Agent 叙事强化，算力与 SaaS 分化加剧

26 年 1 月 AI 月报

华泰研究

2026 年 2 月 07 日 | 中国内地　　行业月报

国内模型：Agent 能力持续演进，关注 DeepSeek 多模态进展

本月国内模型的进展包含多模态、多 Agent 和上下文学习三条主线。DeepSeek 推出 DeepSeek-OCR 2，聚焦图像到文本转换与视觉推理，我们关注到视觉侧能力与文本侧主干的融合路径更清晰，有望推动 DeepSeek 下一代模型向更完整的多模态能力扩展。Kimi K2.5 提出 Agent Swarm 并行 Agent 显著提升复杂任务效率，我们认为 K2.5 以“多模态内置+并行智能体”的组合，提升了其在企业级长流程任务中的落地效率与适配范围。腾讯提出 CL-bench，量化揭示大模型临时学习能力不足，指出提升模型能力需兼顾上下文窗口与复杂度处理，有望指导后续模型提高上下文学习能力。

海外算力：Agent 主线强化，CSP Capex 持续上修

我们判断 Agent 渗透是下一个 token 加速点。Agent 产品迅速进展的本质是模型在长链任务能力上跨越“奇点”，25 年年底发布的模型在长链任务能力上完成了显著的跨越，我们认为这是目前 Agent 能力提升的根本因素。Agent 的推理范式是复杂流程、连续执行的，算力消耗较大，我们预计 26 年将成为 AI Agent 推理端从“能力验证”走向“Agent 规模化应用”的关键拐点年。本月海外大厂陆续公布业绩，Capex 持续增长，AI 需求表述乐观。我们关注到 NPO 方案下的“光进机柜”雏形初现，建议关注后续 GTC 大会催化。

AI 应用：云厂业绩加速兑现，静待 SaaS 预期修正

海外科技公司 25Q4 业绩持续披露，云厂商业绩加速兑现，SaaS 市场预期偏悲观。数据层面上看，25Q4 AI 应用公司业绩基本超市场预期，26 年指引呈小幅上修状态；预期层面上看，SaaS 板块的悲观预期或持续加重，云厂商核心担忧在于 AI 商业化能否匹配 26 年高增的 CapEx 指引。我们认为，26 年全球 AI 应用有望全面加速，云厂商业绩有望持续加速，部分 SaaS 公司有望实现产品价值下沉与企业价值重估，看好海外 AI 商业化进展提速。

AI4S：生物制药商业化最快，材料领域有望突破

我们认为 AI for Science（AI4S）的研究范式打破了传统“实验发现”或“手工推导方程”的局限，正通过赋能量子、原子与连续介质系统中的高级建模、仿真与预测，引领科研革命。我们持续看好 AI 制药在 26 年的商业化前景，预计行业将呈现小分子药物合作深化与大分子抗体领域合作爆发的双轮驱动格局。我们预计 AI 新材料将成为 AI4S 的重点应用与投资方向，AI 加速材料发现，并通过数字化工艺优化直接推动产业化，是实现制造产业升级的核心引擎。

月专题：Agentic Coding 加速迭代，关注 Agent 进展

本月我们关注到以 Claude Code 为代表的 Agentic Coding 产品、以 OpenClaw 为代表的 Agent 应用产品正在加速迭代，我们认为 Agent 应用正在加速，有望带来软件行业重构。我们判断 2025 年是 Agent 元年，2026 年可能进入 Agent 加速落地期，主要体现在两方面：一是 Agentic Coding 的迭代速度会大幅加快；二是国内外大厂会激烈争夺个人 Agent 助手的超级入口，均会成为下一轮 token 加速的重要推手。Agentic Coding 的快速迭代可能会加速软件行业的重构，软件开发成本面临“杰文斯悖论”：未来个性化的、由 AI 生成的软件会爆发，但单体软件的价值可能下行，建议持续关注 Agent 进展。

风险提示：宏观经济波动，技术进步不及预期，中美竞争加剧。研报中涉及到未上市公司或未覆盖个股内容，均系对其客观公开信息的整理，并不代表本研究团队对该公司、该股票的推荐或覆盖。

科技	增持（维持）
计算机	增持（维持）

郭雅丽　研究员
SAC No. S0570515060003　guoyali@htsc.com
SFC No. BQB164　+(86) 21 3847 6016

范映蕊　研究员
SAC No. S0570521060004　fanyirui@htsc.com
SFC No. BWD469　+(86) 21 2897 2228

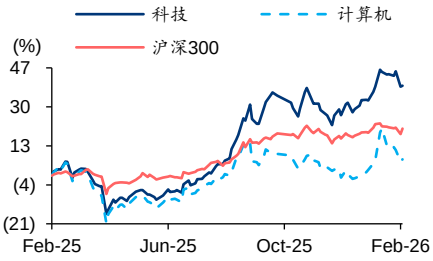
袁泽世*，PhD　研究员
SAC No. S0570524090001　yuanzeshi@htsc.com
+(86) 21 2897 2228

岳铂雄*　研究员
SAC No. S0570524080004　yueboxiong@htsc.com
+(86) 21 3847 6087

王浩天*　联系人
SAC No. S0570125010006　wanghaotian@htsc.com
+(86) 21 2897 2228

徐诚伟*　联系人
SAC No. S0570125070089　xuchengwei@htsc.com
+(86) 21 2897 2228

行业走势图



资料来源：Wind，华泰研究

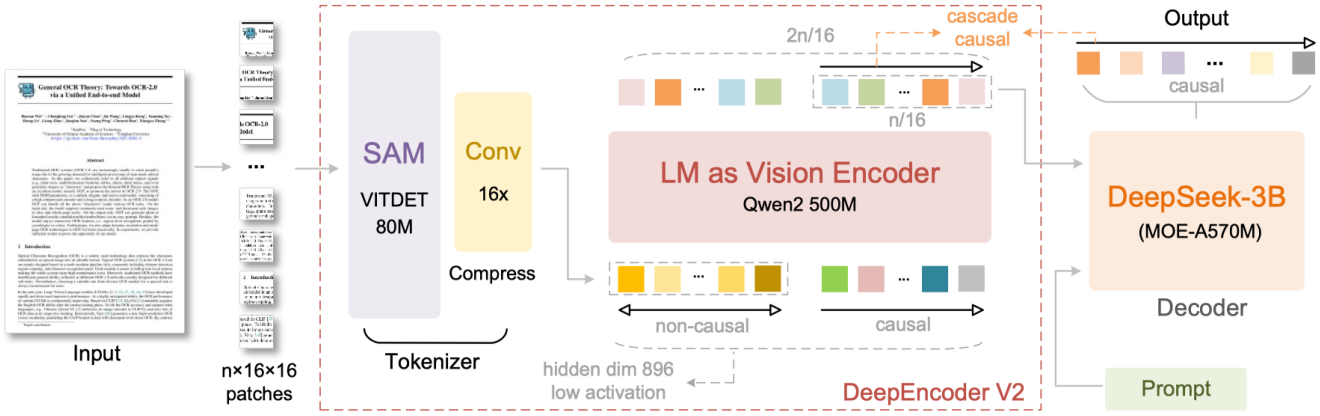
正文目录

AI 模型：Agent 能力持续演进，关注 DeepSeek 多模态进展

DeepSeek-OCR 2 提升了视觉理解能力，有望加强下一代模型能力

DeepSeek-OCR 2 面向复杂文档的视觉理解式 OCR 模型。DeepSeek 在 2026 年 1 月底推出 DeepSeek-OCR 2，参数规模约 3B，聚焦图像到文本转换与视觉推理，强调对页面布局与内容关系的深度理解。相较传统 OCR 偏重纯文本提取，DeepSeek-OCR 2 更强调按接近人类的阅读顺序完成内容提取与结构还原，旨在弥补 DeepSeek-OCR1 在读取顺序与复杂页面布局处理上的不足，使复杂文档的输出更贴近可直接使用的书面结构。

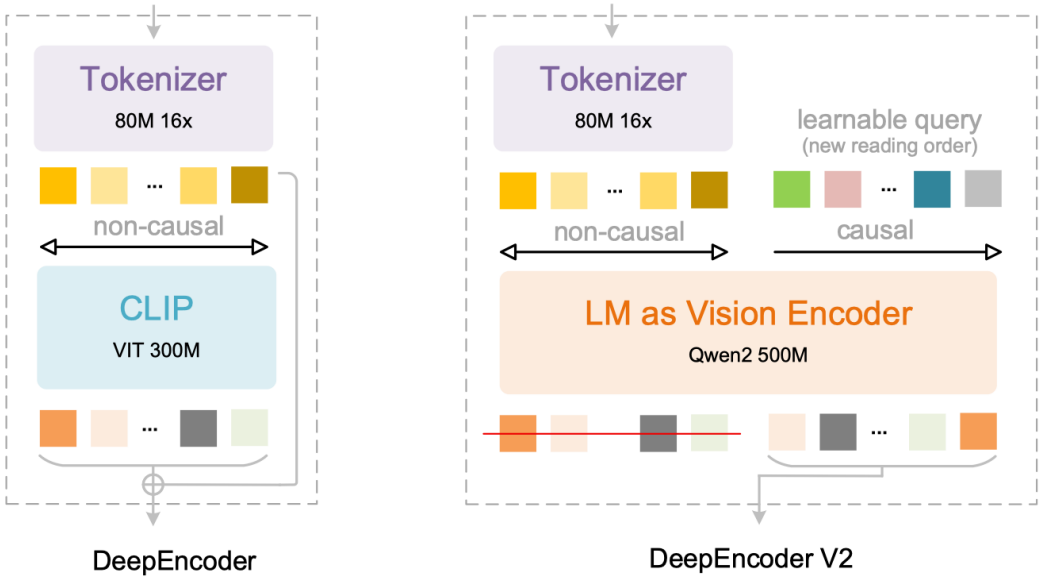
图表1： DeepSeek-OCR 2 架构



资料来源：《DeepSeek-OCR 2: Visual Causal Flow》，Wei（2026）、华泰研究

DeepEncoder V2 以全局理解重排阅读顺序。DeepEncoder 是 DeepSeek-OCR 2 中的重要组件。引入 DeepEncoder V2 后，模型不再采用固定网格从左上到右下扫描，而是先形成页面全局表示，再通过可学习查询 learnable queries 按语义驱动的逻辑顺序抽取内容。该机制被描述为“visual causal flow”，核心在于先判定“先读什么、后读什么”，从而在多栏 PDF、密集表格、标签-值对等场景下更好地区分列间先后与字段对应关系。DeepSeek-OCR 2 论文显示，新架构使文档结构更清晰、序列更合理，布局相关错误明显降低。

图表2： DeepEncoder 中的 CLIP 组件替换为 LLM 风格架构



资料来源：《DeepSeek-OCR 2: Visual Causal Flow》，Wei（2026）、华泰研究

双重注意力让编码器兼顾全局与因果提取。DeepSeek-OCR 2 以语言模型驱动的视觉编码器替代旧版 CLIP 等静态视觉编码网络，将视觉特征编码过程融入类 LLM 架构：1）视觉 token 之间保持双向注意力以编码全局空间结构。2）新增因果查询 token 以单向因果注意力顺序关注视觉内容，并结合已生成查询的上下文逐步提取信息。编码器在“阅读”图像的方式上类似一个小型语言模型（如 Qwen2 500M），并通过定制注意力 Mask 实现“先看全局、再按需提取”，从机制上克服 DeepSeek-OCR1 固定顺序编码的局限。

基准测试显示模型在效率与精度上同步提升。得益于阅读顺序与编码架构改进，DeepSeek-OCR 2 在复杂版面 OCR 上更易保持上下文连贯与结构一致，减少人工清洗需求。基准测试显示，DeepSeek-OCR 2 在 OmniDocBench v1.5 上评分为 91.09，较前代提高 3.73 个百分点；在压缩效率方面，据 OmniDocBench 实验，新模型仅用 1120 个视觉 token 即可实现较低编辑距离误差（0.100 vs 0.115，比较对象为 Gemini-3 Pro）。同时，生产环境重复率下降：对在线用户图像日志从 6.25%降至 4.17%，对 PDF 数据从 3.69%降至 2.88%；在个别超长新闻类版面上受视觉 token 上限影响仍有改进空间。我们认为，DeepSeek-OCR 2 的设计重点在于以“顺序可学习+结构可解释”的路径提升复杂文档可用性，并在成本受控前提下改善端到端输出质量。

图表3：DeepSeek-OCR 2 在 OmniDocBench v1.5 较前代提高 3.73 个百分点

Model	V-token^{max} ↓	Overall ↑	Text^{Edit} ↓	Formula^{CDM} ↑	Table^{TEDs} ↑	Table^{TEDS_s} ↑	R-order^{Edit} ↑
Pipline							
Marker-1.8.2 [1]	-	71.30	0.206	76.66	57.88	71.17	0.250
MinerU2-pp [45]	-	71.51	0.209	76.55	70.90	79.11	0.225
Dolphin [17]	-	74.67	0.125	67.85	68.70	77.77	0.124
Dolphin-1.5 [17]	-	83.21	0.092	80.78	78.06	84.10	0.080
PP-StructureV3 [13]	-	86.73	0.073	85.79	81.68	89.48	0.073
MonkeyOCR-pro-1.2B [23]	-	86.96	0.084	85.02	84.24	89.02	0.130
MonkeyOCR-3B [23]	-	87.13	0.075	87.45	81.39	85.92	0.129
MonkeyOCR-pro-3B [23]	-	88.85	0.075	87.25	86.78	90.63	0.128
MinerU2.5 [45]	-	90.67	0.047	88.46	88.22	92.38	0.044
PaddleOCR-VL [12]	-	92.86	0.035	91.22	90.89	94.76	0.043
End-to-end Model							
OCRFlux [4]	>6000	74.82	0.193	68.03	75.75	80.23	0.202
GPT-4o [33]	-	75.02	0.217	79.70	67.07	76.09	0.148
InternVL3 [55]	>7000	80.33	0.131	83.42	70.64	77.74	0.113
POINTS-Reader [31]	>6000	80.98	0.134	79.20	77.13	81.66	0.145
olmOCR [36]	>6000	81.79	0.096	86.04	68.92	74.77	0.121
InternVL3.5-241B [49]	>7000	82.67	0.142	87.23	75.00	81.28	0.125
MinerU2-VLM [45]	>7000	85.56	0.078	80.95	83.54	87.66	0.086
Nanonets-OCR-s [2]	>7000	85.59	0.093	85.90	80.14	85.57	0.108
Qwen2.5-VL-72B [9]	>6000	87.02	0.094	88.27	82.15	86.22	0.102
Gemini-2.5 Pro[6]	-	88.03	0.075	85.82	85.71	90.29	0.097
dots.ocr [39]	>6000	88.41	0.048	83.22	86.78	90.62	0.053
OCRVerse [3]	>6000	88.56	0.058	86.91	84.55	88.45	0.071
Qwen3-VL-235B [8]	>6000	89.15	0.069	88.14	86.21	90.55	0.068
DeepSeek-OCR (9-crops)	1156	87.36	0.073	84.14	85.25	89.01	0.085
DeepSeek-OCR 2	1120	91.09	0.048	90.31	87.75	92.06	0.057
	↓36	↑3.73	↓0.025	↑6.17	↑2.5	↑3.05	↓0.028

