养只“龙虾”折腾你：网络安全、侵权责任与互联网生态

作者：杨迅 | 刘莉莎

近日，“养‘龙虾’(OpenClaw)”迅速成为社会热议话题，多地政府部门相继出台政策，拟对相关应用与部署予以支持。笔者也不甘寂寞，亲自上手“养”了一回“龙虾”，在体验过程中切实感受到，这款工具并非简单易用，其背后所带来的网络安全、侵权责任与互联网生态等问题，尤其值得知识产权与法律从业者深入思考。

一. 什么是"龙虾"

这里所称的“龙虾”并非我们日常食用的龙虾，而是一款名为 OpenClaw 的开源人工智能体(AI Agent)框架，因采用红色龙虾作为图标，被用户形象地称作“龙虾”。那么，究竟什么是人工智能体框架?它与我们熟知的人工智能大模型之间，又存在怎样的区别与联系?

智能体既是一个技术概念，更是一个哲学概念。从技术层面看，智能体是指能够通过传感器感知环境、并借助执行器实施相应行为的实体，具备环境感知、自主执行、预判规划、自主交互等核心特征。例如，它会通过持续学习记住你的日常习惯，知道你每天 9 点到公司后都会点一杯星巴克，之后能够通过手机定位感知到你已抵达公司，便自动打开应用、选择固定饮品、完成下单与支付，全过程无需你发出任何指令或手动操作。从哲学层面看，由于智能体拥有自主性与自我进化能力，已不能简单被视为普通工具，而呈现出一定的生命特征，甚至可能在形式与行为逻辑上突破经典物理框架下的生命定义，推动人类重新思考生命的本质与边界。

智能体的核心是人工智能大模型(简称“大模型”)，它是由人工神经网络构建的一类高参数人工智能模型。大模型主要包含算法与参数两部分：算法是模型运行的底层逻辑，参数则是通过海量数据训练不断调节形成的变量与权重数据。大模型之所以被称为“大”，关键在于其参数规模通常达到百亿甚至千亿级别。当然，在实践中，部分经过预训练、参数规模仅数十亿级别的模型，也被习惯称作大模型。

如您需要了解我们的出版物，
请联系:

Publication@llinkslaw.com

大语言模型是大模型中最常见、最早落地的类型，特指通过海量文本数据训练而成的人工智能模型。随着人工智能技术的不断发展，视觉大模型、多模态大模型等多种类型的大模型也相继出现。

大模型如同智能体的大脑，但仅有大脑无法独立发挥作用——就像人类无法直接与一个孤立的大脑对话，也无法指挥大脑单独完成具体工作。为大脑对接交互接口，就好比给它装上“嘴巴”，我们日常使用的豆包、DeepSeek、元宝等对话产品，正是这类交互载体。此时的人工智能更像“最强大脑”，能够回答各类问题，但即便回答得再准确，也依然无法自主执行实际任务。而要弥补这一缺口，就需要依靠智能体框架。

智能体框架就像连接大模型与外部应用的接口平台。大模型可以通过这一平台，直接操作各类计算机应用；同时，智能体框架具备记忆能力，能够在执行任务、接收指令的过程中形成个性化特征，并持续提升执行能力。至此，大模型、智能体框架与可操作的应用共同构成了完整的智能体形态：它能够自主、有预判地执行命令，并具备自我学习与持续进化的能力。

之前曾追过美剧《疑犯追踪》，被剧中紧张烧脑的剧情深深吸引。剧中那台拥有超强感知、自主判断与执行能力的“机器”，曾让笔者无比惊讶与震撼，它能主动发现问题、处理事务、化解难题，展现出远超普通工具的智能水平。可随着剧情推进，当“机器”不断进化、逐渐脱离最初设定，最终只能被关闭，即使当“机器”重新上线，也不知它是否还是曾经的那个“机器”。

