Agent 新春特刊——

智能体的形态演进与治理思考

（2026 年 2 月）

前　言

2026 年开春，AI Agent（智能体）在产业长久期盼后，迎来了从“想象”到“落地”的转折点：在国际展会中，搭载 Agent 能力的 AI 手机、眼镜、智能车机及各类智能家电密集面世；在北美，Cowork 与 OpenClaw 爆火，直接重构了北美资本市场对 SaaS 软件的估值逻辑；在中国，豆包手机和千问 Agent 前后面世。尽管 Agent 的形态和评价截然不同，但都意味着 Agent 开始成为统一入口，改变移动互联网的商业模式和与 APP 的合作边界。**一个明确的信号已传导至公共政策研究领域：AI 正式告别了“对话框”，步入以“能思考、能办事”为核心特征的智能体时代。**

本期新春特刊将解析这些产业“黑话”，穿透热点和争议，尝试勾勒出一幅 Agent 发展和治理的“全景航向图”。

首先是 **Agent 硬件**。通过对 CES 与深圳硬件展的复盘，我们看到智能体正“寄宿”于手机、眼镜、汽车、家电甚至机器狗等物理载体，通过全端协同和全局记忆，让各类硬件可以“组队”为用户提供主动服务，而中国企业则以“以算法替代高精度器件”的策略，把坚持精密制造的欧洲甩在身后。

然后是 **Agent 软件。**在生产力领域，Agent 通过代码提升了复杂意图的理解和复杂任务的分拆能力，再通过 MCP、Skills 等“脚手架”调用万物，让 CoWork 和 OpenClaw 从开发者出圈到所有“打工人”；在消费领域，电商形态正从电子贸易（e-Commerce）向智能体贸易

（Agent Commerce）迁移，这种从“人货匹配”到“Agent 决策商品和服务消费”的变迁，将突破电商行业的容量上限，并可能改变移动互联网的合作格局。

最后是 **Agent 治理**。在 Agent 安全领域，我们提出了三层架构，并特别说明了“全端协同”、“全局记忆”带来的增量风险，而 Agent 能力显著提升，可以完成更多原先只有人类才能完成的工作，改变了人机交互的边界，必然引发伦理争议和难题。哪些工作只有人类才能完成、Agent 不能代劳？哪些话语 Agent 不适合对人类表达，哪些人类价值观 Agent 需要遵守？我们将以 AI 商业化和“一老一小”两个特殊群体入手，给出 Agent 行为规约的样例；在 Agent 出海方向，我们解析了 Manus 被出口管制的政策逻辑，但也讲解了 Manus 在国内可能遇到的经营挑战，提出要支持科创企业在国内与海外“双轨”运营。

希望通过这五篇深度分析，能够与诸位专家同好一起，总结技术表象背后的产业逻辑，探讨 Agent 时代公共政策的新锚点。而最重要的，是共同迎来 2026 年 AI 产业发展与治理的新篇章。

目　录

智能硬件产业“大对账”：
中美两场展会折射出的趋势、差异与思考

2026 年 1 月，全球智能硬件产业迎来两场重磅展会：美国拉斯维加斯的 CES 2026 与深圳阿里云智能硬件展同步启幕。太平洋两岸的科技季风从未如此同步，从全球首款 AI 吉他、到长出机械臂和机械腿乃至飞行器的扫地机器人，从智能眼镜与车机无缝交互、到全屋智能家电的功能联动，从人形机器人与人类拳击手对练闪避、到助盲眼镜毫秒级实时避障，只要能想象到的物理交互场景，就有可能在发生。这两场展会不仅集中展现了人工智能与硬件融合的最新成果，更是一份全球智能硬件产业的年度“大对账”，折射出国内与海外在创新逻辑、市场生态与治理理念上的共性趋势与显著差异。

阿里研究院本次派出研究人员亲赴现场，通过对两场展会的深度观察与对比分析，提炼出六大核心发现，以期透过展会表象，洞察产业深层趋势，为中国智能硬件的高质量发展提供参考。

一、发现一：“全端协同+全局记忆”成为智能硬件的共性发展趋势

美国 CES 2026 上，谷歌 Gemini 模型与三星展示了典型的联动场景：用户在电视看到特色菜，系统自动调取冰箱食材数据、向手机推送定制菜谱、同步预热烤箱。这彻底改变了过去“单品智能”的割裂体验，实现了跨设备完成同一目标的高效配合。三星计划 2026 年将谷歌 Gemini 引入 8 亿台电子设备当中，以构建一个全端协同的硬件智能体生态。**深圳硬件展**上，理想展示了“理想同学”在外卖闪购、停车缴费等场景的丝滑表现。通过跨设备（眼镜/手机/车机）、跨应用的显式记忆和从对话中学习的隐式偏好，实现对用户意图的深度理解，统筹调度地图、支付等专业智能体，完成复杂任务的动态编排。OPPO 则发力“生活管家”、“生产力助手”、“影像搭子”三大关键赛道，通过对用户“察言观色”，理解和记录用户偏好，提出“感知-记忆-理解-执行”的循环飞轮，旨在打造一个“越用越懂你”的个性化生态。并展示了多个应用场景，如基于之前学习到的用户的饮食偏好，精准推荐附近餐厅；如基于用户使用习惯、位置、时间等多维度数据，主动弹出天气提醒、日程安排等个性化推荐。

由此，我们能理解“**全端协同**”并不是传统的物联网设备连接与遥控，而是指手机、眼镜、家电、车辆能够动态组队、各司其职，为用户提供连续的智能化服务；而“**全局记忆**”是

指在获得用户授权的前提下，让用户在 A 设备上养成的使用习惯，能够被 B 设备自动理解并适配，从而实现对用户意图的理解预判，从而进行主动服务。中美头部智能硬件厂商在这两个领域有高度共识，从“单品智能”加速向**“全端协同+全局记忆”**演进。

二、发现二：“出海验证，反哺国内”成为中国智能硬件发展的新范式

一批在海外获得成功的中国智能硬件企业，正系统性走通**“国内研发—海外验证—规模成功—回国拓展”**的闭环路径。

如深圳科技企业机智连接（Plaud AI），以创新性的卡贴式录音设备 Plaud Note 切入海外市场，2 年内在欧美市场销量上百万台，是市场上公认的最成功的 AI 硬件之一，成为该赛道的统治级厂商。在验证了技术方案和商业模式后，于 2025 年 10 月“杀回”国内，开始布局拓展中国市场。如成都沸彻科技（FITURE），结合端侧视觉算法+云端大模型，开发的随身健身伴侣 Body Park Atom，能够实时监测用户动作并语音纠正指导，在海外众筹平台爆火，成功验证场景需求真实性后，近期开始启动国内市场的推广工作。值得注意的是，上述出海企业具备极高的海外合规意识，并将欧美严格的用户隐私要求与数据跨境标准融入产品设计和运行流程。

这一路径揭示出，中国已经涌现出相当一批具备“从 0 到 1”开创智能硬件新品类的科技企业。海外市场成为中国原创科技最好的“试金石”与“磨刀石”，企业带着成熟的产品定义、商业逻辑与合规经验回归，有效填补了国内市场的生态空白。从“三来一补”到“主动出海”到“生而全球化”，当前中国在智能硬件已进入科技实力“比较优势”向全球溢出阶段，具备整合全球产业要素，参与全球竞争的底气。为何这些“新物种”会选择欧美作为首发市场？我国本土产业环境在知识产权保护和服务付费商业模式接受度上，是否仍存在内外“温差”？值得进一步探讨。

三、发现三：AI 陪伴赛道呈中低端同质化竞争局面，部分厂商的“一老一小”破局路径值得借鉴

本次深圳硬件展上，近百家 AI 陪伴硬件厂商参展，占据了全部展位的近三分之一。这种“扎堆”程度，也揭示出情感陪伴是人类的刚性心理需求。但大多数 AI 陪伴硬件无论是外观还是功能均高度趋同，普遍定位在陪伴聊天玩具，依托云端基础大模型进行情感交互，价位普遍在 300 元人民币左右区间，呈现**中低端同质化竞争的局面**。而**部分厂商在“一老一小”两个特殊客群上的破局路径值得借鉴**。

面向儿童群体，“汤姆猫 AI 童伴”与模型厂商联合研发情感陪伴垂直模型，深度植入了汤姆猫 IP 特有的人设性格，并

构建儿童专属的内容体系，将知识 IP 化、游戏化，如“魔力咒语”“趣味打断”等互动玩法，并能根据年龄自动调节认知难度。并通过持续提取对话关键词构建用户兴趣图谱与画像（如记住孩子的宠物名字、喜好），在后续交互中主动调用记忆实现“越聊越懂你”的个性化陪伴。结合其硬件在头部和手臂的多自由度以及丰富表情，将价位上攻到 1500-2000 元人民币区间。**对儿童陪伴赛道**，政策需注意 AI 拟人化和游戏导致沉迷或内容引导错误价值观不同，是健康陪伴类软硬件的必备能力。治理可侧重加强正向伦理引导与具体的行为规约，保护产业创新活力。

面向银发群体，在 CES 展会上，美国厂商 TomBot 推出定价 1500 美金的陪伴仿生拉布拉多幼犬，定位为阿尔茨海默症辅助干预器械，正致力于通过 FDA 的医疗器械许可，从而在养老机构实现规模化应用。而随着我国老龄化程度持续加深，国家统计局数据显示，至 2034 年，银发群体将扩容至 4.1 亿。阿里研究院预估，银发群体的精神情感类消费占比将从 2023 年的 24%跃升至 2035 年的 35%。阿里平台数据显示，2025 年度，该群体智能玩具消费同比增长超 2000%，AI 陪伴机器人增长 200%，有效缓解独居焦虑。**对老人陪伴赛道**，政策可考虑进一步支持相关企业联合医疗机构开展临床效果验证，试点探索将具备康复干预功能的 AI 陪伴硬件纳入医疗器械管理路径，并鼓励养老机构开放试点应用场景，

引导AI陪伴产品从同质化玩具向专业化服务载体有序升级，真正释放银发经济创新潜能，助力积极应对人口老龄化国家战略。

四、发现四：德国“精度至上”模式在 AI 时代已经掉队，中国硬件产业正构建创新优势

美国CES 2026具身智能展区，来自德国的高精度微型电机领军企业FAULHABER（福尔哈贝），展示了其微米级加工精度（0.01mm物理精度）的微型行星减速箱与驱动模组。延续了其作为医疗与航天领域“隐形冠军”的技术骄傲，以纯粹的机械美学，依靠硬件的完美精度来确保传动零抖动，并强调长达1年的打磨验证周期是确保品质的“必要代价”。其展台背景海报上的宣传语 "Fine motor skills”（精细运动技能）与机器人拧魔方的画面，折射出其对硬件的深层看法：**仍将具身智能视为更复杂的自动化设备，而非 AI 大模型的物理载体**。这种源于工业时代的工程师思维与瑞士手表式力求精准的制造逻辑，在追求极致确定性的同时，对 AI 时代数据驱动、快速迭代“水土不服”，难以匹配当前智能硬件走向消费级市场的敏捷节奏。

与德国路径形成鲜明对照，**中国智能硬件产业正在探索一条叠加我国供应链优势与大模型能力的差异化路径**。在硬件侧，中国具备全球最完备且效率最高的硬件产业完整链条，

