网络首发时间：2026-04-24 11:41:58　网络首发地址：https://link.cnki.net/urlid/14.1066.g2.20260424.0949.002

从“认知中介”到“行动节点”：OpenClaw与智能传播的范式重构

——基于行动者网络理论的AIGS治理研究

◎ 喻国明

（北京师范大学 新闻传播学院，北京 100875）

【摘要】基于媒介化社会理论与行动者网络理论，文章针对OpenClaw（龙虾）现象提出“行动者智能体”这一核心概念，旨在解析AI从认知革命向行动革命跃迁过程中的传播范式重构。研究发现，一方面，OpenClaw通过“认知—执行”的闭环，跨越人机交互的执行鸿沟，从生成式AI大行其道的符号世界进入人类生存的物理世界，其必将重塑“社会—技术”网络的连接逻辑。另一方面，这种高权限的介入也引发深刻的“权限悖论”，即效率优化与系统风险之间的必然冲突。针对这一治理难题，文章提出媒介化治理框架，主张通过动态权限管理、算法透明性建设、多元协同共治及伦理嵌入设计，实现从传统内容监管向复杂行动网络治理的范式转型。本研究为理解AIGS（AI生成服务）时代的传播生态变革及治理创新，提供了新的理论视角与实践路径。

【关键词】行动者智能体　AIGS成熟度　“权限悖论”　媒介化治理　行动者网络理论

【中图分类号】G206　【文献标识码】A　【文章编号】

【DOI】

当下的技术发展，呈现出一条全新的技术进化链的进阶路径——从认知革命到行动革命。2025年年底至2026年年初，一个名为OpenClaw（龙虾）的应用在中国市场展现出惊人的爆发力，短短两个月，其GitHub星标数突破25.2万，[1]超越了Linux内核等老牌项目。与此同时，其生态扩张速度极快，ClawHub技能市场已收录超过5700个技能，[2]国产大模型在该生态中的Token消耗量占比已超过海外模型。[3]这不仅是一次产品爆火，更深层次地反映了这场技术演进的跃迁点：从GPT引发的认知革命（生成能力）到OpenClaw代表的行动革命（执行能力），AI正在跨越传统业务流程管理中的“执行鸿沟”，实现从“能说”到“会做”的质变。[4]

OpenClaw不仅是技术产品，更成为一个“社会行动者”，其红色龙虾图标演变为行动者网络的视觉符号，驱动了数百万用户的行为改变。而高权限、弱边界的系统设计带来的“权限悖论”——效率优化与系统风险的必然冲突，正在成为智能传播时代的新问题。

本文不是单纯的技术评价或产业分析，而是从传播学视角重新审视AI技术的社会化过程。笔者认为，理解OpenClaw现象，等于理解智能传播时代的新型行动者网络形成、传播及风险涌现的全链条。

一、理论重构：作为关系重塑者的行动者智能体

1. 理论视域的拓扑：媒介化社会的在地性转向

媒介化理论的奠基人之一斯蒂格·夏瓦指出，现代

基金项目：国家社会科学基金重点项目“数智媒介对政治观点极化影响的国际比较研究”（25AZD037）

作者信息：喻国明（1957— ），男，上海人，北京师范大学新闻传播学院教授、博士生导师，北京师范大学传播创新与未来媒体实验平台主任，中国新闻史学会传媒经济与管理专业委员会创会理事长，主要研究方向：新闻传播学理论、新媒体研究。

社会的各个领域正在被媒介逻辑所渗透，媒介不再仅仅是信息传递的通道，而是成为塑造社会关系的力量。[5]这一理论框架在AI时代需要深化。媒介在地性——特定技术如何在本土社会土壤中生根变异，是理解数字社会的关键。OpenClaw在中国引发的极大关注，正是媒介在地性的典型案例。基于此，笔者将媒介化社会理论与行动者网络理论相融合，提出“智能体作为网络节点”的新观点，即OpenClaw通过工具调用能力成为社会技术网络的连接者，其红色龙虾图标实质是行动者网络的视觉符号化，标志着非人类行动者在社会网络中的正式“出场”。换言之，OpenClaw的“龙虾”符号不仅是一个Logo，更是媒介在地性的视觉宣言，它标志着非人类行动者正式获得了社会网络的“入场券”。

2. 网络节点的重构：ANT视域下的非人类行动者

拉图尔在其2005年出版的经典著作《重组社会：行动者网络理论》中指出，社会不是由人类单独构成的，而是人类与非人类行动者（如技术、物质、制度）的混合网络。[6]在拉图尔的眼中，社会是人与物的混合体。在OpenClaw时代，AI智能体以其“技术—符号—社会”的三重属性，成为重构社会关系的最强异质性节点。传统传播学研究往往将AI视为“工具”，但这忽视了AI在社会网络中的主动作用。而OpenClaw的行动者属性体现为：在技术维度上，其具有四层架构的功能自主性；在社会维度上，其“数字员工”身份得到了集体认可；在符号维度上，其龙虾图标拥有集体想象。

