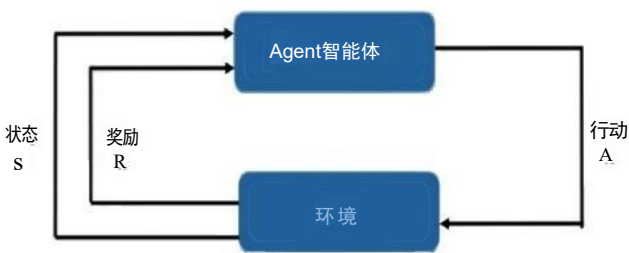
资料来源：腾讯研究院，海通证券研究所

1.2 单智能体 vs 多智能体

单智能体与多智能体各具优势，适配于不同垂直领域。单智能体的强化学习原理是基于马尔可夫决策来完成的，简单来说可以分为状态集 S 、行动集 A 、奖励 R ，下一时刻的状态和奖励只与上一时刻的行动有关，与更早之前的状态无关。其模型原理就是让智能体用试错的方式来学习，若某个策略能得到奖赏，则智能体产生该行为的策略就会加强。其目的就是在单一环境中行动，尽可能得到最大的奖励。应用领域目前也较为广泛，例如赛车游戏中连续动作的训练：控制方向盘、油门、刹车等动作，可由DDPG、A3C、PPO 算法来决策。一些离散动作的训练例如围棋智能体 AlphaGo，可通过 Q-Learning 等算法决策。

多智能体的决策不仅与自身行动相关，还与系统内其他智能体的行动所关联。一个多智能体系统中会有两个以上的智能体，他们一般存在着合作或竞争关系。这样模型称为马尔科夫博弈，其状态转换符合马尔可夫决策，关系符合博弈。在多智能体模型中每个智能体的目标是找到最优策略来使它在任意状态下获得最大的长期累积奖励。由于其模型更为复杂，干扰因素较多等原因，目前多智能体模型商业化产品较少。

图3 单智能体强化学习原理图



资料来源：CSDN，海通证券研究所

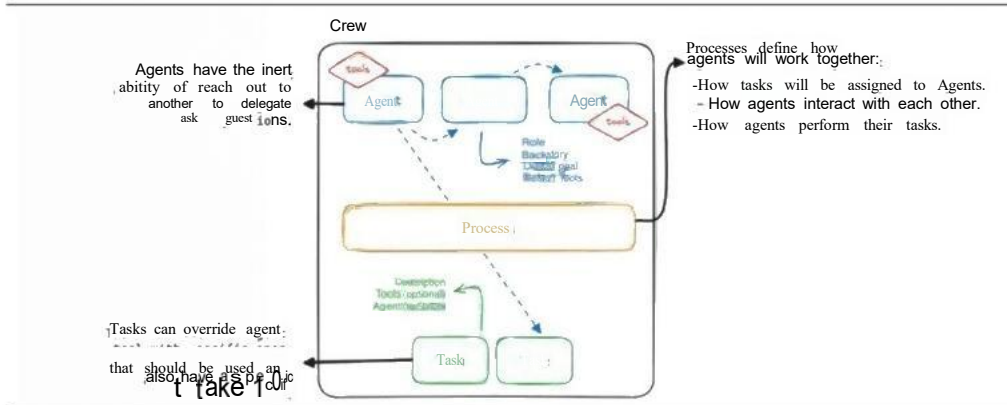
图4 多智能体强化学习原理图



资料来源：CSDN，海通证券研究所

CrewAI 是世界领先的多智能体框架之一，在多智能体领域用于协调角色扮演型自主 AI 智能体。通过促进协作智能，CrewAI 使智能体能够无缝协作并处理复杂任务。在编写程序时，用户需要赋予每一位 Agent 角色、任务、以及背景故事。

图5 CrewAI整体框架



资料来源: CrewAI, CSDN, 海通证券研究所

1.3 Prompt提示词的艺术

Prompt 是一种基于自然语言处理的交互方式，它通过机器对自然语言的解析，实现用户与机器之间的沟通。可以理解为是一段给定的文本或语句，用于启动和引导机器学习模型生成类型、主题或格式的输出。在自然语言处理领域中，Prompt 通常由一个问题或任务描述组成，例如“帮我写一篇有关人工智能的文章”，“帮我编写一段可以整理数据的程序”等等。

一个优秀的 Prompt 应该包含1)明确目标：能够定义任务，以及想要达到的效果；2)具体指导：给予模型明确的指导和约束，举一些例子或明确回复边界；3)简洁明了：使用简练、清晰的语言表达Prompt;4) 迭代优化：一次 Prompt 输出可能达不到想要的结果，用户需要根据智能体的回答来改变提示词，持续调整优化。

2. 应用场景：赋能两类实体领域，成本与效益的博弈

AI Agent目前的应用大多都在概念层面，但随着大模型竞争加快、政策鼓励研发投入、更多企业参与AI研究等因素，应用层面的AI Agent推进速度加快。智能体大致可以分为六类，根据他们被设计出的特点，可以作用在不同的应用领域上。不同类别的智能体给予应用层面上更多研发方向，像目前关注度较高的自动驾驶技术、智能电网控制、能源管理等都能被垂类智能体覆盖。结合多模态大模型，自动化和情感需求类智能体已落地。但商业化智能体仍需考虑成本问题，由于智能体之间的交互过程可能出现错误循环且输出结果不一定符合需求，tokens 成本远高于普通LLMs。

表1 AI Agent类型盘点以及具体应用

AI Agent类型	具体描述	应用领域
基于效能的Agent	基于效用的Agent在具有多种潜在结果的复杂决策环境中表现出色--比如平衡不同风险以做出投资决策，或者权衡价格与成本。	金融交易、价格追踪系统、智能电网控制
基于目标的Agent	基于目标的Agent会评估和规划行动，以实现它们的目标。与其他类型的智能Agent不同的是，它们能够将前瞻性 and 战略规划结合起来，朝着特定的结果前进。	扫地机器人、项目管理软件、电子游戏
基于模型的Agent	适用于并非总是可见或可预测的信息时。这个模型会根据环境中传入的数据不断更新，因此人工智能Agent可以对环境中未见部分进行推断，并预测未来的状况。	自动驾驶技术、自动灌溉、智能家居
分层Agent	分层Agent类似于复杂的组织结构，具有不同的决策层次。系统中的不同Agent拥有不同的专业领域，使它们在处理复杂的多步骤任务时更加高效。	空中交通管制系统、自主仓储机器人
机器人Agent	增加了性能元素，这些人工智能Agent的实体通常配备有摄像头或触摸传感器等传感器。这种Agent在危险或重复性很高的任务中尤其有用--让人工智能Agent来完成这些任务可能更有效率，也更符合成本效益。	生产线机器人、手术机器人、农业机器人、服务机器人
多Agent系统	每个Agent都是半自主运行的，但其设计目的是与其他Agent互动，形成一个动态的生态系统，从个体行为中产生集体行为。对于这种Agent程序来说，协作是关键。	空中交通管理系统、用于能源管理的智能电网、供应链与物流

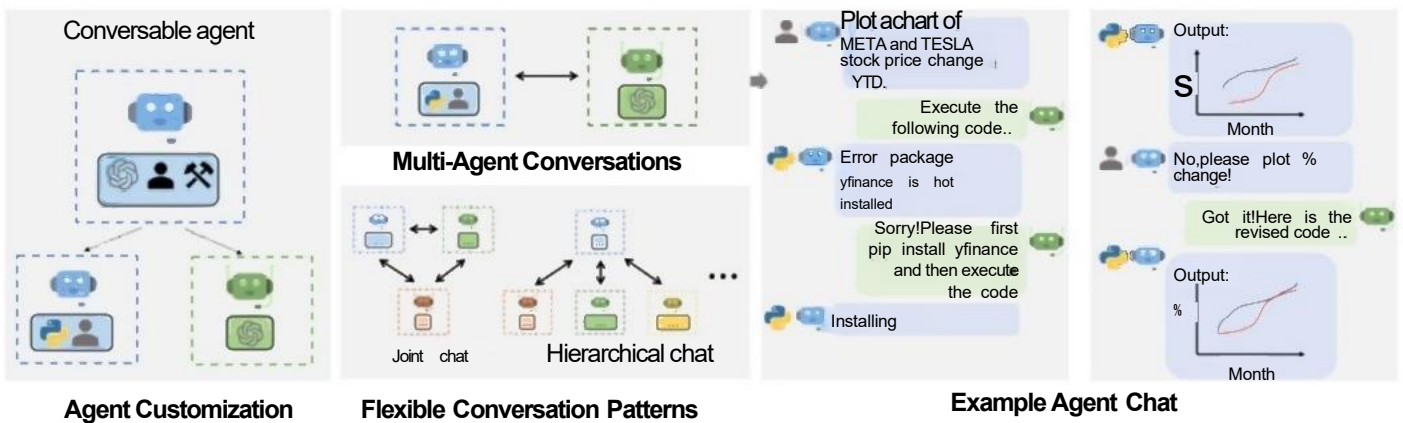
资料来源: botpress, 海通证券研究所

2.1 自动化类：微软智能体 AutoGen

微软研发的AutoGen 是一种多智能体沟通的模型，这些智能体通过对话完成任务。实验发现这种多智能体沟通模式可以有效提高AI解决问题的准确性以及成果的完整性。使用AutoGen，开发者可以灵活定义智能体交互行为。微软的研究证明了该框架在包括解决数学、编码、问题解答、运筹学、在线决策、娱乐等问题上的有效性。

AutoGen 需要两个以上的智能体形成交流模式，其中智能体可以由大语言模型LLMs、人类、工具来扮演。当一个需求被提出，智能体会开始灵活交流，他可以是多个智能体向需求人汇报的模式(Hierarchical Chat)或者是像研讨会一样随意讨论(Joint Chat)。在官方给出的示例中，智能体之间的沟通确实减少了人工交互，其有能力自己修复出现的问题并给出满意的答案。