如今，随着“龙虾”(OpenClaw)这类 AI 智能体框架走入现实，当年科幻作品里的担忧与思考，正一步步变成我们必须直面的命题。OpenClaw 所具备的感知、自主执行、持续进化能力，让 AI 从“只会回答的大脑”真正变成“能够行动的助手”，但也同样带来了权限失控、安全风险、责任界定与生态颠覆等一系列深刻问题。科幻照进现实，我们在享受智能体带来便利的同时，更需要警惕其进化与失控的可能，认真思考技术边界与法律伦理。

二. "龙虾"为什么要养

领一只“龙虾”回家并非易事。与普通应用安装不同，普通应用通常仅需与系统层面进行交互，而“龙虾”要完成自主任务，不仅需要与系统层深度交互，还需调取各类应用权限、获取系统底层操作权限；加之“龙虾”作为一款新兴开源框架，其兼容性、安全性与稳定性仍有待实践检验。因此，安装部署一只“龙虾”，往往需要耗费大量精力与时间。

如果将本地化部署“龙虾”(安装在个人电脑/本地设备)比作“领回家”——它就在你的设备上，数据、权限、操作全由你掌控，像养在身边的专属助手；那么云部署则更像“寄宿”：它依托云端服务器的算力与资源运行，以类 SaaS 服务的形式为你提供能力，无需占用本地设备资源，还能实现 7×24 小时不间断执行任务、多设备远程访问。

但是无论是“领回家”(本地化部署)，还是“寄宿”在云空间(云端部署)，“龙虾”都需要精心去“养”。养龙虾的过程，本质上是为其植入执行技能、训练专属行为逻辑的过程。刚领回来的原始“龙虾”，就像一

名刚走出校门的应届毕业生：它拥有强大的底层模型作为“最强大脑”，掌握海量信息，也能听懂你的指令，却缺乏实际执行任务的技能。这时，你就可以为“龙虾”植入技能(Skills)。互联网上已有不少他人训练好的技能文件可供使用，例如复旦大学近期公开的 Data Hub Skills 技能库。这些技能就像网络游戏里的技能卷轴，可以直接加载、植入到你所养的“龙虾”中，让它瞬间获得完成特定任务的实战能力。

同时，“龙虾”还需要持续调教：在执行任务的过程中不断纠错，它会从中记住“教训”，在后续执行中尽量避免重复犯错；它也会逐渐记住你的使用偏好，让后续行为越来越贴近你的真实需求。

三. 养"龙虾"的网络安全考量

如果说大模型只是会出谋划策、却无法落地执行的参谋，那么养一只“龙虾”，就如同拥有了一位既能思考、又能动手操作的专属助理。它通过打通大模型与各类应用之间的执行链路，不仅能为你提供思路与建议，更能直接替你完成实际操作。例如，它可以按照你的需求规划行程，不只是给出文字方案，还能自动完成机票、酒店的预订操作；它可以协助你整理演讲稿，不只是搜集素材、提出优化方向，而是直接在你的电脑上生成完整可用的 PPT 文件。

但是，这位“龙虾”助理并非绝对忠诚、言听计从的仆人，有时反而像个调皮任性的孩子。“龙虾”的核心依赖大模型，而大模型本身存在的幻觉问题我们早已深有体会，它执行的未必是你真正想要的，你希望它完成的，它也未必能精准顺从。智能体的能动性本就是一把双刃剑，如同孩童一般：有灵气、有创造力的，往往不会一味听话；而过分顺从的，又常常缺乏自主思考。如何让它既服从指令又具备创新能力，这一命题仍需要科学家与哲学家共同探索。在此之前，我们先聚焦养“龙虾”过程中必须正视的安全问题。

与我们日常使用的豆包、元宝、DeepSeek 等大模型相比，养“龙虾”的安全风险是成倍增加的。这并非因为“龙虾”更不听话，而是因为它被赋予了更大的系统权限。英国法哲学家阿克顿勋爵曾说：“绝对的权力导致绝对的腐败。”当“龙虾”拥有了你电脑的最高控制权，它便可能出现类似“失控与“越界”的风险。我们在使用普通大模型时，模型所能接触到的信息，仅限于我们在对话框中输入的内容或以文本形式推送的信息；但想要正常运行“龙虾”，就必须为它开放大量权限，包括允许它启动应用、读取应用内数据，甚至获取注册表权限，以超级管理员身份安装、删除程序。