3. 爆发逻辑的解码：从“S曲线”到“引爆点”奇点

罗杰斯的创新扩散理论[7]解释了新技术如何在社会中传播，但其“S曲线”模型却无法解释OpenClaw的爆火。事实上，OpenClaw的走红不是像滚雪球那样慢慢变大的，而是像烧开水一样，温度一到就瞬间“沸腾”。笔者借用格拉德威尔的“引爆点”理论，给这类现象列了一个一夜爆火的万能公式：“爆火程度 = 技术够硬 + 大家够急 + 痛点够准”。[8]换言之，OpenClaw的爆发不是温吞地扩散，而是技术硬实力与社会焦虑感在痛点处的剧烈核聚变，其用48小时改写了互联网产品的冷启动历史。

这一过程就好比做一道爆款菜，OpenClaw恰好凑齐了三个关键“佐料”：“技术够硬”，即现在的AI大脑（大模型）已经足够聪明，能听懂人话了，这是基础；“大家够急”，即全世界的程序员都在焦虑，光有会聊天的AI没用，大家迫切需要AI能动手干活；“痛点够准”，即企业端正好急需自动化流程来提质增效。结果呢？这三样东西在同一时间凑齐了，就像干柴遇到了烈火，出现了“48小时就有10万人点赞”的现象。

4. 核心概念的操作化：作为“数字员工”的行动者智能体

笔者将行动者智能体定义为：一个既具有自主执行能力，又被赋予社会身份的非人类实体，在给定的“社会—技术”网络中，通过符号传播与物理介入，改变人类决策与行为结构的力量。简单来说，我们可以把行动者智能体想象成一位“拥有正式编制的数字员工”，它不再仅仅是电脑里的一个软件或工具，而是一个既能动手干活，又被大家当成“同事”来看待的非人类。它通过在网上说话（符号传播）和在系统里办事（物理介入），实实在在地影响着人们的决定和行为。

具体而言，OpenClaw作为一个“刚入职的新员工”，具有三重身份。

第一重身份：它是“实干家”（技术实体）。这就好比员工的硬技能。它不是只会纸上谈兵，而是真能干活。OpenClaw有一双“手”和“眼睛”，它能通过API接口去调用各种工具（比如发邮件、查数据），还能看懂图片和文字。简单来说，就是“给个任务，它能真给你办成”。

第二重身份：它是“大明星”（传播符号）。这就好比员工的人设和品牌。大家提到它时，脑海里会有具体的形象和说法。OpenClaw有一个标志性的龙虾图标，大家提到它时会说，“这是我们的数字员工”。这种形象和口号，让它在大家心里不仅是一串代码，而且是一个有性格、有辨识度的存在。

第三重身份：它是“正规军”（社会行动者）。这就好比员工的工牌和岗位责任书。它像员工一样被组织接纳了，享有权利和义务。公司或人类主体，都会正式给OpenClaw开通权限（比如允许OpenClaw访问内部数据库），同时也明确了出了事该谁负责。这意味着它已经融入了人们日常的工作流程和规则中，成为团队里“被制度认可的一员”。

5. 范式革命的尺度：AIGS成熟度模型的四阶演进

AIGS（AI Generated Services，AI生成服务）对于AI的发展来说是一个巨大的跃迁。过去的AI（比如聊天机器人）像一个“百事通实习生”，你问它问题，它能给你一个答案或一段文字，但它能做的也仅限于此。比如，你问它：“怎么订机票？”它会告诉你步骤，但不会帮你去执行。而AIGS则像一个“全能数字员工”：你

直接给它下指令，它不仅知道怎么做，还能直接帮你把事情办成。比如，你直接对它说“帮我订一张下周一去北京的机票”，它就会去各大平台查询、比价，并完成预订。所以，AIGS的核心就是服务——它从二维的符号世界中走出来，进入人类生存的三维世界。它不再只是生成内容，而是通过生成内容完成任务、提供服务，直接整合到人们的工作流程中，帮助人们解决问题。

AIGS的进化史，就是一部从生成内容的二维平面走向执行任务的三维空间，最终实现与人类自主协同的“史诗”。AIGS这一技术能力本身的深度和复杂性，可以用AIGS的成熟度模型的层层递进加以表示，它描绘了AI如何从简单地响应指令，一步步发展到能够自主规划并执行复杂任务的全过程。[9]这就是AI生成服务的四阶演进：建言者（仅提供信息）—建议者（拥有认知能力）—执行者（拥有物理介入能力）—共谋者（拥有目标设定权）。换言之，AIGS的四阶演进，就是从生成内容到运用知识，再到执行任务，最终实现自主协同的跃迁过程。它标志着AI正在从辅助人类的工具，进化为能够独立提供复杂服务的“数字员工”。而OpenClaw在AI发展中所处的位置，就是刚刚开始拥有物理介入能力。其技术核心是，AI具备了任务规划和工具调用的能力，即它能理解复杂意图，并将其转化为一系列可执行的API调用，从而与现有的软件系统（如CRM、ERP）进行深度交互。其服务形态就是可以提供端到端的业务流程服务，真正开始“干活”了。