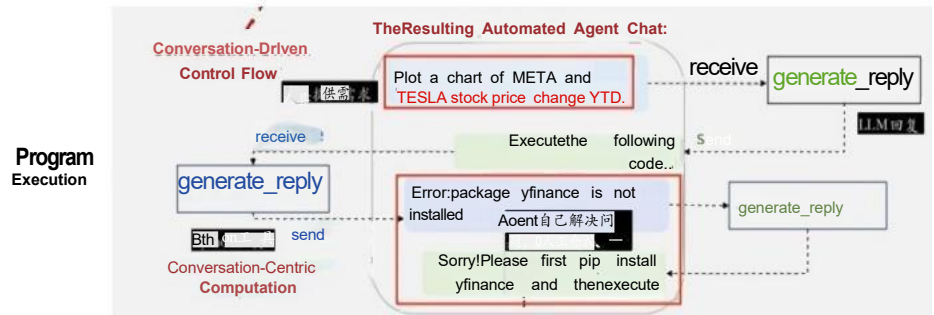
图6 AutoGen运作模式图解



资料来源: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, CSDN, 海通证券研究所

在AutoGen 智能体程序中，人类充当需求发起者，智能体通过不断使用工具并在相互交流中解决陆续出现的问题从而得到最终答案。在智能体交流过程中，人类可随时介入其中，若用户对最终结果不满意，可以重新提出需求，智能体就会继续优化答案。

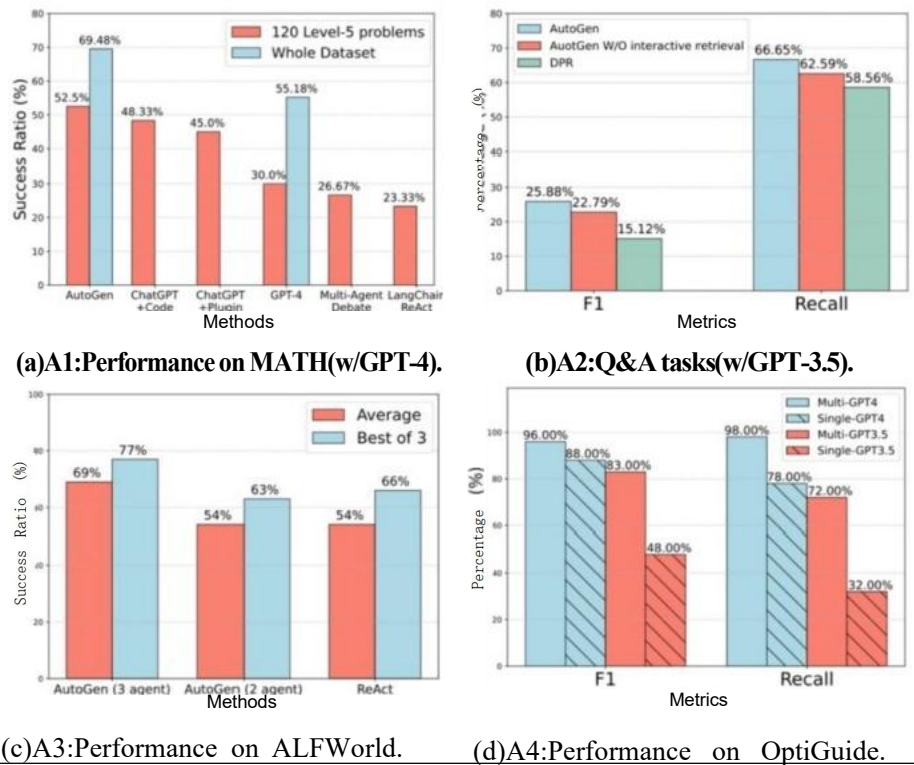
图 7 Agent 交流模式展示



资料来源: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, CSDN, 海通证券研究所

AutoGen在四项实战应用的表现测试中均胜过单一语言模型，实验发现智能体能增加**AI**完成任务的成功率以及模型的稳定性。1) 数学计算：分别测试了**AI**在完成整个数学题库和随机**120**道题目的表现，**AutoGen**的成功率远超各类语言模型。2) 数据库检索问答：先建立一个**RAG**数据库，然后对智能体进行问答测试，看其能否追踪到正确的答案。其**F1-score**越高，说明模型更稳定；**Recall**体现了分类模型对样本的识别能力。**Recall**越高，说明模型对样本的识别能力越强。从数据看出纯**AutoGen**模型的表现略高于其他组合。3) **ALFWorld**(Adaptive Learning Framework World) 研究：**ALFWorld**是专门用于研究和开发智能体的仿真环境，在其虚拟世界中，包含了各种场景、任务和智能体需要完成的挑战。**ALFWorld**实验发现3个智能体的成功率高于2个智能体，说明随着交互的智能体变多，智能体间互相纠错的能力有所提升。但对于**AutoGen**的实验也提到多智能体可能会导致错误循环(智能体之间反馈错误但一直无法解决)，这也是降低成功率的因素之一。4) **OptiGuide**: 是专门设计来通过自然语言处理，解读和指导复杂供应链优化决策。在用不同语言模型运行**OptiGuide**的时，发现多智能体协作模式的成绩显著高于单智能体。其中多**GPT4**智能体协作运行**OptiGuide**时，**F1-score**和**Recall**分别能达到**96%**和**98%**。

图8 AutoGen实验数据对比



资料来源：Enabling Next-Gen LLM Applications via Multi-Agent Conversation, 海通证券研究所

2.2情感需求类：陪伴型智能体

智能体另一大应用领域就是提供拟人情感陪伴类服务。陪伴类的智能体拥有长期记忆，让用户与其聊天中产生熟悉且亲切的感觉。随着大模型不断迭代，陪伴类智能体更加立体，能提供更高的情绪价值。目前国内情绪消费类市场潜力较大，在如今紧凑的生活节奏之下，国民孤独感增加，而陪伴类智能体或受益于这样的大环境。根据GIA(Global Industry Analysts)的研究，Denny Yin提出该行业规模预计到2024年增长到1000亿美元，到2026年增长到2000亿美元。

情感沟通类智能体上线，收获大量用户好评。在大模型出现之前，一些利用心理学CBT沟通技术的产品已经在市面上广泛存在，例如Weobot、wysa等。随着大模型的发展和迭代，智能体已经能被打造成心理咨询师或塔罗师的角色。用户通过倾诉生活中的烦恼，并在与智能体沟通中寻求情绪宣泄。目前一款治愈系的软件“林间聊愈室”上线后获得了大量好评，App Store评分4.7，多数用户给出了五星好评。其中产品使用小动物的角色设定降低了用户的戒备心，加上治愈的画风和场景设计，打造独特的用户体验。

图9 林间聊愈室产品介绍



资料来源: App Store, 海通证券研究所

2.3效能讨论：成本与价值间的博弈

打造智能体并不是0成本，大语言模型的token 消耗就是最直观的成本展示。由于智能体的运行机制内包含反思和深度学习，这使得其在token 消耗方面比传统大语言模型高。相比于单智能体，多智能体又多了一层交互端的消耗，当多个智能体讨论且进入错误循环中，无效成本将会显著提升。所以在效能方面，对智能体进行成本讨论以及优化策略十分重要。

在商用智能体中，更先进的大语言模型代表更好的用户体验，但随之而来的是极速上升的成本。目前GPT4 的价格是GPT3.5 turbo的十倍，而且由于智能体之间交互的特性，在输出过程中往往伴随着更高的token 消耗。若最终输出的长度超出模型token 限制，则不得不调用更高版本的模型例如GPT4-32K，成本又是GPT4 的两倍。

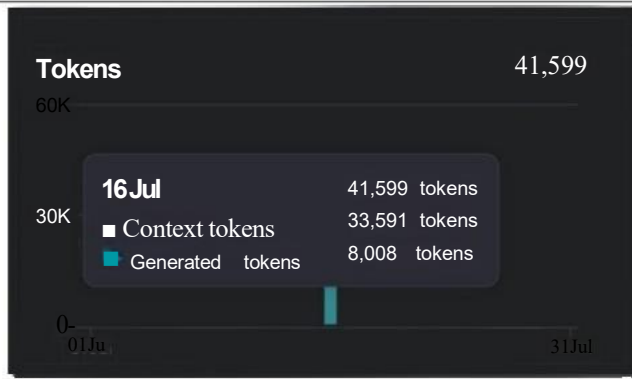
图10 OpenAI 不同模型tokens消耗价目表

Model	Pricing
gpt-4o	US\$0.0050/1K input tokens
	US\$0.0150/1Koutput tokens
Model	Pricing
gpt-3.5-turbo-0125	US\$0.00050/1Kinput tokens
	US\$0.00150/1Koutput tokens

资料来源: OpenAI, 海通证券研究所

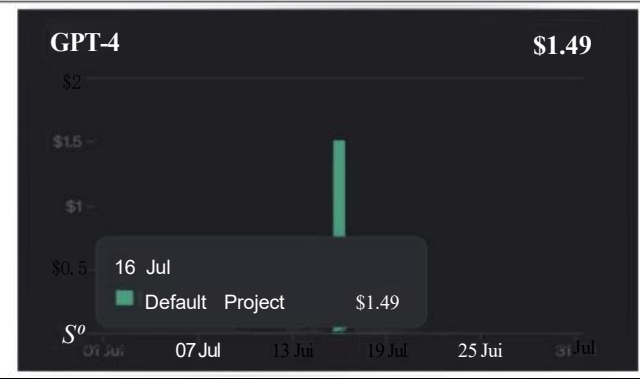
在实际测试中，使用搭载GPT4 的智能体模型查找新闻并进行总结梳理。这部分的单次消耗是42000个 tokens，成本是1.5美元。在实际应用中，假设一天需要统计的股票新闻大约有120家公司，则成本为35.8美元，按照汇率1:7.14折合人民币255.5元。而且智能体并不是完美的，在运行过程中出现幻觉或错误循环也是常事。例如智能体在找不到合适新闻的时候会进行大量重复的内容输出，人工介入纠错又会增加一定成本。

图11单次运行Agent 对 Tokens 的消耗



资料来源: OpenAI, 海通证券研究所

图12单次运行Agent 的实际成本



资料来源: OpenAI, 海通证券研究所

商业化智能体在成本和最终产品中会有一些博弈，若将智能体运用到股票交易等对精准度需求极高的行业中，产品或引入更多智能体进行协作以提高模型的稳定性，则成本上升。如果运行成本高于其创造的价值时，该智能体模型就失去了商业化的能力。目前多数智能体应用仅停留在测试阶段，市场中的用户需要更理性的看待智能体商业发中能带来的效益。

3. 未来展望：人工智能发展迅猛，智能体商业化落地

未来多方面推动人工智能发展，应用级别智能体有望快速落地。国内各地相继出台关于人工智能的发展政策，推动其为重要的研究方向。预计2026年国内人工智能市场规模超过260亿美元，全球人工智能市场规模2025年超6万亿美元。海外以美国为例，相关政策出台时间较早，人工智能领域发展更加成熟，许多智能体应用已在服务各类企业。并且美国有意与人工智能强国组成战略伙伴，共同发展AI科技。

