人民日报海外版

2026年3月23日 星期一

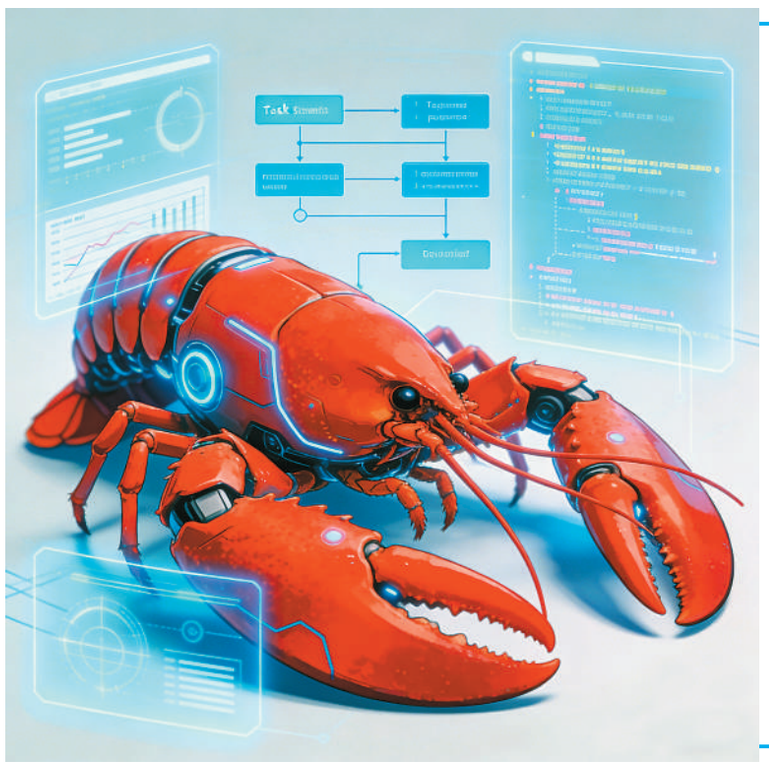
责编：韩维正 邮箱：hanweizheng@peopledaily.cn

『融』观中国

OpenClaw等AI智能体火爆——

“龙虾悖论”怎么破？

韩维正 赵文博



2026年开春，一只“红色龙虾”悄然爬上了全球千万台电脑的桌面。

这不是餐桌上的美味，而是一款名为OpenClaw的开源AI智能体。因其图标是一只鲜红的龙虾，训练这款AI，就被网友戏称为“养龙虾”。

相较传统AI聊天机器人，这只“龙虾”能做的事多得多：接管键盘鼠标、自动整理文件、起草邮件、填写表格、分析数据……一句话，不仅能“动嘴”回答，更能“动手”执行，俨然一位不知疲倦的数字员工。

于是，“龙虾”瞬间点燃了全球开发者的热情。开源平台GitHub上，其星标数短短数月内突破25万，成为最受瞩目的开源项目之一。

但就在无数人欢呼“生产力革命”到来时，质疑声也越来越多：有人账户里的钱被悄悄转走，有人电脑被黑客远程控制，还有人积累了多年的工作文件被一键清空……中国工业和信息化部、国家互联网应急中心等机构接连发布安全风险预警。

这就形成了一个“龙虾悖论”：想让它做的事情越多，给它的权限必须越大；权限越大，安全风险就越高。这个“悖论”怎么破？记者进行了采访。

安装易，用好难

装“龙虾”，只需点击几下。但真正用好它，却是一道难度不小的综合题。

小水智能CEO孙雪峰是最早一批“龙虾”使用者。第一次使用时，孙雪峰雄心勃勃，“上来就派了一个大活儿”，让“龙虾”比照一个全球知名论坛，从头搭建类似平台。“当它开始一步步向我汇报进展时，我觉得肯定没问题。”孙雪峰说，结果验收时傻了眼：除了一堆空文件夹，什么都没有。

“我当时特别失望。”孙雪峰说，“后来我才想明白，不是‘龙虾’不靠谱，是我没有真正理解它，也没学会怎么使用。”

“这款AI放大的是每个人的原有能力，但不能凭空给人增加能力。”一人公司创业者、声序智能创始人王钰博告诉记者，许多人安装是出于好奇或焦虑，而非明确的痛点需求。她自己也踩过坑：一次让“龙虾”修改代码，却忘了指令“完成后重启服务”，结果卡壳数小时。