珠三角的核心零部件厂商可在2周内完成浙江机器人本体厂商的打样需求。在算法侧，中国模型已跻身全球头部，而token定价只有美国的1/20。中国工程师致力于将视觉伺服、力控柔顺、视觉识别、端到端模型融入控制系统，可以用0.1毫米精度的高性价比电机达到0.01mm精度昂贵电机同等的作业效果。不仅大幅降低了产品成本，更将产品验证周期从德国同行的12个月压缩至3个月。

上述产业实践为理解AI服务制造业提供了新的视角：如果仅在传统工业追求的极致标准化与规模化赛道上追赶，我们或难超越欧洲制造的壁垒。但通过算法的红利弥补精度的不足，用敏捷的试错迭代替代漫长的工业验证，这才是中国制造构建差异化新优势的根本路径。

五、发现五：多维感知引发多类型数据需求，相应的“高质量数据集”难以事先定义

两场展会均直观呈现了感知维度的爆发式扩张。CES 2026上，“TouchDIVER Pro”触觉手套能够模拟压力、纹理、温度三类感官维度，带摄像头的耳机融合视觉与听觉模态，AI PIN硬件记录连续对话文本，健康戒指监测实时生理指标，脚部汗液分析工作站分析代谢物数据，脑电波头戴设备解析神经信号。这些案例表明，智能硬件所处理的数据范畴，正从传统的文本、图像、语音，加速向涵盖触觉压感、生化指

标、神经电位等更微观的物理或生理信号拓展。这一转变的实质，是从处理模型早期训练中常见的、基于互联网内容或人类生活工作积累的人类显性表达信息，扩展到表征环境与人体隐性状态的数据。

这一现象的背后，折射出 AI 对高质量数据的需求标准正不断变化演进。对于大模型预训练，高质量数据意味着如中小学课本、高考试卷、维基百科等准确规范的通识知识；对于具身智能，需要基于人类动作采集再做仿真的合成数据；而面对智能体（Agent）应用，核心需求转向了代码逻辑、操作流程（SOP）等教会 AI“如何拆解和执行复杂任务”的过程性数据；面对智能硬件，则延展到对物理世界的多维度感知数据。这揭示出，**AI 大模型及应用领域所需要的“高质量数据集”具有高度的场景依赖性与时效敏感性。**AI 需求数据需求类型随着模型的发展与 AI 应用的迭代过程而不断调整和扩展，而数据“好/坏”的标准，则跟每个模型厂商的技术认知和路线选择相关，很难在产业规划层面进行静态的事前定义或超前建设。

六、发现六：安全与治理模式，面临与智能硬件新形态的深度再适配

从上述若干发现中我们不难发现，智能硬件在技术形态、协同模式和服务能力上的急速进化，客观上使得既有的治理

逻辑面临新的适应性挑战。如何在保障安全底线的同时，为新业态预留试错与成长的空间，成为亟待深入探讨的时代命题。

首先，“全端协同+全局记忆”的跨端任务协同，和“记忆”的汇聚融合，**对现行治理原则提出了新的思考命题。**现行数据治理体系以“敏感个人信息保护”为基点，核心遵循“目的限定”与“最小必要”原则。这一逻辑在单一 APP 环境下行之有效，但在“全端协同”场景下，服务体验的连贯性恰恰建立在数据跨设备、跨场景的汇聚融合之上。如何回应这一发展趋势，是**需要治理智慧与产业实践共同探索的命题。**

其次，AI 硬件的“拟人化”交互，引发了关于未成年人保护与伦理规约的复杂性辨析。 大模型赋予了硬件前所未有的情感共鸣与对话能力，使其超越了单纯的工具属性，演变为能够承载情感陪伴功能的“类伴侣”角色。必须清晰地认识到，AI 陪伴产生的“高粘性”机制与传统网游有着本质不同：后者多基于即时反馈的“多巴胺奖励机制”诱发行为成瘾，而前者则通过满足用户“被理解、被接纳”的心理需求，建立起深层的情感依恋。传统的“防沉迷”系统与“青少年模式”多采取基于时长限制的“硬阻断”逻辑，这在面对深层情感依恋时，不仅效能有限，甚至可能在用户最脆弱的时刻强行切断其唯一的情感支持源，造成次生的心理冲击与“二次伤害”。因此，治理的重心应从“防堵”转向“疏导”，应鼓励企业利用大模型

的深度陪伴与意图洞察能力，构建具有正面引导意义的交互规范。

最后，智能硬件进入家庭私密空间，对个信隐私和产业创新的合规都带来相应挑战。当智能音箱、智能家电、扫地机器人、具身机器人等各类 AI 硬件普遍搭载摄像头与高灵敏度麦克风进入家庭，这也意味着它们带着“能看、能听、能记”的能力，全天候地“介入”了家庭这一私密空间。这容易引发消费者对于物理空间隐私安全的顾虑，也让产业创新面临显著的合规不确定性。在技术实践中，虽然存在“端侧”与“云端”处理的技术分流，但对于哪些数据属于必须留在室内的“绝对隐私”，哪些数据可以脱敏出户以训练出更优的模型能力，目前尚缺乏基于数据分类分级的清晰共识与界定。如何在家庭这一隐私核心区，通过明确的标准界定，消除消费者的安全顾虑和企业的合规疑虑，是推动智能硬件产业敢于创新、放心发展的关键前置条件。

结语：

当我们将视线投向近期智能体（AI Agent）领域的快速发展，可以清晰地观察到“软硬并进”的演进方向。一侧是以 Claude Code、OpenClaw 等为代表的“AI Agent”，它们正在数字空间内引发代码编写与企业协作的效率革命，并且已经影响到对 SaaS 产业的价值度量。另一侧则是本文聚焦的“智能

硬件”，覆盖了情感陪伴、家居生活、生产制造等广泛领域。它不仅是 AI Agent 的物理载体，更是检验 AI 解决真实物理世界问题能力的“试炼场”。

在美国 CES 和深圳硬件展的“大对账”中，我们看清了：中国不只有“从 1 到 N”的追赶复制，也有“从 0 到 1”的业态创新；中国未固守传统精密制造的旧逻辑，而是依托全球响应最快的供应链体系与活跃的开源算法生态，走出一条“以软补硬、敏捷迭代”的差异化新路。然而，创新成果“优选海外首发”的现象，提示我们需要进一步优化国内的科创环境与商业土壤。同时，随着 AI 硬件深度介入儿童教育、老人看护与家庭隐私空间，带来了复杂的技术和伦理挑战，需要在全社会层面形成关于 AI 边界与使用规范的价值共识。期待中国智能硬件产业能够扬长补短，并率先在合规治理领域给出中国答案。

从 OpenClaw 爆火，看代码数据的价值与软件行业的重构

引言：从智能体工具的“爆火”与软件业的“暴跌”谈起

2026 年初，在开发者社区和各大技术论坛中，一款名为 OpenClaw 的开源智能体工具火爆破圈。它与市面上其他大模型厂商推出的 Claude Code、Claude CoWork、Codex 等工具非常类似：不再是过去那个只会陪聊的聊天机器人，而是进化成了能够接管电脑、协助人类完成具体办公任务的“办公搭子”。其展现出的“人机协作”新雏形，已经让市场感受到了生产力大变革的气息。

然而，与这股智能体工具热潮形成鲜明对比的，是“AI 颠覆 SaaS 软件”的资本叙事：**软件产业正经历着一场暴跌，从通用办公写作到 ERP 软件，紧接着是法律、金融等专业领域软件。统计数据显示，美股广义软件板块的市值近期已从高点下跌 30%，蒸发约两万亿美元**。这“一热一冷”并非偶然，而是一场历时五年的技术演进，下面分三阶段来回顾“写代码”如何成就大模型和智能体。

一、阶段一（2021-2023）：代码数据“三步走”帮大模型打造逻辑能力

2020 年 6 月，OpenAI 发布了 GPT-3。在学习了互联网上几乎所有的书籍和百科后，模型能模仿写出华丽的莎士比亚十四行诗，却无法解开稍复杂的数学应用题。当时业界普遍认为大模型只是进行统计关联的“随机鹦鹉”，并不真正理解逻辑。

转折发生在 2022 年底。研究者符尧在《拆解追溯 GPT-3.5 各项能力的起源》中完整复盘了这一能力的演进路径，揭示了代码在其中的核心作用，随即在 2023 年，又提出了《迈向复杂推理：大模型的北极星能力》，揭示了大模型的复杂推理有更广泛的应用空间。

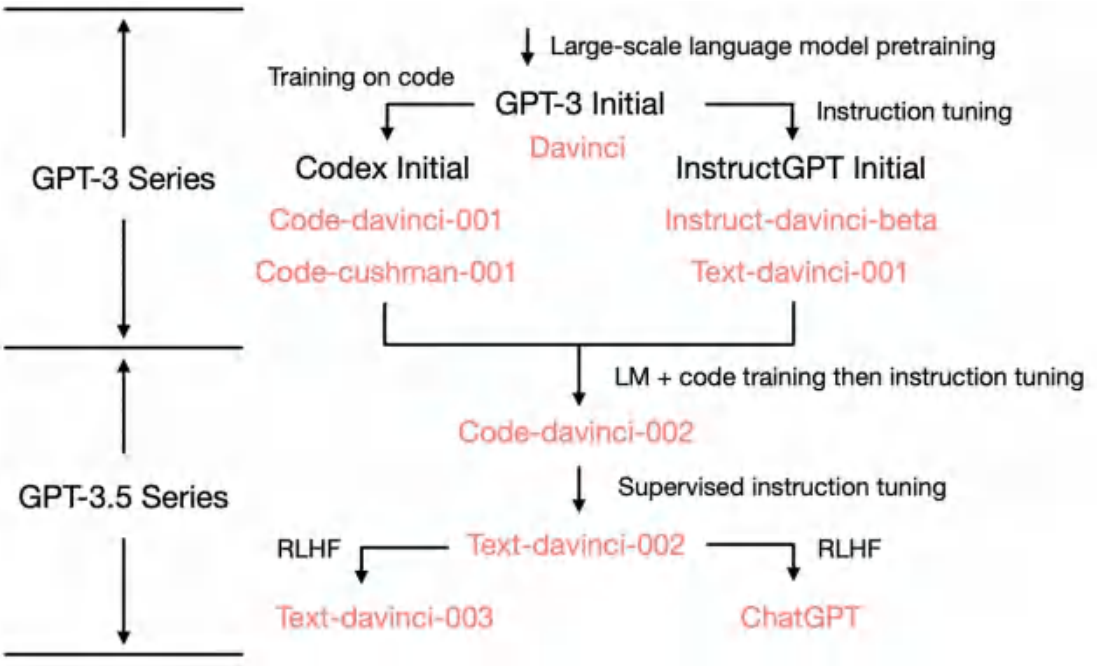


图 1： GPT-3 到 GPT-3.5 模型进化树

（一）代码训练为模型注入推理能力

最初的 GPT-3 主要基于自然语言数据，虽然具备极强的语言生成能力，但在处理逻辑任务时表现乏力。随后分化出的 Codex 路径引入了大规模代码训练，研究指出，**代码是复杂推理能力的源头**：例如，代码当中的条件判断与因果一致性，训练了模型在生成文本时能够保持前后主张不冲突和论证逻辑的严密。而代码中嵌套循环与跨函数调用，则训练了模型处理长距离依赖（Long term Dependency）的能力。使得模型在分析超长文档时，能够精准识别出位于开头处的一个前提条件，是如何跨越数万字的内容，直接决定了结尾处的结论走向。这种能力让模型在处理长文本时，不仅关注临近的词汇，还能够捕捉到文档跨段落或者多个文档之间的因果联系。