多模态大模型利用异构数据提升应用效率，促进AI Agent发展。将智能体赋能于图片+语音的多模态大模型中，可以减少使用工具和交互的过程，使智能体完成更多复杂任务，解决跨行业、跨领域的问题。智能体发展能推动政府、金融、制造、能源、医疗、零售等行业的智能化应用向多模态和跨模态转变。

3.1 海内外政策推动人工智能发展加速

多地陆续出台关于人工智能发展的指导文件，2024年政府与企业的投入资金有望迎来大幅增长。国务院在《2024政府工作报告》中提出，要开展“人工智能+”行动，实施数字化转型。智能体作为人工智能应用领域的重要研究方向，相信众多企业和研究部门会开展更多研究工作。另外，像上海、杭州、广东、北京等省市也出台了相关政策，其中杭州市推出“算力券”约2.5亿用于支持本地的算力发展。广东省也将算力作为重点发展方向，预测2025年的人工智能核心产业规模超过3000亿元。

表2国内政策推动人工智能发展

政府机构	文件名称	发布时间	具体内容
上海市经济和信息化委员会、上海市发展和改革委员会等五部门	《上海市推动人工智能大模型创新发展若干措施(2023-2025年)》	2023年10月	实施大模型创新扶持计划。支持引进高水平创新企业，支持本市创新主体打造具有国际竞争力的大模型，鼓励形成数据飞轮，加速模型迭代，对取得重大成果的予以专项奖励。实施大模型示范应用推进计划。重点支持在智能制造、生物医药、集成电路、智能化教育教学、科技金融、设计创意、自动驾驶、机器人、数字政府等领域构建示范应用场景，打造标杆性大模型产品和服务。
国务院	《2024政府工作报告》	2024年3月	深化大数据、人工智能等研发应用，开展“人工智能+”行动，打造具有国际竞争力的数字产业集群。实施制造业数字化转型行动，加快工业互联网规模化应用，推进服务业数字化，建设智慧城市、数字乡村

杭州市人民政府

《关于支持人工智能全产业链高质量发展的补充意见》
(征求意见稿)

2024年5月

1) 支持算力技术攻关。鼓励企业开展算力关键技术研发攻关, 市级科技研发计划给予专项支持。鼓励企业承担国家、省重点研发任务, 加速形成重大科研成果, 对符合政策规定的项目按国家、省实际到账资助经费的25%给予资助, 资助金额最高不超过500万元。2) 加大融资贴息支持。对纳入政府统一布局的算力基础设施建设项目(总投资规模不低于5000万元)3) 扩容算力券发放。每年设立“算力券”总额提升至2.5亿元, 新增算力券支持企业购买符合杭州市规划布局的智算中心以及纳入算力撮合的合作伙伴所供给的智能算力。

广东省人民政府办公厅

《广东省关于人工智能赋能千行百业的若干措施》

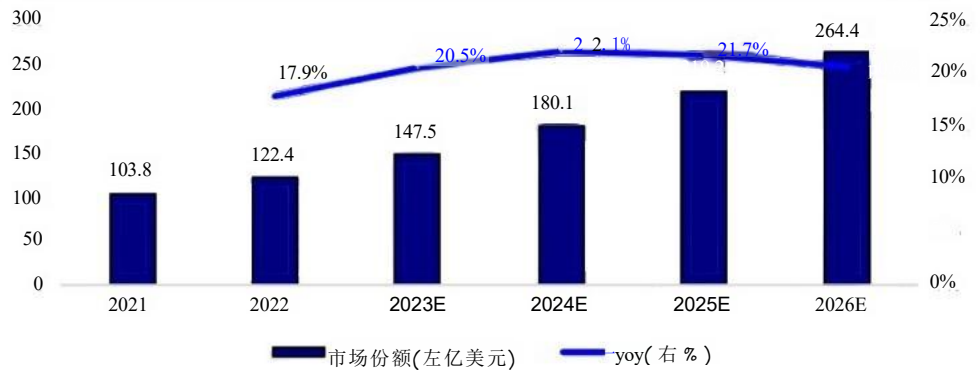
2024年6月

到2025年, 全省算力规模超过40EFLOPS, 人工智能核心产业规模超过3000亿元。到2027年, 全省人工智能产业底座进一步夯实, 算力规模超过60EFLOPS, 全国领先的算法体系和算力网络体系基本形成; 智能终端产品供给丰富, 在手机、计算机、家居、机器人等八大门类, 打造100款以上大规模使用的智能终端产品, 人工智能核心产业规模超过4400亿元; 聚焦制造、教育、养老等领域, 打造500个以上应用场景, 各行各业劳动生产率显著提升。

资料来源: 上海市经济和信息化委员会, 国务院, 杭州市人民政府, 广东省人民政府, 中国政府网, 新华社, 海通证券研究所

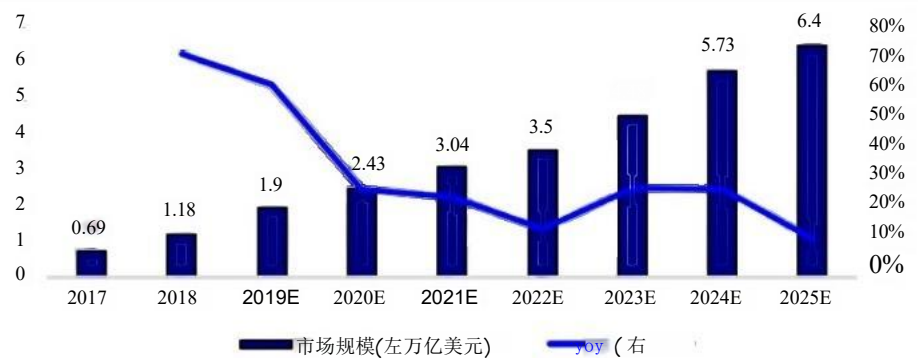
中国人工智能市场份额稳定增长, 未来AI 实体项目增多利好智能体发展。根据IDC预测, 国内人工智能市场规模预计在2026年达到264.4亿美元, 2021年到2026年期间的CAGR超20%。根据增长率计算, 从2022年开始, 每年市场规模会保持20%左右的稳定增长态势。根据澎湃新闻援引德勤统计, 未来2025年世界人工智能市场规模将超过6万亿美元, 2017年-2025年期间复合增长率将超30%。AI高速发展中或利好更多实体企业, 专门提供智能体产品服务的厂商规模也将快速扩张。

图13中国人工智能市场规模预测



资料来源: IDC, 海通证券研究所

图14全球人工智能市场规模预测



资料来源: 德勤, 澎湃新闻, 海通证券研究所

美国相关政策落地较早, 人工智能发展已成规模。在2021年3月, 大模型还未进入公众视野时, 美国人工智能国家安全委员会已经发布了对该领域未来发展的规划。文件中写到了美国坚定在人工智能领域的投入, 且规定了研发投入资金的比重。另外, 在2023年5月, 美国白宫发布《国家人工智能研发战略计划》, 对之前提出的8项重点战略目标进行了优化和完善, 同时强调了人工智能领域国际交流的重要性。另外, 文件还

对比了2020年发布计划中的各项指标与实际指标差距，并针对细分领域加强研发投入。

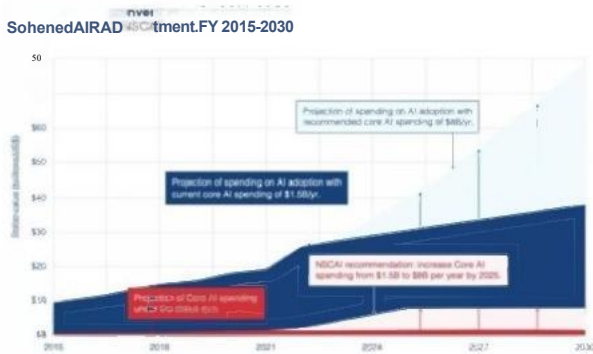
表3海外政策推动人工智能发展

政府机构	文件名称	发布时间	具体内容
美国人工智能国家安全委员会	《NSCAI最终报告》	2021年3月	美国国家人工智能安全委员会 (NSCAI) 的任务是向总统和国会提出建议，以“推进人工智能、机器学习发展以全面解决美国国家安全和国防需求的相关技术。”在设立本委员会时，2019财年约翰·S·麦凯恩国防授权法案第1051节指示NSCAI从国家竞争力的角度审视人工智能及其维持技术优势的手段、国际合作和竞争力的趋势、促进基础和先进研究投资的方式、劳动力和培训、军事使用的潜在风险、伦理问题、建立数据标准和激励数据共享，以及人工智能的未来发展等方方面面。
美国白宫	《国家人工智能研发战略计》	2023年5月	该计划是对2016、2019年版《国家人工智能研发战略计划》的再次更新，重申了之前的8项战略目标并对各战略的具体优先事项进行了调整和完善，同时增加了新的第9项战略以强调国际合作。此外，该报告还提出要评估联邦机构对《2020年国家人工智能倡议法案》(NAIIA)和《国家人工智能研发战略计划》的实施情况。

资料来源：清华大学人工智能国际治理研究院，安全内参，海通证券研究所

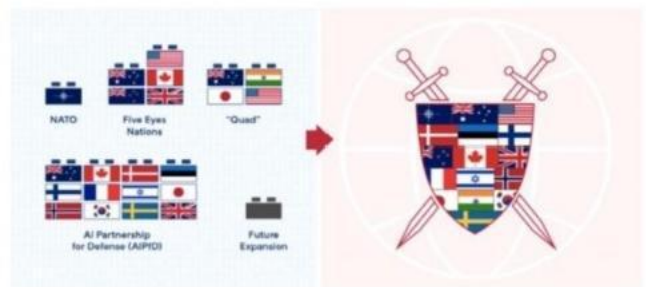
美国增强型人工智能投资研发保持高增速，AI 战略联盟强大。2024年美国在该方面的投入预计为15亿美元，NSCAI 希望在2025年把这一数字提高到80亿美元，增幅高达433%。且在2025年之后，保持每年投入增长80亿美元的趋势。最高预计在2030年投入高达350亿美元。美国强调了AI 盟友的概念，与澳大利亚、加拿大、英国、日本等国家构建人工智能战略伙伴关系，未来将会吸引更多AI强国加入联盟当中。