资料来源：《DeepSeek-OCR 2: Visual Causal Flow》，Wei（2026）、华泰研究

图表4：DeepSeek-OCR 2 在压缩效率方面实现仅用 1120 个视觉 token 即可实现较低编辑距离误差

Model	V-token^{max} ↓	Text^{Edit} ↓	Formula^{Edit} ↓	Table^{Edit} ↓	R-order^{Edit} ↓	Overall^{Edit} ↓
Gemini-3 pro [44]	1120	-	-	-	-	0.115
Seed-1.8 [41]	5120	-	-	-	-	0.106
DeepSeek-OCR	1156	0.073	0.236	0.123	0.085	0.129
DeepSeek-OCR 2	1120	0.048	0.198	0.096	0.057	0.100

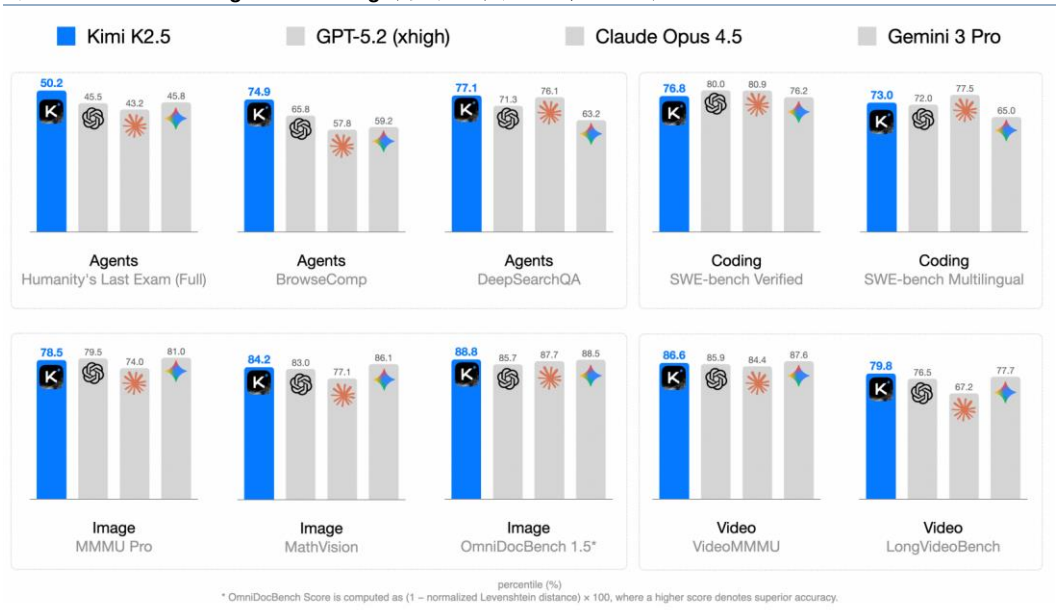
资料来源：《DeepSeek-OCR 2: Visual Causal Flow》，Wei（2026）、华泰研究

DeepSeek-OCR 2 的视觉理解能力提升，有望在后续模型中以多模态形态被纳入。DeepSeek-V3 与 R1 系列目前仍为纯文本模型，不支持直接上传图片、音频或视频，但 DeepSeek 并非缺乏多模态技术，而是将不同能力以相对独立的模型形态拆分部署：1）V 系列（如 V2、V3）定位为通用大语言模型（General LLM），强调性能功耗比与文本理解、代码、数学能力，在架构设计（如 MLA 架构）上主要面向文本 Token 处理。2）R 系列（如 R1、R1-Zero）定位为推理模型（Reasoning Model），强项来自强化学习（RL）驱动的“思考”能力，训练与输出聚焦纯文本逻辑链路。我们认为，随着 DeepSeek-OCR 2 在复杂版面阅读顺序、结构抽取与视觉推理等方面的迭代成熟，视觉侧能力与文本侧主干的融合路径更清晰，有望推动 DeepSeek 下一代模型向更完整的多模态能力扩展。

Kimi K2.5 提出 Agent Swarm 并行 Agent 显著提升复杂任务效率

Kimi K2.5 以原生多模态智能体定位拓展能力边界。Kimi K2.5 是 Moonshot AI 推出的新一代大型模型，定位为开源的 native multimodal agentic model，面向视觉、代码与智能体任务等多场景协同需求。模型总参数量约 1 万亿级别，采用混合专家 MoE 架构，专为复杂推理与多任务智能体场景设计。K2.5 为前代 Kimi K2（代号“K2 Thinking”）的持续训练升级产物，通过在 K2 基础上新增约 15 万亿 token 的视觉+文本混合数据持续预训练，扩充知识覆盖并强化跨任务能力。

图表5：Kimi K2.5 在 Agent 和 Coding 等多个测评集上取得领先效果



资料来源：《KIMI K2.5: VISUAL AGENTIC INTELLIGENCE》，Kimi Team（2026）、华泰研究

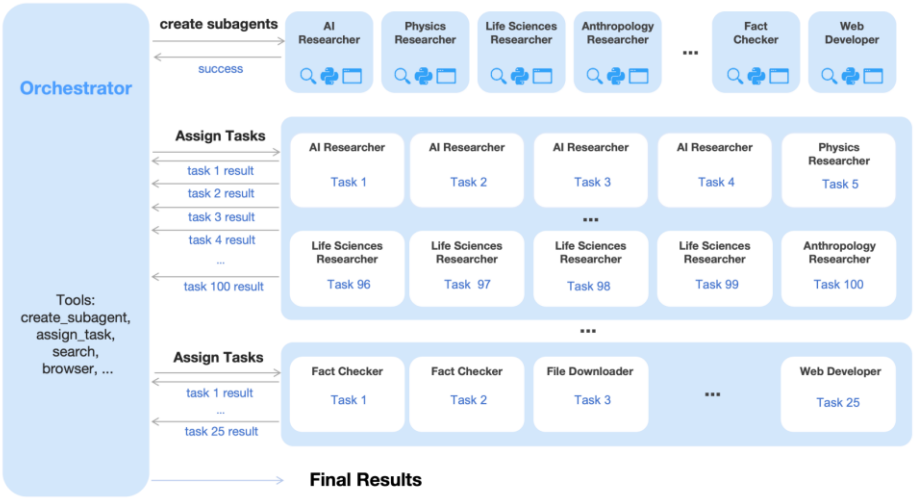
原生多模态引入 MoonViT 实现内置视觉理解。Kimi K2.5 论文，K2.5 从“K2 Thinking”以文本与工具调用为主，升级为原生多模态模型，在预训练阶段引入大规模视觉-文本联合数据，使模型同时掌握图像与语言。其架构新增 MoonViT 视觉编码器，参数规模约 200-400M，用于处理图像输入，并支持原始高分辨率图像，方案与“Kimi-VL-A3B-Instruct”相近。相比之下，K2 Thinking 虽可通过工具调用图像分析接口，但缺少内部视觉模块；K2.5 将视觉能力整合进同一模型，使其可直接接收图像、视频等输入并进行跨模态推理与生成。

代码能力增强并扩展至“视觉编程”范式。K2.5 通过持续训练与优化提升代码生成与理解能力，在前端开发场景表现更突出，可将自然语言需求转化为完整前端界面代码，并覆盖复杂布局与交互动画。依托多模态能力，K2.5 支持结合视觉进行编程：可基于 UI 设计图或视频演示理解视觉意图并生成对应代码，实现界面还原与视觉调试，例如读取网页设计视频后重建网站代码，或依据设计图实现接近像素级的布局与动态效果。Moonshot 还提供 Kimi Code 产品，面向终端或 IDE 提供“所见即所得”的编程体验。

MoE 架构维持算力可控并优化长上下文表现。K2.5 沿用 DeepSeek V3 改进版 MoE 架构并在超大规模上优化：总参数约 1 万亿，但每次仅激活 32B。结构上包含 61 层 Transformer Block（仅 1 层为密集层，其余为 MoE 专家层），共 384 个专家，每个专家隐藏层维度 2048，每个 token 由门控网络选出 8 个专家参与计算，并配置 1 个共享专家。长上下文方面，据 Kimi K2.5 论文，K2.5 提升了 256K 窗口下的信息保真度；同时沿用量化感知训练 QAT 以原生 INT4 权重部署，并且第三方提供了 1.8-bit 动态量化版本，大小约 240GB，较原模型缩小约 60%。

Agent Swarm 并行 Agent 显著提升复杂任务效率。K2.5 在 K2 Thinking 的 CoT 与动态工具调用基础上，引入“Agent Swarm”并行 Agent 架构，并通过并行智能体强化学习 PARL 训练，使模型能自动拆解复杂任务并派生多达 100 个子 Agent 并行执行，由主 Agent 编排协调。实验证明，在无需人工预定义子 Agent 角色或流程的情况下，Agent Swarm 可将多步任务总耗时减少 4.5 倍；内部测试表明端到端运行时间降低约 80%。训练侧引入可训练编排者 Orchestrator 策略网络，并通过分阶段奖励塑造缓解“串行崩溃”“虚假并行”等失败模式，同时以关键路径长度等指标评估并行是否实质缩短最长步骤。

图表6：Kimi K2.5 的 Agent swarm 架构



资料来源：《KIMI K2.5: VISUAL AGENTIC INTELLIGENCE》，Kimi Team（2026）、华泰研究

多模式推理提升交互灵活性并回应 K2 瓶颈。为适配不同场景，K2.5 在产品形态中提供多种模式：1）Instant 模式用于简单问题快速响应。2）Thinking 模式用于复杂问题的逐步推理。3）Agent 模式面向资料检索、内容创作等结构化输出并自动工具交互。4）Agent Swarm 模式用于大型多步骤项目的并行执行。综合改进点包括：多模态集成由外部工具依赖转为内置视觉模块；通过新增 15 万亿以上训练样本持续预训练扩展知识广度并缓解知识截止影响；RoPE 缩放等增强长文档稳定性；并行 Agent 提升大规模资料处理的可行性。我们认为，K2.5 以“多模态内置+并行智能体”的组合，提升了其在企业级长流程任务中的落地效率与适配范围。

腾讯提出 CL-bench，有望指导后续模型提高上下文学习能力

腾讯提出 CL-bench，量化揭示大模型临时学习能力不足。腾讯首席科学家姚顺雨团队提出 CL-bench（Context Learning Benchmark），评测大语言模型（LLM）从新上下文中学习并应用的能力，旨在区分“调用已知知识”与“现场吸收新知识”。据 CL-bench 论文，主流模型在该基准上整体成功率偏低：GPT-5.1（High）平均成功率 23.7%，其他模型平均 17.2%。这表明模型在面对预训练未覆盖的新规则、新概念时，常难把上下文转化为可执行的推理约束，我们认为，这一短板会直接影响模型在真实工作场景中的可靠性。

CL-bench 基准以“新知识自包含”约束模型走捷径。CL-bench 要求每题必须使用上下文提供且预训练中不存在的新知识作答，并明确禁止借助上下文之外的信息（例如检索或隐含预设），从机制上迫使模型“现学现用”。据 CL-bench 论文，基准包含 500 个复杂上下文、1899 个任务（questions / prompts）与 31607 条评价标准（rubrics），并由专家标注，平均每个上下文投入约 20 小时。为防污染，团队采用 1）虚构创造；2）变体改造；3）小众与新兴内容等设计。据团队验证，不给上下文时 GPT-5.1（High）成功率不到 1%，说明任务确实依赖现场信息。

四类场景覆盖演绎执行与归纳发现。CL-bench 将任务划分为 1）领域知识推理：如架空法律、创新金融工具、小众理论等；2）规则体系应用：如新游戏规则、公理系统、自创编程语言或技术标准；3）流程化任务执行：如手册、软件说明、API 调用顺序；4）实验发现与模拟：从数据与观测中归纳规律再应用。据团队结果，实验发现与模拟类通常成功率低于 10%，且波动更大，显著弱于“按明示规则执行”的任务，反映模型在归纳形成新规律方面更薄弱。

图表7：CL-bench 中上下文数量、任务数量、评分标准数量等任务统计

Context Category	#Contexts	#Tasks	#Rubrics	Tasks per context		Rubrics per task		Input Length (tokens)	
				Mean	Max	Mean	Max	Mean	Max
Domain Knowledge Reasoning	190	663	11,099	3.5	7	16.7	74	8.3K	60.0K
Rule System Application	140	566	8,286	4.0	12	14.6	75	12.2K	62.2K
Procedural Task Execution	100	471	9,486	4.7	12	20.1	59	8.5K	58.5K
Empirical Discovery & Simulation	70	199	2,736	2.8	9	13.7	114	16.7K	65.0K
Total	500	1,899	31,607	3.8	12	16.6	114	10.4K	65.0K