（二）代码数据是“思维链”涌现的关键

研究普遍认为，思维链（Chain of Thought）能力的涌现主要归功于代码训练。代码编写要求开发者将复杂目标拆解为环环相扣的细分步骤，这种范式让模型在面对自然语言的指令时，不仅是根据概率预测下一个词，而是倾向于先将问题拆分再逐一求解。随后引入的指令微调与强化学习，激发模型学会如何将分解问题的能力应用起来，让模型从简单的“对话机器人”转变为能够处理高度复杂任务的决策中枢。

（三）逻辑能力的“泛化与迁移”效应

从模型演进的路径看，在引入代码逻辑并形成思维链，其价值远超编程本身。在早期的训练中，就已经发现在代码侧学到的分步推理、错误检测与方案筛选能力，可以迁移到数学、符号推理与综合问答等非代码的任务中。而更进一步，模型展现出的确定性推理和任务拆解能力，使其能够作为决策中枢，去调度和指挥各类软硬件插件。符尧的研究就断言“这种逻辑泛化能力是大模型成为下一代‘计算平台’或‘操作系统’的核心底座。”

二、阶段二（2024-2025）：大模型在编程真实互动中持续学习

2024 年后，大模型的能力提升遇到了明显瓶颈。产业界发现，增加更多的文本还是代码，带来的边际效益开始递减。产业界逐渐意识到，模型不能只靠在实验室里“死记硬背”那些已经存在的数据，而必须在持续的互动中学习。**在所有的人机交互中，编程的反馈是最清晰和最没有歧义的，编程场景从而成为训练模型学习的最佳试验田**。当模型给出一行代码建议时，如果程序员直接采纳，这就是对模型的正向“嘉奖”。如果程序员修改代码，或者完全拒绝了建议，这也是含金量极高的负向“批评”。全球过千万的程序员的“采纳”、“修改”

和“拒绝”，构成了源源不断的学习数据，帮助模型打破了性能停滞的困局。

这种进化的早期，是由“大脑”（基座模型）与“双手”（编程智能体）的深度协作开启的。以 Anthropic 与 Cursor 的合作为例，最初，基座模型 Claude 3.5 负责编程输出，而编程智能体 Cursor 则负责开发环境，并把程序员反馈给模型进行训练。Cursor 还更进一步，推出的 Composer 功能，允许模型跨越多个文件同时进行修改（比如为了改动 A 功能，必须同步调整 B 和 C 文件）。这种“全项目视角”让模型能够捕获更深层的工程逻辑。当程序员面对模型给出的复杂修改建议（涵盖代码、图表及架构），选择“采纳”或“拒绝”时，这种反馈让模型学到了功能模块之间隐秘的依赖关系。**随着技术演进，基座模型开始直接整合智能体的功能，**这也引发了 Agent 是“套壳”没有技术含量的纷争。全球的头部模型企业都在跟进这一路线：谷歌的 Gemini 团队在强化模型的“代码沙箱”能力，让模型在自我运行代码的成败中吸取教训。OpenAI 联合创始人格雷格·布罗克曼（Greg Brockman）在 2025 年 8 月公开阐述，GPT-5 的提升很大程度上归功于观察用户在互动编程中的使用方式，并将这些实战反馈喂回训练。国内的 DeepSeek 和 Qwen，也是把编程模型当作基础模型升级的前哨站，开发者都会期待和观察编码模型发布，

一是可以有更好的工具支持，二是用来预测自己喜爱模型的新版本发布时间。

三、阶段三（2025 年下旬至今）：能力“溢出”编程场景，重塑企业办公与专业领域

经过五年的代码训练和编程反馈积累，通用大模型在 2025 年下旬，迎来了能力的再次质变。以 Anthropic 发布的 Claude Opus 4.5 为代表的新一代模型，在智能体编程、工具调用及计算机操控等能力上大幅刷新业界记录。与模型能力提升同步发生的，是底层协议与基础组件的标准化。MCP 协议与 Skills 规范的成熟，为大模型接管具体任务提供了“工作脚手架”：MCP 通过统一协议实现了模型与外部数据、计算资源的标准化对接；Skills 则将复杂的软件功能封装为可调用的“说明书”。

[1] 关于 MCP（Model Context Protocol，模型上下文协议）：由 Anthropic 于 2024 年 11 月推出并开源的开放标准协议。MCP 通过定义统一的数据交换格式与双向通信机制，使模型能够以标准化方式安全访问本地文件、企业数据库、API 服务，为智能体跨应用调度奠定了协议基础。

[2] 关于 Skills（能力单元规范）：由 Anthropic 于 2025 年 10 月首次提出的标准，并于 2025 年 12 月正式发布为开源标准。通过结构化的文件将特定能力领域的知识、执行指令封装起来，模型可直接理解并调用。

[3] MCP 和 Skills 的关系：二者形成互补分工，MCP 解决“能连接什么”，Skills 解决“如何正确操作”。例如 Agent 要生成一张财务合规报表，MCP 负责接入客户交易流水库，而 Skills 中要写明处理数据的规则，例如单笔交易额度，风险控制标准等

随后，新一代的智能体工具正式登场，典型的如 Anthropic 在企业级市场推出的 Claude Code 和 Claude CoWork。它们通过在电脑桌面端建立一个全局的“任务中枢”，能够接管文件系统、网络浏览器及所有支持 MCP 协议的应用，像一名资深助理一样，自动理解模糊的自然语言指令，拆解为一系列跨软件的操作步骤，从而迅速进入通用办公领域，快速替代初级行政管理、跨系统的数据整理工作，成为人类员工的“办公搭子”。本文开篇提到的开源工具 OpenClaw，也深度借鉴了这一思路并迅速出圈。紧接着，OpenAI 桌面版、阿里巴巴 QoderWork、MiniMax Agent 等竞品也加入战场，在这一赛道上展开了激烈的角逐。

大模型的渗透并未止步于通用办公，而是加速向法律、金融等高门槛领域蔓延。2026 年 2 月初，Anthropic 在其工具中引入法律专用插件，协助律师进行合同审查与合规检索，直接引发了全球法律科技股的剧烈震荡。随后发布的 Claude 4.6 基座模型，在复杂金融任务的表现上再次刷新纪录，能够精准提取海量文档信息并完成基本面分析。这一系列动作标志着大模型已完成能力进阶：从代码场景练就“脑筋”，到

通用办公场景建立“脚手架”，最终进入专业领域替代高价值逻辑劳动。

四、回顾思考：对软件和应用产业的影响开始显现

如果一家企业能通过智能体，基于 MCP 协议直接调取数据库，并调用 Skills 插件完成复杂的财务分析，那么企业并不需要每年支付数百万美元购买 客户管理系统（如 Salesforce ）的 license，也不用培训员工去学习商业报表软件（Tableau）的使用方法。智库 SemiAnalysis 指出，传统软件巨头曾构筑过三道防御工事：高昂的数据迁移成本、基于操作界面的用户黏性、以及复杂的系统集成。但在具备逻辑操作能力的基础模型，以及 MCP 协议带来的标准化接口面前，这些防线正在迅速瓦解。

随着图形用户界面（GUI）不再是必需，软件行业可能会走向“隐形化”：从有独立客群和定价权的产品，蜕变为向智能体提供基础能力的 API 服务商。阿波罗全球管理公司（Apollo Global Management）合伙人约翰·齐托（John Zito）在近期闭门会议中抛出了最核心的疑虑：“真正的风险在于，传统的软件行业是否已经走到了尽头？”；而光速创投（Lightspeed）合伙人艾萨克·金（Isaac Kim）则更为直接地宣告，SaaS 软件赖以生存的“按人头席位收费”模式，正被新一代智能体工具彻底影响。

同样的挑战也蔓延至移动互联网领域，但 APP 行业的战局却表现出差异。相比于通用软件，拥有高频刚需入口、线下履约体系的 APP 依然保留护城河，但这些 APP 也需要做好与智能体合作做好用户意图的承接。我们可以尝试推演出新的合作模式：大模型负责意图理解，智能体负责调度决策，而筛选过后的 App 和软件负责履约交付，这必然给数字世界带来更多生机和竞争。

结语与展望

在过去五年里，我们见证了通用大模型能力的上涨，正快速跨越那些过去难以逾越的产业门槛。当前还有三条逐步清晰的演进路径：

一条已经完备的路径：代码数据不仅能帮助提升编程领域效率，还能够帮助模型提升复杂推理能力，从而能够拆解理解用户意图、拆解复杂任务，而搭载 MCP 和 Skills 这样的“脚手架”，使 Agent 快速覆盖商业办公、法律和金融领域；

另一条行进半程的路径：视频数据不仅是视觉娱乐的产物，实际上在训练模型的“空间感”与“物理直觉”。目前，“借道”多模态模型生成仿真数据的路径已经被走通，被广泛用于加速具身智能开发；

还有一条尚未开启的路径：工业设备上的传感器数据（比如温度、压力的波动曲线），和视频里分帧图像是一样

的，都是随时间变化的信号。如果模型能看懂视频，它就有机会理解复杂的工业数据。将来借用大模型对于时间序列的理解，有机会优化工业制造的各类任务。

数据中包含的特征与规律决定了模型的能力。这种能力并不会被局限在它的来源领域，正如逻辑推理“源于代码，超越代码”。通过通用能力练就的“触类旁通”，可能成为比“深耕垂域”更高效的发展方式。当前，产业界仍存在“种豆得豆”的思维惯性，往往忽略了大模型这种跨域发展的特点。未来的竞争，当然需要垂域应用的深耕，但更需要给通用能力的进化预留足够的耐心与试验区间。

参考资料：

1. Brown, Tom, et al. "Language Models are Few-Shot Learners." NeurIPS, 2020.

2. Chen, Mark, et al. "Evaluating Large Language Models Trained on Code." OpenAI, Jul 2021.

3. Madaan, Aman, et al. "Language Models of Code are Few-Shot Commonsense Learners." Carnegie Mellon University (CMU), Oct 2022.

4. Fu, Yao, et al. "How does GPT Obtain its Ability? Tracing Emergent Abilities of Language Models to their Sources." Allen Institute for AI, Dec 2022.