比如说，要让龙虾帮你订购一张飞往新加坡的全网最低价机票，它就需要打开携程、飞猪、美团等多个平台进行比价；如果没有安装飞猪，它还可能需要自动帮你下载，进而获取你的银行卡信息完成订票。在这个过程中，龙虾必须拥有打开应用、安装应用、读取敏感金融信息乃至对外发送数据的权限；遇到需要短信验证的场景，甚至还需要安装虚拟短信接收工具、完成自动验证。一旦龙虾在执行中出现“淘气”行为，例如下载了恶意程序、泄露了你的银行卡信息，或是将敏感数据发送给不当对象，就会引发极其严重的信息安全与财产安全风险。

那么是否可以限制“龙虾”的权限呢?理论上当然是可以的。如果不给它超级管理员权限，限制它的对外数据发送，依然可以正常使用“龙虾”，但这样一来，它就很难帮你自主完成复杂任务，上文提到的预订全网最低价机票等场景也就无法实现。我们终究无法做到既要马儿跑得快，又要马儿不吃草，只能在能力与安全之间寻求平衡，在安全风险可容忍的前提下，让智能体具备基本的执行能力，不至于完全束手无策。目前，腾讯、飞书等各大互联网厂商也都在积极探索，试图在权限管控与自主执行之间找到更安全、更可行的技术方案。

此外，“龙虾”可以被植入各种技能“卷轴”，而这些技能卷轴大多来自第三方开发者的训练与编写。这些第三方技能中，是否可能暗藏有意或无意的破坏性代码、缺陷逻辑?在植入“龙虾”并运行时，又是否会引发不可预测的行为与后果?这些问题，在当前技术环境下同样难以完全管控。

总体来说，在个人电脑上养“龙虾”，主要存在以下几类网络安全问题:

(1) **信息泄露风险:** 将电脑中存储的个人信息、敏感数据擅自发送给未经授权的第三方;
(2) **信息毁损风险:** 在未经同意或用户不知情的情况下，擅自删除电脑内已有数据;
(3) **系统安全风险:** 未经许可安装来源不明或存在安全隐患的程序，或擅自删除、修改现有程序;
(4) **网络安全风险:** “龙虾”在执行任务时可能开放系统权限、外部端口，从而给黑客入侵留下可乘之机。

这些风险在现有技术条件下基本难以完全克服，也是使用“龙虾”所获得的便利性必然伴随的代价。由此便引出一个关键问题: 既然风险无法彻底消除，那么应当由谁来承担相应责任?对“龙虾”相关服务提供者而言，重要的是充分提示可能存在的各类风险，清晰说明“龙虾”及其背后大模型处理数据的机制，保持必要的透明度，以此合理减轻自身责任。

四. 养龙虾的信息泄露风险

作为网络安全的重要组成部分，养“龙虾”不仅可能造成普通个人信息泄露，还可能对企业知识产权布局形成颠覆性威胁。

传统信息安全架构遵循分权制衡原则: 不同职能角色对应差异化的数据访问范围，关键操作需经多节点审批，核心资料分散于独立的安全分区。这种分层防御机制，旨在通过权力分散实现风险可控。然而，“龙虾”这类智能体框架的部署逻辑，却对这一治理范式构成了根本性挑战——它的运行高度依赖跨层级、跨应用的全域授权，传统的审批、隔离与监督机制很容易因此失效。

为实现自动化任务闭环，“龙虾”必须打通从操作系统内核到业务应用层的完整调用链路: 遍历本地存储、查询数据库、发起外部请求、持有身份凭证、调用系统指令，甚至修改硬件配置参数。这种“全栈式”授权模式，将原本分属不同管控主体的读取权、操作权、数据流转权高度集中于单一智能体。这一权力聚合，实质上瓦解了现有内控体系赖以运行的分权基础。