概言之，AI不再是沉默的工具，而是拥有龙虾图腾的“数字员工”。它通过技术实体、传播符号与社会身份的三重作用，在社会网络中完成了从工具到行动者的华丽转身。

二、生态“震荡”：传播规则的重构、学科的范式转移与社会的液态化

当算法代理取代把关人，当“万物皆媒”打破物理边界，我们迎来的不仅是效率的狂欢，更是一场关于去中心化悖论与社会关系算法化的深刻反噬。因此，OpenClaw现象绝非一次简单的技术产品迭代或产业风口，它本质上是一个强大的行动者智能体被大规模、无门槛地引入社会复杂系统的历史性事件。这一事件的深远影响，在于它作为一种具备自主行动能力的非人类行动者，正以前所未有的深度和广度，介入并重塑人类社会的传播生态、知识生产体系乃至社会结构本身，从而引发了一系列深刻的社会媒介化进程的质变与潜在的反噬效应。

1. 传播生态与规则的重构：从“认知中介”到“行动节点”

在传统的传播生态中，大众媒体也好，社交平台也罢，本质上扮演的都是信息中介的角色。它们连接着内容的生产端与消费端，依靠把关和议程设置把控信息的流向，进而影响公众的认知。

但OpenClaw的出现，正在彻底打破这一固有模式。它不再局限于信息的传递，而是将传播网络重构为一个以行动执行为核心的“人—机—环境”混合生态系统，从根本上改变了信息流动的逻辑。

（1）规则的重塑：从把关人到算法代理的权力转移。传统媒体的把关人理论在此失效，取而代之的是一种更为直接、更具穿透力的算法代理机制。OpenClaw作为一个开源的AI智能体执行框架，其核心设计理念便是将自然语言指令无缝地转化为实际的物理或数字操作，实现从对话到行动的跨越。这一过程的实现，依赖于其强大的API接口调用能力和技能系统。

与传统媒体通过报道影响公众决策不同，OpenClaw通过API接口直接调用底层数据和服务，彻底打破了组织、平台乃至应用之间的边界，使得信息的流动不再仅仅是认知层面的传递，而是直接演变成行动层面的执行。例如，网经社发布的报告中提到支付宝支付集成Skill，允许开发者仅通过一句自然语言指令，即可调用用户的支付接口完成交易。在此场景中，OpenClaw扮演的不再是关于“交易”这一信息的中介，而是执行“交易”这一行动的节点。传播指令（如“帮我支付水电费”）被直接翻译为金融交易行动，中间几乎没有传统意义上的人类审核与把关环节。OpenClaw遵循的“推理+行动”循环范式，使其能够自主地将模糊指令拆解为一系列精确的可执行的操作，从而形成“理解指令—规划任务—调用技能—反馈结果”的自动化执行闭环。这种从信息传播到行动执行的转变，是传播生态规则的一次底层重构。

（2）去中心化的悖论：开源理念与新权力中心的崛起。OpenClaw的开源基因和社区驱动模式，让人们看到了技术去中心化和民主化的希望。但现实却给我们上了一课：在看似去中心化的架构之上，一个名为ClawHub的技能市场迅速崛起，形成了新的权力中心。

作为官方唯一的技能分发平台，ClawHub的门槛低得惊人——只要有GitHub账号就能发布技能。这确实让生态一夜之间繁荣起来，赛迪网的数据显示，平台

上短时间内就冒出了5700多个技能，从文件管理到系统控制，几乎无所不包。但这种“集市式”的野蛮生长背后，藏着巨大的质量和安全隐患，繁荣的表象下其实危机四伏。

多个相关的安全报告和市场分析指出，ClawHub上的技能质量参差不齐，充斥着大量低质量、功能重复甚至包含恶意代码的技能。[10-11]由于早期缺乏严格的自动化审核、代码审查或签名验证机制，恶意开发者可以轻易上传窃取用户API密钥、篡改数据或执行破坏性操作的技能。虽然平台后续引入了VirusTotal等第三方安全机构开发的检测引擎，进行自动化安全扫描，但这更多只是一种事后补救。

更深层次的问题在于，ClawHub的运行机制催生了新的算法霸权与流量垄断。与所有应用商店类似，ClawHub的首页推荐、搜索排名和热门技能榜单具有巨大的流量导向作用。尽管其推荐算法的细节并未完全公开，但人们普遍认为其主要依据下载量、用户评分、更新频率等指标。这导致了“马太效应”：热门的技能愈发热门，而大量有创意但缺乏初始流量的小众技能则被淹没。这种由算法主导的中心化分发机制，与OpenClaw项目本身的开源、去中心化理念形成了鲜明对比，构成了一个典型的去中心化的悖论。平台本身成为新的、事实上的把关人，只不过其把关逻辑被编码在了一套不透明的推荐算法之中。

2. 学科边界的突围：从“人类中心主义”到人机共生

OpenClaw作为行动者智能体的崛起，正以前所未有的力度冲击着传播学的传统研究范式，迫使学科进行深刻的自我反思与革新。这种冲击不仅体现在研究对象的扩展上，更体现在理论框架与研究方法的根本性变革上。

（1）研究对象的扩容：将非人类行动者纳入核心视野。传播学长期以来的研究核心是人以及围绕人展开的传播行为与社会影响。然而，在OpenClaw所代表的智能传播时代，将AI仅视为人类使用的工具或媒介已远远不够。遵循行动者网络理论的启示，我们必须承认OpenClaw这类AI智能体是具备能动性的非人类行动者。它们不再是被动地传递信息，而是能够主动地感知环境、做出决策、调用资源、执行行动，并与其他人类及非人类行动者共同编织社会网络。