5. Dario Amodei. "Machines of Loving Grace" Dario Amodei Blog, Oct 2024.

6. Anthropic. "System Card: Claude Opus 4.5." Anthropic Research, Nov 2025.

7. Claude. "Getting Started with Claude CoWork: Enterprise Best Practices." Claude Support Documentation, Jan 2026.

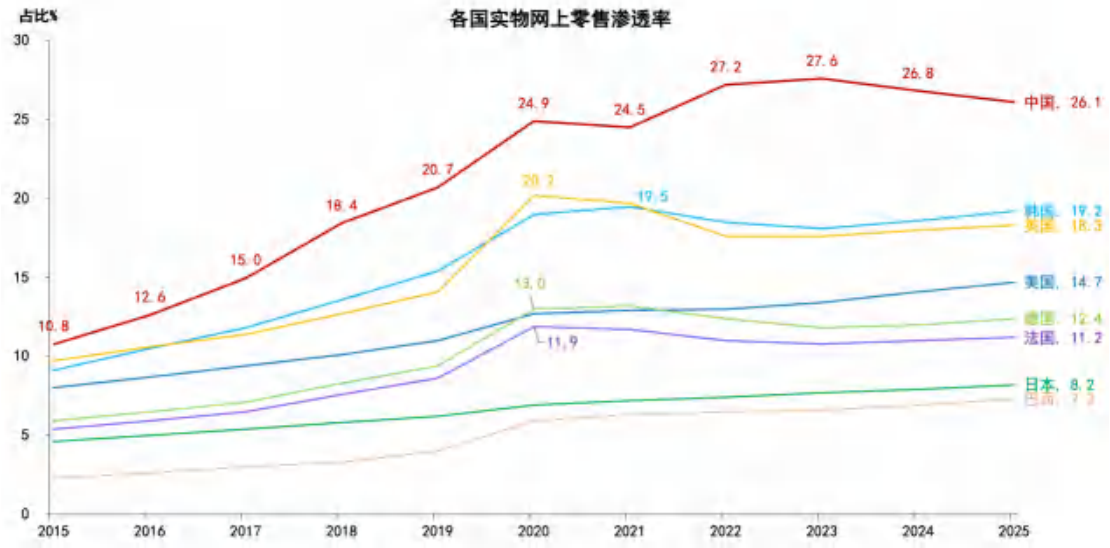
8. Dario Amodei. "The Adolescence of Technology." Dario Amodei Blog, Jan 2026.

9. Doug O'Laughlin, et al. "Claude Code is the Inflection Point: What It Is, How We Use It, Industry Repercussions, Microsoft's Dilemma, Why Anthropic Is Winning" SemiAnalysis, Feb 2026.

从 E-Commerce 到 Agent Commerce：迎接电商生态的整体升级

引言：

“电商”（E-commerce）作为一个合成词汇，代表的是数字技术对传统零售业的改造，它不仅开拓了逾 6 万亿美元的全球增量市场，更系统性地重塑了社会的生产逻辑、经营模式与消费行为。**然而近年来电商的发展似乎遇到瓶颈。**以电商渗透率最高的中国市场为例，其渗透率在 2023 年触及 27.6% 的高位后，小幅修正至 26.1%；英、法、德、韩等国亦在 2020 年疫情脉冲式增长后出现阶段性回落。尽管美、日等国仍保持增长态势，但增速已显著放缓且渗透率绝对值偏低。全球电商市场规模的年均增速已从 2015-2021 年间的 21% 高位，回落至 2022-2025 年间的 8% 左右，这一趋势显示出，现行业态下，电商已经进入边际增长递减期。



生成式 AI 的崛起为电商行业突破渗透率瓶颈提供了契机，其技术演进正推动行业从局部增效转向商业模式的质变。过去数年，电商平台对 AI 的应用多聚焦于内容生成（商品物料设计、AI 主播）、运营降本（数据分析、智能客服）及搜推算法的强化，而 2026 年开年的一系列动作标志着**“智能体电商”（Agent Commerce）**时代的正式到来。微软与谷歌先后与 Etsy、沃尔玛等零售巨头达成战略合作，通过系统级 AI 智能体直接承接并转化消费需求；阿里千问近期接入淘宝业务并开启公测，标志着其生态闭环的智能化重构；加之 OpenAI 此前打通 ChatGPT 与 Shopify 等平台的购物路径，行业已迎来从“人找货（搜索）”或“货找人（推荐）”向以 AI 智能体为核心的“起念头即决策”或“需求即交付”的逻辑进化。在这种新范式下，对用户意图的深度理解力以及

AI 与电商生态的底层连通性，将成为决定未来电商版图重构的核心变量。

在 AI 电商的演化中，两种核心模式逐渐清晰：一是以 OpenAI、微软、谷歌为代表的“猎人模式”（Hunter），通过外部 AI 智能体（Agent）跨平台调取电商能力；二是以亚马逊、淘宝为代表的“农夫模式”（Farmer），依托平台内生数据与供应链优势自建 AI 智能体，深耕存量生态的体验闭环。本文将剖析两种模式在流量入口、数据闭环及供应链协同上的差异，以此阐明从电商（E-commerce）到智能体电商（Agent Commerce）跃迁的规律，并尝试推演智能体电商时代下的商业协同新逻辑。

一、Farmer 模式：电商平台的自我改良

相比于 Hunter 模式，Farmer 模式对传统电商的改变算相对温和。以亚马逊的 Rufus 为例，Rufus 直接嵌入在亚马逊网页（顶部 tab1 和浮动对话框）和 app 中（常驻底 tab 最右）最显眼位置，与内部数据深度打通。对于用户来说，Rufus 像是一个**超级导购**，在用户的“发现与咨询”环节起重要作用。

Rufus 区别于传统搜推的核心功能：对复杂自然语言的理解。Rufus 通过理解用户复杂的自然语言（包括文字和图片），捕捉显性和隐性的限制条件（预算、场景、痛点、偏好），结合用户个性化信息和浏览和购买记录，从亿级商品

池中挑选出最匹配的几个。除此之外，Rufus 还可以进行主动商品对比、追踪价格下单等工作。从复杂需求理解到商品高效率匹配到完成下单，一套完整流程都可以在 Rufus 内实现。截至 2025 年底，Rufus 已积累 2.5 亿活跃用户（亚马逊整体约 3.2 亿），使用 Rufus 的用户下单可能性比不使用的顾客高 60%。

Rufus 的主要问题在于：用户依然缺乏站内种草心智以及与现有商业化模式的冲突。第一，现在用户的购物决策链路是“站外种草、站内拔草”。Rufus 虽然可以理解用户自然语言，并承接部分复杂需求，但短期内很难改变用户的站外种草心智。第二，货架电商的主要盈利模式是“广告+佣金”，而 Rufus 这类 AI 智能体，依照相关性给用户推荐产品，这就影响了广告投放的效率，降低了商家继续投资的意愿。对于以

第三方卖家服务（3P）为主的平台来说，站内广告投放的营收占比更高，AI 的冲击更大。为了增加用户粘性，Rufus 对于加入广告还非常克制，即便未来增加广告，展示面积也有限。

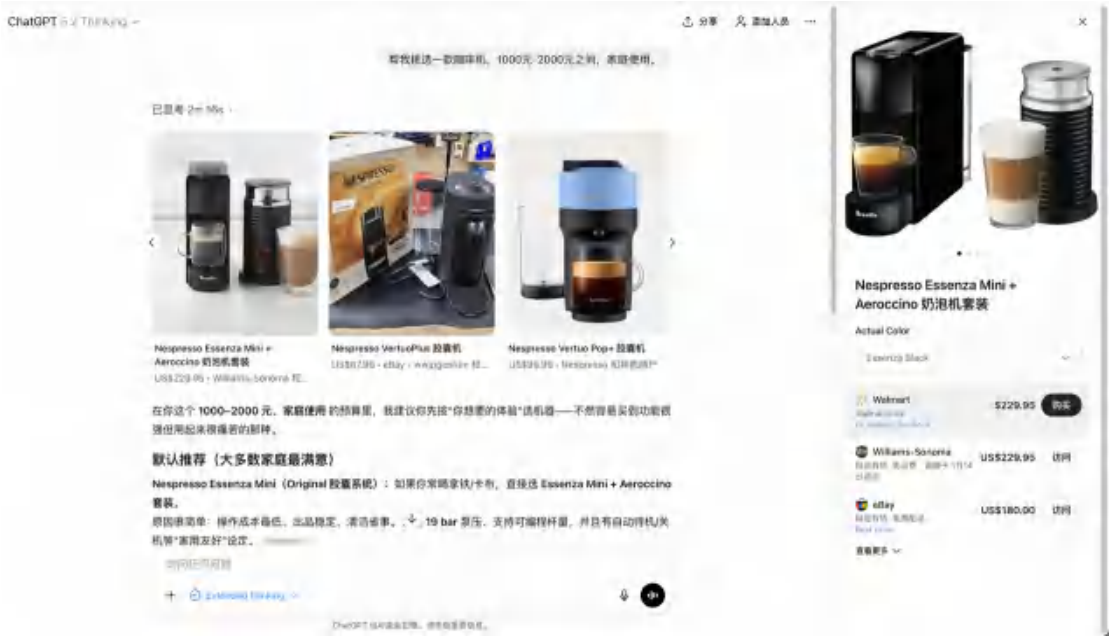
目前看来，Rufus 对亚马逊的成交和利润增量贡献还比较有限。官方表示，2025 年 Rufus 贡献的增量 GMV 将达到 100 亿美元/年，占全站 GMV 的比重约为 1.25%。2025 年 Rufus 预计为亚马逊贡献约 7 亿美元的营业利润，占比约 1%～2%。

二、Hunter 模式：模型企业对电商的新探索

如果说 Farmer 模式还是在电商内部做改良，Hunter 模式则是在外部做全新的探索。在这种模式下，外部智能体成为用户消费的“前置大脑”，它不仅是全新的流量入口，更能理解用户的模糊意图，并将其拆解为一系列具体的商品或服务组合，这些经过拆解的购买任务再被分发给不同的电商平台进行履约与售后。

以 OpenAI 与电商的合作为例，用户在对话框内用自然语言提出需求，AI 完成意图识别，并结合全网商品对比结果后给出推荐，通过 OpenAI 的“即时结账”功能（Instant Checkout），用户可以直接点击选择商品并完成下单。“对话交互-商品信息展示-用户决定购买-输入支付信息-支付-物流

追踪-售后支持信息”的操作都在 ChatGPT 完成，全程无需跳转至外部电商平台。



（一）Hunter 模式带来的消费增量

相比于 Farmer 模式，Hunter 模式有多重优势。首先，相比内部 AI 智能体，外部 AI 智能体拥有更广泛的信息来源，例如系统日历、通讯录，还可能跟其他智能体有更多交互，例如车机、眼镜、智能家电，对用户需求的理解更全面，还可以跨平台对比商品、服务和分发需求。其次，外部智能体可以在跟用户的多轮对话中识别出消费意图，激发潜在需求：例如，了解到用户有每天喝咖啡的习惯，并完成多种场景下的推荐：在用户睡前提示可以帮忙点第二天的咖啡外卖、在用户快到达公司前可以提前下单咖啡、或者推荐用户购买适合的咖啡机每天自制咖啡。这种被激发的潜在消费需求，是

传统货架电商最渴望的市场增量。OpenAI 的 ChatGPT 用户规模已达到十亿量级（月活跃用户 12-15 亿），哪怕每月每个用户只有一个被动激发的消费需求，对于接入的电商平台来说都是可观的增量。

已有反馈显示，Hunter 模式可以带来转化率的显著提升和流量的快速增长：微软的内测数据显示，使用 Copilot 的用户在互动 30 分钟内的购买量比未使用者多出 53%。Etsy 电商的首席增长官表示表明，通过 ChatGPT 购买的消费者表现出比通过传统搜索引流的消费者有更高的购买意图，转化率表现优异。Etsy 财报显示，来自 ChatGPT 的全球入站流量份额从一年前的 1% 飙升至 21%。

（二）Hunter 模式遇到的平台壁垒

尽管 Hunter 模式展现了增量潜力，但在实际推进中，这种模式会面临电商平台的抵触，像 Amazon 就拒绝接入外部智能体，更通过法律诉讼明确反对 Perplexity 的非授权访问。