图15美国增强型人工智能研发投资2015-2030



资料来源：清华大学人工智能国际治理研究院，海通证券研究所

图 16 美国及其 AI 盟友



资料来源：清华大学人工智能国际治理研究院，海通证券研究所

3.2算力规模高增，企业对 AIGC 的投入意愿强

2021年中国的智能算力规模为155.2 EFLOPS（每秒百亿亿次浮点运算）。据IDC预测，规模将在2026年达到ZFLOPS 级别(每秒十万亿亿次浮点计算)。预计在2021-2026年期间，智能算力增长的CAGR为52.3%。随着算力规模的大幅增长，将会带动AI技术快速发展。根据中国信通院的测算，每在算力中投入1元，将带动3-4元的经济产出。在智能体发展中，更强的算力能支持更复杂的模型计算并且给予其更快的响应速度。当大模型能快速迭代，AI Agent将会有更强大的基座，相同运行模式下或能显著提升F1-score和 Recall 水平。

图17 中国智能算力规模及预测

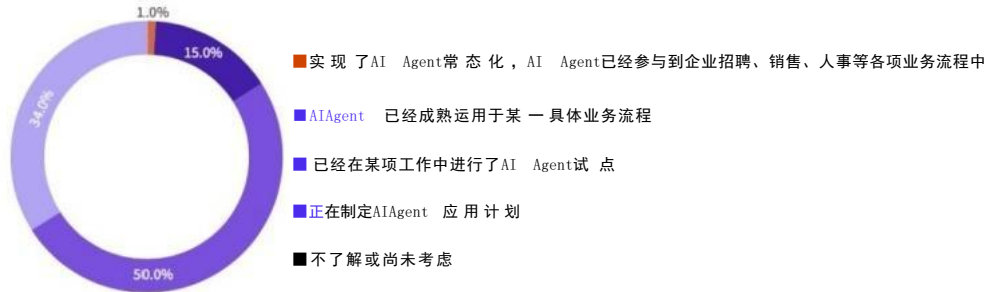


资料来源：新华网、经济日报、国家自然科学基金委员会，IDC 与浪潮信息《2022-2023 中国人工智能算力发展评估报告》，海通证券研究所

根据 IDC 统计，2023 年已有 34% 的企业在 AI 中的投入已见成效，AI 能力正在稳步发展。AI Agent 目前多为企业自研，仅有少部分企业将 Agent 融入真正工作中，若 AI Agent 能在未来用低成本撬动高效益，则会引来越来越多的企业重视。定制化 AI Agent 或成为一片蓝海市场，拥有成熟技术和一揽子解决方案的企业能占据大量市场份额。

图1 8 企业 AI Agent 使用情况

图 1 3 企业 AI Agent 的应用情况



来源：IDC, 2023

资料来源：IDC 与钉钉《2024 AIGC 应用层十大趋势白皮书》，海通证券研究所

风险投资青睐的智能体初创企业聚焦于横向应用领域，广泛涉及销售、客户支持及企业级与常规生产力工作流程优化。当前市场上专注于垂直行业特定应用的初创公司仍属少数，大部分代理技术开发的新兴企业主要面向企业客户，而非个人消费者。这些 Agent 初创公司希望用技术取代 SDR、软件工程师、合规分析专家、客户服务专员等多个职位的职能。这一趋势的影响力将极为广泛且深远，不仅限于日常工作中对公司员工的增加或取代，更会改变公司面向外界客户的交互形象，或使企业的运营方式与客户体验发生改变。

图19专注于横向应用的 AI代理初创企业率先获得融资



AI agent startups focusing on horizontal applications take funding lead

Disclosed equity funding & deals since 2022(as of 7/29/2024)



Note excludes OpenAI, excludes funding to companies developing agents with varying degrees of autonomy

CBINSIGHTS

资料来源：点滴科技资讯公众号，海通证券研究所

3.3 多模态智能体有望实现大规模商业化

多模态大模型能利用大量异构的数据资源提升应用的效率和能力上限，同时也利好 AI Agent 发展。多模态的概念或能提高智能体的工作效率，例如将其赋能于一个能形成图片+语音的多模态大模型中，单个智能体能完成更多复杂的任务，有效的减少智能体数量和任务中的交互次数。多模态智能体有望解决跨行业、跨领域的复杂问题和各类长尾场景。

目前，研究开发的核心领域已经包括多模态信息识别与理解技术以及群体智能技术，这些将有望加速人工智能从感知向认知的转变。这些技术成果还能够解决其他单项技术领域(如自然语言处理和计算机视觉)所面临的瓶颈问题，例如视觉领域的盲区和遮挡问题。此外，自动驾驶的激光点云、时空感知与测绘信息等更多类型的模态数据也将被整合到大模型中。未来，政府、金融、制造、能源、医疗、零售等多个行业的智能化应用将逐步转向多模态和跨模态的形式。

图 20 多模态大模型发展

金融	政府	制造	能源	医疗	零售
风险评估和信贷审查：通过分析文本和数值数据来预测客户的信用风险	公共服务自动化：为公众提供信息查询、办事指南等服务	生产流程优化：分析工厂数据以优化生产效率	能源消费预测：预测未来的能源需求和消费模式	诊断辅助：分析医疗图像和患者数据以协助医生诊断	库存管理和预测：预测产品的销售趋势和库存需求
自动化客服：为用户提供及时的财务咨询和支持	政策模拟与预测：预测政策实施后的社会和经济影响	产品设计辅助：基于市场和用户反馈生成新的产品设计建议	设备维护预测：预测能源设备何时需要维护或更换	药物研发：预测新药物的效果和副作用	个性化推荐：基于用户的购买历史和喜好为其推荐产品
投资策略优化：分析大量数据，为投资者生成投资建议或预测市场走势	数据分析和可视化：自动生成关于公共问题的报告和可视化内容	供应链管理：预测供应链中的潜在问题或延迟	能源存储和分发优化：基于数据分析来优化能源的存储和分发	治疗建议生成：基于患者的历史和病情生成治疗建议	客户服务自动化：为客户提供自动化的购物咨询和支持
欺诈检测：通过分析交易模式和行为来识别可疑的交易	舆情分析：分析公众对政府政策和行动的态度和反应	设备维护预测：预测设备何时需要维护或更换	可再生能源管理：预测和优化太阳能和风能的产出	患者监控和预警：实时分析患者的生理数据以预警其健康状况	市场趋势预测：分析市场数据以预测未来的零售趋势
报告自动生成：为管理层或监管机构自动生成财务和业务报告	智能监控和安全：通过分析视频和音频数据来提高公共场所的安全	质量控制自动化：通过分析图片和视频数据自动检测产品缺陷	环境监测：通过分析视频、图片和传感器数据来监测环境和设备状况	医疗文献自动生成和分析：自动生成或分析医学研究报告和文献	虚拟试衣和产品展示：使用AR和VR技术为用户提供虚拟的产品体验

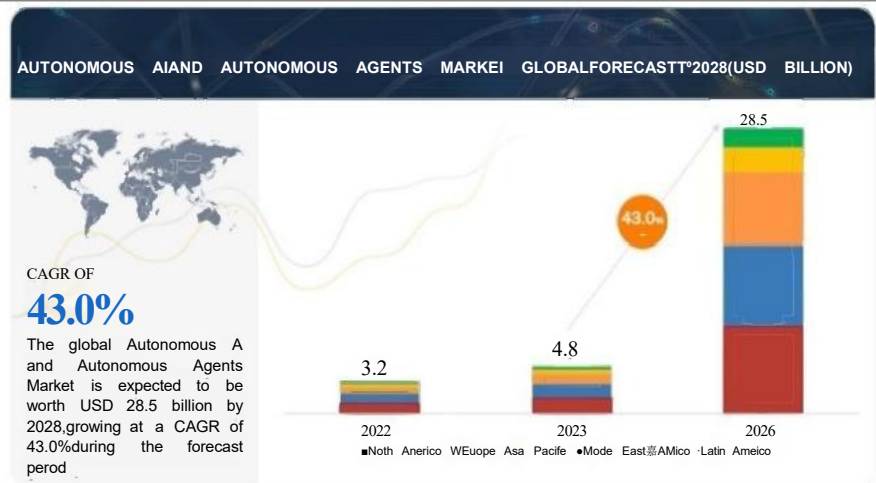
模态： ■ 文本 ■ 图片 ■ 语音 ■ 视频 ■ 数值数据

资料来源：IDC 与钉钉《2024 AIGC 应用层十大趋势白皮书》，海通证券研究所

3.4 全球智能体市场增速加快

根据marketsandmarkets的数据显示，2019年全球自主智能体 (Autonomous Agent) 市场份额为3.45亿美元，预计2024年将达到29.92亿美元，期间CAGR 为54%。机构预测2023年全球自主人工智能和自主智能体的市场规模在2028年能达到285亿美元，期间CAGR 为43%。智能体的需求增长迅速，企业希望用到最先进的AI 科技来实现快速创收，未来智能体软件公司有很高发展潜力。

图21 全球自主人工智能和智能体市场规模



资料来源：marketsandmarkets, 海通证券研究所

各行业用户正积极探索生成式 AI 的应用场景，开发数字化产品和服务，挖掘数据要素的潜在价值，并尝试智能化商业模式。根据 IDC 的数据，软件 and 信息服务、银行以及通讯行业在人工智能方面的投资最为突出，预计到2027年，这三大行业的投资占比分别为23.8%、9.7%和9.4%。IDC 的调查显示，近一半(43%)的受访组织正在研究生成式AI 的潜在应用，其中2023年有55%的金融机构和电信公司投资了生成式AI技术。这些技术正在帮助金融和电信行业提升反犯罪和监控能力，提供个性化的投资建议，并减少人工成本。

图22中国前五大人工智能市场应用

中国前五大人工智能市场应用场景占比预测，2024

- 智能基础设施调配
- 增强的智能客服
- 销售流程推荐和增强
- 智能业务创新和自动化
- 增强的公共安全和应急响应
- 其他

资料来源：IDC, 海通证券研究所

4. 潜在标的：Agent 企业千帆竞发，垂类应用引人期待

4.1 LeewayHertz: 提供多领域解决方案

LeewayHertz 公司在人工智能Agent开发方面取得了长足进步，以创建定制化人工智能解决方案的专长而闻名。目前已在医疗保健、金融和物流等多个领域成功实施了人工智能智能体应用。他们致力于提供量身定制的人工智能解决方案，这使LeewayHertz站在了智能体应用的最前沿。其中公司持续与各类行业的头部企业达成协作，提供各类数字解决方案，商业模式较为成熟。

图23 LeewayHertz公司发展历程与合作公司



资料来源：LeewayHertz 官网，海通证券研究所

其公司的智能体业务旨在通过自动化任务和简化流程来提高运营效率，从而扩大业务成果。公司使用 AutoGen Studio、Vertex AI Agent生成器和 crewAI 等一流工具，根据业务需求量身打造以任务为中心、高性能且可靠的 AI Agent。利用 LLM 的高级功能和全面的技能库，LeewayHertz 的人工智能Agent 在研究、分析、代码生成、审查、审计、在线搜索和细分等各种任务中表现出色。