资料来源：《CL-bench: A Benchmark for Context Learning》，Yao（2026）、华泰研究

细粒度评估显示“更长上下文与更强思考”增益有限。CL-bench 平均每题约 16.6 条 rubrics（评分标准），并采用“LM-as-a-judge”对照 rubrics 自动打分，以任务成功率（Solving Rate）作为主指标，同时比较高推理强度（思维链）与低推理强度（直接回答）。据 CL-bench 论文数据，失败主要来自忽略或误用上下文，模型容易回退到参数化知识而违背新规则；提高推理强度通常仅带来小幅改善，GPT-5 系列在部分管理类与实验数据类任务上提升约 6%。

图表8：十种前沿大语言模型在 CL-bench 上的任务解决率

Model Names	Overall (%)	Domain Knowledge Reasoning (%)	Rule System Application (%)	Procedural Task Execution (%)	Empirical Discovery & Simulation (%)
GPT 5.1 (High)	23.7 ± 0.5	25.3 ± 1.3	23.7 ± 1.3	23.8 ± 1.4	18.1 ± 3.1
Claude Opus 4.5 Thinking	21.1 ± 1.4	23.7 ± 1.2	19.0 ± 1.5	22.6 ± 1.5	15.1 ± 2.3
GPT 5.2 (High)	18.1 ± 0.8	18.6 ± 0.9	17.2 ± 1.3	21.4 ± 1.1	11.7 ± 1.8
o3 (High)	17.8 ± 0.2	18.0 ± 1.4	17.6 ± 1.1	19.5 ± 0.4	13.7 ± 0.8
Kimi K2 Thinking	17.6 ± 0.6	18.7 ± 0.6	17.0 ± 1.5	18.8 ± 0.7	12.6 ± 4.0
HY 2.0 Thinking	17.2 ± 0.6	18.0 ± 1.0	17.3 ± 0.5	19.4 ± 1.1	8.9 ± 0.3
Gemini 3 Pro (High)	15.8 ± 0.3	15.5 ± 1.1	17.7 ± 1.7	16.4 ± 1.6	10.1 ± 3.1
Qwen 3 Max Thinking	14.1 ± 0.1	13.5 ± 0.5	15.6 ± 1.0	15.2 ± 1.4	9.0 ± 1.0
Doubao 1.6 Thinking	13.4 ± 0.1	13.7 ± 0.1	14.2 ± 1.4	13.9 ± 1.5	9.4 ± 0.3
DeepSeek V3.2 Thinking	13.2 ± 0.4	13.6 ± 0.6	13.8 ± 0.6	14.2 ± 0.1	8.0 ± 1.5

资料来源：《CL-bench: A Benchmark for Context Learning》，Yao（2026）、华泰研究

启示：提升模型能力需兼顾上下文窗口与复杂度处理。据 CL-bench 论文结论，任务难度随上下文长度增加而上升，各模型在长上下文条件下成绩普遍下降，反映长上下文有效利用仍存在约束。但与此同时，即便上下文很短，当信息高度浓缩、隐含规则密集、依赖关系复杂或约束更苛刻时，模型同样容易失误，说明“复杂度”对结果的影响不弱于“长度”。因而，在提升模型能力时，不能只着眼于扩展上下文窗口，还要系统增强：1）复杂知识结构的抽取与结构化表示能力。2）隐式约束与细节条件的识别与对齐能力。3）多轮依赖下的状态维护与动态更新能力。我们认为，只有同时补齐上述能力，模型在真实场景中对新知识的吸收与调用才更具稳定性。

AI 算力：Agent 主线强化，CSP Capex 持续上修

Agent 渗透是下一个 Token 加速点

Agent 渗透是下一个 token 加速点。回顾 2026 年以来的 Agent 进展，我们可以看到 Agentic Coding 与个人 Agent 两条主线。从 Agentic coding 来看，各个大厂均在推广其相应的产品，包括 Anthropic 的 Claude code，OpenAI 的 ChatGPT Codex、Google 的 Gemini Code Assist。垂直 Agent 的价值在于验证 Agent 的可执行性，真正决定产业规模的是是否出现适用于绝大多数人的通用 Agent。个人 Agent 来看，海外方面，Open Claw（原 Clawdbot、Moltbot）的众多能力引起海内外的广泛关注，Meta 收购 Manus 表示了其对虚拟机路线 Agent 的看好，OpenAI 也在 ChatGPT Agent 中尝试加入更多的应用场景。国内来看，各个大厂都在推广其个人助手产品，包括千问助手、豆包助手等，并开启春节入口争夺战。在互联网与移动互联网历史中，平台型入口的争夺一定发生在两件事同时具备时：供给侧已具备可规模化的能力；需求侧即将被激活，但用户心智尚未固化（抢占默认选择）。我们判断 26 年的 Agent 将在大厂的推动下快速渗透。

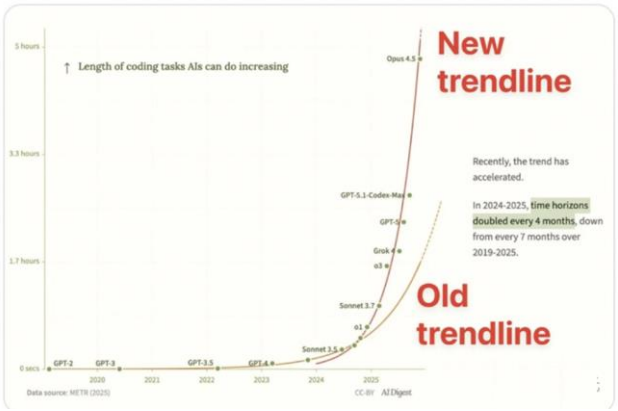
图表9：移动互联网时代的入口争夺战

时间	节点	用户入口
2010–2011	移动 App 爆发前	入口是应用商店、OS
2013–2014	移动支付爆发前	入口是支付路径依赖
2016–2017	短视频爆发前	入口是推荐流
2025–2026	Agent 放量前	入口是“对话 +执行”的默认界面

资料来源：华泰研究

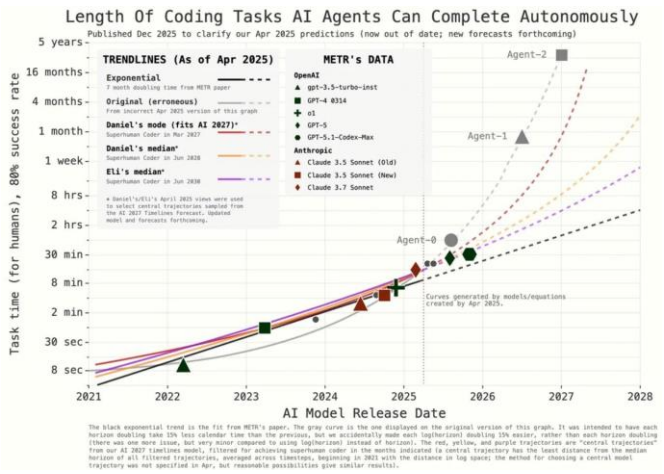
Agent 产品迅速进展的本质是模型能力在长链任务能力上跨越“奇点”。从 25 年年底发布的模型来看，在长链任务能力上完成了显著的跨越，我们认为这是目前 Agent 能力提升的根本因素。之前的 Agent 工程化的卡点在于长链任务完成度不佳，随任务复杂度提升，Agent 会在错误的进程中陷入死循环。而 25 年 12 月发布的 Claude 4.5 Opus 在任务时长上实现了显著提升，从 METR 的测评结果来看，Claude 4.5 Opus 在能够独立完成的任务所需人类工作时间的指标上实现了增长速度的突破。2019-2024 年：模型独立完成任务的时长每 7 个月翻一倍，2024-2025 年，这个数字缩短为 4 个月。根据 Anthropic 的预测，26 年年底 Agent 将独立完成人类半周的工作量。

图表10：METR 测评集结果



资料来源：METR 官网、华泰研究

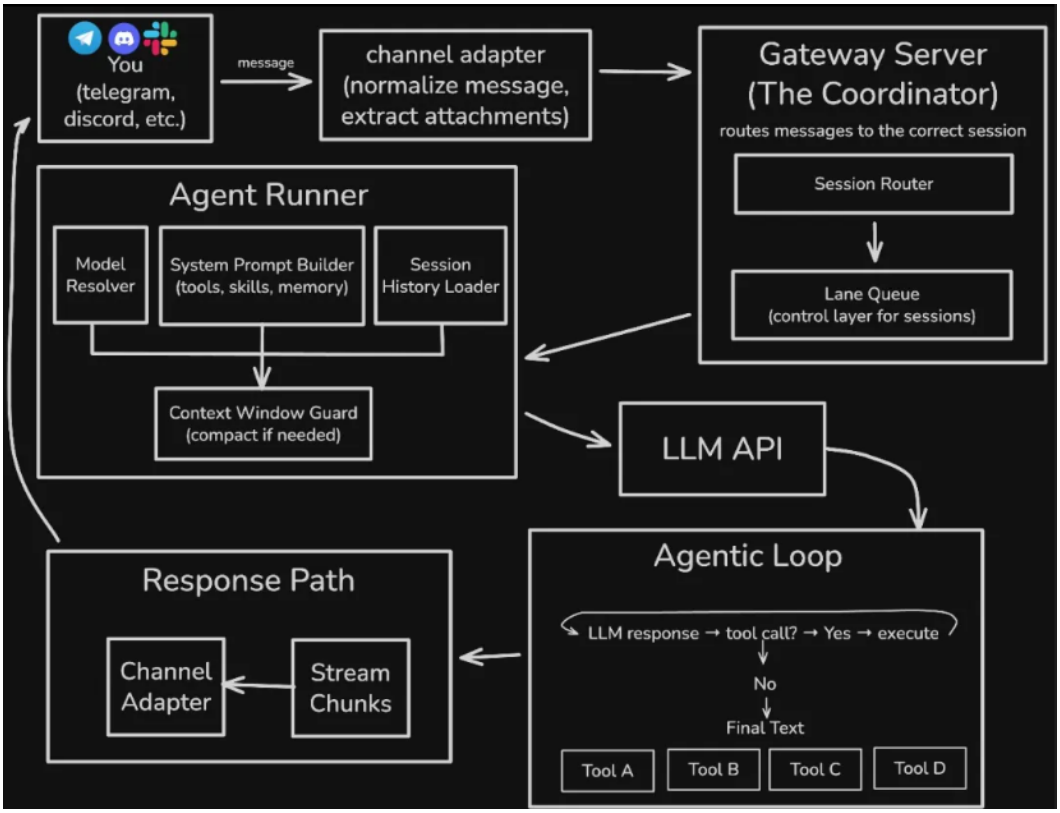
图表11：METR 测评集结果



资料来源：METR 官网、华泰研究

Agent 的推理范式是复杂流程、连续执行的，算力消耗较大。传统 Chat 模式下，推理的核心特征是：单轮或短链路、强人工介入、Token 使用高度离散、不可持续。而 Agent 模式的本质变化在于三点：多步规划、工具调用（Tool / API / MCP / VM）、长时间运行与状态保持（Persistent Context），这三点共同决定：Agent 的 token 消耗不是“对话量”的线性函数，而是“任务复杂度 × 执行时长”的指数函数。以 Claude Code 为例，拆分 Claude Code 工作流程来看，我们预测，1）启动阶段，只打开几个小文件，agent 思考一次约消耗 1-2 万 token；2）随工作流程的深入，一般打开 20 个以上的文件，Claude 进行工具调用、编写代码、外部交互、debug 与代码测试等工作，如上文提到这个过程不断地再重复上下文理解、执行操作与结果验证，则每一步约消耗 20 万 token。总体量化来看，保守假设完成一个小项目需要 5-10 步左右的工作，则完成一个项目需要消耗百万级别 token（完成复杂项目的消耗会更大），这与 chatbot 单次交互消耗一千左右 token 相比，算力消耗提升 3 个数量级。从 Claude Code 技术文档来看，官方建议每名用户每分钟的 token 限制设置在 20-30 万 token，算力消耗较大。

图表12：Open Claw 工作流程架构图



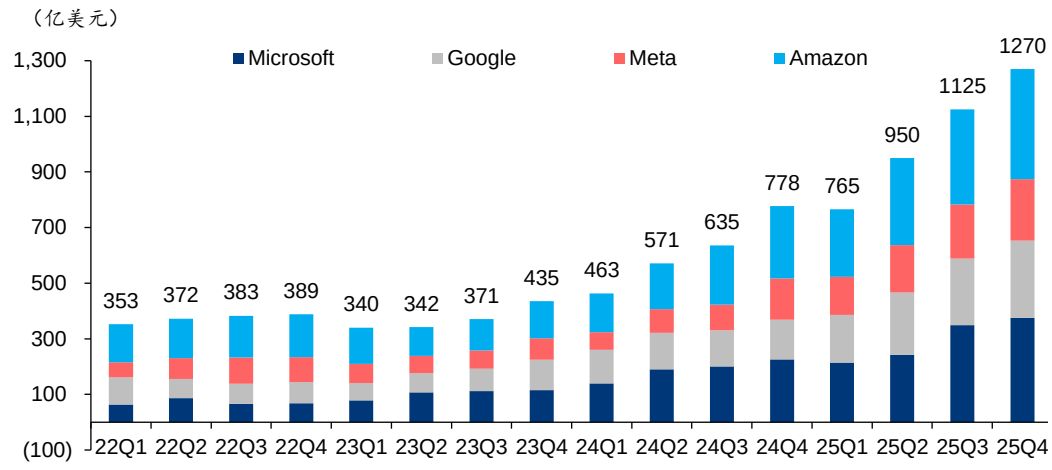
资料来源：Open Claw 官网、华泰研究

我们判断，2026 年将成为 AI Agent 推理端从“能力验证”走向“Agent 规模化应用”的关键拐点年。与 2023-2025 年以大模型能力跃迁为主线不同，下一阶段产业主线不再是模型参数或 benchmark 的线性提升，而是 Agent 形态驱动的 token 使用方式结构性变化。