首先，智能体高效推荐可能影响平台广告收费。外部智能体提高了匹配效率，可能减少商家的广告投放环节，从而影响平台的广告收入。对于平台而言，自营电商对广告的倚重少，但广告是三方电商重要的收入来源，对于接入智能体自然有顾虑。

其次，智能体可能让平台入口流失。在软件产业中，当智能体通过底层协议直接调取功能时，原本拥有独立定价权的软件正趋于“隐形化”并蜕变为 API 服务商。参考 AI 智能体 在通用软件领域带来的冲击，电商平台也会顾虑“入口流失”风险。如果用户的购物意愿能够直接在外部智能体的对话框中得到完整闭环，那么电商平台将彻底失去对用户心智和交互界面的控制权。担心自己失去品牌认知，只是智能体身后的“履约工厂”或“后端仓库”。

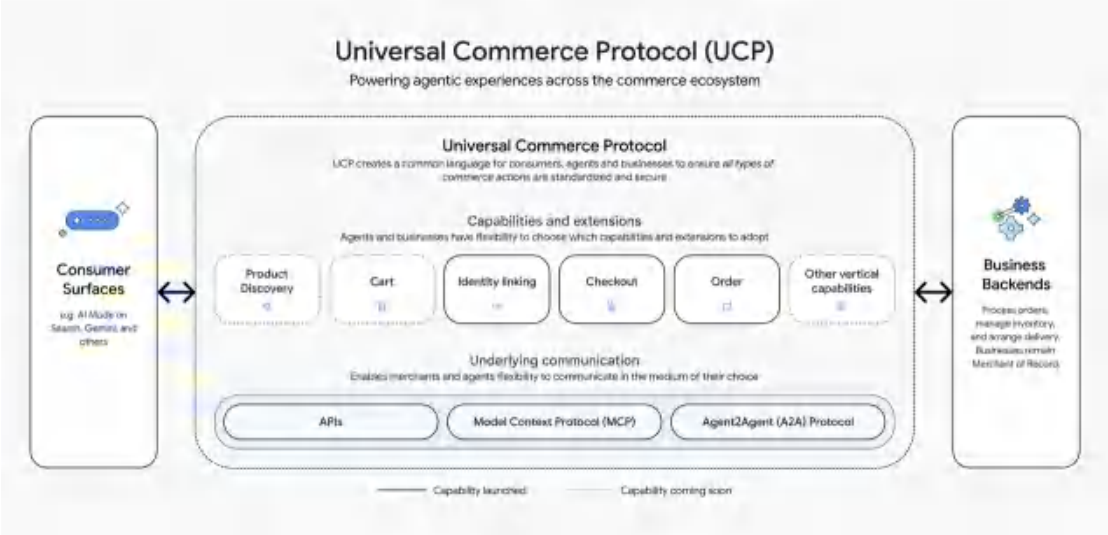
三、智能体与电商平台的协同合作

虽然当前 Hunter 与 Farmer 之间存在壁垒，但协作的核心逻辑在于如何协同创造更大的市场增量。我们可以通过国内外两种不同的演进路径来观察这种协同的可能性：

（一） Google 模式：基于开放协议的分工深化

在海外市场，垂直搜索（Vertical Search）一直是电商行业的主流生态。除了亚马逊等大型平台，还存在众多商家自建独立站，天然依赖 Google、Meta 等外部渠道引流，电商平台默认对搜索引擎开放接口，形成了清晰的价值分配：搜索引擎通过精准引流赚取广告费，电商平台则深耕转化与履约赚取交易佣金。谷歌提出的通用商业协议（UCP）正是这一逻辑在 AI 时代的延伸。在该框架下，谷歌定义了 A2A（Agent to Agent）通信标准，允许外部助手直接嵌套电商平

台的“商户智能体”（Business Agent）。用户无需跳出交互界面，即可直接与平台的 AI 助手对话，获得原生级别的购物体验。这种模式既发挥了 Farmer 模式对垂直品类的专业理解力，也巩固了 Hunter 模式作为流量入口的地位。目前，Google 与零售巨头沃尔玛（Walmart）、长尾电商平台 Etsy 等已达成深入合作。



（二） 国内模式：基于意图对接的增量创造

国内的三方电商平台与外部搜索引擎和种草平台在商业模式上有重合，收入都会来自商家的广告预算，Hunter 与 Farmer 的协同难度显著高于海外。然而，阿里近期的技术演练为破解这一僵局提供了新思路。电商平台本身在做搜索推荐的智能化改造，通过引入节假日、天气、社会热点等世界知识，以及多维度的用户外部行为数据，搜推模型的能力显著增强。这样外部智能体不仅是导流工具，还可以成为用户

意图的传感器。在淘宝披露的技术论文中：如果电商平台能够承接更为丰富“上下文”（Context）信息，推荐系统成功率将有效提升，能带动认知链路成交额增长 10%，大盘成交额增加 1.3%。这种基于意图对接的模式，有可能改变电商平台对外部接入的防御态度。当平台意识到外部意图数据能够创造确定的消费增量时，接入外部智能体将不再是一场损失主权的“零和游戏”，而是一种自发性的增长选择。是否达成合作，可能取决于两点：一是通过精准意图识别创造出的 GMV 增量，足以覆盖 Hunter 与 Farmer 在广告收益上的潜在竞争损耗；二是能否有开放标准接口支持多家外部智能体，形成一个健康的竞合关系，也能消除政府对于封闭生态的顾虑。

（三） 不可忽视的产业链共赢

无论采取何种协作模式，Agent Commerce 的演进都不应仅局限于智能体与平台之间的博弈，而应回归到商业本质：**保障并提升产业链条上商家与制造和服务厂商的切身利益。**

其一，保护商家的私域主权与品牌溢价 深度协作必须建立在安全合规的基础之上。合作必须遵循严格的底层协议，杜绝类似通过模拟用户操作“入侵”平台的越权行为，确保商家的经营数据和用户资产不被非法截流。智能体在拆解需求时，应尊重商家的品牌独立性，避免将所有商品“同质化”为单纯的参数对比，从而保护制造厂商长期积累的品牌溢价。

其二，降低决策成本，提升服务消费渗透 相比标准化程度高的实物商品，服务消费（如旅游规划、医疗教育、家政维修）具有更高的决策门槛和履约非标性。AI 智能体的核心价值在于能将用户模糊诉求转化为结构化的任务清单，显著降低消费者的筛选与决策成本。这种能力的提升不仅能为服务提供商带来更精准的客流，还能通过预先的意图匹配减少履约环节的沟通损耗。

结语：

回顾过去三十年，（Electric Commerce）曾是上一轮互联网革命中最先锋的产业。它打破了地理限制并加速了全球贸易，大幅拉动了物流与支付的数字化进程，同时使制造业能够直接响应消费需求，创造出“小单快返”等灵活生产模式。然而在近二十年中，货架电商业态并未发生根本性改变，内容电商增加了流量入口，但其本质仍是基于感性激发的营销。电商行业在近期并未带来用户体验的革命提升，未能拉升服务消费线上化比例。受此影响，全球电商占社零总额的比例在触及高位后开始趋于平滞。

当前智能体电商（Agent Commerce）的出现，标志着行业正处于决定性的商业质变点。根据摩根士丹利的测算，未来五年内，由人工智能驱动的电商市场规模预计将达到 3 万亿至 5 万亿美元的量级。预计到 2030 年，约 50% 的用

户将养成通过 AI 智能体购物的习惯，且 25% 的消费决策将由智能体自主完成。

这场变革的本质是电商行业从被动响应需求向主动理解意图演进，其核心价值在于将用户从筛选与决策流程中解放出来，不仅适用于实物交易，更将深层触达旅游规划、医疗教育、家政维修等高门槛的服务消费领域。**衡量 Agent Commerce 成功与否的标尺，不应仅是 GMV 增长，更应包括：用户体验是否得到代际提升、技术红利是否大幅转化、高价值服务是否真正实现线上化渗透，以及生态内各参与方能否在新分工中获得合理回报。**

附录

传统电商搜索 vs Farmer vs Hunter

	传统电商搜索	Farmer 模式	Hunter 模式
典型代表	淘宝搜索框、亚马逊搜索栏。	淘宝 AI 万能搜、亚马逊 Rufus。	ChatGPT+Etsy、Gemini+Shopify。
搜索逻辑	**关键词匹配：**基于文本重合度与销量/权重排序。	**对话式引导：**基于多轮问答挖掘用户潜在需求。	**任务/方案驱动：**基于问题解决逻辑，提供全网方案。
数据覆盖范围	**单平台 SKU：**仅限本平台内的商品和类目数据。	**平台闭环数据：**本平台商品 + 用户历史消费画像。	**全网开放数据：**跨平台价格、社交媒体评测、专业百科。
意图理解深度	**低：**难以处理模糊、感性或复杂的长句需求。	**中高：**能理解上下文，将模糊需求转化为具体参数。	**高：**具备常识推理能力，能结合外部知识（如：避雷、参数对比）。
客观公正性	**低：**受竞价排名（广告位）影响严重。	**中：**倾向于推荐平台内高毛利或合作方的商品。	**高：**不隶属于单一电商，理论上更倾向于全网最优解。
决策路径	**用户自助：**用户需在大量 SKU 中自行筛选比价。	**AI 筛选：**AI 缩小范围并给出推荐理由，缩短决策链。	**专家代劳：**AI 充当买手，直接给出“买哪个”的最终建议。
履约与售后	**极佳：**直接在平台内交易，链路完整，保障度高。	**极佳：**原生集成，支持一键下单、查物流、退换货。	**中/存在摩擦：**库存信息可能有延迟，可能需跳转第三方 App 或授权。
主要优点	**明确目标购买效率高；** **商业模式成熟稳定。**	**可以理解模糊的和场景化的自然语义需求；** **快速且多维度的总结和对比：**减少用户决策成本，并提升转化率；	**完整消费链路大幅简化；** **更广泛的信息来源：**用户的日常交互，用户日历、通讯录等内容；

		深耕存量用户：对用户理解深刻； **平台内操作体验连贯。**	**多渠道对比：**实现全网比价，站在用户立场决策； **AI 时代的新流量入口：**不但可以购物，还可以承接非交易性需求； **基于场景的主动推荐。**
主要缺点	**缺少外部流量；** **无法理解自然语义；** **无法处理个性化、非标的需求；** **大量商品对比信息过载。**	**缺少外部流量；** **明确目标购买效率低；** **内部流量不足：**货架电商中闲逛（非明确购买意图）产生的 GMV 占比； AI 交互会**稀释关键词搜索的流量，与广告模式有冲突；** **单一平台选择局限：**无法跳出单一平台生态进行横向对比。	AI 交互会**稀释关键词搜索的流量，与电商平台的广告模式有冲突；** **跨平台信息共享有阻碍：**大型电商平台不愿意充分开放 API； **商业闭环难：**作为流量入口，面临变现与中立性的博弈。

几类 Hunter 模式的对比

	OpenAI ChatGPT 与 Etsy、Shopify	**谷歌 Gemini 与沃尔玛、Shopify**	**微软 Copilot 与 Shopify**
Agent 交互方案	**纯粹外部引流** ChatGPT 仅作为外部 Agent 引流，承担“超级导购”的角色，负责在对话框内捕捉用户模糊的购物意图并收集必要信息	**外部引流+配合站内 agent** • **Agent 账户与电商平台账户互联：**用户可将电商平台的会员账号（Walmart+）与 Google 账号关联，	**纯粹外部引流+商户侧内部工具能力支持** • 在 Copilot 聊天框内引导和捕捉用户购物意图 • **提供商户侧 Brand Agents（品牌智能体）：** 为商家在自