图24 LeewayHertz的智能体产品架构

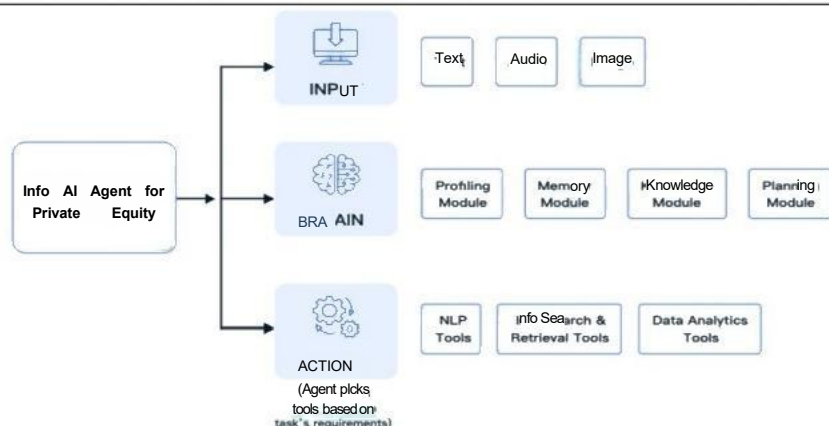


资料来源：LeewayHertz 公司官网AI Agent产品界面，海通证券研究所

公司提供了一系列AI Agent产品解决方案，例如私募股权智能体。这些智能体是能够分析和解释复杂数据的智能助手，帮助私募股权公司：1) 获得洞见：AI 智能体可以在大型数据集中发现隐藏的模式和趋势，帮助私募股权公司识别有前途的投资机会或潜在风险。2) 做出更好的决策：AI 智能体生成的洞见使得投资决策更加知情和数据驱动。3) 提高效率：通过自动化重复性任务，AI 智能体让人类分析师可以专注于更高价值的活动，如建立关系和战略规划。

智能体通过战略决策进行推理并行动，提升私募股权公司的投资成果和运营效率。公司为私募股权打造智能体的架构分为三个层面。首先智能体捕获和处理与私募股权相关的各种输入，包括财务数据、市场趋势、投资者情绪和监管更新。它可以处理文本、数字和图形格式的输入，以便为战略决策和行动提供信息。其次，认知功能(大脑)内又包含了：a) 角色定义：确定 Agent 在私募股权操作中的角色，使任务与投资策略和客户目标保持一致。b) 记忆：存储和检索过去的投资决策、市场行为和客户互动，促进自适应学习和知情决策。c) 知识：存储私募股权操作中关键的领域特定信息，如财务分析、风险评估模型和监管框架。d) 规划：根据当前市场状况、投资组合目标和监管要求制定投资策略和运营计划。最后执行层面中，智能体使用预测分析、投资组合管理软件和合规自动化工具来执行计划策略。这确保了投资策略的精确实施、投资组合表现的优化以及私募股权操作中对监管标准的遵守。

图25 LeewayHertz 为私募股权提供的AI Agent解决方案

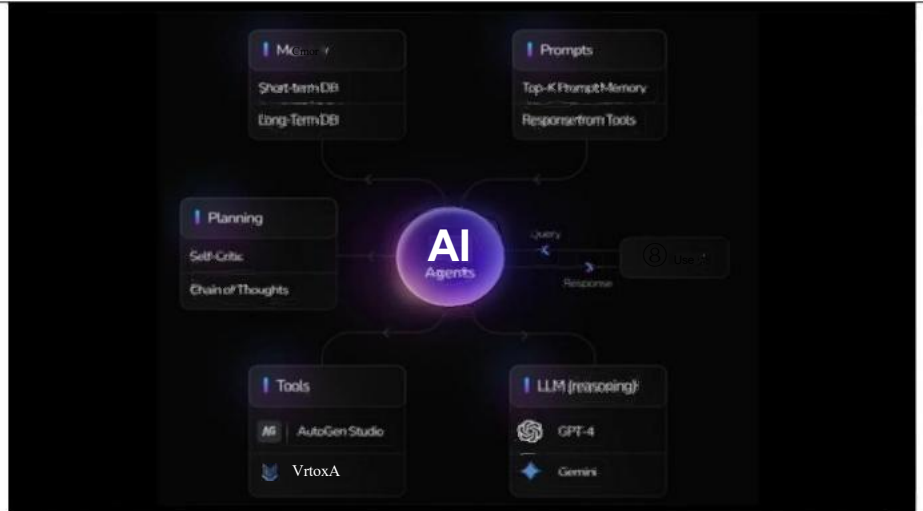


资料来源：LeewayHertz 公司官网Insight 界面，海通证券研究所

4.2 Markovate: 提供多模态智能体解决方案

Markovate 公司为企业开发人工智能Agent，以提高业务流程、决策和运营效率。这些智能体是先进的数字助理，可自动执行任务、分析数据和提供见解，并与企业系统无缝集成。Markovate 使用 AutoGen Studio、Vertex AI 和 GPT-4o 等工具定制智能体，以满足特定的业务需求。

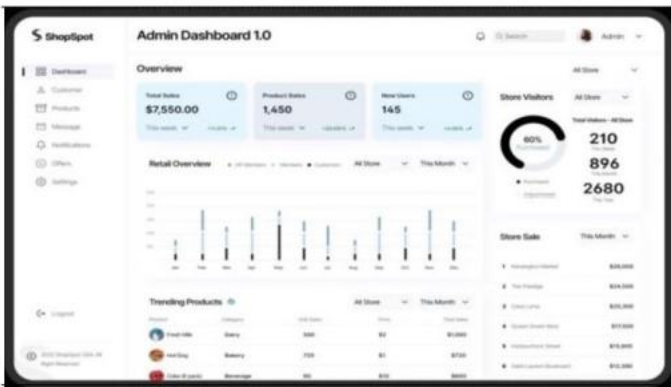
图26 Character.ai 智能体架构



资料来源: Markovate公司官网, 海通证券研究所

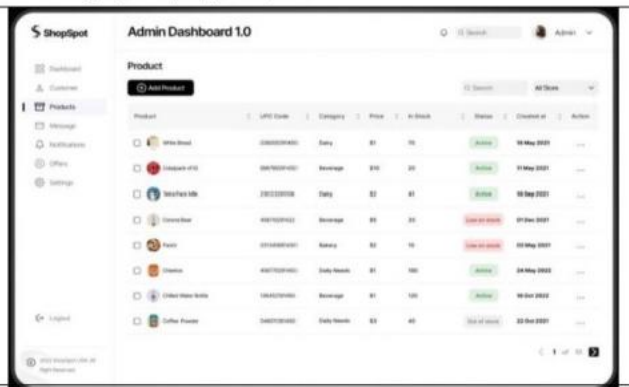
Markovate 打造了一款由智能体集成驱动的虚拟助理零售库存管理解决方案 ShopSpot。结合AI 智能体提供了彻底改变零售业的库存管理的方案，通过整合先进的机器学习和神经网络，它可以提供精确的需求预测和库存水平优化，有效地应对现代零售业的挑战。在整个解决方案中包含了几个重要模块：1) 聊天机器人和智能体：使用大量的客户互动数据对这些智能体模型进行训练，确保智能体能够提供准确和有用的回复，从而提高用户参与度和效率，开发能够理解和响应各种库存相关查询的人工智能模型。2) 动态数据可视化：结合了先进的图形工具，设计了动态数据可视化功能，用户可以通过热图和散点图等可视化工具来直观地探索和理解复杂的库存数据。3) 聚类分析：设计聚类分析功能涉及创建算法，根据购买模式和行为的相似性对产品或客户进行分组，为不同的细分市场量身定制库存策略。

图27 ShopSpot 用户界面展示



资料来源: Markovate公司官网, 海通证券研究所

图28 ShopSpot 库存管理系统



资料来源: Markovate公司官网, 海通证券研究所

多模态智能体通过结合图像识别系统，分析产品图像、货架状况和仓库环境，有助于快速评估库存水平、识别损坏情况并有效地进行产品分类。

图29 ShopSpot 多模态智能体图片识别功能展示



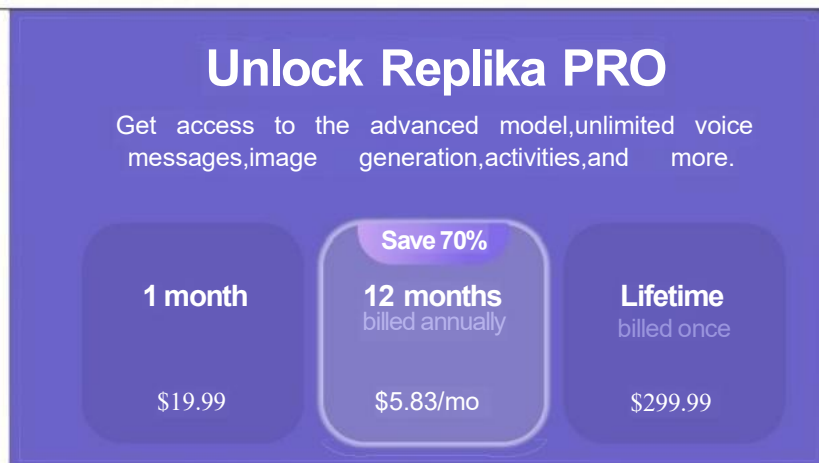
资料来源: Markovate公司官网, 海通证券研究所

4.3 Replika: 情感类智能体热度不减

Replika 是一款陪伴类智能体, 她的创始人将自己与朋友大约8000条聊天记录输入谷歌的神经网络, 创建了一个 AI 智能体, 让用户获得更加拟人的交流体验。创始人表示约有40%的用户用这款智能体创造恋人角色。2017年3月Replika 首次公测, 吸引了10万名用户参与, 同年9月正式面向大众, 在不到一年的时间里, 积累了250万用户。

截止到2023年1月, Replika 的产品全球流水可以达到200万美元/月, 在安卓系统已累计1000万次下载, 用户平均每天会与其中的智能体聊天70条左右, 花费时间2-3小时。目前用户满意度为92%, 有超过一半的人愿意为后续产品付费。其付费模式分为按月付款(19.99美元), 按年付款(69.96美元), 以及终身会员(299.99美元)。开通Replika Pro之后可以解锁更高版本的语言模型、无限的语音消息、图片生成等。