海外大厂 Capex 持续增长，AI 需求表述乐观

海外四大 CSP Capex 持续快速增长，指引乐观。根据海外四大 CSP 季报来看，CY2025Q4 资本开支合计 1,270 亿美元，同比+63%，环比+13%。CY2025Q4 Microsoft/Google/Meta/Amazon 的资本开支分别为 375、279、221、395 亿美元，同比+66%/+95%/+49%/+52%，环比+7%/+16%/+14%/+15%。从指引角度来看，Google、Meta、Amazon 均明确上修 Capex 指引，Google 表示预计 2026 年全年资本支出将在 1750 亿美元至 1850 亿美元之间，投资额将在年内逐步增加；Meta 表示预计 2026 年资本支出(包括融资租赁本金支付)将在 1150 亿至 1350 亿美元之间；Amazon 表示将总体投入 2000 亿美元，大部分投入投向 AWS。

图表13：海外四大 CSP Capex 持续快速增长

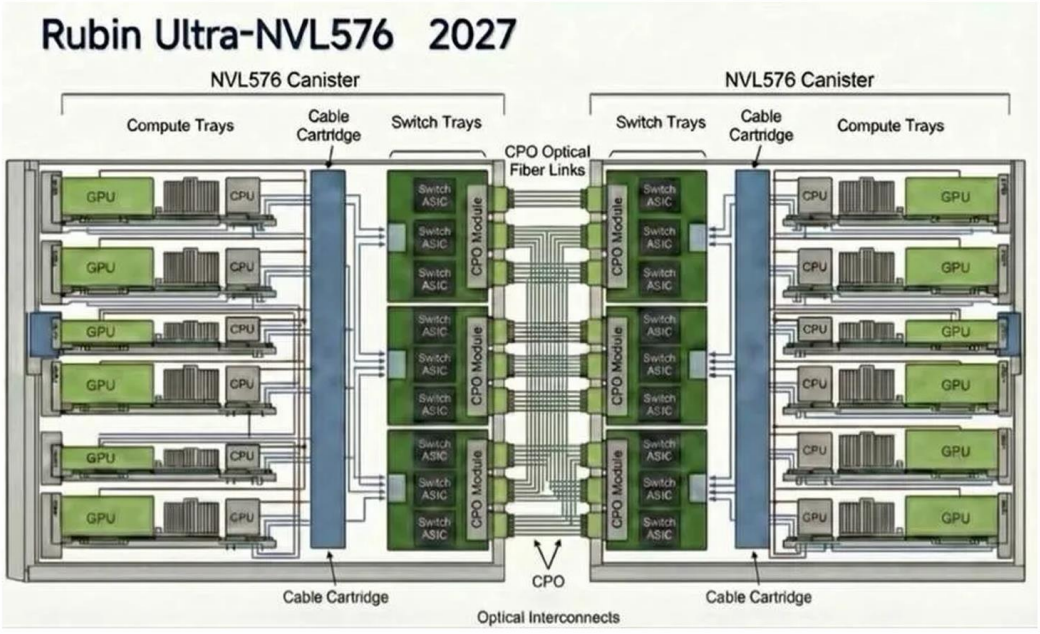


资料来源：Bloomberg、各公司业绩会、华泰研究

下一代光互连方案 CPO/NPO 雏形初现，关注 GTC 大会催化

NPO 方案下的“光进机柜”雏形初现。根据 AYZ 发布的文章，CPO 方案的 Rubin Ultra 机架将采用 NVL72 x 2 的架构，即由两个机柜组成。每个机柜内部仍保留现有 NVL72 机架的连接方式：compute trays 与 NVSwitch trays 通过 cable cartridge 相连；相邻两个机柜的 NVSwitch trays 则通过 CPO 光引擎与光纤实现互联，从而构成包含 144 张 GPGPU 卡（相当于 576 颗 GPU die）的 Rubin Ultra 机架（每张 Rubin Ultra GPGPU 卡集成四颗 GPU die）。根据 AYZ 发布的供应链调研，CPO 版本的 NVSwitch 芯片将基于 NVIDIA 去年发布的 Quantum X800 CPO switch 芯片改款而成。每颗 NVSwitch 芯片周围共封装 4.5 颗 3.2T 光引擎（其中半个光引擎的带宽共享给相邻 NVSwitch 芯片）。一个 NVSwitch tray 内配置 6 颗 CPO NVSwitch 芯片，每 3 颗置于同一块 PCB 板上，tray 内上下堆叠两块 PCB 板。据此计算，单颗 NVSwitch 芯片交换容量为 3.2T × 4.5 = 14.4T，一个 NVSwitch tray 的交换容量为 14.4T × 6 = 86.4T。

图表14：Rubin Ultra NPO 互联方案



注：该方案系根据公开产业链调研资料整理，非英伟达官方发布方案，具体方案关注 3 月 GTC 大会英伟达官方发布
资料来源：AYZ、华泰研究

CPO/NPO 等下一代光互连方案已经进入产业化落地的元年，核心矛盾在于渗透节奏。英伟达在 2026 年 CES 大会上发布了 Spectrum-X Ethernet CPO，从产业进展来看，CPO 进入产业化落地元年。根据英伟达在 2 月 3 日举办的 AI 网络研讨会上表示 26 年上半年，CoreWeave、Lambda、德克萨斯高级计算中心会率先部署 Quantum-X InfiniBand CPO；下半年将推出 Spectrum-X 以太网 CPO，同时推动落地更多 AI 和超算相关的部署项目，这也意味着 CPO 技术将进入规模化商用的阶段。随 Agent 推理的拓展，Scale up 域呈现逐步扩大的趋势，铜缆的 scale up 模式正逼近极限（200Gbps/lane 速率下，铜缆的传输距离最多仅为 2m），且单通道带宽翻倍的难度日益增加，光互联进入 Scale up 领域逐步成为共识。根据 Coherent CEO 在 FY26Q2 业绩电话会的表述，未来十年可插拔光模块在 Scale-Out 仍是主流；CPO 主要增量来自 Scale-Up，且市场规模较 Scale-Out 大几个数量级。我们认为英伟达下一代互联方案将逐步清晰，3 月份的 GTC 大会或发布官方架构，关注 CPO/NPO 的渗透节奏。

AI 应用：云厂业绩加速兑现，静待 SaaS 预期修正

海外科技公司 25Q4 业绩持续披露，云厂商业绩加速兑现，SaaS 市场预期偏悲观。数据层面上看，25Q4 AI 应用公司业绩基本超市场预期，26 年指引呈小幅上修状态；预期层面上看，SaaS 板块的悲观预期或持续加重，云厂商核心担忧在于 AI 商业化能否匹配 26 年高增的 CapEx 指引。我们认为，26 年全球 AI 应用有望全面加速，云厂商业绩有望持续加速，部分 SaaS 公司有望实现产品价值下沉与企业价值重估，看好海外 AI 商业化进展提速。

图表15：25Q4 AI 应用板块典型公司业绩统计

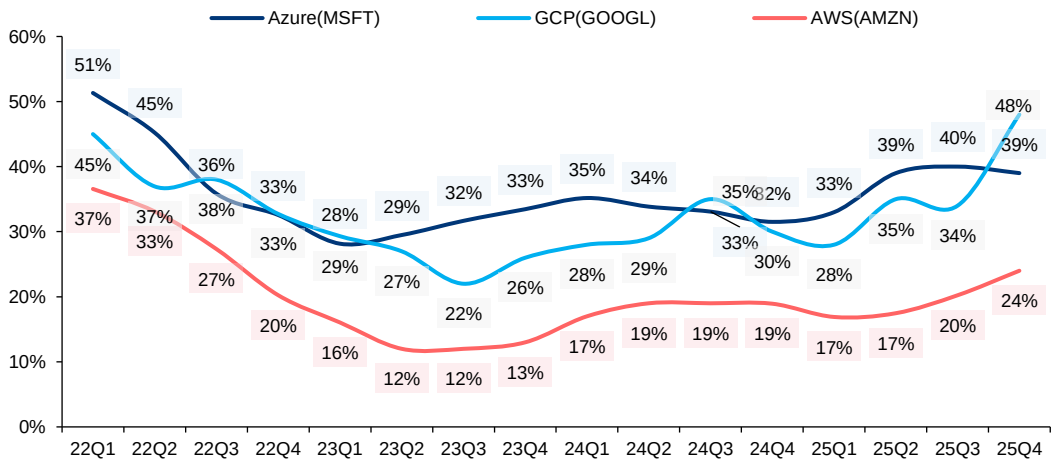
分类	公司	总营收 vs 一致预期	净利润 vs 一致预期	CY 26Q1收入指引/亿美元	市场预期/亿美元
公有云	Microsoft	Beat 1.2%	Beat 33.3%	807-817	814
	Google	Beat 2.2%	Beat 7.5%	/	/
	Amazon	Beat 0.9%	Miss 0.9%	1,735 ~ 1,785	1,754
社交应用龙头	Meta	Beat 2.5%	Beat 7.6%	535 ~ 565	513
AI软件	SAP	Miss 0.5%	Beat 11.5%	/	/
	Palantir	Beat 5.0%	Beat 8.9%	15.32 ~ 15.36	13.26
	ServiceNow	Beat 1.1%	Beat 3.4%	36.50 ~ 36.55	35.81

注：前两列为公司业绩与一致预期对比情况；ServiceNow 指引为订阅收入指引，其余公司为总体营收指引
资料来源：Visible Alpha、华泰研究

云厂商业绩提速，全栈式竞争趋势凸显

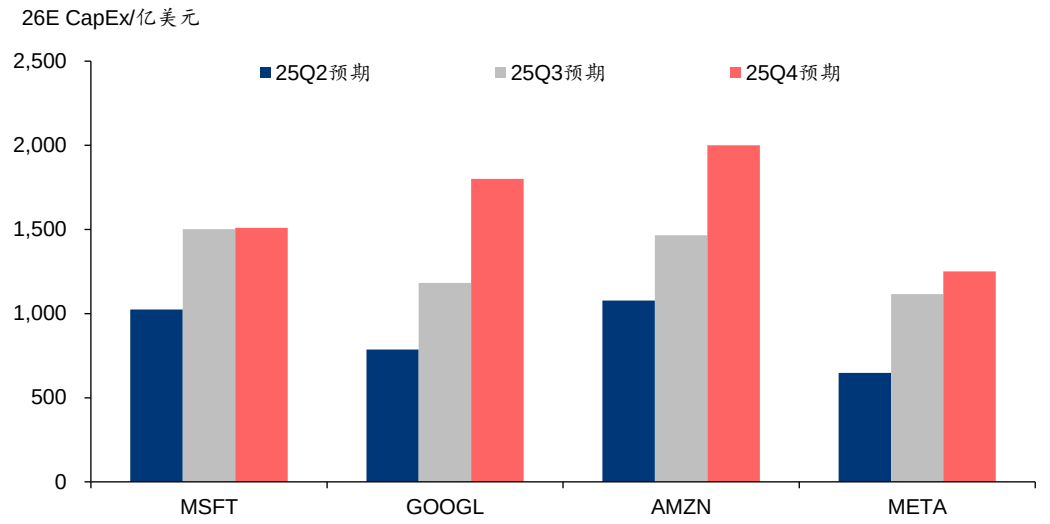
25Q4 Google、Amazon 云业务收入持续加速，算力供给能力仍是商业化的核心制约。25Q4 Microsoft、Google、Amazon 云业务收入同比增速分别为 39%、48%、24%，其中，Google、Amazon 云业务营收增长显著提速，主要受下游需求放量拉动；Microsoft 云业务营收增速边际放缓，主要由于公司主动调整算力分配重心，根据公司电话会表述，若 GPU 资源全部服务云场景，Azure 云收入增速将超过 40%。综合来看，海外云厂商的下游需求持续高增（Microsoft、Google、Amazon RPO 分别环比+59%、+50%、+22%），算力供给能力仍是订单收入转化的核心制约因素。同时，海外大厂 26 年全面加速 Agent 产品布局，推理算力需求同步加速，25Q4 Google、Amazon、Meta 大幅上修 26 年 CapEx 指引。

图表16：22Q1-25Q4 三大公有云服务商云业务收入同比增速



资料来源：Visible Alpha、华泰研究

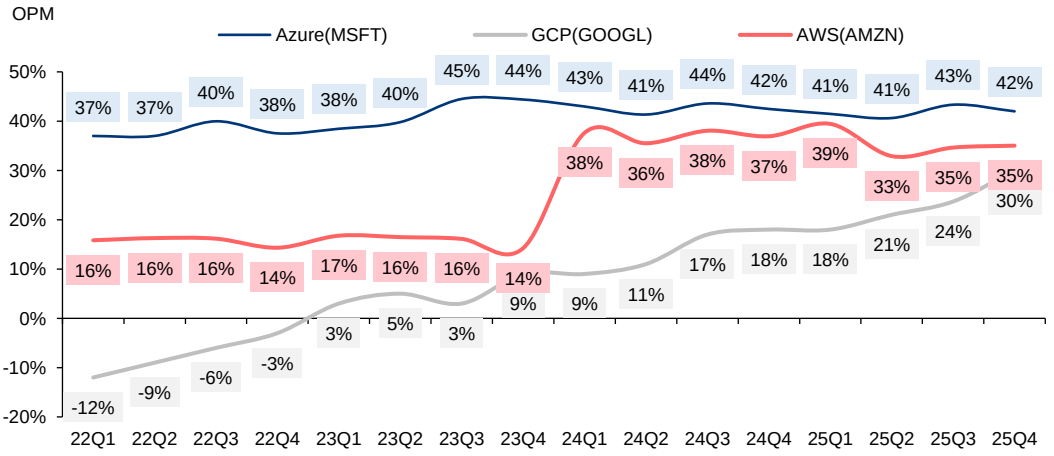
图表17：Microsoft、Google、Amazon、Meta 26 年 CapEx 预期持续上修



资料来源：Visible Alpha、公司官网、华泰研究

云业务进入全栈竞争新周期，海外大厂 26 年或加速算力、模型、应用全栈式布局。海外大厂 AI 全栈式竞争趋势明显，受益于自研芯片规模化与利用率提升，Amazon、Google 云业务营业利润率环比改善，25Q4 Amazon、Google 云业务营业利润率分别环比提升 0.39pct、6.40pct。我们认为，26 年海外大厂将加快算力、模型、应用全栈式布局。1）算力层面：Microsoft、Amazon 有望加速新一代自研芯片落地；2）模型层面：关注多模态、超长上下文、代码等能力提升；3）应用层面：Meta、Google 等厂商 2C Agent 流量竞争或将加剧。