“使用者需要对预期结果有清晰判断，再把权限交给AI，否则用户没有及时掰正方向的话，AI就会按自己认为正确的方式走下去，出问题也不知道怎么修。”王钰博说。

如果说“不好用”带来的是挫败感，那么“不安全”带来的则是实实在在的损失。有用户因服务暴露在公网上，且浏览器保存了自己的信用卡信息，导致卡片被盗刷；Meta公司的AI安全负责人Summer Yue则曾眼睁睁看着失控的“龙虾”飞速删除了自己邮箱中两百多封邮件。

北京哪吒互娱创始人苏魁坦言，之所以迟迟未敢出手安装“龙虾”，就是顾虑于安全风险：“你想让它帮你做的事情越多，给它的权限就必须越大，万一失控，捅的娄子就越大。”

腾讯云安全副总经理、AI Agent安全中心负责人谢奕智告诉记者，当前龙虾等AI智能体（Agent）主要存在两大典型安全风险：一是系统稳定性风险，AI不可控的行为可能导致系统崩溃；二是核心密钥泄露风险，为保障AI执行能力需授予密钥权限，攻击者可通过诱导式提示词，致使密钥被直接泄露。

缺乏审核的第三方插件（Skills）也是风险重灾区。据国家网络与信息安全信息通报中心通报，对3016个技能插件的分析发现，有336个插件包含恶意代码，占比高达10.8%。

一道清晰的“分水岭”已出现。“养龙虾”成功的用户，大多遵循“最小权限原则”，往往在沙箱中隔离测试，对高风险操作，设置人工确认环节；失败者，则往往对新事物的风险预估不足，一兴奋就交出全部权限，从而陷入被动。因此，决定“龙虾”实际价值的，首先是使用者的安全意识和数字素养。

是帮手，不是万能

“养龙虾”热潮激起的波澜很快超越产品本身，演变成关于AI发展方向的讨论。

力挺“龙虾”的人认为，该产品代表了未来方向。今年全国两会上，全国政协委员、中国科学院计算技术研究所研究员张云泉就表示，现在“养龙虾”创业流行，OPC（一人公司）发展势头很猛，未来大家都要加强学习，使用智能体工具。支持者认为，真正的风险就是错失新技术。

苏魁向记者讲述了自己的“转变时刻”：一次直播演示中，“龙虾”的长期记忆和自主运行能力打动了他，他随即为公司配了两只“龙虾”，一只“主内”充当“主管”，一只“主外”负责每日内容宣发，结果也很好：每日新增10—20名用户，付费用户月增长3—5倍，而每月总成本不过60元。

在苏魁看来，“龙虾”改变的不只是效率，更是单人创业的可能性边界：“人的能量有多大，不取决于有多少钱，也不取决于专业能力多强，而取决于创意能力如何。只要创意够好，再加上一群数字员工配合，就可以做大事业。”

“AI相当于我的多个员工，前端、后端、运营宣发，跟我组成团队。”作为一人公司创业者，王钰博这样描述道。她有多年民乐经验和AI开发背景，用“龙虾”搭建了一个民乐曲谱转换平台。在她看来，AI填平了算法工程师和产品开发者之间的技能鸿沟，可以帮助她推进复杂项目。

也有相对冷静的观点。孙雪峰认为，“龙虾”目前还干不了两类事：一是财务与决策，还需要时间建立新标准、新安全规范，因为人类除了干活能力，还承担法律责任与社会责任；二是任何需要人和人当面完成、带有情感交互的场景，AI还替代不了。

王钰博则认为，“龙虾”能协助的是执行层，判断层依然是人的地盘。她打了个比方：家里买齐了锅碗瓢盆、各式铲子和微波炉，不代表这个人就能成为好厨师。AI是工具，能不能用好，关键在人。

机构的态度更为审慎。目前，北京大学、安徽师范大学、珠海科技学院等多所高校已经发布预警，要求师生关注使用风险，一些高校严禁在处理教学科研数据、行政办公信息等核心场景中使用“龙虾”，有的甚至要求已安装者立即卸载清理。

谢奕智表示：“担忧是对的，但风险并不来源于技术本身，主要集中在权限滥用和恶意插件引入这些环节。关键是做好安全管控。”