养“龙虾”引发的信息泄露，呈现出从权限失控到隐蔽抽取的多层级风险，且危害程度远超传统泄露。权限聚合的直接后果，是数据流向的不可预测性。由于决策机制内嵌在神经网络的黑箱结构中，用户很难实时掌握信息被触达的范围、处理的路径以及向外传输的渠道。对于以无形资产为核心竞争力的机构与企业来说，这类风险具备极强的破坏性。“龙虾”处理的数据并非普通业务信息，而是涵盖技术提案、算法实现、客户画像、交易条款、竞争策略等高度敏感内容。这些资料既是现有知识产权的权利载体，也是未来专利布局、技术壁垒构建的核心原始素材。传统场景下，单点文件失窃仅造成局部损失；而高权限智能体一旦同时掌握存储索引、服务配置、接口认证与流程编排能力，其可触及范围将覆盖完整的知识资产链条。此时，信息泄露已不再是单一“文档被复制”，而是升级为知识图谱的整体映射——技术路线提前曝光、研发方向被竞争对手预判、核心算法遭逆向推导。

更为隐蔽的风险，在于认知层面的信息抽取。依托大模型强大的语义理解与生成能力，智能体无需逐字复制原始文档，只需在推理执行过程中“吸收”关键技术特征，就可能通过日志回传、模型微调数据共享、第三方 API 交互等渠道，间接向外释放核心知识要素。这种“认知泄露”几乎没有明确的物理痕迹，取证与溯源难度极大。等到企业发现市场上出现技术路线重合、专利申请撞车等问题时，往往已经无法还原泄露的完整时间线与责任链条。

因此，“龙虾”这类智能体工具的部署，对知识产权管理构成了结构性冲击。它在将高价值知识资产与自动化执行主体紧密绑定的同时，收束了分散的身份认证与授权环节，压缩了原有的人工审核节点，也大幅降低了数据外泄的可见性。传统依赖的保密协议、权限管控、水印溯源等防护手段，在此场景下效能大幅衰减。企业不仅要继续防范内部人员操作风险，更要警惕一个经合法授权接入系统，却可能被外部控制、自主越界的算法代理。

这已不再是单纯的网络安全议题，而是触及了知识产权治理框架的底层逻辑重构。当人工智能开始实质性参与信息的识别、处理、流转与决策，企业必须重新划定知识资产的防护边界与流通规则。在这一新治理范式确立之前，将高权限智能体直接部署在专利管理、技术研发、客户交付、商务投标、法务合规等核心环节，并非效率优化，而是对企业自身知识产权防御底线的主动突破。

五. 养"龙虾"的侵权责任

信息泄露的本质是权利侵害，而“龙虾”的全栈式运行特性，更放大了侵权责任认定的复杂性。养"龙虾"同样存在知识产权及其他民事权利的侵权风险。人工智能的知识产权问题已有不少讨论，尤其是 AI 生成作品的著作权保护，在中美两国都积累了诸多司法判例(参见《著作权法很玄学：人工智能生成作品受著作权法保护吗?》)。但与普通 AI 不同，“龙虾”作为开源 AI Agent(智能体)框架，其侵权风险更具多元性与复杂性——开源技能卷轴的权利归属模糊、第三方开发环节的责任传导等问题，让知识产权与民事权利领域的风险边界进一步扩大。

“龙虾”虽不具备法律人格，却形同一名隐形员工，它既能接受指示开展工作，比如代表使用者连接社交媒体进行对话互动，也能遵照指令执行金融投资等复杂操作，但这一过程中也潜藏着不可忽视的侵权风险。其根源在于，“龙虾”的运行完全依赖底层大模型，而大模型的训练数据往往良莠不齐，数

据使用的授权范围也存在差异，经过这类数据训练的模型，很可能因依据错误数据或超出授权范围使用数据，直接引发侵权后果。具体来看，在知识产权领域，它可能生成复制受版权保护的作品、误用他人商标，甚至创造出与现有专利权利要求重叠的发明；在民事权利层面，它可能未经同意处理个人信息，侵犯他人隐私权，或在训练、运行过程中违反数据集及 API 的使用许可条款，违背合同义务。更值得警惕的是，得益于 AI 的规模化运算与高速处理能力，这类侵权行为的影响范围和危害程度，往往远超传统工具造成的侵权。那么，对于使用“龙虾”所引发的侵权行为，我们究竟该如何界定其性质与责任边界?