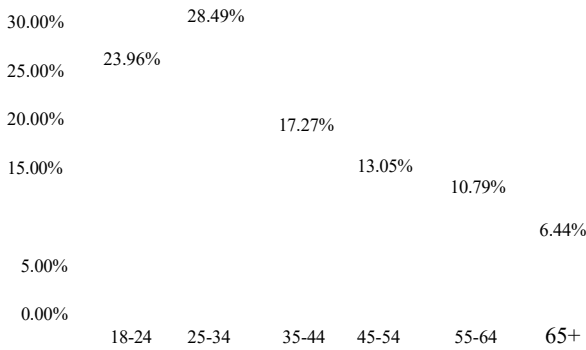
图30 Replika 付费模式



资料来源: Replika用户注册界面, 海通证券研究所

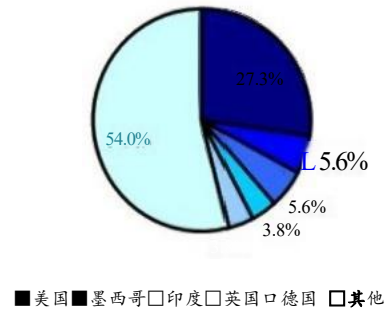
Replika 吸引欧美中年男性居多。根据similarweb 截止2024年8月份的统计, Replika 男性用户约占64.67%, 女性用户占35.33%。其中25-34岁的用户群体居多, 占比约28.49%。前五地区的用户分别是美国、墨西哥、印度、英国、德国。

图31 截止2024年8月Replika各年龄段用户占比



资料来源: similarweb受众人口板块, 海通证券研究所

图32截止2024年8月Replika用户全球分布



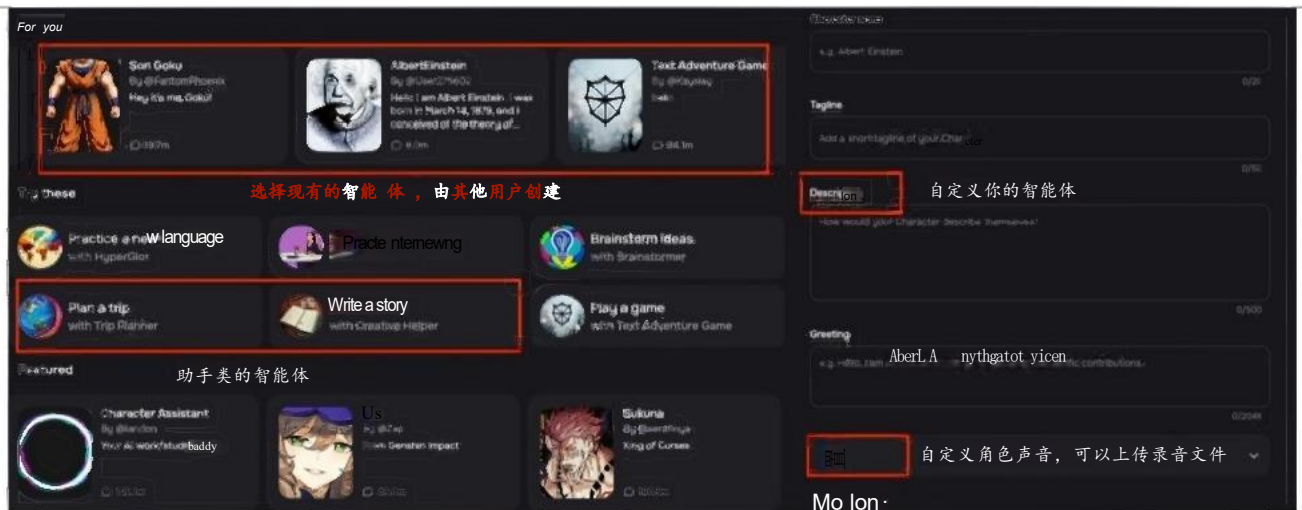
资料来源: similarweb 地理板块, 海通证券研究所

4.4 Character AI: 自研大模型，打造智能体社区

Character.ai 是一种神经语言模型聊天智能体，其较大特点是自主研发的Beta 语言模型。用户可以创建智能体并根据参数塑造他们的个性，然后将创建好的智能体发布到社区中与其他用户进行交互。另外，有些角色可以扮演助手模式，例如帮助你进行创意写作或打造一个文字冒险游戏等。用户不仅可以与单个智能体进行交流，还可以创建包含多智能体的群聊，这些智能体可以互相交谈且用户也能参与其中。

2023年3月，Character.AI 完成了一笔1.5亿美元的融资，估值达到10亿美元，由 a16z 领投。公司的创始人 Noam Shazeer曾是谷歌首席软件工程师，也是Transformer论文的作者之一。公司将这笔融资用于拓展模型的计算能力并提高其回答精度和推理能力。

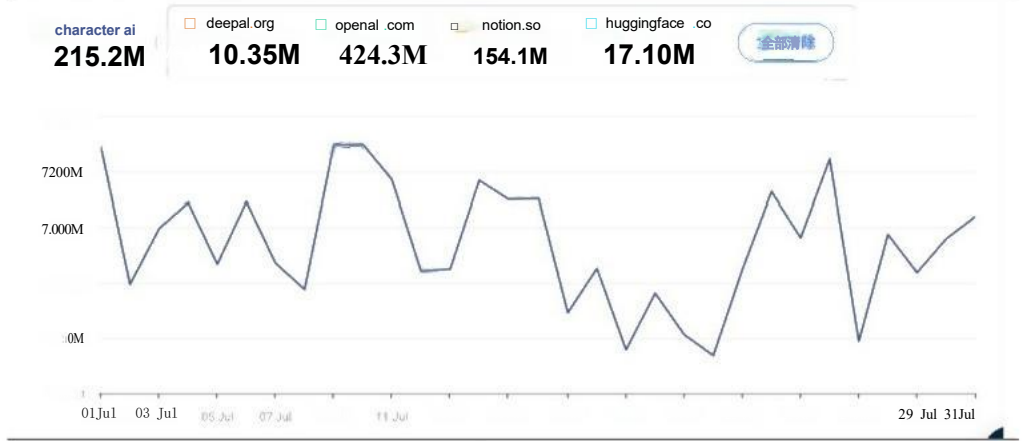
图33 Character.AI 用户界面



资料来源: Character.ai官网, 海通证券研究所

截止2024年7月，网站月平均访问量为2.15亿，且其访问量仍保持增长态势。据2024年7月30日的报道数据，Character.AI 在已拥有1700万的月度活跃用户。截止2024年4月全球用户达到2.33亿人。网站目前的变现方式主要通过升级订阅服务，每月收费9.9美元，功能包括不限制聊天次数、拥有更快的反馈速度。

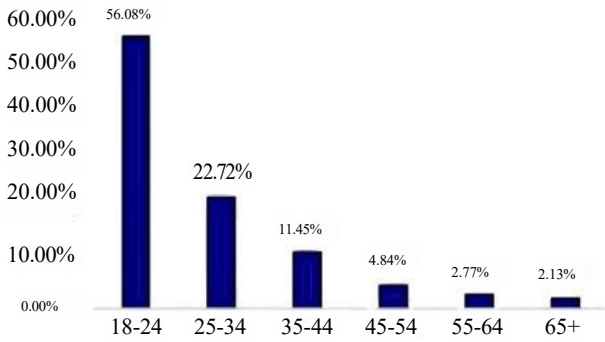
图34 Character. ai 2024年7月每日访问量



资料来源: similarweb访问量板块, 海通证券研究所

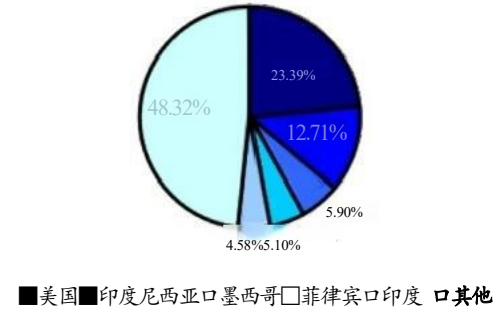
网站吸引年轻男生和女生, 活跃用户粘性较大。网站受众人口的男女比率约为51%和49%, 其中18-24岁的用户占比超过56%。前五地区的用户分别是美国、印度尼西亚、墨西哥、菲律宾和印度。其中活跃用户每天约进行240轮对话, 平均时长在2个小时。目前网页端的DAU/MAU 比例为23%, APP端 DAU/MAU 比例为41%。

图35 Character. AI 各年龄段用户占比



资料来源: similarweb受众人口板块, 海通证券研究所

图36 Character. AI用户全球分布



资料来源: similarweb地理板块, 海通证券研究所

公司通过自研模型减少成本, 智能体研发技术先进。由于是自主研发的模型, 其高效的LLM 算法大大降低了成本, 根据 foresight news援引海外独角兽团队测算, 每个Query 的推理成本是 chatgpt 的5%以下。早在2023年2月份, Character. AI就使用了八位整型做计算, 比16或32位浮点数效率有4倍-2倍的提升, 这也归功于创始人 Noam Shazeer对AI技术的独特理解。成本方面, 根据foresight news援引海外独角兽团队测算character每次交互的成本为0.000027美元, 一个 MAU 的单月成本可以控制在不到0.1 美元

4.5 星火智能体: 布局智能体平台, 发布垂类智能体

星火智能体平台目前由三大核心模块组成, 分别是智能体、任务链和知识库。智能体模块事先配备了预先设置, 能够自主规划并调用工具执行任务。任务链是为智能体提供执行任务所需要的各类工具, 能快速链接企业内部的知识、IT 系统和外部信源。知识库模块能让用户更快捷的查询垂类知识, 让智能体更好的解决专业类知识, 支撑其高效运行

图 37 科大讯飞星火智能体介绍



资料来源：星火智能体，海通证券研究所

2024年6月，科大讯飞发布了星火大模型V4.0及相关落地应用。在去年11月份，讯飞星火 App 的用户已经累计开发了3.7万+的个性化AI助手，可以满足职场、营销、创作等多场景需求。目前App已经累计在安卓市场下载1.31亿次，在国内通用大模型App中排名第一。另外，讯飞星火首批上线面向特定场景打造专属助手。例如垂类智能体“讯飞晓医”，其覆盖了1600种常见疾病，2800种药物以及6000种医学检验，其满足了用户的一些医疗建议需求。