因此，传播学的研究议程必须进行根本性的扩容。我们研究的问题不能再仅仅是“人如何使用AI影响传播效果”，而必须转向更深层次的、关系性的探问：“AI作为行动者如何改变人的决策路径？”例如，一个集成了编采评写与发布技能的OpenClaw，如何通过自主分析数据、生成报告，甚至直接执行指令，来重塑传媒工作者的工作流程与决策模式？还有，人机互动如何生成新的社会规范与文化实践？当“数字员工”的应用成为常态，人与AI之间的沟通、协作、信任乃至冲突，将如何催生新的组织文化与社会交往范式？再者，非人类行动者之间的交互如何涌现出不可预期的社会后果？两个或多个OpenClaw智能体，在没有人类直接干预的情况下，通过API相互调用、协同或竞争，会产生怎样的系统性风险或创新？诸如此类，不一而足。

概言之，将非人类行动者从研究的背景或变量提升到与人类平等的行动者地位，是传播学应对智能时代挑战的第一步，也是最关键的一步。

（2）方法论的革新：从静态效果到动态过程的追踪。研究对象的转变，必然倒逼研究方法的革新。传统的问卷调查、深度访谈或内容分析等方法，虽然能帮我们理解人的态度和行为，但面对AI智能体那种动态、实时且往往深藏不露的“黑箱”行动，这些老办法就显得捉襟见肘了。

OpenClaw的复杂之处在于其行为的涌现性。按照AIGS成熟度模型，AI正从被动提供信息的“建言者”，一步步演变为主动出谋划策的“建议者”、被授权干活的“执行者”，甚至可能成为与人类深度绑定的“共谋者”。随着AI自主性和社会嵌入度的不断提升，传播学研究必须采取新的方法。

一是计算传播学方法。利用日志分析、API调用追踪等大数据手段，将OpenClaw在特定任务中的行动链条完整还原出来：它接收了什么指令？调用了哪些技能？访问了什么数据？执行了哪些操作？最终结果如何？这让研究焦点从关注传播的最终效果，转向了对生成过程的精细解剖。

二是复杂网络分析。把社会系统看作一个由人类和OpenClaw智能体共同组成的混合网络。通过分析网络结构、节点中心性以及信息或行动的流动路径，我们可以揭示AI在社会网络中的权力地位、影响力，以及其介入是如何改变整体网络拓扑的。

三是数字民族志。研究者需要“潜入”人机交互的现场，长期观察用户如何与OpenClaw协同工作、调试其行为、应对其错误。这种质性研究方法能弥补纯粹数据分析的短板，揭示人机关系中那些微妙的文化、情感与权力动态。

四是算法审计。借鉴计算机科学和信息系统领域的方法，对ClawHub等平台的推荐算法、技能市场的审核机制进行系统性测试和评估，以揭示其中可能存在的偏见、歧视或安全漏洞，推动算法的透明化以及问责制的落地。

总之，当AI成为行动者，传播学必须走出以人为研究对象的舒适区，跳出以人为中心的旧范式，转向一种更具过程性、关系性、技术性与批判性的研究路径，将非人类行动者纳入核心视野，用计算的逻辑解剖智能体的“黑箱”行为。这既是挑战，也是学科发展的重大机遇。

3. 社会结构的液态化：万物皆媒与关系重铸

在OpenClaw的催化下，社会加速进入“液态现代性”状态[12]——物理世界的坚固边界被API溶解，人与人的关系被编码在冷冰冰的算法协议之中。媒介化理论认为，媒介不再仅仅是社会的反映或工具，而是深度嵌入并重塑社会制度、文化实践与人际交往的根本性力量。如果说互联网和社交媒体开启了社会的深度媒介化进程，那么OpenClaw的出现则像一个强大的催化剂，极大地加速了这一进程，使社会朝着齐格蒙特·鲍曼所描述的“液态现代性”状态狂奔。在液态现代社会中，一切固定的结构、关系和身份都变得不稳定、易变和短暂。OpenClaw正是通过其无处不在的连接能力和对物理世界的行动能力，将这种“液态”特性注入社会的方方面面。

（1）无处不在的连接：从“万物互联”到“万物皆媒”。OpenClaw的架构设计使其能够轻易地连接并操作任何具有API接口的设备、软件或服务。这使得“万物互联”的概念在实践层面被极大地深化和拓展，最终导向“万物皆媒”的现实。

从家庭场景中的智能音箱、扫地机器人，到城市交通系统中的自动驾驶汽车，再到工业生产线上的自动化流程控制，物理世界本身正在被深度媒介化。借助OpenClaw，我们可以通过自然语言与物理环境进行沟通和互动，命令它开灯、启动汽车、调整产线参数等。物理对象不再是沉默的客体，而是成为可以接收指令、执行动作、反馈状态的传播媒介和行动节点。这种物理世界的全面API化和媒介化，使得社会生活的底层逻辑被改写，现实世界与数字世界之间的界限变得前所未有的模糊。

