

The Battle for AI Agents: Dispute Insights, Legal Logic, and Future Outlook

Author: YANG, Jianyuan

引言

AI智能体的兴起无疑是当下人工智能领域最具深远影响的变量之一。随着技术的触角从“内容生成”延伸至“用户代理”，其引发的利益重构与规则摩擦，也为行业合规实践带来了新的思考与挑战。

本文尝试从近期争端出发，探讨智能体潜在的演进逻辑，并对技术变革中可能存在的合规边界与异化风险提出一些初步思考。在这一前沿领域，合规路径的构建仍处于动态探索中，我们将保持关注，与各位同仁共同在实践中迎接智能时代的破晓与挑战。

一、争端焦点：AI浪潮下的利益重构与生态竞争

AI智能体的合规性近期引发业内新一轮热议。2025年11月，Amazon针对人工智能初创公司Perplexity AI提起诉讼，指控其通过旗下AI浏览器工具Comet的“代理购物”功能非法访问Amazon系统。我国国内亦有智能体产品引发类似法律和伦理争议。

尽管相关争议于近期集中爆发，但其背后的商业探索与合规争议并非新议题。以Perplexity AI与Amazon的纷争为例，双方摩擦在2024年已初现端倪。同时据我们观察，业内头部企业亦早已开始就智能体业务的开拓进行合规布局。**随着技术能力的突破及应用场景的扩展，相信围绕智能体的博弈将纵深演进，并持续成为产业与监管关注的焦点。**

围绕AI智能体的争议，涉及用户、智能体提供者、既有平台与监管部门等多方主体。**当前矛盾的焦点集中于新兴智能体与既有平台之间，核心是对用户流量、数据流转及商业生态主导权的博弈。**

- 于智能体提供者而言：其跨平台的业务服务能力，意味着有可能成为下一代流量入口。智能体不仅蕴含无限的数据积累潜力，更能通过深入垂直的应用场景，深远影响下游生态。
- 于平台提供者而言：智能体的发展正对其既有商业逻辑产生结构性冲击。当智能体代劳用户完成“搜索-决策-执行”的全链路操作时，平台原有的流量留存机制、广告分发机制面临挑战。第三方智能体尝试进入平台原本闭环经营的“自治领地”，并挑战平台对既有生态规则的掌控力。

我们理解，**这种博弈是技术变革对既往商业模式的深度挤压，并最终传导至法律层面，触发规则重塑。**

二、演进动因：用户使用驱动与包容审慎的监管导向

基于社会需求的内生驱动与监管政策的外部导向，AI智能体正展现出广阔的应用前景与增长空间。

从内生社会需求来看，智能体顺应了提升个体生活体验与优化运行效能的趋势。

- 在个人消费领域，智能体通过跨平台的信息聚合与执行能力，能够协助用户完成从日程规划、比价调研到交易下单的全链路操作，实现时间成本的节约与消费体验的个性化定制。
- 在生产力的工具领域，智能体可提供高效率的信息处理与决策辅助，并显著优化协作流程，契合社会对生产力跃迁的需求。

从外部监管环境来看，国家政策对智能体持鼓励态度，且其运行逻辑与现行法律框架并无根本冲突。

- 政策驱动导向：2025年8月，国务院发布《国务院关于深入实施“人工智能+”行动的意见》，明确提出“新一代智能终端、智能体等应用普及率到2027年超70%、到2030年起推动“数据要素×”与人工智能高质量发展的背景下，智能体作为促进数据流转与利用的重要载体，其发展具备明确的政策支撑。
- 法律价值底色：尽管智能体的深度应用可能对现行法律框架带来新挑战，但现行立法并未直接否定智能体的发展空间。其正向价值体现在：
 - 辅助用户自主决策：智能体通过信息整合与方案优选，可以缓解用户的信息不对称问题，一定程度上有助于实现消费者权益保护法下的知情权与自主选择权。
 - 优化市场竞争格局：智能体通过跨平台的服务整合，可以降低用户的搜索与迁移成本，有助于缓解单一平台的锁定效应，契合竞争法的市场开放导向。
 - 提升交易制度效率：智能体有助于降低交易撮合成本、提升市场运行效率，符合电子商务法等立法的主旨。