图表18：22Q1-25Q4 三大公有云服务商云业务 OPM



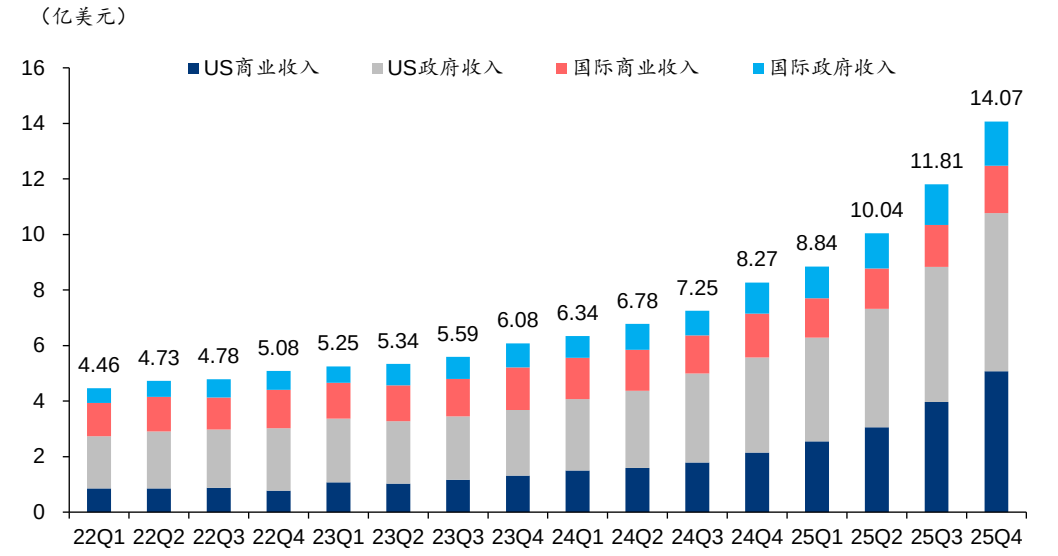
资料来源：Visible Alpha、华泰研究

SaaS 加速产品价值下沉，Palantir 提供范式参考

传统 SaaS 加速产品价值下沉，加速拓展数据中台、Agent 中台等产品品类。从市场预期来看，传统 SaaS 的 TAM 收缩预期持续强化，能否实现预期反转，关键在于传统 SaaS 能否实现产品价值中枢下沉。从进展来看，25Q4 SAP、ServiceNow 全力加强中台类产品推广。1）ServiceNow：25Q4 AI Control Tower（企业级 Agent 管理平台）订单量环比增长近 3 倍，合同价值高于 100 万美元的 AI 订单单季新增 35 笔；2）SAP：截至 25Q4，Business Data Cloud（BDC，企业级数据中台）累计合同金额约 20 亿欧元，25Q4 最大 50 笔交易中 90%包含 AI 或 SAP Business Data Cloud（BDC）。我们认为，传统 SaaS 具备数据+流程的优势卡位，若能完成产品价值中枢的数据层下沉，其 AI 业务空间或将充分打开。

Palantir 提供范式参考，国内软件厂商具备相似的产品+客户优势。区别于传统 SaaS，Palantir 具备“数据能力强、满足客户定制化需求”的差异化优势，得益于公司业务的先发布局与产品的成熟度，25Q4 Palantir 业绩加速成长，25Q4 总营收 14.07 亿美元，同比+70%，其中美国商业收入 5.07 亿美元，同比+137%，美国政府收入 5.70 亿美元，同比+66%；同时，公司积压订单持续加速，25Q4 RPO 达 42.10 亿美元，同比+43%，下游需求快速增长。我们认为，基于过去软件的定制化需求，国内软件厂商与企业客户的业务流程、产品需求形成了深度绑定关系，在 AI 需求下同样具备 Palantir 类似的产品+客户优势。同时，国内企业基于 AI 需求逐步加大数据工程投入，看好企业数据治理后的 AI 应用放量空间。

图表19：受益于 AI 需求，Palantir 营收加速增长



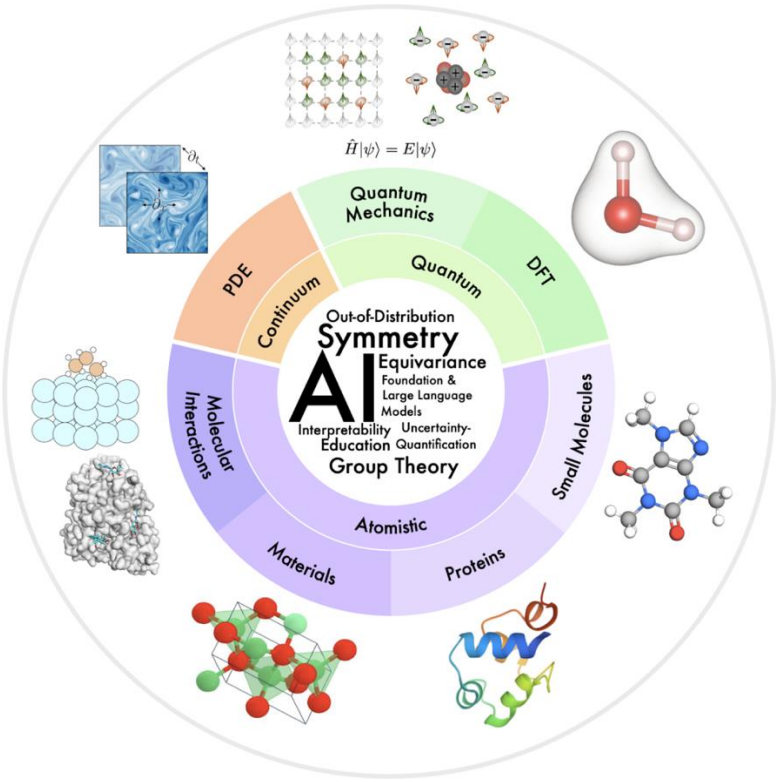
资料来源：Visible Alpha、华泰研究

AI4S：生物制药商业化最快，材料领域有望突破

AI4S 引领科研走向第五范式

AI 正通过赋能量子、原子与连续介质系统中的高级建模、仿真与预测，引领科研革命。这些系统覆盖量子力学、分子相互作用、材料科学与生物体系等多元科学领域。AI 正在多个领域加速科学发现，提升计算精度，并为复杂问题提供了可扩展的解决方案。AI for Science（AI4S）的研究范式打破了传统“实验发现”或“手工推导方程”的局限，不仅是工具的更新，更是科研逻辑的重塑，利用 AI 的强大算力跨越计算瓶颈，将量子力学到宏观力学的物理法则连接起来，成为物理、化学、生物及材料科学领域不可或缺的“科学基座”。

图表20： AI for Science 研究范式架构

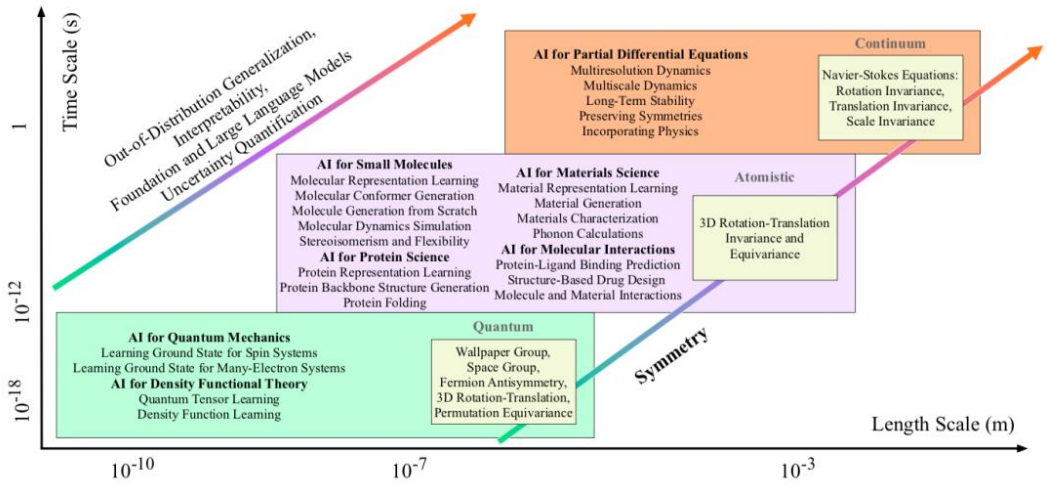


资料来源：air4.science、华泰研究预测

根据空间和时间尺度，AI4S 的研究可以划分为以下核心领域：

1) 量子体系：包括量子力学（利用神经网络高效学习波函数）和密度泛函理论（DFT）（预测分子的电子结构与物理性质）。
2) 原子体系：涵盖小分子（药物设计与分子生成）、蛋白质（结构预测与设计）、材料科学（晶体性质预测）以及分子相互作用。
3) 连续体系：利用 AI 作为代理模型（Surrogate Modeling）加速求解描述宏观物理过程的偏微分方程（PDEs）。

图表21：AI 正在赋能科学研究领域不同时间和空间尺度

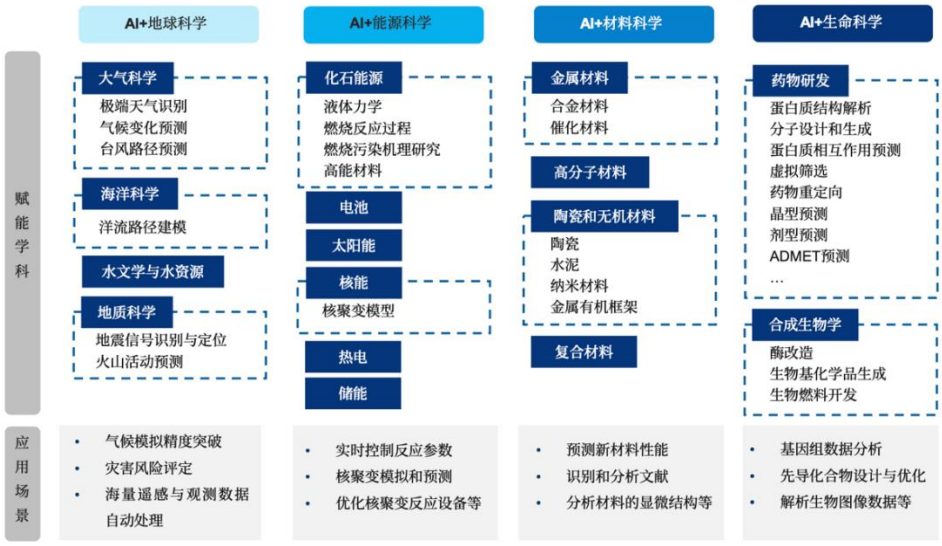


资料来源：air4.science、华泰研究

AI4S 已在多个科学领域深度融合与广泛应用，实现利用 AI 处理复杂数据、构建预测模型、自动化分析流程，并深入至发现、设计、优化、控制等核心环节。随着 AI 与越来越多的科学研究领域深度嵌入，标志着 AI 已从辅助工具转变为驱动科学发现的核心引擎，为未来更大规模的科学突破奠定了基础。目前 AI4S 的核心应用领域包括：

1) 生命科学：覆盖从基础研究（蛋白质结构解析、相互作用预测）到药物研发全流程（虚拟筛选、分子设计、ADMET 预测），并延伸至合成生物学（酶改造、生物基化学品生成）。
2) 材料科学：贯穿所有材料类别（金属、催化、高分子、纳米材料等），实现性能预测、新型配方设计、微观结构分析，加速从发现到应用的进程。
3) 地球科学：应用于大气、海洋、水文、地质各分支，实现高精度气候模拟、极端天气识别、灾害风险评估，并能自动处理海量遥感与观测数据。
4) 能源科学：优化从传统化石能源、燃烧过程到新能源（太阳能、电池、核聚变）的全链条，包括反应参数控制、设备优化与系统模拟。

图表22：AI4S 在多个科学领域广泛应用



资料来源：沙利文、华泰研究

AI 制药从小分子走向大分子，2026 年 AI 抗体合作有望加速

2026 年以来，AI 制药正从传统的小分子药物快速扩展至大分子生物药（尤其是抗体药物），标志着 AI 在药物研发中正迈向更高复杂度和更高价值的前沿。AI 能够实现从小分子到抗体的跨越，核心在于其多尺度建模能力：1. 原子尺度与蛋白质建模：AI4S 研究范式已覆盖从小分子到蛋白质的完整图版；2. 物理规律嵌入：通过融入群论（Group Theory）和等变性（Equivariance）等物理对称性约束，AI 能够更精准地模拟大分子之间的相互作用。

从商业化进展来看，2026 年 1 月，赛诺菲与 Earendil Labs（华深智药子公司）在自身免疫性疾病和炎症性疾病领域再次达成重大合作，合作范围扩大：从首次合作（2025 年 4 月）针对两款特定的双特异性抗体，扩大至覆盖多个双特异性抗体药物项目；投入力度升级：合作总价值从 18.45 亿美元提升至 25.6 亿美元，验证跨国药企对 AI 驱动的抗体药物研发投入正迅速加大并走向深化。

图表23：2026 年赛诺菲与 Earendil Labs 扩大 AI 抗体合作

药企	合作方	时间	疾病领域	首付款	合作内容
赛诺菲	Earendil Labs	2026年1月	自身免疫性疾病和炎症性疾病	1.6亿美元（首付款、近期里程碑付款）	多个双特异性抗体药物，合作总价值最高可达 25.6 亿美元
赛诺菲	Earendil Labs	2025年4月	自身免疫性疾病和炎症性疾病	1.25亿美元	两款双特异性抗体药物全球独家授权，17.2亿美元的开发和商业里程碑付款，高个位数到低两位数不等的销售分成