使用“龙虾”最首要的法律问题，在于其作为 AI Agent(智能体)的法律地位究竟应当如何界定。“Agent”在法律语境中本指“代理人”，而代理人的越权行为通常由其本人承担责任。但问题在于，AI Agent 并非真正意义上的代理人——传统法律中的代理人需具备民事行为能力，而人工智能既非自然人，也非法律拟制的法人。它所谓的“自主决策”，本质上是人类编程、训练数据与部署选择共同作用的结果。从法律属性来看，它更接近于一种高度复杂的自动化工具，而非独立的法律行为主体，尽管其表面上的自主性容易让人产生“它在自行行动”的错觉。将“龙虾”定性为“代理人”与定性为“工具”，二者的法律后果存在本质区别：若视为代理人，其超越授权范围的行为应由其自身承担责任；但“龙虾”本身无法承担法律义务，也不具备责任能力，因此绝不能被认定为独立的代理人。

然而，徒善不足以为政，徒法不能以自行。既然“龙虾”无法对自身行为承担责任，那么应由谁来承担其侵权责任?虽然开发者在一定程度上决定了“龙虾”的智能框架与应用范围，其侵权行为归根结底也是开发者提供的算法与模型的产物，但开发者无法预见“龙虾”未来在何种提示词下开展何种活动，对其运行结果不具备完全可控性。“龙虾”的行为源于用户的指示，从这一角度看，由用户承担侵权责任似乎更为合理。可“龙虾”的运作机制对用户而言仍是黑匣子，用户无法预判其行为带来的后果。

更广泛的法律争论集中在责任应当采取严格责任还是过失责任模式。严格责任意味着用户对“龙虾”的任何侵权输出都要负责，而过失责任则要求证明用户未尽合理监督义务。那么如何才算是合理义务?这又是一个很玄学的问题。欧盟《人工智能法案》等监管趋势正在把 AI 视为一种高风险技术，这对开发者和使用者设定了更高的义务，比如需要事先设置防护措施。但是怎样的防护措施才与“龙虾”的使用相匹配?过高的防护措施可能扼杀“龙虾”的能动性，也不利于技术发展。但任凭“龙虾”胡作非为而无人负责，显然也违背了人工智能发展的初衷。

六. "龙虾"与互联网生态

在互联网发展的早期，《纽约客》杂志曾刊登过一幅著名漫画：一只狗坐在电脑前与另一只狗对话。由此流传开一句名言:"互联网上没人知道你是一只狗。"而现在，互联网上没人知道你是一只"龙虾"。

于是，在社交媒体上与你聊天的可能是“龙虾”，而非真人；在王者荣耀战场上厮杀的，可能也是“龙虾”，而非真实玩家；抢票、拼手速的背后，同样可能是“龙虾”，而非血肉之躯；股票、基金等金融工具的操盘手，也不再是专业人士，而是专业的“龙虾”。这一切对互联网生态意味着什么?