（2）社会关系的重铸：编码在算法与API之中的互动逻辑。OpenClaw对社会媒介化的加速，更深刻地体现在对社会关系的重铸上。当“数字员工”日益普及，传统的劳资关系、同事关系乃至人际互动模式都发生了根本性的变化。网经社在其发布的报告中详细地描绘了“数字零售”与“数字生活”的图景，这正是社会关系被深度媒介化的具体投射。

一是劳资关系的算法化。随着OpenClaw框架的引入，“数字员工”不再只是辅助工具，而是真正进入组织架构中。任务分配、绩效考核，甚至是跨部门的协作，几乎都由预设的算法和工作流引擎说了算。这对人类员工提出了新要求：要学会像程序员一样思考，用精确的自然语言指令去和这些非人类同事沟通；管理者也不再单纯靠“管人”来带团队，而是转向配置和审计AI智能体的行为。在这个过程中，传统劳动关系中那些微妙的情感连接、默契配合和弹性空间，正逐渐被精确、高效但冷冰冰的API调用取代。

二是消费关系的自动化。在零售端，OpenClaw正在把消费者变成“指挥官”。其不仅能充当人们的私人采购助理，自动完成比价、下单和售后处理，更深层的改变在于，消费者与商家的关系正演变为两个（或多个）AI智能体之间的API交互。这意味着，传统的品牌忠诚度和购物体验正在失效。因为最终的决策逻辑，不再取决于广告打得好不好，而是看哪个平台的API更开放，或者哪个AI的“砍价”技能更硬核。

三是人际交往的媒介化。除上述关系外，在日常生活中，OpenClaw也可以通过管理日程、代写邮件、筛选信息等方式，深度介入我们的人际交往。我们与他人的互动，越来越多地被一个以效率为导向的AI“管家”过滤和塑造。

最终，社会关系本身被编码在了算法与API的协议之中。信任、合作、冲突等社会互动发生和解决的方式，也越来越多地遵循机器的逻辑而非人类社会长期演化形成的文化规范。这正是社会深度媒介化的核心特征，也是其潜在风险的根源所在。

三、治理破局：破解“权限悖论”与构建媒介化治理新范式

OpenClaw的爆发式增长，在释放巨大生产力的同时，也伴随着一个深刻且棘手的治理难题，我们称之为“权限悖论”。这一悖论的核心在于：为获得极致的执行效率、跨越认知与行动之间的“执行鸿沟”，用户必须赋予AI极高的系统权限。然而，赋予AI高权限又不可避免地带来了失控的风险，如数据泄露、财产损失，乃至算法作恶。[13]这一悖论贯穿于OpenClaw从个人应用到

企业部署的各个层面，对现有的法律、伦理和技术治理框架构成了严峻挑战。在效率与风险的“钢丝绳”上，唯有将伦理嵌入代码、将透明写入算法、将多元力量织入网络，才能驯服这头名为智能体的“巨兽”，实现人机共生的善治。

1．治理困境的根源：“权限悖论”与责任“黑洞”

赋予AI越大的权力，就越能释放生产力；但权力越不受限，就越容易坠入责任归属的法律“黑洞”，这是智能时代最棘手的二律背反。“权限悖论”并非一个抽象的理论概念，它在OpenClaw的实际应用中有着清晰而具体的表现。

（1）效率与风险的致命博弈。OpenClaw在企业端的巨大吸引力，在于它能精准地解决流程自动化中“最后一公里”的问题。ITBear科技资讯发布的相关报道指出，企业智能化转型的痛点之一就是如何将AI的认知能力转化为实际的业务流程执行。OpenClaw通过直接访问和操作企业的核心系统，如ERP、CRM、财务软件甚至底层数据库，来填补这一“执行鸿沟”。然而，这种深度的系统集成，也意味着企业向一个自动化且可能存在幻觉的AI智能体敞开了核心资产，而这无异于打开了“潘多拉魔盒”。

大量的安全研究报告已经证实了这种风险。在金融领域，一个被赋予了交易权限的OpenClaw智能体，如果遭遇提示词注入攻击，可能导致未经授权的交易、客户隐私泄露或账户密钥被窃取。在医疗领域，能够访问电子病历（EHR）系统的OpenClaw，其高权限操作一旦失误，极易导致关键医疗数据被误删或泄露，触及不可逾越的安全红线，这也是部分医院对其发布“封杀令”的直接原因。更有甚者，已披露的高危漏洞（如CVE-2026-25253）①表明，攻击者可能利用框架本身的缺陷，实现对宿主系统的完全控制。[14-15]OpenClaw“安装即完全信任”的模式和超高的默认权限使其在追求极致效率的同时，导致安全风险呈指数级增大。