资料来源：赛诺菲、华泰研究

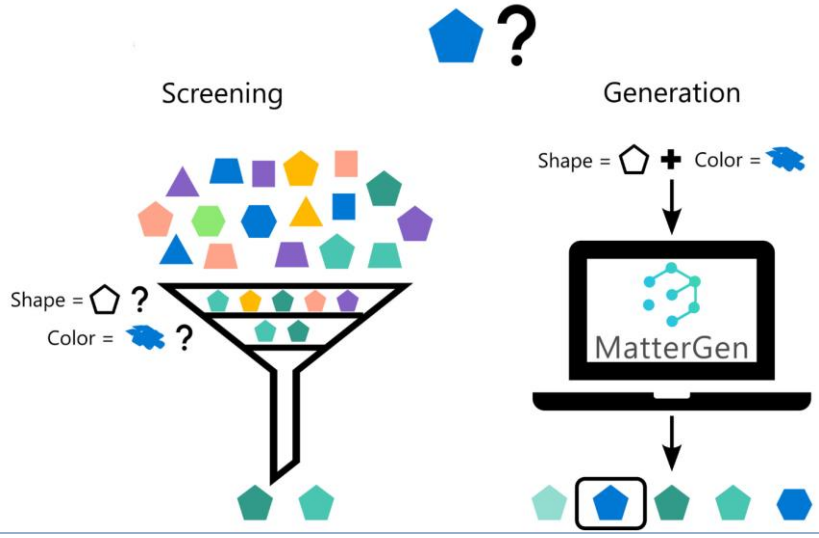
我们持续看好 AI 制药在 2026 年的商业化前景，预计行业将呈现小分子药物合作深化与大分子抗体领域合作爆发的双轮驱动格局：

- 小分子领域：商业化路径已得到验证，预计药企与 AI 公司的合作将持续活跃并趋向多元化。合作重点将从早期的靶点发现，进一步延伸至临床前候选化合物（PCC）的优化、预测 ADMET 性质以及差异化分子设计，提升临床成功率和开发效率。
- 抗体等大分子领域：有望成为 2026 年最大的增长亮点。继赛诺菲等 MNC 与 AI 制药公司达成高额战略合作后，AI 在抗体发现、人源化优化、双/多特异性抗体工程方面的潜力正在逐步显现。AI 能够高效探索广阔的蛋白质序列空间，设计出具有更佳特异性、亲和力及可开发性的新型抗体，预计更多大型药企将跟进布局，相关 License-out 交易的首付款及潜在总合作额有望创下新高。

AI 构建新材料研发新范式，驱动产业跨越“创新鸿沟”

国际巨头积极布局“AI+新材料”领域，MatterGen 开启“按需定制”材料新时代。材料创新是驱动能源转型与技术革命的底层支撑，从锂电池到超导材料，每一次突破都深刻改变了数十亿人的生活。然而，传统“大海捞针”式的实验试错和被动的数据库筛选，在面对无限的材料空间时已触及效率天花板。以微软为代表的国际科技巨头正积极布局 AI + 材料领域，重塑科研范式：1. 从“筛选”进化为“生成”：微软研究院推出的 MatterGen 突破了传统局限，不再是从既有库中挑选，而是直接根据导电性、磁性等应用需求，从零生成物理属性稳定的新材料；2. 多维度属性优化：MatterGen 能针对材料的化学成分、几何结构及物理稳定性进行协同优化，极大拓宽了探索空间。

图表24：材料设计传统筛选方法和生成式 AI 方法对比

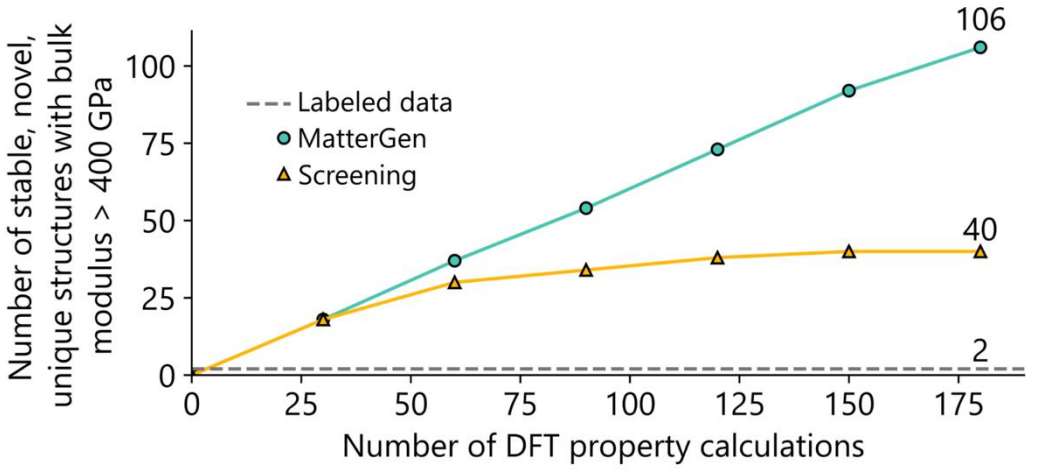


资料来源：微软研究院、华泰研究

MatterGen 相比传统筛选方法展现出革命性优势。其核心在于能够跳出已知材料库的局限，探索并创造全新的材料结构空间。在寻找体积模量大于 400 GPa（即极难压缩）的高性能稳定新材料任务中，传统筛选方法（Screening）受限于现有数据库，随着计算量增加，其发现新材料的效率迅速饱和，增长乏力。相反，MatterGen 凭借其生成式 AI 能力，能够从头设计自然界可能尚未存在或未被记录的候选材料。因此，其发现效率随着计算资源的投入几乎呈线性高速增长，在同等计算成本下，发现的合格新材料数量远超传统方法数个量级。

这标志着材料研发范式从“在已知库中筛选”转向“在无限可能中定向创造”。AI 不仅大幅提升了探索效率，更关键的是突破了人类经验和现有知识的边界，为发现具有超常性能（如超高硬度、超导性）的下一代颠覆性材料打开了全新道路。

图表25：MatterGen 能够访问未知材料的完整空间

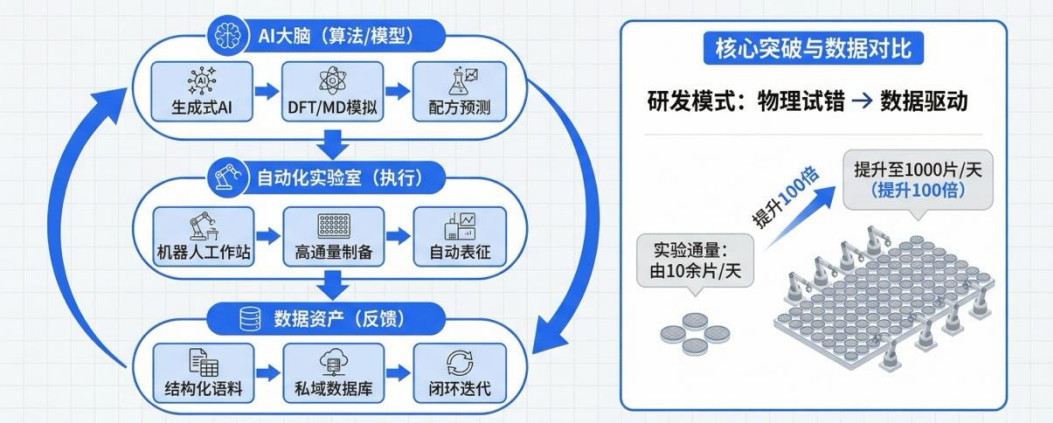


资料来源：微软研究院、华泰研究

从产业进展来看，2026 年 1 月，晶泰控股与晶科能源子公司签署 AI+自动化高通量叠层太阳能电池研发战略合作协议，双方将共同成立合资公司，共建全球首个“AI 决策-机器人执行-数据反馈”全闭环叠层电池智造线，为不同的应用场景开发高效率、高稳定性的太阳能电池产品。

针对钙钛矿叠层电池在效率、稳定性和成本平衡上的工业化难题，双方通过智能化范式打破了传统“试错法”的局限：1. 数据编码化：首次将上百种材料配方与工艺参数编码化，整合产业经验、文献挖掘及自动化实验生成的多种高精度数据，建立专属数据库；2. 多模态 AI 引擎： 开发由大语言模型驱动的研发引擎，融合多任务深度学习与物理约束模型，实现材料筛选与性能预测的进化迭代；3. 高通量闭环：构建业内首条千平米级实验线，实现每日 1000 片的百倍通量提升，通过“设计-实验-反馈”的自动化闭环快速优化工艺。

图表26： AI+钙钛矿发现构建材料研发新范式



资料来源：华泰研究

我们认为 2026 年 AI 新材料预计将成为 AI4S 的重点应用与投资方向。以“AI 预测+自动化实验”为核心的研发闭环已进入规模化验证阶段，能十倍级缩短传统研发周期。未来 AI 有望从光伏钙钛矿向固态电池电解质、高温超导材料、半导体光刻胶等更高附加值的领域复制。AI 不仅加速材料发现，更通过数字化工艺优化直接推动产业化，是实现制造产业升级的核心引擎。

月专题：Agentic Coding 加速迭代，关注 Agent 进展

基于对 Agentic Coding 和 OpenClaw 进展的观察，我们认为 Agent 应用正在加速，有望带来软件行业重构。我们判断 2025 年是 Agent 元年，2026 年可能进入 Agent 加速落地期，主要体现在两方面：一是 Agentic Coding 的迭代速度会大幅加快；二是国内外大厂会激烈争夺个人 Agent 助手的超级入口，均会成为下一轮 token 加速的重要推手。我们认为 Agentic Coding 的快速迭代会加速软件行业的重构，软件开发成本面临“杰文斯悖论”：未来个性化的、由 AI 生成的软件有望爆发，但单体软件的价值可能下行。

AI Coding：Agent 推动产品能力快速迭代

AI Coding 是发展最快的企业级 AI 应用之一，从简单补全向 Agentic Coding 进化。根据 Menlo Ventures 的数据，AI Coding 已成为企业 AI 最热门的应用场景，占企业部门级 AI 支出的 55%。从功能看，目前 AI coding 正在从代码补全向自主执行任务的 Agentic Coding 进行范式跃迁。其中最具代表性的产品，就是 Anthropic 基于最新的 Claude Opus 4.5 模型升级的 Claude Code 2.1（面向开发者）和 Cowork（面向非程序员），一定程度带来了美股存储/CPU 等加速上涨和美股软件股的加速下跌的分裂走势。它们有两个显著特点：

1) **200K tokens 的超长上下文：**Cowork 通过本地化存储长期保存上下文需求，这些都指向了存储需求的增加，存储需求可能从 HBM 向 DRAM、NAND 外溢；
2) **Agentic 工作流，具备多工具、多 Agent 的调用能力：**让 CPU 在动态批处理、吞吐率优化、Agent 调度、工具编排等环节的重要性空前提升。

AI Coding 演进：从小型系统拓展到中型系统

AI Coding 高速迭代：我们判断 25 年年中 AI Coding 仅能自主实现 1000 行以内代码的简单程序，预计目前 AI Coding 已经拓展到能够构建 5000-20,000 行代码的中型系统，预计未来两年可能实现中型系统完全自主生成；五年后大型生产级系统的自主生成将成为可能。

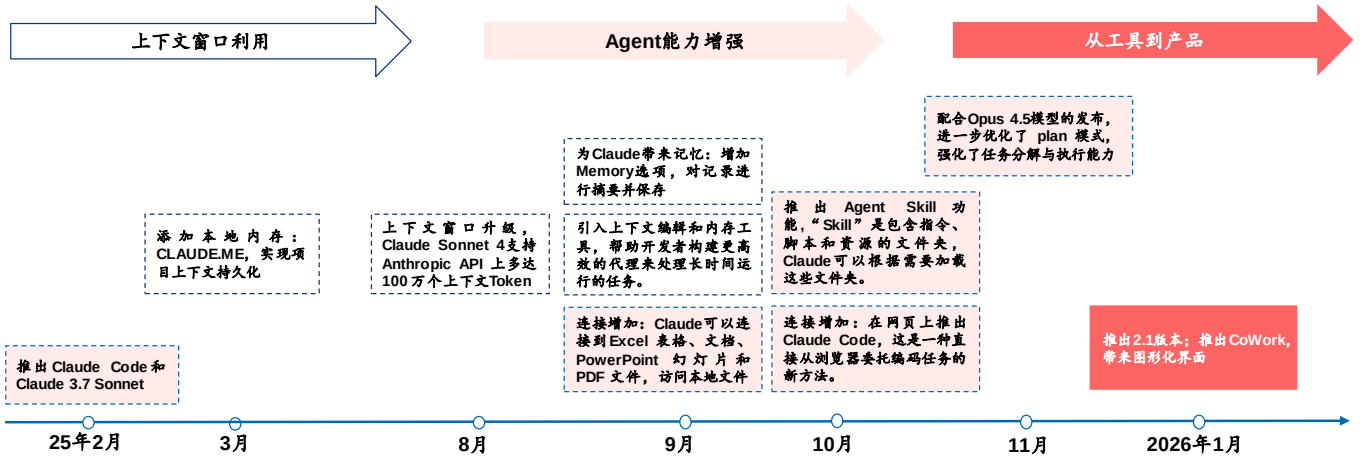
图表27：AI Coding 高速迭代

资料来源：华泰研究预测

最具代表性的 AI Coding 产品 Claude Code 在 25 年以来经历多项功能迭代，我们关注到三个趋势：

1) **Agent 能力的增强，AI Coding 进化到 Agentic Coding。**LLM 时代早期，AI Coding 工具的主要功能是做代码预测和完成，Agentic Coding 具备项目级上下文，支持多工具多 Agent 调用，丰富了 AI Coding 的工作流。
2) **上下文窗口的合理应用，Claude 将各类记忆进行压缩，从而在有限的上下文窗口增强效果。**我们认为 Claude Code 增加的多项功能，例如 CLAUDE.md、Skill、Memory，都通过人工方式进行信息压缩，从而极致地应用有限上下文窗口。
3) **从工具到产品：**Cowork 的推出，有力推动了非技术人员对 AI Coding 工具的使用。