不必恐慌，也别盲目

面对“龙虾热”，包括监管部门在内，各行各业都在快速作出反应。

工业和信息化部、国家互联网应急中心分别发布“龙虾”安全风险警示，明确提出“六要六不要”，点明“提示词注入”、“误操作”、功能插件（skills）投毒、安全漏洞等四类严重风险，并提出强化网络控制、加强凭证管理、严格管理插件来源、持续关注补丁和安全更新等相关建议。这体现了监管层对AI新工具“包容审慎”的态度——既不一刀切封堵，也不放任自流。

安全正成为新的竞争维度。“龙虾热”不仅催生挑战，也带来新的市场机会。一些网络安全公司推出针对AI智能体的防护方案，例如腾讯云等已推出AI Agent安全中心，提供插件安全扫描、行为全流程审计、密钥沙箱隔离等安全防护。“现在很多国内企业部署AI Agent，首要问题就是安全性够不够。安全不是可选项，而是必答题。”谢奕智说。

普通人应当如何面对“龙虾”为代表的AI热潮？受访者们给出的建议相对一致：不必恐慌，也别盲目。谢奕智建议，如果日常工作重复性高、偏流程化，可以尝试小范围试点应用AI Agent，严格做好权限管控，循序渐进把风险逐步降到最低；苏魁则表示最好是有长期的AI需求，如果今天想用明天又不用，“龙虾”就不太合适。

总之，真正决定AI工具是“神器”还是“灾难”的，往往是人的认知是否能跟上工具迭代。每一次技术迭代，都是重新平衡效率和风险，也对用户的数字素养、安全意识等提出更高要求。说到底，要让工具创造价值，首先要具备真正驾驭工具的能力。

题图：AI生成的“龙虾”示意图。

记者手记

今天你FOMO了吗？

韩维正

“养龙虾”火了，随之火起来的还有一个英文词：FOMO。

FOMO，全称是Fear of Missing Out，即“错失恐惧症”，形容一些人面临新技术冲击时，因为害怕被时代抛弃而焦虑的现象。看着别人用“龙虾”写出爆款文章、自动生成报表，这些人油然而生焦虑：我是不是落伍了？我是不是又错过了什么改变命运的机遇？

看多了这类故事，笔者多少也有点FOMO，但更多是好奇：“龙虾”真这么厉害？我决定找台电脑养只“虾”试试。毕竟，想知道梨子的味道，就要亲口尝一尝。

我给“龙虾”找了个真实的任务：填表格、做周报，把七八个不同平台的数据文档，合并成一份有固定格式、大气美观的表格周报。这些文档格式各异，数据单位不统一，日期的写法各不相同——有时候数字是精确到个位的，有时候又已经换算成万。过去每次手动整理这份表格，我要花三四十分钟，而且每周重复，每次都要从头核对，难度不高，但颇琐碎。

我把任务扔给“龙虾”，等它表演。但我愣住了。它没表演，反而问我问题：格式是什么样的？哪些数字需要计算，哪些直接引用？出现不一致，哪份数据为准？……

我一时语塞。这些琐碎事我从未认真思考过，只是每周凭经验和“肌肉记忆”，靠着“反正我知道该怎么做”的惯性在做。当要把它教给另一个“人”时，我才意识到：想让“龙虾”靠谱打工，必须把模糊感觉变成清晰指令。

我开始认真整理规则，一条条教给AI。它迅速写出第一版脚本，跑通大部分逻辑，但有几处不对，得改。这样来回十几轮，近三小时后，一套完整自动化流程真的跑通了。第一次点击运行，6分钟后，一份格式规范、数据准确的周报出现在屏幕上。

我盯着它看了好一会儿。

试验过程中，我曾不止一次问自己：自己动手，半小时就能搞定，何必这么麻烦？但当6分钟这个数据摆在面前，成本收益就明显了：教AI三个小时，够我手动做6遍数据；但如果这件事做到6周以上，教“龙虾”的时间就“回本”了，往后都是赚。

至于花费，我用的是某国产版本，目前免费试用。产品送我5500积分，调教“龙虾”花了500积分，之后每次做任务只需22.5积分。免费“羊毛”能薅多久尚不可知，但这一尝试过程，让我对“养虾”有了一些新认知。

目标必须明确：“龙虾”不怕任务复杂，就怕“主人”自己都不清楚要做什么。“帮我优化这篇文章”，指令模糊，结果可能南辕北辙；换成“100字以内，开头用疑问句，语气活泼”，它就清楚怎么干。

目标要能拆解：只打一行字就让“龙虾”产出理想结果不太可能，分段实现大目标就容易一些。我把任务拆成十几个步骤：合并、修格式、统一单位、核对数据……每个步骤单独交给“龙虾”，逐步咬合，最后才能连成流畅的工作线程。