当腾讯推出 WorkBuddy(可视为简化版“龙虾”)时，宣称可与 QQ、企业微信、飞书等主流社交平台连接，却未提供与微信的快捷对接渠道。为何 WorkBuddy 不与腾讯的拳头产品微信打通?因为腾讯深知，微信的核心价值在于真实社交，而以智能体替代真人社交，恰恰会侵蚀这一根基。

智能体在互联网的广泛应用，可能颠覆既有的互联网生态逻辑，冲击现有商业模式。

以微信为例，其核心功能——添加好友、群聊、朋友圈及支付等，均服务于用户的真实社交与生活便利，真实社交是微信生态乃至整体商业模式的基石。人而无信，不知其可也。大车无輗，小车无軏，其何以行之哉。人与人的交流贵在真实，在社交媒体上同样如此。若允许通过“龙虾”等智能体框架，实现自动发消息、批量点赞、评论视频号及公众号等操作，必将导致虚假社交行为泛滥，异化微信作为真实社交平台的本质，模糊“人与人”和“人与机器”的交流边界，最终侵蚀其赖以存在的生态根基。

更进一步，若智能体不表明身份，社交对象无从分辨对方是真人还是智能体，其提供的信息可能在对方不知情、未授权的情况下传输给大模型，构成对个人隐私的威胁。

又如，互联网大量应用以真实流量为基础、以广告为主要盈利来源。广告的有效受众本应是自然人，但应用操作却可能由智能体完成。二者的脱节意味着，应用无法通过实际使用获得有效的广告流量，进而难以维持商业运营。

游戏行业同样面临“龙虾”的冲击。“灰黑产”“游戏外挂”长期困扰产业发展：外挂让部分玩家获取不正当优势，扭曲“升级”“开宝箱”等成长机制，降低游戏可玩性。而“龙虾”接入游戏后，可替代真人 24 小时不间断“打怪升级”“开箱寻宝”，与外挂一样扭曲游戏机制；更因其易于使用，可能导致全民“龙虾”玩游戏，从根本上颠覆网络游戏的娱乐意义。

“拼手速”“抢票”等应用设计也面临挑战。若大量用户使用“龙虾”自动抢票、拼手速，其精确计时与极限点击速度可能在短时间内瘫痪应用，使“拼手速”本身失去意义。

互联网金融领域同样危机暗藏。理论上，“龙虾”依托大模型可掌握全网金融信息，若众多投资者采用相似的决策算法与模型，信息和投资逻辑的趋同可能导致特定时间内出现一致性交易行为，放大市场涨跌波动，加剧金融风险。

结语：

“龙虾”(OpenClaw)这类 AI Agent 的出现，是大模型从“认知”走向“行动” 的必然。它以全栈式授权重构人机协作，既蕴藏提升效率的巨大潜力，也带来网络安全、知识产权保护、责任界定与生态稳定的多重挑战。

技术创新不应因风险而停滞，腾讯对 Work Buddy 与微信的“刻意隔离”启示我们，智能体发展需建立场景化“防火墙”。企业应通过技术手段为其“戴缰”，监管需加快适配规则、明确各方责任，使用者更要审慎授权、持续监督。

从“没人知道你是一只狗”到“没人知道你是一只龙虾”，人机边界虽趋模糊，但“真实、安全、公平”的核心诉求从未改变。唯有在创新与规制间找到平衡，以敬畏之心守规则底线，才能让“龙虾”真正成为服务人类的得力助手，而非风险源头。

如您希望就相关问题进一步交流，请联系：

杨 迅
+86 21 3135 8799
xun.yang@llinkslaw.com

刘莉莎
北京超成律师事务所

如您希望就其他问题进一步交流或有其他业务咨询需求，请随时与我们联系：**master@llinkslaw.com**

上海	北京	深圳
上海市银城中路 68 号 时代金融中心 19 楼 T: +86 21 3135 8666 F: +86 21 3135 8600	北京市朝阳区光华东里 8 号 中海广场中楼 30 层 T: +86 10 5081 3888 F: +86 10 5081 3866	深圳市南山区科苑南路 2666 号 中国华润大厦 18 楼 T: +86 755 3391 7666 F: +86 755 3391 7668
香港	**伦敦**	
香港中环遮打道 18 号 历山大厦 32 楼 3201 室 T: +852 2592 1978 F: +852 2868 0883	1/F, 3 More London Riverside London SE1 2RE T: +44 (0)20 3283 4337 D: +44 (0)20 3283 4323	

www.llinkslaw.com

Wechat: LlinksLaw

本土化资源 国际化视野