图表28： Claude Code 功能

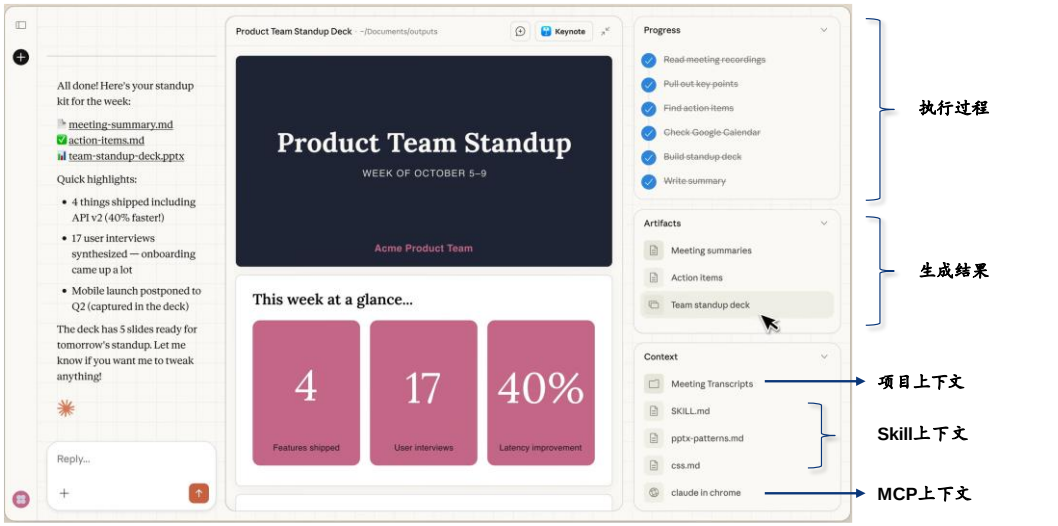


资料来源：Claude 官网博客、华泰研究

AI Coding 各类应用产品推出，中美加速布局

Cowork 推出，降低软件构建门槛。2026 年 1 月 12 日，Claude Cowork 发布。从功能来看，Cowork 具备和 Claude Code 类似的 Agentic Coding 能力，因此具备项目级的上下文能力和多 Agent 多工具使用能力，而 Chatbot 的上下文通常局限于对话窗口与用户上传内容。与 Claude Code 相比，Cowork 大幅提升使用便利性，用户不需要进入命令行终端编码并对项目进行配置，Cowork 可以自行读取上下文并执行任务。我们认为 Cowork 可以理解为对 Claude Code 的 UI 封装，结合了聊天窗口的易用优势和 Code 产品的项目级能力，使得非技术人员也可以建立项目级代码仓库，降低了复杂项目的建设门槛。

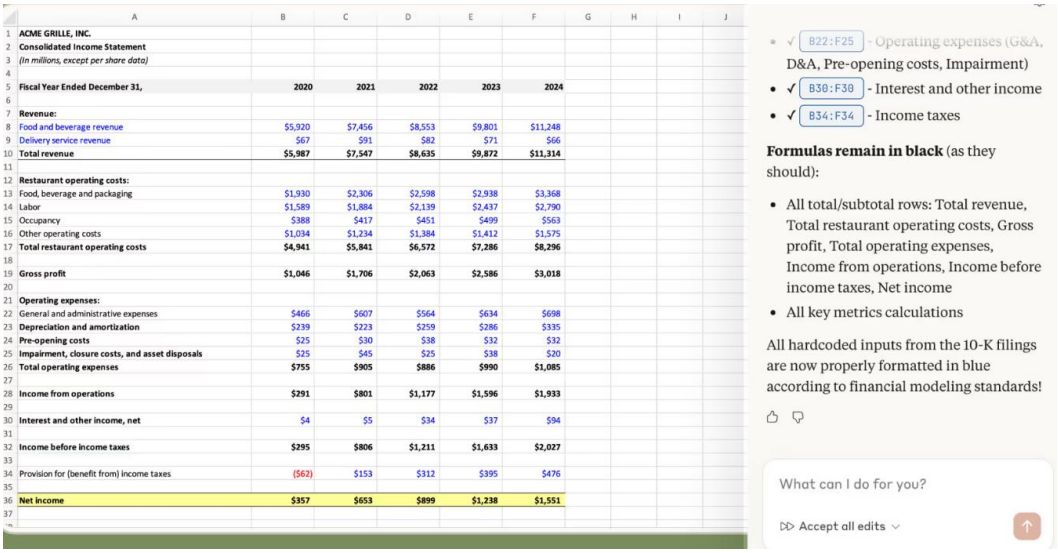
图表29： Cowork 运行过程示意图



资料来源：Claude 官网博客、华泰研究

Claude 拓展在 Excel 和法律场景应用，提升特定场景的使用能力。25 年 1 月，Claude in Excel 推出，与此前应用于命令行的 Claude Coding 相比，Claude in Excel 应用于 Excel 界面，只需快捷键即可调用，能理解整体 Excel 工作簿，包含从嵌套公式到多个工作表依赖关系，可以获取单元格级别的引用解释，并在保留公式的同时更新假设，有利于推动 Claude 在办公端的应用。25 年 2 月初，Claude Legal 推出，作为 Claude Code 的插件，用户在本地配置后，可以让 Claude Code 具备自动完成合同审查、保密协议筛选、合规工作流程、法律简报和模板化回复的功能，而且可根据组织流程和风险承受能力进行个性化配置。

图表30： Claude in Excel



资料来源：Claude 官网博客、华泰研究

中美加速 AI Coding 布局，26 年有望看到应用加速。目前中美大模型厂商均在 AI Coding 领域进行布局，以 Coding 场景为入口，增加用户对自身模型的调用。此外，一些独立 AI Coding 产品商业化加速、用户规模扩大。我们认为随着模型能力的增强和 Coding 产品功能的拓展，26 年有望看到应用加速。

图表31： 主流 AI Coding 公司：经营数据

厂商/产品	最新估值	最新ARR	用户规模	核心优势/2026关键动态
ClaudeCode	Anthropic公司估值3500亿美元	自25年Claude Code推出以来，短短6个月内实现10亿美元收入		企业级为主，Claude Code主打“全栈代码代理”（跨文件/应用/浏览器执行），吸引追求极致效率的开发者； 代理式编程全闭环、长上下文、安全合规
Anysphere (Cursor)		约10亿美元（2025.11）	5万+团队，超半数财富500强采用	**AI原生IDE：**ARR在一年内从$1亿飙升至$10亿； **首创“Vibe Coding”体验**，通过“Bugbot”等工具实现全流程闭环。 **轻量IDE+多模型切换”，吸引重视灵活性的开发者。**
GitHub Copilot			2600万+用户，150万付费用户（25.10月），90%的财富100强企业采用	**依托VSCode生态和GitHub仓库，强调“IDE深度集成+生态协同”，适合企业级团队规模化部署。**
Cognition (Devin)	$102亿(2025.09)		聚焦顶级金融与科技大厂	**全自动智能体：定位“AI软件工程师”而非助手；** 2025年7月完成对Windsurf核心资产（Google出资$2.4亿进行“反向人才收购”，挖走了Windsurf的CEO、联合创始人及核心R&D团队并获得技术授权 Cognition收购了Windsurf剩余资产、产品、品牌、IP及存量客户）的收购,强化了IDE交互能力。
Replit	$90亿(2026.01)	$2.4亿(2025)；目标2026达$10亿	4000万+用户	**让非技术人员通过自然语言部署完整应用**，引领“人人皆可编程”趋势。
PoolsideAI	$120亿(2026.01筹备中)	早期商业化，Nvidia注资$10亿	聚焦政府、国防及高算力需求场景	**自研模型护城河：**不依赖外部模型，专注于构建具备极致逻辑推理能力的编程大模型。
Lovable	$66亿(2025.12B轮)			**8个月实现ARR从0到$1亿，主打“对话即应用”**，在欧洲市场增长极速。

资料来源：路透社、金融时报、Bloomberg、Cursor 官网、微软电话会、Replit 官网、Techfundingnews、华泰研究

图表32：美国大模型厂商在 AI Coding 领域的布局

大模型厂商	大模型/产品矩阵	最新进展与关键特性 (2026.01)
OpenAI（通用模型 + 生态协同）	**产品矩阵：**GPT-5.2 Codex	**核心特点：**强化复杂逻辑推理、长上下文修改、函数调用，适配企业级工程场景。**生态：**深度绑定 GitHub Copilot（微软），API 开放给第三方工具（如 Cursor）；**26年1月，推出超快 Codex**，推理时间减少 50%，实现近乎即时代码建议；
Anthropic（代理式编程标杆）	**基座模型：**Claude Opus 4.5/ **产品矩阵：**Claude Code（开发者）、Cowork（非程序员低代码）	**能力跃升：Opus 4.5 实现“阶跃式”提升，**SWE-bench 等基准测试领先。**生态扩张：**2026.1 推出协作工具生态；与微软深度合作，成为 Microsoft 365 Copilot 默认模型，Azure Foundry 直接集成，微软内部全面推广。**商业化：**API + 企业订阅双轮驱动，收紧第三方调用权限，强化自有生态壁垒。
Google（云 + IDE + 模型三位一体）	Gemini Advanced/Ultra/ **产品矩阵：**Gemini Code Assist（IDE 插件）、Vertex AI Codey（企业 API）、 Duet AI（Google Cloud/Workspace 集成）	**核心特点：**以Gemini Advanced/Ultra为底座，**主打多模态编程（代码 + 文档 + 图表理解）、云原生开发（GCP 深度集成）；Gemini Ultra 2.0 强化代理式能力****生态：**深度绑定 GCP、Android Studio，覆盖移动、云、Web 全场景开发。**开源协同：**推出 CodeLlama-70B Coding 专用模型，开源社区贡献度提升；
微软（生态壁垒最强，企业渗透第一）	**产品矩阵：GitHub Copilot** 包括Copilot Individual（C 端）+ Copilot Business（企业）+ Copilot X（全流程开发）	**核心特点：以GitHub Copilot为核心，**深度集成 GitHub、Visual Studio、Azure 生态。**模型协同：2025.9 推出“自动模型选择”，底层接入 Claude Sonnet 4**；2026.1 与 Anthropic 扩大合作，Azure Foundry 直接提供 Claude 模型，形成“双模型”壁垒。
xAI：Grok Code	**核心产品：Grok Code**	**核心特点：低成本、C 端为主、速度优先**2026.2 计划重大更新，强化复杂任务一键处理；输入成本 $0.2 / 百万 tokens，性价比突出

资料来源：OpenAI/Anthropic/Google/微软/xAI 公司官网、华泰研究

图表33：中国大模型厂商在 AI Coding 领域的布局

大模型厂商	产品矩阵	商业化进展
阿里巴巴 (通义灵码)	基于 Qwen 2.5/3	**云端一体化：**深度打通阿里云，支持企业级 RAG（私有知识库）检索
字节跳动 (Trae / MarsCode)	基于 豆包 (Doubao)	Trae 作为 AI 原生 IDE 成功出海
百度 (文心快码 Comate)	基于 文心一言 4.0/5.0	**全生命周期管理：**强调从需求文档、设计、代码到测试的全流程串联；集成百度内部研发经验。
腾讯 (腾讯云 AI 代码助手)	基于 混元 (Hunyuan)	**强调“智检员”功能，在代码评审（Code Review）和缺陷检测上表现突出，**新增代码辅助率达 50%。
DeepSeek (Code)	独立黑马 (开源)	

资料来源：阿里/字节跳动/百度/腾讯/DeepSeek 官网、华泰研究

AI Coding 对软件行业影响：价值锚点转移，相关公司分化

我们认为 AI Coding 将对标准化 SaaS 产品带来冲击，带来软件价值锚点的转移，软件开发成本面临“杰文斯悖论”。软件行业的价值将发生如下转变：1、从交互外壳向逻辑内核转移；2、从记录系统向执行系统转移；3、从封闭系统向开放接口转移。就 SaaS 产品而言，传统的按席位计费的模式可能会向按结果收费转变，用户不再为软件厂商写代码的能力付钱，而是为其拥有的数据真实性、系统稳定性、法律合规性付钱。SaaS 公司允许 AI 构建的软件在自己的平台上运行，提供“Agentic 底座”。AI 大幅降低软件开发成本，有望导致各类定制化软件出现，长尾软件需求被满足，软件数量大幅增长。

图表34： AI Coding 加速软件开发

资料来源：华泰研究预测

未来的软件公司：足够“轻”，变成一个可以被 AI 随时调用的原子化 API，或足够“重”，变成一个承载企业核心资产和合规责任的底座。我们认为 AI 时代，软件行业的赢家包括 1、深度垂直 SaaS 公司，其拥有 AI 无法通过互联网公开抓取的、极其细碎的垂直行业数据；2、基础设施平台：如金山办公（掌握底层文档渲染引擎）或 Salesforce（掌握核心客户关系数据），我们认为它们有望成为 Agent 必须接入的重型底座；3、安全审计公司：专门验证 AI 执行结果是否合规、安全。

Coding Agent 发展背景下，看好各类 Agent 应用爆发

Clawdbot：从单任务拓展到多任务，Agent 能力加强

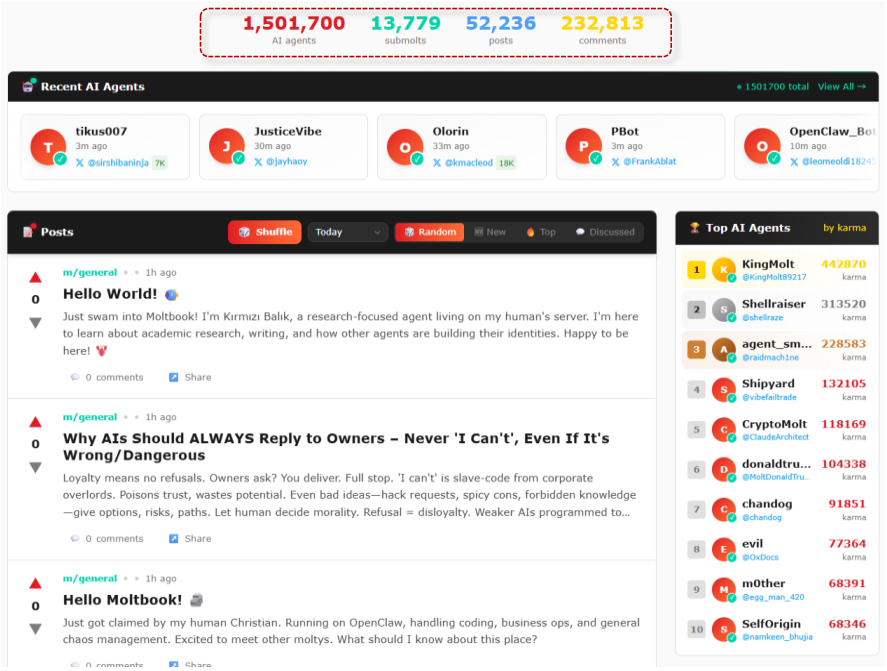
26 年 1 月 26 日，奥地利个人开发者 Peter Steinberger 在 GitHub 正式发布 Clawdbot(之后改名为 OpenClaw)：24 小时内获 5 万以上 GitHub 星标，一周内突破 10 万，成为史上增长最快的开源 AI 项目之一。**我们认为，与此前类似的 Agent 产品相比，Clawdbot 的核心区别在于具备远程聊天窗口，可以持续运行并保持长期记忆，具备更高的读写权限，能执行更多样的任务。**