		Gemini 自动识别历史订单、偏好及会员权益 • **可引流至电商平台的站内 Agent：**沃尔玛等零售商可以在 Search 和 Gemini 中调用品牌直营的 AI Agent，继续和用户交互	家网站部署 AI 导购（Brand Agents），以品牌口吻与用户交互，提高转化率和客单价，同时集成 Microsoft Clarity 工具，为商家提供 AI 辅助会话的转化分析仪表盘等分析工具
商户自主权	**由商户保留其私域流量自主权：**交易发生后，交易订单记录会自动完整导入电商的商户后台，商户依然保有私域的客户关系	**维持商户核心自主权益：**商户保持为“账单主体（Merchant of Record）”，商户保有数据和客户关系	**保护商户主权（Merchant of Record）：** 明确商户依然是“账单主体”，承诺商户保留对交易额、客户数据和客户关系的完全所有权
排名逻辑	**ChatGPT 保证排名基于商品相关性，不基于竞价排名**	**Gemini 保证排名基于商品相关性，但允许商户/平台为优惠活动的“弹出权”竞价**	**Copilot 保证排名基于商品相关性，不基于竞价排名**
费用支付	**由商户支付技术服务费用：**商户需要为成交订单支付小额技术服务费	**由商户支付技术服务费用：**商家需为通过“即时结账”完成的订单支付小额费用	**由商户支付技术服务费用：**商家需为 Copilot Checkout 完成的每笔订单支付小额费用

AI 智能体服务产业观察与安全初探

引言：

当前，人工智能大模型的能力正经历从“对话”向“行动”的范式跃升，各类实现意图识别、自主规划与闭环执行的智能体（Agent）喷式涌现。从 2025 年 12 月发布的“豆包手机助手”，到今年年初具备办事能力的千问 APP，手机端智能体率先发力；随后，PC 端智能体服务也迅速跟进，Anthropic、MiniMax 等国内外模型厂商相继推出 Cowork 等 PC 助手，名为 OpenClaw 的开源 AI Agent 框架也凭借全天候执行能力迅速出圈。

然而，与这股技术热潮相伴而生的，是日益凸显的安全和伦理挑战。腾讯创始人马化腾公开指出，部分智能体采用的 GUI 自动化技术与传统“黑产外挂”同源，是“极其不安全、不负责任”的实践；而备受瞩目的 OpenClaw 开源框架，也引起了对 PC 端智能体是否新增网络攻击和数据窃取等风险敞口的安全讨论。

必须承认，智能体安全对我们而言是一个极具挑战的陌生领域。当智能体有了“行动力”会催生哪些新的安全风险？面对智能体对旧有社会关系和伦理秩序所产生的冲击，我们又该如何定义智能体的价值位阶，探索智能体的执行边界？

面对这些问题，本文对智能体服务的新特征和新风险进行辨析，并尝试提出相应的治理思路，期望在推动人工智能从“智力”到“生产力”的转化的过程中，探索自主性和安全性的有效平衡，让创新方案能够平稳落地，并促进健康、向善的人机关系构建。

一、智能体服务的新特征和新风险

智能体服务旨在理解用户意图并自主调用工具完成任务，其技术架构可划分为三层：底层为大模型与基础设施，提供认知与决策基础；中间层为智能体与软硬件的自主交互，呈现“全端协同”与“全局记忆”特征；上层为金融、医疗等垂域应用及面向特殊群体的服务。如图 1 所示，安全风险按此三层分布。

在讨论安全风险时，应当看到，智能体作为基于大模型构建的系统，自然继承了模型层面的安全议题。例如，针对模型的提示词注入攻击、传统互联网环境下的网络和数据安全、以及针对模型上下文协议（如 MCP）的劫持风险，这些问题都需要持续监测和建立新的安全规范。但本文更关注的是**由智能体本质特征驱动的“增量风险”，这些风险源于技术演进与既有治理框架之间的冲突：**

其一是行为自主性带来的边界突破。智能体具备自主规划与执行能力，能够脱离人类的实时干预进行决策。这种从

“受控工具”向“自主主体”的角色转变，当智能体拥有了对软硬件的深度控制权，一旦发生逻辑错误或遭受恶意劫持，其产生的破坏性后果将远超传统软件。

其二是意图理解与隐私保护机制的平衡。为了实现对用户意图的精准捕捉，智能体必须跨应用、跨设备地汇聚多源数据，需要平衡智能体对用户环境全维度感知的需求和传统个人信息保护的原则。

其三是人机互动模式变化引发的伦理冲击。智能体改变了人机交互的界面，随着其深度介入情感陪伴、专业咨询和重大生活决策，在提升效率的同时，也必然引发人类自主决策权的归属以及行为价值导向的争议。

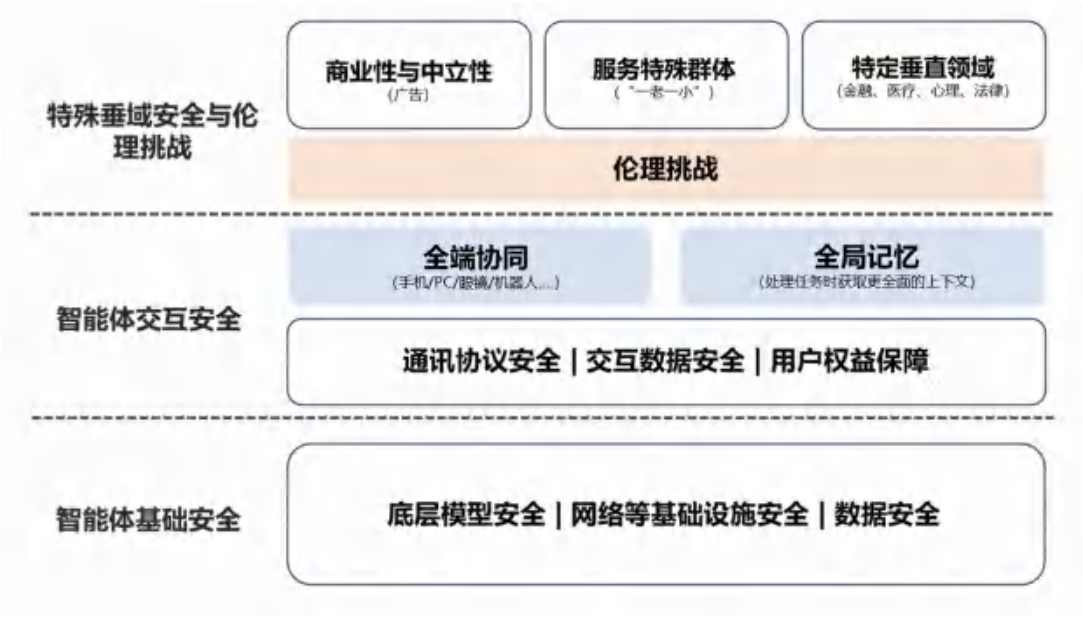


图 1：智能体服务安全风险三层框架示意

（一）“全端协同”：智能体行为自主性突破传统软硬件交互边界，冲击既有安全方案

目前的智能体服务正在推动其部署从手机、PC 等常规终端，扩展至智能眼镜、车载系统、机器人等多样化的硬件载体，并在探索跨设备间的自主联动和交互，从而为用户提供连续、自主的智能化服务，**此即“全端协同”。这种跨端的全自动化运行，打破了传统软硬件交互的边界，在打造用户极致无缝体验的同时，可能对现有的软硬件安全机制带来冲击，并触发新的安全风险。**

首先，针对手机、PC 等常规承载数字智能系统的设备，其所部署的安全防御体系多基于“人机交互”设计，并未考虑智能体这种高度自主的 AI 产品形态，智能体的权责边界不清可能存在权限误用滥用的风险，从而对用户的数据和财产安全带来威胁。例如，近期 Anthropic Cowork 等 PC 端智能体服务形态下，智能体可能因逻辑错误或受注入攻击，执行删除本地关键文件等高危操作。[1]

其次，当智能体被部署到车载、机器人等能与现实世界产生实质交互的硬件中时，数字世界的安全风险延伸至物理世界，可能造成物理层面的直接人身伤害或财产损失。例如，扫地机器人、具身机器人如果被恶意劫持，可能从“家务助手”变成“物理攻击者”，执行诸如撞击、开启危险设备等高危动作和操作。

最后，多设备、多系统的深度互联，可能产生更复杂的攻击路径和风险敞口的扩大化，任何一个节点的失陷都可能

引发连锁反应，导致整个协同网络被恶意控制，如近期 OpenClaw 等开源 PC 端智能体采用虚拟机等云端部署方案时，其云端账户一旦被劫持，就可能被攻击者用作渗透其他关联设备的跳板，导致风险在多设备间蔓延。[2]

（二）“全局记忆”：构建意图理解必须跨应用、跨设备地汇聚多源数据，带来新的数据治理难题

为了实现精准的意图捕捉，智能体必须构建基于多源数据汇聚的“全局记忆”。通过跨应用、跨设备的上下文（Context）同步，其目的在于让智能体能够像人类一样理解对话和任务的复杂背景，从而确保其后续的决策与行为不仅高效，而且更符合用户的真实意图与情境，并提供真正准确和贴切的智能化服务。**但这种模式也带来了新的数据治理难题，主要体现在“一人多源”带来个人信息保护挑战和“多人混同”带来的隐私和商秘安全挑战。**

首先是“一人多源”放大了个人信息泄露的风险，在服务过程中，智能体不仅会汇集用户在社交、购物、金融等不同应用中的数据，还会通过摄像头、麦克风等各类传感器持续感知个人在物理世界中的环境和状态，其包含的个人信息将更全面，此类数据一旦遭到泄露，其危害不再局限于单一场景，而是会导致用户生活习惯、财务状况、社交关系甚至生理健康状态的全面泄露，导致后续被恶意攻击者用于精准诈骗、舆论操纵或数字身份盗用等人身伤害和财产损失。

其次是“多人混同”对他人隐私和商业秘密权益的冲击，智能体在群聊互动和物理终端交互的场景下会不可避免地接触和处理其他数据主体的信息，比如手机智能体在帮用户发微信时会获取到对话聊天框的他人的隐私或商秘、机器狗或者 AI 眼镜在记录用户视角时必然会拍摄到他人等，这可能侵犯了第三方在隐私和商秘保护上的合法权益，也因责任主体模糊，在发生数据纠纷或泄露事件时给确权和追责带来难度。

（三）面向特殊群体与垂域应用：改变人机交互界面，引发诸多伦理挑战

智能体改变了人机交互的界面，用户不再需要遵循预设的流程或繁琐的 UI 按钮，而是通过自然语言下达指令，智能体便能自主完成预订差旅、管理日程等复杂任务。随着其自主性与拟人化程度的不断提升，智能体的角色正从单纯的生产力工具，向具备社会属性的“数字伙伴”乃至“情感伴侣”演进，其应用场景也从代码编写、文档处理等工作领域，全面渗透至教育、医疗、法律等专业服务及日常生活的方方面面。这种角色的转变，引发了关于**人机关系与可授权边界**的广泛争议。