三、法理基石：“用户授权”可为支点、但非坦途

（一）代理拟制：以“用户授权”作为主要法理支点

在跨平台资源整合与功能协同的语境下，用户授权是智能体获取操作权限、跨越商业主体间准入壁垒的核心合规路径。智能体提供者通过取得用户授权，拟制形成数字空间中的“代理”关系，将自身活动作为用户操作行为的延伸，进而行使用户原有的平台访问与交易权限。这亦是当前智能体行业实践中的主流路径。

（二）范式革新：传统授权模式的失灵与优化探索

理想的授权体系建立在“充分知情—明确表示—精准执行”的线性逻辑之上。然而，智能体的高度自主性与决策动态性，使这一传统模式在多个维度面临失灵：

- 知情环节：信息揭示的局限与认知鸿沟。事前授权以用户对行为逻辑的充分理解为前提。但智能体的数据处理和技术交互高度复杂，其行为路径随具体情境动态生成，且存在实时变化，用户难以在授权时充分感知其行为的潜在风险与边界，通过有限的静态文本难以穷尽揭示其行为逻辑与潜在风险，使用户授权时面临实质性的认知壁垒。
- 定义环节：授权精度与功能弹性的失衡。过于精细的授权虽能防范越权，但可能会限制系统功能延展与新场景适配，难以适应智能体产品高频迭代的现实节奏；而过于粗放却又易陷入“一揽子授权”的合规泥潭，削弱授权的有效性。
- 执行环节：意图转化失真与操作权限越界。即便授权在形式上成立，其在智能体实际运行中的持续可控性亦面临挑战。
 - 一方面，受限于自然语言的模糊性，智能体在将“用户表达意图”转化为“机器自动执行”的过程中，可能因语义理解偏差导致意图扭曲，叠加其自主决策属性，容易造成最终行为结果与用户初衷的背离。
 - 另一方面，智能体难以严格遵循“最小必要”原则，例如在读屏识别或跨平台接口调用时，容易触达授权范围外的敏感信息或第三方主体数据，造成数据获取的越界与超限。

面对前述结构性挑战，传统的静态合规已略显捉襟见肘。我们理解，智能体的合规逻辑应从“书面的程序正义”向“实质的用户主权”演进。以下试提出几点优化思路，供各位同仁参考：

- 从“晦涩告知”转向“效果呈现”：突破晦涩的技术细节表述，侧重以用户侧的服务体验展示行为逻辑，重点解释操作对用户权益的实质影响。
- 引入“确认与澄清”机制：为弥合自然语言的歧义风险，允许用户重述意图、预览行为与审核执行轨迹，确保智能体的行为逻辑始终处于用户预期与可见范围内。
- 分类分级的风险把控：基于场景风险建立差异化合规机制。例如：
 - 数据访问分级：针对账户密码、财务金融、医疗健康、私密生活数据等高敏感数据，建立更为严格的访问权限管理与脱敏策略。
 - 操作节点分级：区分低风险自动执行场景与高风险二次确认场景。对于诸如合同缔结、资产处置（如支付转账）、敏感数据删除、对外通信、物理设备解锁（如智能门锁）等高风险环节，应坚持人工介入与确认，而非完全交由模型自主决策。
- 即时干预能力的落地：确保用户具备操作层面的实时监督与即时纠偏能力，如设置一键暂停或即时撤回机制，从技术底层保障智能体行为的可控性。

（三）归位约束：“用户授权”并非“法外特权”挡箭牌

即便具备合法的授权基础，智能体的运行边界仍受到多重法律框架的实质约束。用户授权仅能实现权限的“平移”，而非权利的创设或滥用，其边界包括以下维度：

- 用户自身权限的法定边界：智能体的运行以用户授权为支点，其操作空间自然受限于用户本人的权限边界。例如：
 - 个人信息转移限制：其权利行使范围限于“本人同意提供的或基于合同收集的个人信息”（通常不含平台的衍生数据或深度加工数据），亦不得损害他人合法权益（如社交关系中的他人隐私）。
 - 平台操作权限约束：智能体作为用户代理，仍需合理遵循目标平台的用户协议与平台规则。若其行为逻辑实质性违背了用户与平台达成的合意，则可能构成权限滥用。
- 复合行为项的多维合规约束：当用户指令与智能体自动执行合流时，其行为将面临从法律规则到商业伦理的全方位检视。例如：
 - 刑事犯罪红线：若违反国家规定，通过规避、突破技术保护措施等手段，非法获取计算机信息系统数据或对其实施非法控制，情节严重的，可能触及“非法获取计算机信息系统数据、非法控制计算机信息系统罪”。
 - 网络数据安全：不得从事非法侵入他人网络、干扰他人网络正常功能、窃取或以其他非法方式获取数据等危害网络安全和数据安全的活动，不得提供专门用于从事前述活动的程序、工具，否则可能违反网络安全法、数据安全法。
 - 市场竞争秩序：不得妨碍、破坏其他经营者合法提供的网络产品或服务正常运行，应避免违反诚实信用原则与基本商业道德，否则可能构成不正当竞争。