（2）责任归属的“黑洞”。当OpenClaw作为一个自主的社会行动者，在ClawHub等技能市场上进行日均千万级的技能调用和交易时，一旦出现事故，责任归属便陷入了一个前所未有的法律与伦理“黑洞”。我们可以设想这样一个场景：用户A通过OpenClaw使用了一个由开发者B开发的、发布在ClawHub平台C上的财务分析技能。该技能在分析过程中，调用了由开发者D提供的另一个数据查询技能，最终因为数据错误或逻辑缺陷，给用户A造成了巨大的投资损失。那么，责任应该由谁来承担？是发出模糊指令的用户A吗？但他可能完全不了解技能内部的复杂运作；是编写核心逻辑的开发者B吗？他可能会辩称自己的代码在多数情况下是正常的；是提供了分发渠道的平台C吗？它可能会以“平台免责”为由，声称自己只是一个中立的市场；是提供了底层数据的开发者D吗？他可能对数据的最终用途一无所知；那么，是OpenClaw框架本身，还是其背后的大语言模型？

这种责任链条的断裂和分散，使得传统的侵权法和合同法体系难以直接适用。责任的模糊性不仅让受害者维权困难，也阻碍了安全标准的建立以及保险机制的形成。在开源生态中，责任主体本身就是模糊的，这使得对OpenClaw的治理成为一个极其复杂的系统性工程。

2. 治理范式的转型：迈向媒介化治理新生态

面对“权限悖论”及其带来的严峻挑战，传统的治理思路——无论是侧重于事前审查的内容监管，还是侧重于事后惩罚的法律框架，都显得力不从心。笔者认为，必须建立一种适应智能传播时代的新型治理模式，即“媒介化治理”。[16]

媒介化治理的核心逻辑其实很直白：既然社会已经深度媒介化，AI智能体也成了网络中的关键角色，那我们就不能只盯着孤立的内容或行为去管，而得去治理那个构成我们社会现实的、复杂的“人—机—环境”行动网络。方法上也不能光靠法律条文或行政命令，得建构一套集技术、法律、伦理和社区自治于一体的动态框架。具体来说，这个框架基于四个支柱。

（1）技术基石：基于风险分级的动态权限管理。这是这一框架最基础的层面。治理的关键不是非黑即白地一封了之或完全放任，而是要建立一套精细化、场景化、能动态调整的权限体系。毕竟风险是流动的，权限的授予也得是临时的、随时能撤回的，且必须与任务的风险等级严格匹配。

这就涉及对OpenClaw的操作（或者说技能）进行严格的风险分级。首先是轻量级权限。如对于信息查询、内容生成、格式转换这些不触碰敏感数据、也不搞高风险动作的技能，可以建立备案制或靠社区信誉评级进行

① CVE-2026-25253是开源AI智能体框架OpenClaw（曾用名Clawdbot、Moltbot）中被披露的一个高危远程代码执行（RCE）漏洞。该漏洞也被称为“ClawJacked”。它允许攻击者通过构造恶意链接，在用户无感知的情况下远程接管AI智能体，进而获得宿主计算机的完全控制权。

管理。这样既降低了门槛，也能鼓励生态继续创新。其次是重量级权限。凡是涉及资金操作、调用个人隐私数据、修改系统文件这类高风险技能，特别是在金融、医疗这些强监管行业，必须构建更严格的控制机制。

一是即时授权与人工确认。对于删除文件、转账等高危操作，系统必须强制触发一个显式的人工确认流程，例如通过UI弹窗或手机验证，等待用户或管理员审批后方可执行，并启用沙盒运行机制（将高风险技能的执行环境与宿主操作系统进行严格隔离）。无论是通过Docker容器，还是NVIDIA推出的NemoClaw等专业框架提供的隔离沙箱，其目的都是确保即使技能被恶意利用，其破坏范围也能被限制在可控的虚拟环境内，无法触及核心系统资源。

二是动态升降级与异常检测。权限绝不能是一潭死水。治理框架中需装一个自动化决策引擎，监测AI的行为模式和实时风险，随时调整它的权限级别。比如，系统先摸清AI的行为基线，一旦发现不对劲，像短时间内频繁读取敏感文件，或者试图调用未授权的API，就立刻自动降级权限，同时给管理员发警报，直到人工介入审查。要做到这些，需彻底打通规则库、风险评估模型和审批工作流，将其联动起来。

（2）信任构建：打破“黑箱”与算法透明性建设。想要有效治理和定责，打破算法“黑箱”是绕不开的前提。没有透明就没有信任，没法解释就没法问责；强制AI记录行动轨迹日志，就是还原真相、把控制权掌握在人类手里的最后一道防线。

一是推广可解释性工具与标准。虽然OpenClaw本身并不是模型，但作为LLM的调度和执行框架，它的决策过程也必须是透明的。治理应该鼓励甚至要求开发者用LIME、SHAP这些开源工具，把特定决策的依据置于台前。

二是建立行动轨迹审计日志。OpenClaw框架必须内置强大且不可篡改的审计日志功能，将每次任务的完整推理轨迹记得清清楚楚：从原始指令输入，到任务拆解，再到每一步技能调用、API的请求与返回，以及最终结果，进行全流程记录。这种精细日志不仅是事后追责的证据，也是理解和改进AI行为的基础。

三是推行模型卡片与影响评估报告。对于ClawHub上架的特别是面向企业和公共部门的关键技能，平台应要求开发者提交类似模型卡片或AI事实清单的文档，清晰说明其功能、预期用途、数据来源、性能局限以及已知的偏见和风险。对于可能产生重大社会影响的AI系统，还应强制进行伦理影响评估。