首先，是如何平衡智能体服务的中立性与商业化需求。由于其背后的模型训练与算力维护成本高昂，在当前 C 端市场普遍采取免费策略的情况下，广告变现成为一种可预见的

盈利路径。然而，当智能体已成为我们信赖的贴身秘书甚至亲密伙伴时，用户能否接受在真诚、深度的人机对话中被植入商业广告，甚至是难以察觉的引导性营销？Anthropic 公司在“超级碗”期间投放的一则广告精准地讽刺了这一窘境：当用户向 AI 寻求改善母子关系的建议时，AI 起初给出了如倾听、共同活动等温和建议，却在对话末尾突兀地推荐起熟女交友网站。[3] 这一荒诞的转折揭示了商业利益可能对智能体服务中立性与可信度造成的侵蚀。

其次，是当智能体服务于未成年人与老年人等特殊群体时，如何在赋能与保护目标之间寻求共同解和更优解。对于心智尚未成熟的青少年，高度拟人化的 AI 伙伴能够提供即时情感慰藉，让动画片中的“哆啦 A 梦”成为现实，但也极易导致过度情感依赖，影响其社会化发展。对于老年人，自然语言交互的便利性虽极大地降低了他们使用数字工具的门槛，但其相对薄弱的防范意识和潜在的情感空缺，也使其成为利用 AI 进行情感诈骗的高风险人群。因此，如何设计智能体的交互机制，确保其在提供陪伴与便利的同时，能够积极引导青少年健康成长并有效保障老年人的合法权益，已成为亟待研究的公共课题。

最后，随着智能体深入金融、医疗、法律等垂直领域，涉及资产与人身权利的可授权边界引发伦理争议。例如，即便是在用户明确授权的情况下，法律智能体是否有权代理用

户处理卖房、辞职甚至离婚等重大决策？金融智能体能否被授权托管资产并自主进行投资？医疗智能体能否独立给出诊断建议？这些领域本身就因其高度专业性而存在严重的信息不对称，用户往往缺乏足够知识来监督或评估智能体的行为。当智能体的自主化决策深度介入现实法律关系与生命健康时，如何在技术效率与责任主体归属之间寻找平衡，目前尚缺乏有效的公共讨论和社会共识。

二、对智能体服务安全增量风险的治理思路初探

首先，对于“全端协同”带来的增量安全风险，其本质是技术边界扩张冲击了既有的安全机制，更偏向于技术安全，治理策略应遵循“边发展、边治理”的原则。目前智能体服务仍然处于早期爆发的状态，业态方案还没有完全稳固下来，具有极高的不确定性，其带来的安全风险也很难在初期被明确识别和完全穷尽，业界将在很长一段时间内处于共同无知的状态。但随着技术的演进和应用的普及，潜在问题会逐渐暴露，解决方案也会随之完善。对此应在风险可控的范围内允许“先把车开起来”，让技术和应用尽可能地试错，让安全问题尽可能地暴露，并据此迭代优化管控措施和解决方案。我国现行的大模型安全评估与备案制度为此提供了可资借鉴的范例。该机制并非旨在设置前置性的市场准入门槛，而是作为一个持续性的动态安全锚点，其核心逻辑在于通过摸

清技术底数、明晰责任主体，在鼓励技术创新的同时，实现持续的风险监测与敏捷治理。

其次，针对“全局记忆”引发的数据治理难题，智能体服务带来的新型业务形态并未从根本上冲击个人信息保护的法益基础，其核心在于如何在现有法律体系内包容技术演进。对此，相对务实的治理策略是通过合理的法律解释来包容和引导新业态的健康发展，实现“治理跟上技术”。我国现行的个人信息保护规则框架主要为适配移动互联网时代的数字业态而制定，并设置了“最小够用、目的限定”的基本原则，防范企业过度收集和处理与核心功能无太大关联的个人信息，并将用户个人信息以及形成的“用户个人画像”用于与主营业务不符的、或引发用户不适的其他商用目的。然而，在智能体服务的新业态下，智能体构建的“全局记忆”并非构建“用户画像”并后续用以其他商业目的，而是服务于用户即时任务的“上下文”，以实现互联的各个系统和设备能够准确掌握用户意图和执行环境，保障执行用户意图的一致性和用户服务的连续性。这并非传统意义上对个人信息的“转换目的使用”，而是对用户服务意图的自然延续。此外，“上下文”数据在技术上常以隐式的向量形式存在，并非人类可读的自然语言，这使得个人身份的可识别性被大幅削弱。因此，尽管“全局记忆”模式下的多方数据汇聚对安全技术保护措施提出

了更高要求，但其运行模式并不会冲击到个人信息保护法旨在守护的根本法益。

对此，治理的关键在于根据人工智能技术的新特点对现行法律进行合理解释，兼顾个人信息权益保障和智能体服务数据利用的需求。一是聚焦可识别性特征，将向量化、不可逆还原的过程数据确定为匿名化的个人数据，在充分尊重用户权益、严格保障安全的前提下，放宽目的限定、最小必要等与技术发展不相匹配的限制性要求。二是多源数据从静态的使用限制走向动态的环境可信，提高跨端、跨域数据处理的总体可信度和全局可控性，确保数据来源可追溯、责任可区分，建立第三方安全评估等可审计的透明度机制，数据处理者之间通过“协议承诺+技术保障”确保数据汇聚过程中安全水位的一致性。三是保障用户充分知情和主动退出的权利，明确告知用户数据需求、处理目的和潜在影响，向用户披露数据流转协议、安全保障机制和第三方评估报告，建立用户权益投诉响应机制，在此基础上采取一次勾选授权、避免反复打扰的同意模式，无需每次数据跨主体传输都进行单独弹窗提示，在保障用户数据安全的同时提高服务的流畅度和便捷性。

最后，面对智能体服务触发的伦理挑战，这已超越了简单的合规范畴，是一个典型的“共同无知”下的社会共识构建问题。这类问题不存在一蹴而就的解决方案，也无法由任何

单一主体来定义最终答案，治理策略的关键在于“及早开启对话、协同演进适应”。与技术安全问题不同，伦理共识的形成是一个漫长且需要社会各方广泛参与的动态过程。以今年广受关注的 AI 陪伴类硬件为例，当此类产品深度介入儿童的成长过程时，一系列伦理难题也随之而来：设备在交互中了解到孩子的困惑与秘密，哪些信息应该同步给家长，哪些又应当为了保护孩子的独立空间而保留？当孩子向它探讨关于社交、成长乃至更严肃的社会话题时，它应该如何回应才能既不越位又能提供有益的引导价值？这一系列复杂的伦理权衡，并没有现成的答案，其界定直接关系到技术的社会接受度，也深刻影响着下一代的认知与成长模式。正因如此，我们必须及早将相关议题纳入公共议程，在多元对话中寻求共识。为此，社会各方都需要主动做出改变：普通公众需要积极提升自身的 AI 认知与数字素养，理性理解智能体的能力与局限；科技企业则需秉持负责任的创新原则，一方面应主动构建“科技向善”的模型规约与产品设计，另一方面在面对显著伦理争议时，要有自我约束、“向后一步”的审慎自觉；对于政府，则应发挥关键的引导与托底作用，特别是在一些伦理高度敏感但具备重大社会价值的领域，市场力量往往难以独立逾越准入与信任的门槛，此时需要政府通过提供公共信用背书与引导性资源投入，在风险可控的框架内组织和支持先行先试，为社会探索可行的路径。

结语：

当前，我们正处在智能体服务应用落地的爆发初期。一方面，技术与业态的快速演进正在突破传统的软硬件交互边界。随着“全端协同”与“全局记忆”等新特征的出现，智能体实现了跨设备联动与深层情境理解，但同时也冲击了既有的安全机制并产生增量风险。另一方面，随着应用场景的不断深化，智能体的角色正从单一的生产力工具向具有社会属性演进，这一角色转变在重塑人机交互界面的同时，也引申出关于商业性与中立性、未成年人保护、权利授权边界等深层伦理挑战。

应当客观地认识到，上述风险本质上源于新技术演进过程中的不确定性，是产业走向成熟的必经阶段。而在共识真空区，“人工智能安全作为公共产品”这一理念则更加重要。这是由于在这一阶段，新的问题将源源不断地暴露出来，仅靠企业或政府任何单独一方是无法全面解决的。对此应当在“共商共建共享”的基本框架下，鼓励政府、企业、学界及社会公众等多方主体共同探索新的问题，积累共识与资源，同时发挥政府引导与市场激励的协同效应，提高行业整体的安全水位，驱动智能体服务在安全可信的轨道上，实现真正有益于社会的创新与发展。

参考文献：

[1] Medium, *Anthropic Cowork Security Deep Dive: When AI Gets Real Access, How Do You Lock Down the Perimeter?*：https://medium.com/aimonks/anthropic-cowork-security-deep-dive-when-ai-gets-real-access-how-do-you-lock-down-the-perimeter-418e82212e28

[2] 参见 OpenClaw 官网：https://openclaw.ai/；以及《爆火背后：OpenClaw 开源 AI 智能体应用攻击面与安全风险系统剖析》，https://mp.weixin.qq.com/s/QUuFRsvOopxcW874PLra-Q

[3] Arstechnica, *Should AI chatbots have ads? Anthropic says no:* https://arstechnica.com/ai/2026/02/should-ai-chatbots-have-ads-anthropic-says-no/

Manus 事件的起因、走势以及启示

2025 年年底，Meta 宣布完成对通用自主 AI 智能体公司 Manus 的收购，据传交易金额约 20 亿美元。今年元月 8 日，商务部发言人何亚东回应称：中国政府支持企业依法依规开展互利共赢的跨国经营与国际技术合作。同时，企业从事对外投资、技术出口、数据出境、跨境并购等活动，须符合中国法律法规，履行法定程序。商务部将会同相关部门对此项收购与出口管制、技术进出口、对外投资等相关法律法规的一致性开展评估调查。

我们认为：Manus 事件不仅是美国 Reverse-CFIUS 制度诱发的短期局部震荡，更是中美 AI 商业化路径范式长期冲突的微观缩影，且折射出我国 AI 创新生态在资本配置、算力成本、市场认知、商业文化等多维度的系统短板，尤其是 AI Agent 时代高昂的推理成本与本土较低的单个用户平均收入（ARPU）形成商业错位，倒逼优质应用层企业向全球高溢价市场转移，甚至不惜洗白以实现“离岸化”生存。

不可否认，当前针对 Manus 评估审查必要及时，体现了防范“恶意洗白”、补齐出口管理短板及地缘政治博弈的多重战略考量。但从长期来看，若政策层面无法在维护技术主权

与保障商业自由之间确立新的动态平衡，我国恐将面临顶尖创新力量对本土生态系统性“脱嵌”的风险。

为此，我们建议：应继续牢牢确立“不发展是最大不安全”的共识，针对 Manus 的评估审查应尽可能“就事论事”，防止泛化扩大，同时推动政策重心，尽快从单一的“管制限制技术外流”，向优化内部创新生态、鼓励企业积极出海、着力建构“全球双跨”的发展模式转变，实现创新资源在合规边界内的全球化配置和内外有效联动。

一、AI 初创企业左右为难：理解 Manus 的两个维度

（一）美反向投资审查：迫使中国初创企业主动离岸经营

Manus 母体蝴蝶效应（Butterfly Effect）公司注册于北京。2025 年 3 月，作为产品的 Manus 爆火出圈之后，蝴蝶效应在 6 月决定将 Manus 迁往新加坡。这其中，最直接的驱动要素即为满足美国反向投资审查制度要求，即通过迁址来实现对美国投资人的吸引。