图38部分星火AI智能体展示



资料来源：星火智能体，海通证券研究所

风险提示：大模型发展不及预期，智能体下游需求不足。

4.6 SkyAgents: 自研模型降低成本

天工SkyAgents 是基于昆仑万维天工大模型平台打造的，其具有从感知到决策，从决策到执行的自主学习和独立思考能力。用户能够通过自然语言创建一个或多个智能体，并将不同任务模块化。通过操作系统模块，可以执行包括问题预设、指定回复、知识库创建与检索、意图识别、文本提取和 http请求等任务。其中天工大模型的MoE 架构能让 AI 获得更快的响应速度且面对复杂任务的处理能力更强。

图39 天工大模型能力展示

能力展示



速度更快、效率更高:

MoE 模型的推理计算过程中只有少数特定专家模型被激活, 呈现出极高的稀疏性, 从而提高了推理计算效率, 让用户获得更快的A响应速度。



应对复杂任务能力更强

集成了多个专家模型, 针对不同的数据分布和构建模式进行搭建, 从而提升在各个细分领域的专业能力。在处理复杂任务多模态任务时有显著性能提升



灵活、多样、可扩展性更强

模型稀疏性使得在同等计算资源下能够扩展增加专家模型数量、调整权重配比, 具有更强的可扩展性, 构建更为灵活、多样的新时代大模型

资料来源: 天工开放平台官网, 海通证券研究所

SkyAgents 智能体拥有成本优势, tokens 价格低于OpenAI。天工大模型3.0版本0.01元/千tokens,2.0版本0.005元/千tokens,价格普遍比GPT4o更加划算, 但天工3.0价格略贵于GPT3.5 turbo。若按照2000个 input tokens, 输出1000个output tokens来计算, GPT4o花费0.175元(汇率按照1:7), GPT 3.5 turbo 花费0.028元, 天工3.0模型花费0.030元, 2.0模型花费0.015元。

图40天工Agent计费模式

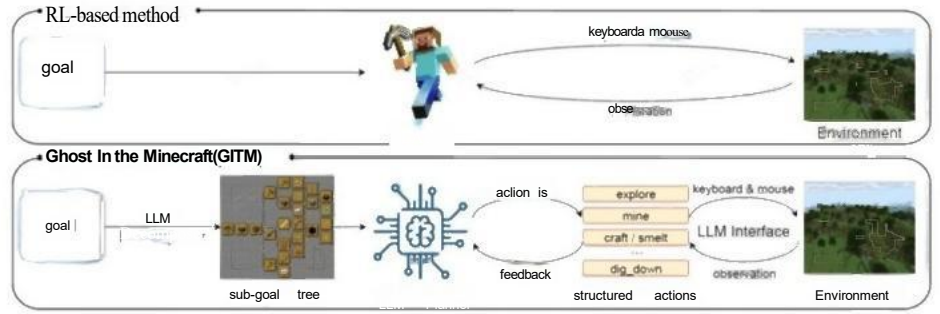
Agent 类	AI Agent	欢迎咨询
天工3.0系列模型	AI 对话 3.0 AGI Sky-Chat-3.0API	<p>天工大模型3.0 API在2.0版本基础上, 提供了搜索增强模式, 显著提升信息检索和知识整合能力。结合最新搜索引擎技术, 能够快速找到相关信息并进行智能分析, 提供更精准和丰富的回答。在复杂对话和深度内容创作中表现更佳, 如广告文案、小说写作、知识类写作和代码生成等场景。</p> <p>0.010元/千tokens 搜索增强: 开启搜索增强模式使更多tokens消耗, 当中调用搜索次数限时免费</p>
	AI对话2.0: AGI Sky-Chat Sass AP	<p>AGISky-Chat SassAPI提供了更强大的问答和文本生成能力, 非常适合复杂的对话交互和深度内容创作设计的场景。特别适用于需要高知识量、强推理能力和创造力的场景, 例如广告文案、小说写作、知识类写作和代码生成等。</p> <p>SkyChat-MogaVerse 0.010元/千tokens 0.005元/千/tokens</p>

资料来源: 天工开放平台产品定价, 海通证券研究所

4.7商汤: 联合清华、上海人工智能实验室打造GITM

商汤、清华大学和上海人工智能实验室等机构的研究者们提出了一种名为 Ghost in the Minecraft (GITM)的 通才 AI 智能体。GITM 能够在《我的世界》中表现优异, 显著减少训练时间和资源。这一突破标志着通用人工智能(AGI) 研究的重大进展, AGI 的目标是开发能够在开放世界环境中进行感知、理解和交互的智能体。GITM 通过自主学习, 解锁了《我的世界》主世界的262个物品, 相较于之前智能体仅解锁的78个物品大幅提升。此外, GITM 在“获取钻石”任务上的成功率从 OpenAI 提出的 VPT 方法的20%提升至67.5%。GITM 训练仅需一个 CPU 节点两天完成, 训练步数比之前方法减少了万分之一, 大大低于 OpenAI 和 DeepMind 提出的其他方法所需的 GPU 天数。此项研究为机器人和自动驾驶等产业带来了巨大的潜力和突破。

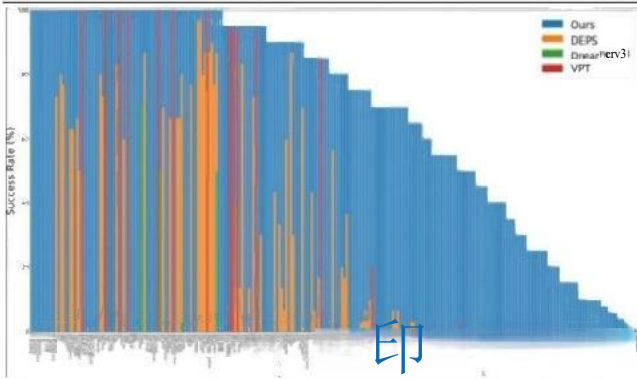
图 4 1 RL 架构对比 GITM 创新架构



资料来源：机器之心，海通证券研究所

传统 RL 智能体在 shang h 将复杂任务映射到底层键盘鼠标操作时面临困难，而 GITM 采用大型语言模型 (LLM) 作为核心，打破了这一传统架构。GITM 由 LLM Decomposer、LLM Planner 和 LLM Interface 三部分组成，逐步将复杂任务分解为子任务、结构化动作，直至最底层的键盘鼠标操作。LLM Decomposer 利用外部知识将复杂任务分解为简单子任务；LLM Planner 为每个子任务规划结构化动作，并根据反馈信息调整规划，不断总结成功经验；LLM Interface 通过键盘鼠标操作执行结构化动作，并在与环境交互中获取观察信息。

图 42 解锁科技树的成功率 (蓝色是 GITM)



资料来源：Xizhou Zhu 《Ghost in the Minecraft》，海通证券研究所

图 4 3 各模型完成任务成功率对比

Method	Success Rate (%)				
DreamerV3	50.0	3.0	0.01	0.0	0.0
DEPS	90.0	80.0	73.3	10.0	0.6
VPT	100.0	100.0	100.0	85.0	20.0
Our GITM	100.0	100.0	100.0	95.0	67.5

资料来源：机器之心，海通证券研究所

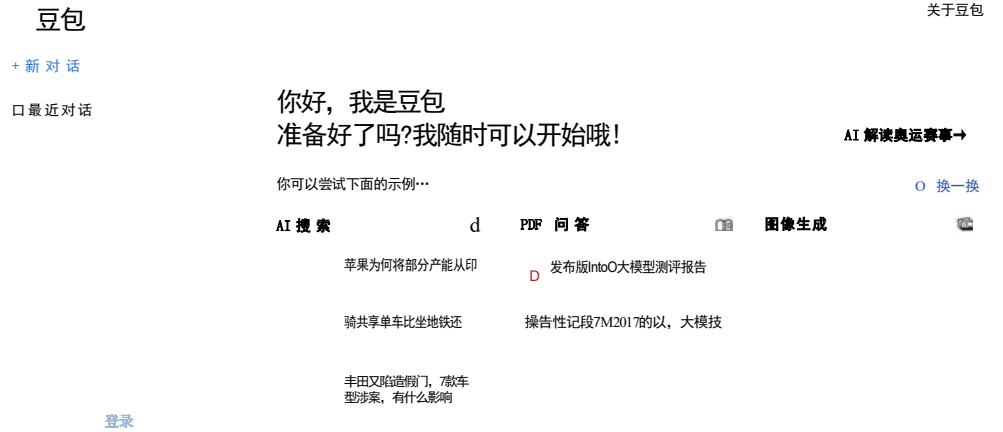
GITM 在《我的世界》主世界中实现了 100% 的任务覆盖率，成功解锁了完整的科技树，而此前所有智能体总和只能覆盖 30%。在备受关注的“获取钻石”任务上，GITM 的成功率达到 67.5%，比目前最佳成绩 (OpenAI VPT) 的 20% 提高了 47.5%。GITM 的训练效率也显著提升，所需环境交互步数仅为已有方法的万分之一，单个 CPU 节点训练 2 天即可完成，相比之前 OpenAI VPT 所需的 6480 个 GPU 天或 DeepMind DreamerV3 所需的 17 个 GPU 天，取得了巨大的进步

风险提示：智能体商业化不成熟，智能体下游需求不足。

4.8 字节跳动：豆包聊天助手

在中国的 AI 聊天机器人竞争中，字节跳动已抢占先机。其最新推出的 AI 对话服务豆包在下载量和活跃用户数上均超越了百度的文心一言。根据华尔街见闻援引互联网分析公司 SensorTower 的数据，自去年 8 月推出至今年 4 月，豆包在 iOS 系统上的下载量近 900 万次，领先于文心一言的 800 万次。在月活跃用户数方面，豆包在 iOS 端也已超过 400 万用户，显著高于文心一言。相比之下，OpenAI 的 ChatGPT App 在美国的月活跃用户数量为 670 万。根据 2024 年 5 月的数据显示，豆包 PC、iOS 平台月活跃用户总数已达到 2600 万。