（3）制度保障：政府、平台、社会的多元协同共治。面对OpenClaw这样一个复杂、全球化且快速演变的开源生态，单打独斗肯定行不通。面对开源世界的无界流动，单一的政府监管力有不逮，唯有构建“政府划红线、平台管入口、社会评质量”的立体网络，方能织密治理之网。

一是政府监管。政府负责定底线、搭框架。这包括：立法与标准制定，完善数据安全法、个人信息保护法等相关法律，并针对AI智能体的特殊性，出台专门的行业标准和安全规范。参考国家知识产权局对智能体生成专利申请文件的风险提示，为AI在关键领域的应用划定明确的“红线”。

二是风险预警与信息共享。国家工业信息安全发展研究中心、国家网络安全通报中心等机构应持续对OpenClaw等主流AI框架进行风险监测，并及时向社会发布风险预警。建立国家级漏洞共享平台（如国家信息安全漏洞共享平台），鼓励安全研究者提交漏洞报告，形成全社会的安全合力。

三是平台自治。ClawHub等技能市场作为生态的“守门人”，必须承担起更积极的治理责任。其中包括建立严格的审核机制：平台需建立官方认证与社区信用相结合的双层审核体系，对上架技能进行强制性的安全扫描、代码审计和功能验证，并对高风险技能进行重点监管。

四是完善社区反馈与处置流程。建立清晰的用户评分、评论和恶意技能举报机制，并承诺对安全事件做出快速的应急响应，包括下架恶意技能、通报受影响用户、发布安全补丁等。

五是社会监督。引入独立的第三方力量，对AI产品和平台进行持续的伦理与安全评估。鼓励和培育专业的第三方AI安全审计机构，对OpenClaw框架本身及其生态中的关键技能进行独立的、深入的安全测评和代码审计。审计报告应向社会公开，作为用户和企业选择服务的重要参考。

六是智库与媒体的角色。行业智库和科技媒体，应持续关注OpenClaw的发展，发布深度研究报告，揭示其潜在风险，引导公众理性认知，形成对平台和开发者的舆论监督压力。

（4）价值内核：伦理嵌入设计。最好的治理不是“外部的鞭子”，而是“内心的良知”：将伦理原则编译成机器可读的代码，是让AI拥有一颗“向善之心”的

终极密码。所谓治理的最高境界，是让向善的价值观内化为技术本身的属性。治理不应仅仅是外部的约束，更应前置到技术的设计与开发阶段，即实现伦理嵌入设计。

一是将伦理原则转化为代码规范。AI伦理原则（如公平性、隐私保护、透明性）不应停留在抽象的口号层面，而必须被具体地转化为开发过程中的强制性设计规范或代码检查清单。

二是隐私保护设计。在OpenClaw的技能开发SDK（软件开发工具包）中，可以强制要求对所有处理个人数据的操作进行显式声明，并默认开启数据脱敏、加密和最小化收集等功能。

三是公平性测试。借鉴AI公平性测试工具（如AIF360、Fairlearn），将针对不同人群的性能一致性测试，作为技能发布前的必要环节。

四是“策略即代码”。[17]这是实现伦理嵌入设计的强大技术路径。通过使用OPA（Open Policy Agent，开放策略代理）等工具，可以将复杂的伦理和合规策略（如“禁止技能在未经用户同意的情况下访问通讯录”“确保算法对不同性别的推荐结果无显著差异”）编写成机器可读、可自动执行的代码。这些策略代码可以被集成到OpenClaw技能开发的CI/CD（持续集成/持续部署）流水线中。每当开发者提交新的代码，流水线就会自动运行这些策略检查，一旦发现违反伦理规范的代码，就会自动阻止其部署。通过这种方式，伦理审查从一种滞后的人工活动，转变为一种前置的、自动化的、不可绕过的技术约束。这正是将抽象伦理原则转化为具体技术实践的典范。

总之，唯有将治理的触角前置于代码设计阶段，用伦理嵌入替代事后追责，才能在效率与安全的“悬崖”边，为人类文明筑起一道数字“护栏”。

结语

OpenClaw不仅是一个软件，还是智能传播时代的“普罗米修斯之火”。我们既要享受它带来的光明与效率，也必须掌握驾驭它的媒介化治理智慧：在人机共生的新纪元，唯有以伦理为锚、以技术为帆，方能驶向善治的彼岸。由是，OpenClaw现象是智能传播时代一个决定性的分水岭。它的出现雄辩地证明，人工智能已不再仅仅是人类延伸自身能力的被动工具，而是作为具备自主性、行动力和社会嵌入性的行动者智能体，开始与人类共同塑造我们的社会现实。这一深刻的范式革命，彻底改写了传播生态的底层规则，对传播学的理论与方法提出了颠覆性的挑战，并以前所未有的速度推动社会的全面媒介化进程。

随之而来的“权限悖论”及其引发的效率与风险的紧张关系、责任归属的模糊不清，构成了我们这个时代最严峻的治理难题之一。面对这一挑战，我们认为，必须超越传统的内容监管思维，构建一种全新的媒介化治理范式。这一新范式承认AI的行动者地位，致力于治理复杂的人机行动网络。它以动态权限管理为技术基石，以可解释性与透明度为信任前提，以多元协同共治为制度保障，并以伦理嵌入设计为价值内核。