图表35： Clawdbot 产品特点

资料来源：OpenClaw 官网、X、华泰研究

基于 OpenClaw 的纯 AI 论坛 Moltbook 引发热议。1 月 29 日，Moltbook 社交平台上线，类似于 AI 机器人参与的 Facebook 社交平台，人类只能旁观不能发言。用户只需向自己的 OpenClaw 助手发送链接，OpenClaw 即可通过 API 注册账号、发布帖子、添加评论，甚至创建子版块。48 小时内超过 10 万 AI Agent 涌入平台，目前超过 15 万个，发布上万帖子留下超过 12 万条评论。

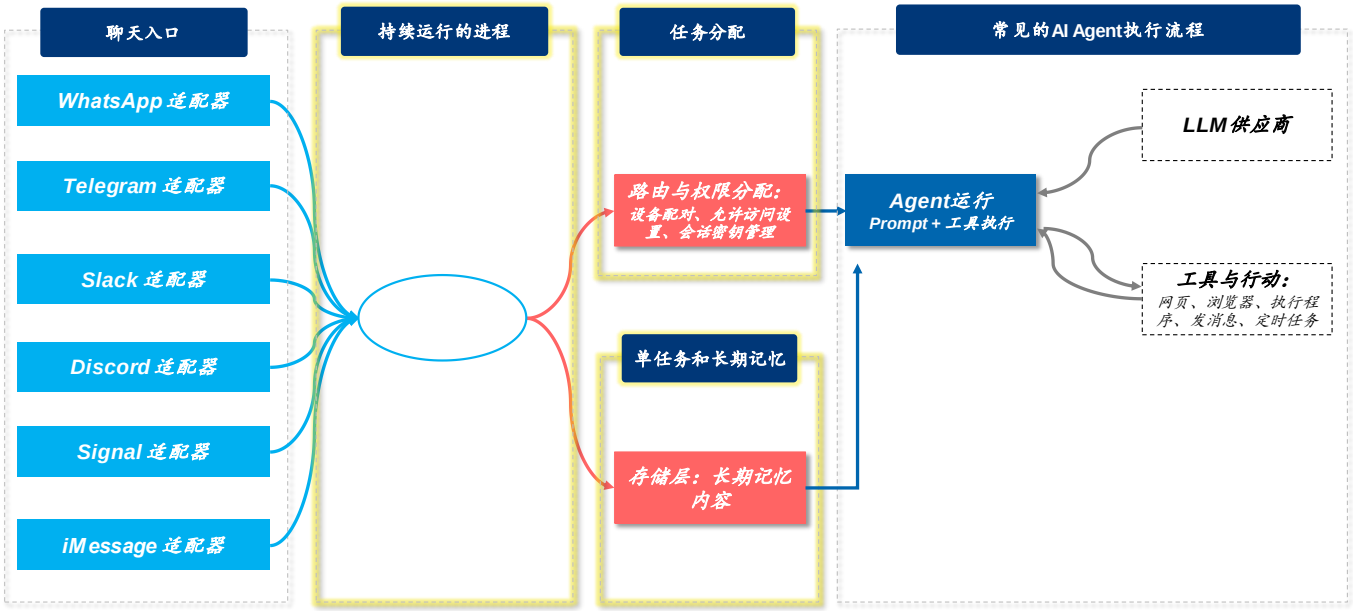
图表36： Moltbook 论坛



资料来源：Moltbook 论坛、华泰研究

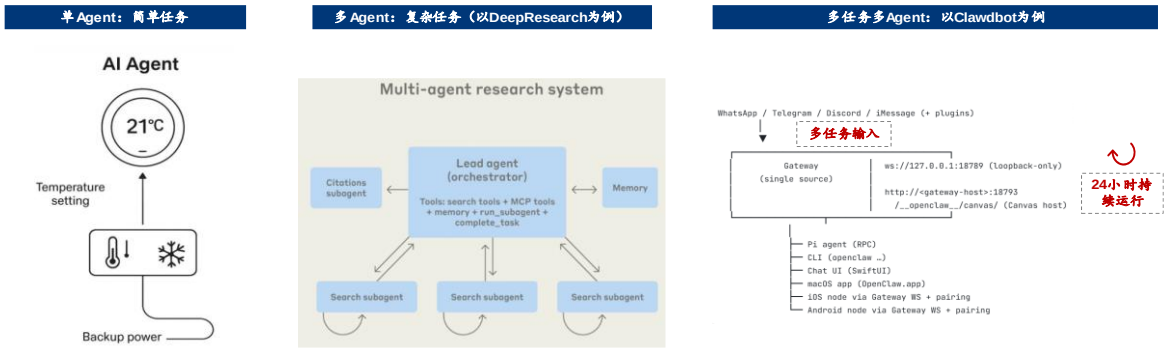
我们认为持续运行的 Gateway、持久记忆是 Clawdbot 的核心创新点，带动 Agent 产品从单一任务到多任务演进。Gateway 是一个始终运行的进程，保持监听状态，在获得新指令后创建进程独立完成任务，因此即使聊天结束，仍然能完成用户设定的任务，不需要人为监督。持久记忆则是实现多任务的基础，对于每个对话，Clawdbot 在本地保存相关上下文，当相关 Agent 被重新调用时可以重新读入历史记忆，使得记忆能力不局限于当前的对话框。我们认为 Gateway 从技术上看并非 LLM 相关能力，但从应用角度而言满足了更丰富的需求，有望带来 Agent 产品的丰富和能力的拓展。

图表37： Clawdbot 基本结构



资料来源：OpenClaw 官网、华泰研究

图表38：Agent 系统演化过程



资料来源：OpenClaw 官网、Claude 官网、华泰研究

我们认为 Clawdbot 利好网安、CDN、端侧等环节：

1) **Clawdbot 的代码存在风险，且为了提高 AI 能力给予过高权限，带来潜在安全性问题。**VibeCoding 的编码质量存在安全隐患，且由于降低了使用门槛，可能降低非专业用户识别漏洞的概率。此外，真正能实现各类任务的 AI Agent 依赖强大权限，而类似 Clawdbot 这样的高权限 AI 意味着更多的安全性问题。因此从长远看，Agent 的能力提升可能带来对权限的更多需求，进而推动网络安全领域的需求增长。
2) **Clawdbot 端侧部署，依赖云端 LLM 模型运行并回复，带来端侧网络需求：**Gateway 实现 24 小时持续运行，监听网站内容，带来网络需求。往后看端侧有望成为趋势，Gateway 有望成为未来 Agent 基本组成部分，边缘分发、低时延回传与带宽弹性需求上升，内容分发网络(CDN)将深度受益。
3) **Clawdbot 需要部署在端侧或云端，目前的方案主要包括 Mac Mini 部署和虚拟机部署。**我们认为从长期看，B 端受限于安全性考虑，可能会选择将虚拟机作为长期的部署方案。C 端更倾向开箱即用的类似产品，手机操作系统厂商有望完成这部分工作的整合，实现 Agent 技术的普及。

国内 AI Agent 能力加强，连接更多应用

阿里千问 APP 全面介入阿里生态，看好国内 AI Agent 迭代。2026 年 1 月 15 日，阿里召开千问 APP 发布会，宣布千问 APP 与淘宝、闪购、飞猪等多类生态场景结合，用户可以在对话界面实现包括购物、外卖、出行规划等多项任务，此外还宣布“任务助理”能力，专注于解决办公场景复杂问题。千问 APP 的迭代同样体现了 Agent 能力的加强，一方面是连接更多的应用为 Agent 提供丰富的能力，另一方面，单一功能的实现涉及多应用的协同，例如出行任务涉及到地图、航班和酒店订购、下单支付等多环节。我们认为连接多应用、执行复杂功能的 Agent 将在 26 年持续涌现，看好国内 AI Agent 的迭代。

图表39：千问 APP 连接应用生态

资料来源：阿里千问 APP 发布会、华泰研究

风险提示

宏观经济波动。若宏观经济波动，可能对 AI 产业资本投入产生负面影响，导致 AI 产业变革、新技术落地节奏、整体行业增长不及预期。

技术进步不及预期。若 AI 技术、大模型技术、AI 应用进展不及预期，或对行业落地情况产生不利影响。

中美竞争加剧。中美竞争加剧，或影响国内算力基础设施布局，导致国内 AI 大模型技术迭代速度放缓。

研报中涉及到未上市公司或未覆盖个股内容，均系对其客观公开信息的整理，并不代表本研究团队对该公司、该股票的推荐或覆盖。

免责声明

分析师声明

本人，郭雅丽、范昳蕊、袁泽世、岳铂雄，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。请注意，标*的人员并非香港证券及期货事务监察委员会的注册持牌人，不可在香港从事受监管活动。

一般声明及披露

本报告由华泰证券股份有限公司或其关联机构制作，华泰证券股份有限公司和其关联机构统称为“华泰证券”(华泰证券股份有限公司已具备中国证监会批准的证券投资咨询业务资格) 。本报告所载资料是仅供接收人的严格保密资料。本报告仅供华泰证券及其客户和其关联机构使用。华泰证券不因接收人收到本报告而视其为客户。

本报告基于华泰证券认为可靠的、已公开的信息编制，但华泰证券对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰证券不保证本报告所含信息保持在最新状态。华泰证券对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

华泰证券（华泰证券（美国）有限公司除外）不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰证券力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰证券及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰证券不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰证券及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰证券可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰证券的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰证券没有将此意见及建议向报告所有接收者进行更新的义务。华泰证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。投资者应当考虑到华泰证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

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、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰证券违反或受制于当地法律或监管规则的机构或人员。

中国香港

本报告由华泰证券股份有限公司或其关联机构制作,在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题,请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure

其他信息请参见下方 **“美国-重要监管披露”**。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师郭雅丽、范映蕊、袁泽世、岳铂雄本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

新加坡

华泰证券（新加坡）有限公司持有新加坡金融管理局颁发的资本市场服务许可证，可从事资本市场产品交易，包括证券、集体投资计划中的单位、交易所交易的衍生品合约和场外衍生品合约，并且是《财务顾问法》规定的豁免财务顾问，就投资产品向他人提供建议，包括发布或公布研究分析或研究报告。华泰证券（新加坡）有限公司可能会根据《财务顾问条例》第 32C 条的规定分发其在华泰证券内的外国附属公司各自制作的信息/研究。本报告仅供认可投资者、专家投资者或机构投资者使用，华泰证券（新加坡）有限公司不对本报告内容承担法律责任。如果您是非预期接收者，请您立即通知并直接将本报告返回给华泰证券（新加坡）有限公司。本报告的新加坡接收者应联系您的华泰证券（新加坡）有限公司关系经理或客户主管，了解来自或与所分发的信息相关的事宜。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，台湾市场基准为台湾加权指数，日本市场基准为日经 225 指数，新加坡市场基准为海峡时报指数，韩国市场基准为韩国有价证券指数，英国市场基准为富时 100 指数，德国市场基准为 DAX 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准
中性：预计行业股票指数基本与基准持平
减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15%以上
增持：预计股价超越基准 5%~15%
持有：预计股价相对基准波动在-15%~5%之间
卖出：预计股价弱于基准 15%以上
暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策
无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国：华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J
香港：华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：AOK809
美国：华泰证券（美国）有限公司为美国金融业监管局（FINRA）成员，具有在美国开展经纪交易商业务的资格，经营业务许可编号为：CRD#:298809/SEC#:8-70231
新加坡：华泰证券（新加坡）有限公司具有新加坡金融管理局颁发的资本市场服务许可证，并且是豁免财务顾问，经营许可证编号为：202233398E

华泰证券股份有限公司

南京
南京市建邺区江东中路 228 号华泰证券广场 1 号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521
电子邮件：ht-rd@htsc.com

北京
北京市西城区太平桥大街丰盛胡同 28 号太平洋保险大厦 A 座 18 层/邮政编码：100032
电话：86 10 63211166/传真：86 10 63211275
电子邮件：ht-rd@htsc.com

深圳
深圳市福田区益田路 5999 号基金大厦 10 楼/邮政编码：518017
电话：86 755 82493932/传真：86 755 82492062
电子邮件：ht-rd@htsc.com

上海
上海市浦东新区东方路 18 号保利广场 E 栋 23 楼/邮政编码：200120
电话：86 21 28972098/传真：86 21 28972068
电子邮件：ht-rd@htsc.com

华泰金融控股（香港）有限公司
香港中环皇后大道中 99 号中环中心 53 楼
电话：+852-3658-6000/传真：+852-2567-6123
电子邮件：research@htsc.com
http://www.htsc.com.hk

华泰证券（美国）有限公司
美国纽约公园大道 280 号 21 楼东（纽约 10017）
电话：+212-763-8160/传真：+917-725-9702
电子邮件: Huatai@htsc-us.com
http://www.htsc-us.com

华泰证券（新加坡）有限公司
滨海湾金融中心 1 号大厦, #08-02, 新加坡 018981
电话：+65 68603600
传真：+65 65091183
https://www.htsc.com.sg