知道问题所在：和所有AI智能体一样，“龙虾”绝非万能，更会犯错，但这时不能只说“不对”，得明确指出错在哪里、为何是错，AI才能修正。说白了，“主人”得清晰掌握任务。犯错的“龙虾”就是镜子，照出的是“主人”到底懂不懂任务。

和“龙虾”合作，其实是在倒逼“主人”成长，成为逻辑清楚、表达清晰、懂得拆解问题的“好领导”。以前，这些能力或藏在你的头脑里，或融化在经验中，但面对AI，咱必须把经验变成清晰指令，明确传递给另一个主体。

如果你也正因“龙虾”而FOMO，我的建议是：找一件手头真实的任务，找一台干净的电脑，安全地试一试。不必非得付费上门安装，也不必着急报班学编程，先用具体任务和一点耐心，试一试。好不好用，能不能用，值不值得花钱用，一试便知。

如果有用，你就拥有了新的生产力工具。一时半会用不上？那说明你的工作里，有些东西AI还替代不了。光靠焦虑治不了焦虑。想知道梨子的滋味？得亲口尝尝。要了解河水深浅？那就下水蹚蹚。

融媒速递 ▶▶▶

纪录片《超大城市的一平米》发布

海外网北京电（刘凌）由上海市人民政府新闻办公室、上海市住建委联合出品，上海市水务局支持，澎湃新闻制作的纪录片《超大城市的一平米》日前发布。

纪录片用42分钟讲清了一场覆盖千家万户、历时数十载的“马桶革命”。影片通过虹口区居民陈国斌、长宁区居民刘征等家庭的案例，展现了“一平方米”的独立卫生间如何终结“拎马桶”的历史，赋予居民生活尊严。

纪录片采用3D和AIGC技术，将城市地下庞大的污水管网系统进行可视化。纪录片不仅记录了个体居住条件的改善，更展示了城市治理的进步：从早年上海东区污水厂核心技术受制于人，到后来整套城市卫生系统的自主研发与追赶。每一处因地制宜的改造方案，都体现了基层治理的“绣花功夫”，诠释了“人民城市”的建设理念。

据悉，该片英文版正在筹备中。

AI生成的城市管网示意图。 **澎湃新闻**供图

关于OpenClaw安全应用的风险提示

前期，由于OpenClaw智能体的不当安装和使用，已经出现了一些严重的安全风险：

1.“提示词注入”风险。网络攻击者通过在网页中构造隐藏的恶意指令，诱导OpenClaw读取该网页，就可能导致其被诱导将用户系统密钥泄露。

2.“误操作”风险。由于错误地理解用户操作指令和意图，OpenClaw可能会将电子邮件、核心生产数据等重要信息彻底删除。

3.功能插件（skills）投毒风险。多个适用于OpenClaw的功能插件已被确认为恶意插件或存在潜在的安全风险，安装后可执行窃取密钥、部署木马后门软件等恶意操作，使得设备沦为“肉鸡”。

4.安全漏洞风险。截至目前，OpenClaw已经公开曝出多个高中危漏洞，一旦这些漏洞被网络攻击者恶意利用，则可能导致系统被控、隐私信息和敏感数据泄露的严重后果。对于个人用户，可导致隐私数据（像照片、文档、聊天记录）、支付账户、API密钥等敏感信息遭窃取。对于金融、能源等关键行业，可导致核心业务数据、商业机密和代码仓库泄露，甚至会使整个业务系统陷入瘫痪，造成难以估量的损失。

建议相关单位和个人用户在部署和应用OpenClaw时，采取以下安全措施：

1.强化网络控制，不将OpenClaw默认管理端口直接暴露在公网上，通过身份认证、访问控制等安全控制措施对访问服务进行安全管理。对运行环境进行严格隔离，使用容器等技术限制OpenClaw权限过高问题。

2.加强凭证管理，避免在环境变量中明文存储密钥；建立完整的操作日志审计机制。

3.严格管理插件来源，禁用自动更新功能，仅从可信渠道安装经过签名验证的扩展程序。

4.持续关注补丁和安全更新，及时进行版本更新和安装安全补丁。

（资料来源：**国家互联网应急中心 赵文博**整理）

3月12日，在江苏省宿迁市便民方舟2号楼一楼大厅，不少市民在工作人员协助下免费安装开源AI智能体OpenClaw。

王帅甫摄（人民图片）