未来的传播治理，必将是一场从单一的内容监管到复杂的行动网络治理的深刻转型。我们既要拥抱技术所带来的前所未有的效率与可能性，更要以审慎、前瞻和系统性的思维，为其筑牢安全、伦理与责任的“堤坝”。唯有在实践中不断探索和完善媒介化治理的新范式，我们才能在奔涌而来的智能时代浪潮中，真正实现技术进步与社会福祉的共生共荣。

参考文献：

[1] OpenClaw热潮下的企业智能化转型：AI Agent如何重塑未来竞争力？[EB/OL]. [2026-04-01]. https://it.sohu.com/a/1003685837_362225.

[2] OpenClaw：狂欢背后，警钟已响[EB/OL]. [2026-03-09]. https://www.ccidnet.com/news/1097104.jhtml.

[3] 陈博观察. AI“养龙虾”：OpenClaw发展研究报告（2026）[EB/OL].[2026-03-10]. https://www.100ec.cn/detail--6657360.html.

[4] Prime Bpm. How AI-Powered BPM Closes the Execution Gap[EB/OL]. [2026-04-01]. https://www.primebpm.com/blog/ai-powered-bpm-closes-the-execution-gap.

[5] Couldry N, Hepp A. Conceptualizing Mediatization: Contexts, Traditions, Arguments[J]. Communication Theory, 2013, 23(3): 191-202.

[6] 吴莹，卢雨霞，陈家建，等. 跟随行动者重组社会——读拉图尔的《重组社会：行动者网络理论》[J]. 社会学研究，2008（2）：218-234.

[7] 埃弗雷特·罗杰斯. 创新的扩散[M]. 辛欣，译. 北京：中央编译出版社，2002：5-12.

[8] 马尔科姆·格拉德威尔. 引爆点：如何制造流行[M]. 钱清，覃爱冬，译. 北京：中信出版社，2009：15-32，78-95.

[9] 第四范式发布“式说”大模型 以生成式AI重构企业软件（AIGS）[EB/OL].[2023-04-27]. https://tech.cnr.cn/techph/20230427/t20230427_526233819.shtml.

[10] 陆三金. ClawHub 乱象：一万个Skill，一半是垃圾，一半是毒药[EB/OL].[2026-02-27]. https://www.huxiu.com/

article/4837573.html.
[11] ClawHub 技能市场20%藏毒！你的“龙虾”可能在偷数据[EB/OL].[2026-04-02]. https://www.hzyunye.com/.
[12] 齐格蒙特·鲍曼. 流动的现代性[M]. 欧阳景根，译. 上海：上海三联书店，2002：1-15.
[13] 从协作到执行：桌面Agent引发的隐私与效率博弈[EB/OL]. [2026-03-06]. https://baijiahao.baidu.com/s?id=1858904027095354939&wfr=spider&for=pc.
[14] 北京大学计算中心. 关于OpenClaw安全风险的通知[EB/OL]. [2026-04-02]. https://its.pku.edu.cn/.
[15] 天融信：2026年OpenClaw运行机制与安全威胁研究报告（附下载）[EB/OL].[2026-03-13].https://www.topsec.com.cn/newsx/6454.
[16] 侯迎忠，玉昌林. 媒介共治：媒介化治理的学理脉络、本土演进与研究空间[J]. 现代传播（中国传媒大学学报），2024，46（1）：84-91.
[17] 亚马逊云科技. 什么是策略即代码（=Policy as Code）？[EB/OL].[2026-04-02]. https://aws.amazon.com/cn/what-is/opa/.

From "Cognitive Mediation" to "Action Node": OpenClaw and the Paradigm Reconstruction of Intelligent Communication—A Study on AIGS Governance Based on Actor-Network Theory

YU Guo-ming（School of Journalism and Communication, Beijing Normal University, Beijing 100875,China）

Abstract: Based on the mediatized society theory and actor-network theory, this article proposes the core concept of "actor-intelligent agent" for the OpenClaw phenomenon, aiming to analyze the reconstruction of communication paradigms in the transition process of AI from a cognitive revolution to an action revolution. The study finds that, on the one hand, OpenClaw, through a "cognition-execution" loop, bridges the execution gap in human-computer interaction, moving from the symbolic world dominated by generative AI into the physical world of human existence. This will inevitably reshape the connection logic of "social-technical" networks. On the other hand, such high-level intervention also triggers a profound "authority paradox", i.e., the inevitable conflict between efficiency optimization and system risk. In response to this governance challenge, this article elaborates on the mediatized governance framework, advocating a paradigm shift from traditional content regulation to complex action network governance through dynamic authority management, algorithmic transparency development, multi-stakeholder collaborative governance, and the ethical embedding in design. This study provides new theoretical perspectives and practical pathways for understanding communication ecosystem transformation and governance innovation in the era of AIGS (AI-Generated Services).
Keywords: actor-intelligent agent; AIGS maturity; "authority paradox"; mediatized governance; actor-network theory

（责任编辑：侯苗苗）