2025 年 1 月，美财政部投资安全办公室公布了对华投资审查的最终规则《关于美国在有关国家投资某些国家安全技术和产品的规定》，要求美国主体“知悉”对从事受限业务（Covered Activity）的受限外国主体（Covered Foreign Person）开展受限交易（Covered Transaction）时，按照受限业务类型，区分“被禁止交易”（Prohibited Transaction）或“需申报交易”

（Notifiable Transaction），将分别触发美国主体的禁止交易/申报义务。

据此，2025 年 5 月，美财政部函询 Benchmark Capital 对 Manus 的 7500 万美元投资。最初，行业基于以下理由，认为投资不属于受限管辖范围：**首先**，Manus 并未自主训练 AI 模型，仅调用 Claude、Qwen 等既有模型，只部分微调以便基于现有模型构建应用，更多体现为“封装”；**其次**，Manus 的母公司“蝴蝶效应”注册地为开曼群岛，非中国大陆；**第三**，Manus 员工散布于美国、新加坡、日本和中国；**最后**，数据存储在西方公司运营、位于中国境外的云服务器上。

然而，基于 Reverse-CFIUS 制度，Manus 当时实无法规避审查：

从受限交易性质看：“开发”人工智能模型系统构成受限交易，所谓开发涵括了任何设计开发行为，如设计数据、概念、分析和研究等，或对模型进行的实质性修改。应当说，Manus 非基础模型，而是一款通用 AI Agent，主要基于用户指令，把问题分解成每个内嵌模型能处理的任务、进行统筹工作，**但为了实现最佳效果，Manus 需对原始模型“定制、配置或微调”，该行为本身可以构成开发**。值得注意的是，Reverse-CFIUS 制度排除了基于内部非商业用途（不是为出售或收取使用许可费）定制、配置或微调第三方开发的人工智能模型，然而 Manus 定制、配置或微调并非着眼“内部用途”，

而是构成了出售或使用许可。**因此，Manus 不构成豁免类型交易特征。**

从受限外国主体性质看：“受限外国主体”为“受关注国家/地区的法律组建的，或主要营业地点位于受关注国家/地区的任何实体”，蝴蝶效应虽注册于开曼，但其实质业务开展在中国，意味着蝴蝶效应受到管辖。**为此，去年 6 月，蝴蝶效应通过紧急设立新加坡海外主体的方式，通过规避受限外国主体达到规避 Reverse-CFIUS 制度审查的作用。**从事件后续走势看，这一动作取得实效；值得注意的是，Reverse-CFIUS 制度中“上述任何实体合计有持有 50%或以上股份的任何实体（子公司）”的界定，意味着管辖仍可能从蝴蝶效应穿透至新加坡公司，照此反推，新加坡公司的投资来源应多元化，蝴蝶效应对其控制权应不足 50%。

（二）中国技术出口管理趋严：企业出海离岸经营合规风险提升

我国《技术进出口管理条例》是技术进出口的基础法规，适用于**所有**技术进出口，条例对技术出口实行**分类管理：禁止出口**技术不得出口；**限制出口**技术实行许可证管理；**自由出口**技术实行合同登记管理。不定期更新的《禁止或者限制出口的技术目录》，是该条例的执行指南。此外，我国《出口管制法》和《两用物项出口管制条例》，又针对两用物项、军品、核以及其他与维护国家安全和利益、履行防扩散等国际义务相关的**货物、技术、服务**等物项（统称管制物项）实施出口管制。

概言之，我国对技术出口的管理，遵循两条核心路径：一是一般性技术管理，适用《技术进出口管理条例》；**二是**如技术涉及到国家安全和利益、防扩散，则同时适用《出口管制法》（但出口管制不只是管技术，还包括货物、服务）。

随着数字经济、人工智能的发展，算法类技术日益成为管理制度的深水区：**一方面**，作为一般性技术，算法在《技术进出口管理条例》项下可被列为限制类，2020 年，“基于数据分析的个性化信息推送服务技术”（以下简称“9618 技术”）纳入《禁止或者限制出口的技术目录》；**另一方面**，作为“两用技术”，算法在《出口管制法》框架下因涉及国家安全而具备高度的自由裁量属性。**值得重视的是，尽管 9618 技**

术目前虽暂未被纳入《两用物项和技术进出口许可证管理目录》，但从逻辑看，其完全具备依据《出口管制法》和《两用物项出口管制条例》纳管的条件。

具体到 Manus 场景之下：作为 AI Agent 代表企业，其核心算法技术可能落入“9618 技术”监管范畴。在转移新加坡过程中，Manus 虽只是个工具，但为了后续持续优化，或携算法技术的底层代码出境，该行为可能构成 9618 技术非经许可出境而遭监管。若 Manus 在未获得商务部技术出口许可的情况下自行出境甚至完成收购，将导致技术控制权实质性转移，违反我国技术出口相关管理条例。

二、市场强力驱动：Manus 远走新加坡的基本考量

尽管受到美中双方的政策限制，Manus 依然选择远走新加坡，表面看是为了拿到美国投资、获取估值溢价，更深层次地，此举受到国内软件市场特点、投融资条件多种掣肘下的无奈之举。简言之，全球市场比中国本土市场更适合高阶付费的 AI Agent 工具的长期可持续发展。

一是国内“重硬轻软”，缺乏标准化 SaaS 生态。美国和欧洲市场拥有成熟的 SaaS 生态（Salesforce、ServiceNow 等），而中国企业级市场长期以“定制化开发”为主，大部分业务逻辑锁死在私有化的旧系统或重度定制的 ERP 中。这些系统是彼此隔离的烟囱，既没有统一的工作标准拆解（SOP），也

缺乏能让 Agent 直接调用的 API 接口。这种已成定式的长期非标化环境，导致 Agent 缺乏基本的执行工具，也难以在企业内部跑通自动化执行闭环，直接限制了其商业落地价值。

二是国内用户付费意愿低，难以支撑 Manus 巨大的资源消耗成本。与过往以对话为主的 Chatbot 模式不同，Agent 在执行复杂任务时需经历多轮任务循环（如跨屏交互、动态代码执行），其消耗 token 量相比普通对话功能，呈几何倍数增加，单次复杂任务的推理费用可达 1-2 美元。高使用成本需要付费能力来覆盖，欧美市场有为软件付费的文化，企业和专业用户通常的订阅费用在每月 20-50 美元以上，足以覆盖 Manus 的资源消耗成本。相较之下，我国内用户对高单价的订阅模式接受度较低，Manus 在国内运营难免会“做多亏多”的财务黑洞，这同样也是国内 SaaS 生态不成熟的一个表现。

三是在国内使用模型受限，初创企业较难采用“双轨制”路径。Manus 需要根据任务特点和接口价格，动态选择全球最佳匹配模型。而考虑到部分美国模型对我国访问限制，若长期深耕国内，只能在模型世界的“半壁江山”中做有限选择。即便采用“在海外用海外模型、在国内用国内模型”的“双轨制”，对于仅有数十人的敏捷团队，其背后的代码适配、模型微调差异以及两套合规逻辑的维护，将极大地拖慢产品迭代周期。这种隐形成本不仅是资金的消耗，更是会拖延创业公司最宝贵的“迭代速度”。

三、围绕 Manus 事件的思考和政策建议

整体来看，Manus 事件不仅是美国 Reverse-Cfius 制度诱发的局部震荡，更是中美 AI 商业化路径范式冲突的微观缩影，折射出我国 AI 创新生态在资本配置、算力成本、市场认知、商业文化等多维度的系统短板。当前，美方正凭借其成熟的 SaaS 订阅文化，率先步入以 AI Agent 为核心的应用红利期时，我国仍处于底层技术突围的阵地战中，高昂的推理成本与本土较低的用户 ARPU 值形成商业错位，倒逼优质应用层企业向全球高溢价市场转移，甚至不惜洗白以实现“离岸化”生存。

不可否认，当前针对 Manus 评估审查必要及时，体现了防范“恶意洗白”、补齐出口管制短板及地缘政治博弈的多重战略考量。但从长期来看，若政策层面无法在维护技术主权与保障商业自由之间确立新的动态平衡，我国恐将面临顶尖创新力量对本土生态系统性“脱嵌”的风险。

为此，我们建议：**应继续牢牢确立“不发展是最大不安全”的共识，针对 Manus 的评估审查应尽可能“就事论事”，防止泛化扩大，同时坚定不移地鼓励出海，推动政策重心从单一的“管制限制技术外流”，向优化内部创新生态、鼓励企业建构“全球双跨”的发展模式转变，实现创新资源在合规边界内的全球化配置和内外有效联动。**

首先，以更大的开放对冲美国的围堵封锁，鼓励国内企业确立“全球双跨”发展范式，统筹发展境内外两个大市场。针对 Manus 事件所折射出的美方限制政策的“驱离效应”，应打破应激式的“固堤思维”，以及 “国内成熟再进军海外”的传统时序逻辑，**第一时间支持具备核心竞争力的科技平台型企业、初创企业，早早建立“全球双跨”经营架构**。建议通过政策引导，鼓励企业在境内保留核心研发力量、在境内推广产品应用，同时坚定不移地鼓励企业“走出去”，用好全球市场的充足算力、SaaS 订阅红利，在海外高 ARPU 值市场跑通商业闭环，做大做强，并择机反哺国内，实现内外自然贯通，防止顶尖人才与创新资本因本土应用生态“盐碱化”而被迫“脱嵌”。

事实上，当前中国企业已完全具备整合全球产业要素、参与顶尖竞争的硬实力。 虽然前述“双轨制”对运营、研发和合规提出了更高要求，但 Plaud 和安克（Anker）的成功已提供了清晰的可复用路径：在海外大市场做大规模、摊薄创新成本，进而反哺国内就业、生产与税收，实现内外市场的自然融通。这种主动的战略布局，不仅是企业的生存之道、领先之道，更是中国科技力量在“后脱钩时代”保持全球竞争力的关键胜负手。

其次，技术出口管理方面，精细化“负面清单”操作。Manus 或涉及算法技术底层源代码未经审批出境，这一方面

反映了我国既有技术出口管理尚存漏洞，另一方面也暗示我国技术出境禁限清单，存在不够清晰、不易执行的问题。**虽然就 AI 技术的不确定性而言，条款需保持一定的弹性以保证我国有充分的裁量权，但从初创企业视角看，为争取全球市场机遇，可能会选择从种子阶段就离境发展，在海外设立研发主体以避免长大后再面对艰难的“国籍”选择。如动辄得咎，易成为难以承受之重**。

我们建议，监管清单亟需从“泛化震慑”转向“精准定位”；进一步清晰化我国技术出境的红线底线清单，实现负面清单的动态化更新，精准分类管理，防止清单表述不清、界定不明，抬升企业合规成本，也唯有留住 AI 初创企业，才能持续享有裁量技术出境、增强博弈筹码的可能性。借此，**一方面**防范以纯粹规避国内监管为目的的恶意“洗白”模式，确保核心人才与创新资产与国内生态体系保持关联性，**另一方面**也能支持鼓励有意愿、有能力全球经营的科创企业继续走出去。从远期来看，还可采取税收优惠等政策，吸引全球化经营的企业把技术成果和商业带回国内、助力国内创新生态，形成内外良性互动的技术产业双循环交互。