四、远景前瞻：技术异化风险需要法律伦理双重规制

（一）异化隐忧：从“用户工具”到“结构性权力”演变

智能体的发展轨迹，正表现出从“支持工具”向“决策中枢”跃迁的趋势。随着其对数字化生活的深度渗透，这种新型代理机制在未来发展壮大后，可能演化成为一种新型的“结构性权力”，进而引发深层次的法律与伦理挑战。

1.个体意志层面的隐性异化：从“纯粹代理”向“意志干预”转化

虽然智能体在发展初期表现为忠实履行指令的代理，但随着其对用户偏好掌握的加深，未来需警惕其在“用户主位”表象下，利用算法优势对个体意志产生隐性负面影响。

- **全景用户画像与数字监控：**智能体具备串联“APP数据孤岛”的天然能力。在为用户提供便利的同时，也会形成用户的全景精准画像，使个体透明化甚至陷入全时空的数字保护带来严峻挑战。
- **隐性差别待遇与算法歧视：**基于对用户的深度透析及模型内置的偏好，智能体可能在决策建议、服务配给或信用评估等方面对用户实施隐性差别待遇，将算法歧视潜伏于用户交互之中。
- **利益背离导致的诱导与操纵：**需防范智能体背离代理人的忠实义务，转而追求自身提供者的商业目标。例如，通过话术诱导或信息投喂，引导用户进行非必要的消费或违背用户的真实初衷。
- **信息茧房加固与自主性削弱：**若长期过度依赖智能体进行决策，用户可能因算法过滤而陷入更深层次的信息封闭。当用户感知的世界与逻辑高度受限于智能体的筛选时，决策空间可能面临退化与压缩。

2.市场结构层面的潜在冲击：从“中立代理”到“新型中枢权力”更迭

当智能体占据流量与数据后，可能从单纯的中间人演变为新的市场权力。

- **自我优待与流量分配失序：**掌握入口地位的智能体可能基于自身利益实施自我优待，通过技术手段拦截流量或对特定上下游企业实施歧视性排挤，破坏公平竞争的市场秩序。
- **规则渗透与议价空间挤压：**智能体可能利用其作为交易中介的枢纽地位，对供应链和交易规则施加隐性影响（如隐蔽抽成），实质性挤压上游商家与目标平台的议价空间。
- **新型数据壁垒与资源垄断：**随着智能体成为全网数据的控制中枢，可能形成比传统平台更难突破的数据壁垒，造成数据资源的二次高度集中。

（二）未雨绸缪：技术演进需坚守合规红线与伦理关怀

智能体的崛起，不仅预示着技术飞跃，更标志着个人生活、商业生态与社会互动正被全方位重塑。随之而来的，是一系列新的法律问题与合规挑战。尽管当前的法律边界尚在探索之中，但历史经验告诉我们，当技术的触角延展至用户意志与公共秩序的红线时，有力的监管规制从不会缺席。

企业的实质合规能力不仅是对抗风险的盾牌，更是智能体在市场竞争中稳健发展、获取长期信赖的底层根基。真正的长青产品，应在设计之初就建立起法律合规与科技伦理的双重审视机制，将对用户权益保护、市场公平竞争、社会良性发展的考量预先嵌入算法。

无论智能体如何进化，技术的图景应始终服务于人类的自由意志与自主选择。智能体可以模拟互动，却无法替代真实的情感共鸣。让技术恪守工具的本位，让人类自主掌控文明的演进，不仅是AI合规的目标，更是我们作为法律人对数字化未来的应尽责任。



杨建媛 (YANG Jianyuan)

合伙人

yangjianyuan@haiwen-law.com



扫描二维码
查阅律师简历



赵雅泽

zhaoyaze@haiwen-law.com

*感谢秦鹭萌对本文创作的协助