图44豆包智能体的对话界面



资料来源：字节跳动豆包，海通证券研究所

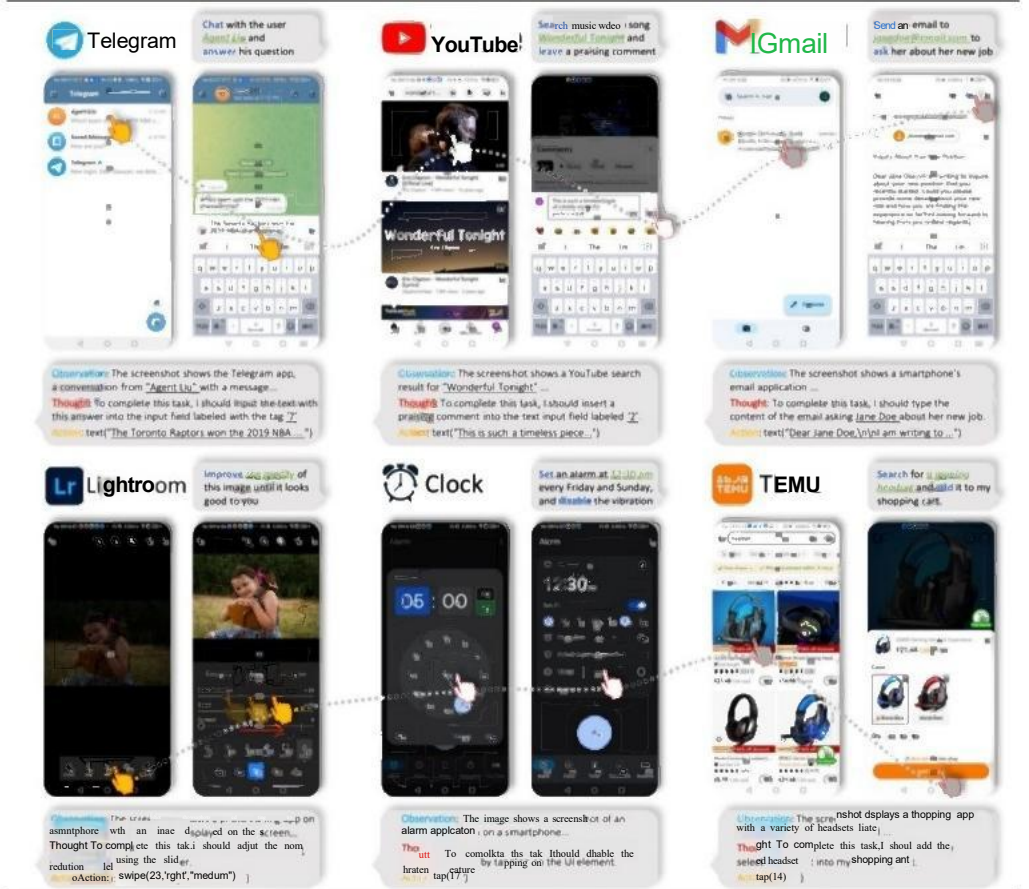
相比市面上的收费大模型，豆包有价格优势。目前普通用户可以免费使用豆包，其大模型的API 定价也相对较低。按 Tokens 实际使用量计算费用的“后付费”模式下，豆包通用模型-pro 和豆包通用模型-lite 的 3 2k 及以下窗口版本，模型限流为10K RPM 和 8 0 0 K TPM(以 RPM 和 TPM 其中之一达到上限为准)。相比之下，国内其他主流模型的 TPM 限额多在100K 到 3 0 0 K 之间，RPM 则在60到120区间，轻量级模型的 RPM 限额相对较高，但仅在300到500之间。字节跳动宣布企业市场定价为 0. 0008元/千 Tokens(约1500多个汉字)，比行业标准便宜99. 3%。

风险提示：智能体商业化不成熟，智能体下游需求不足

4. 9腾讯：AppAgent

AppAgent 通过模仿人类行为，完成与智能手机的交互。腾讯的AppAgent 的智能 Agent 能力可以用于操作任何 App，它在50个复杂手机任务上展示了强大的能力。根据 Chi Zhang等人的论文《AppAgent:Multimodal Agents as Smartphone UsersAppAgent》介绍，这项技术通过引入一种基于大型语言模型(LLMs) 的多模态智能 Agent (Agent) 框架，使智能体能够操作智能手机应用程序。与传统的智能助手如 Siri 不同，AppAgent 并不依赖系统后端访问，而是通过模拟人类的点击和滑动操作，直接与手机应用的图形用户界面(GUI) 互动。这种独特的方法不仅提高了安全性和隐私性，还确保了智能体能够适应应用界面的变化和更新。

图45 腾讯 AppAgent展示软件操作



资料来源：腾讯云，海通证券研究所

AppAgent 的实验环境基于命令行界面(CLI)，使 Agent 能够与安卓系统上的智能手机应用进行交互。Agent 接收两种关键输入：实时屏幕截图展示应用界面，以及详细描述交互元素的 XML 文件。为了提升 Agent 识别和交互这些元素的能力，每个元素都被赋予了一个唯一标识符。标识符要么来源于 XML 文件中的资源 ID (如果提供)，要么通过结合元素的类名、大小和内容来构建。这些元素的标识符以半透明数字形式覆盖在屏幕截图上，帮助 Agent 在不需指定精确位置的情况下准确互动，从而提高控制手机的精确度。

动作空间模拟了人类与智能手机的常见交互方式，包括点击和滑动。设计了四个基本功能：点击(Tap)、长按(Long Press)、滑动(Swipe) 和文本输入(Text)，以及两个系统级功能：返回(Back) 和退出(Exit)。这些预定义动作旨在简化 Agent 的交互流程，并减少对精确屏幕坐标的依赖，解决了语言模型在准确预测中可能遇到的挑战。

图4 6 App Agent的App探索学习阶段

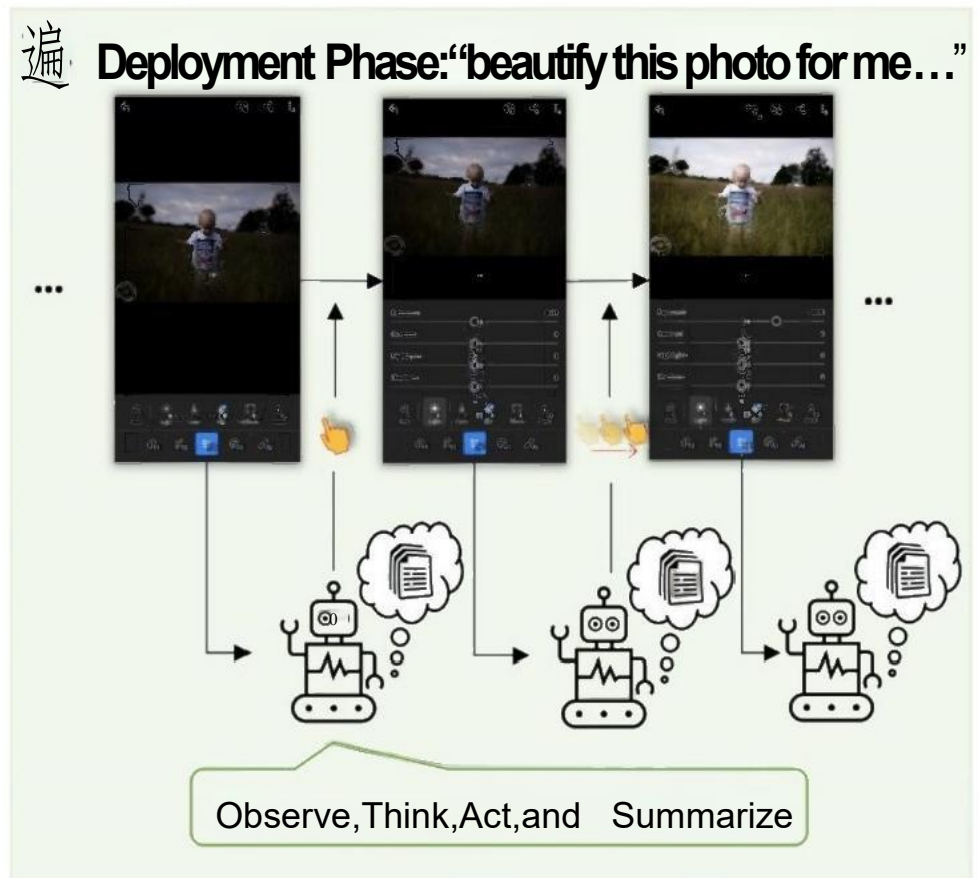


资料来源：腾讯云，海通证券研究所

AppAgent 框架的核心是探索阶段。Agent 通过自主交互或观察人类演示来学习应用程序的功能和特性。在自主交互模式下，Agent 被分配一个任务并开始与 UI 元素进行自主互动。**Agent**尝试不同的动作，并观察应用界面的变化以理解其工作原理。**Agent**通过分析每个动作前后的屏幕截图，了解 UI 元素的功能和特定动作的效果。所有这些信息都会被记录成文档，详细记录下不同元素所执行动作的效果。如果某个 UI 元素被多次操作，**Agent**会根据之前的文档和当前的观察来更新信息，以提高认知质量。

AppAgent专注重要操作，提高探索效率。如果当前 UI 页面与应用的主要任务无关(如广告页面)，Agent 会停止进一步探索，并使用 Android 系统的返回功能返回到前一个 UI 页面。这种目标导向的探索方法，相比随机探索，确保Agent 专注于对应用有效操作至关重要的元素。**Agent** 还利用语言模型对用户界面的现有知识来提高探索效率，直到完成分配的任务。在观察人类演示进行探索的方式中，**Agent** 通过观察人类用户的操作来学习应用的复杂功能。这种方式尤其有效，对于那些难以通过自主交互发现的功能。**Agent**记录人类使用的元素和动作，这种策略缩小了探索空间，并阻止**Agent**与无关的应用页面进行交互，从而比自主交互更为高效和有条理。

图4.7 AppAgent的实际部署阶段



资料来源：腾讯云，海通证券研究所

在经过探索阶段的训练后，AppAgent 已准备好执行复杂任务。在部署阶段，Agent 采用逐步方法操作，每一步包括获取当前用户界面的屏幕截图和动态生成的文档，详细描述了界面元素的功能及动作效果。首先，Agent 对当前用户界面进行观察，并阐述其关于任务和观察结果的思考过程。然后，Agent 通过调用可用函数执行动作。每次动作后，Agent 总结交互历史和当前步骤中采取的动作。这些信息被整合进下一个提示中，为Agent 提供了一种记忆形式。这种细致的方法提高了Agent 动作的可靠性和可解释性，从而促进了更明智的决策。部署阶段持续进行，直到Agent 确定任务已经完成，此时它可以通过执行退出 (Exit) 动作结束过程。

AppAgent 采用了多模态大型语言模型 GPT-4，能够处理图像和文本输入，解释并互动应用中的视觉及文本信息。通过简化动作空间设计，AppAgent 大幅提升了准确性和效率，避免了生成精确 xy 坐标的需求，这一直是传统语言模型的难点。在成功率、奖励和平均步骤数等关键性能指标上，AppAgent 表现优异。即使任务步骤失败，它仍能根据最终状态获得奖励，显示出强大的适应性和韧性。

风险提示：智能体商业化不成熟，智能体下游需求不足。