“OpenClaw像是让人工智能坐上了总裁的位置，手里还拿着公司的钥匙、公章和调度权”

■这是3月11日在浙江省湖州市吴兴区飞英街道拍摄的开源AI智能体“龙虾”手机端页面。 新华社发

最近半个月，IT界最火爆的就是OpenClaw。这只画风极其古典的红色小龙虾，上线不久便迅速成长为2026年度现象级“开源奇迹”。与之相对的，是不断有安全警告发出。昨日，国家安全部在微信公众号上发布了“‘龙虾’(OpenClaw)安全养殖手册”(下称“手册”)，明确表示“龙虾”在创新改变生活的同时，也存在原生风险，用户要理性辨别、规范使用，让“龙虾”成为遵规守纪、产能高效的“数字员工”。

那么，OpenClaw这只“龙虾”，它的安全风险究竟在哪里？

■新快报记者 王敌

隐患 权限过高没边界感

“在完全不考虑隐私和经济的情况下，‘龙虾’展现出的能力，容易让用户产生AI突然变强了的错觉。”载乐网络科技创始人王自如在最近的一次直播中谈到，OpenClaw的亮眼，是建立在两项重大妥协之上，一是权限无限，它几乎没有隐私侵占的限制；二是经济无限，使用时完全不考虑用户成本。王自如说，“事实上，当今的主流大模型本身都具备较强能力，只是受限于规范和成本，没敢这么玩。”

最近一周，国家互联网应急中心与工信部都发布了安全风险警示，许多高校和企业也禁止了其在生产环境中的应用。

对此，华南理工大学未来技术学院教授许言午表示，OpenClaw和传统大模型不一样。传统大模型更像是顾问，只提供建议，采不采纳在用户，风险可控；OpenClaw则像是让人工智能自己坐上了总裁的位置，手里还拿着公司的钥匙、公章和调度权。

许言午指出，智能体通常知道“什么能做、什么不能做”，但这种边界很多时候不是铜墙铁壁，而更像是“写在墙上的规章制度”，智能体经常可以突破。OpenClaw要替人干活，就必须接触大量外部资源，比如OAuth授权、SSH密钥、本地文件权限、浏览器会话等等。“正常情况下，它当然能提高效率，但一旦它被外部劫持或者自身判断失误，它下达的错误命令就不再只是‘建议’，而会直接变成行动。”许言午说。

百川智能CEO王小川预言，今年将会是智能体安全事故集中爆发的一年。他打了一个比方：“权限失控的智能体，就像有一把上了膛的手枪的3岁小孩。”

实操 用“龙虾”花费不菲

OpenClaw火起来后，token就成了高频词汇。中文里token的正式译名叫“词元”。

词元是大模型理解语言的单位。粗略算，1个词元≈1个汉字或0.75个英文单词。你发消息、AI回消息，都按词元收费。一般情况下，AI回复消耗的词元是用户提问的2到5倍——AI说得越多，你花费得也就越多。如果选了比较高端的模型，词元消耗还会指数级增长。

有些企业推广OpenClaw时宣称“赠送千万token”，结果用户只用了几天，8位数的词元就见了底。

不过，这笔钱不是OpenClaw收的，而是模型公司按用量收的。也就是说，如果真的想好好养“龙虾”，用户必然要开许多会员服务，以保证有足够多的指令来调教它。

成功调教了“龙虾”的极客小胡透露，他的“养虾”过程很顺利，对小龙虾的调教也很到位。不过，在“养虾”的过程中，小胡开了许多会员，其中智谱订阅会员大约1000元，MiniMax的订阅约500元，OpenAI的会员也要几百元，Silicon-Flow(硅基流动)会员也要几百元。“养‘龙虾’这一个多月，购买各种会员和服务器，已经花了数千元。”小胡说。

由于使用OpenClaw需要使用命令提示符，很多时候词元都被浪费在了不精确的指令上。“养虾”用户大壮表示，大家用电脑早已习惯了图形界面，结果一碰OpenClaw，直接被拽回DOS时代。“突然要用命令提示符，整个人都不好了。”大壮坦言，要精确地用英语输入计算机能理解的指令，多一个空格、少一个斜杠，“龙虾”都学不会技能。

国安部 明确隐患规范使用

国家安全部发布的手册中，确认了OpenClaw具备强大的功能，让AI从“给出方案”变成了能“落地执行”，而且具备自我优化的能力。

不过，手册提醒用户，要认清OpenClaw的真面目，给自己的电脑做好防护。

为实现OpenClaw强大功能，用户常赋予“龙虾”最高系统权限，这既可能导致AI误操作造成数据损失，更严重的是，一旦被攻击者渗透，设备管理权限可能被悄然窃取，主机遭远程操控，资源被非法占用。使用中，部分用户将个人敏感信息交由“龙虾”处理，一旦系统被攻破，用户数据可能被窃取，将面临隐私泄露乃至财产安全的巨大威胁。有些“龙虾”可在社交网络自主发声，若被不法分子接管，用户言论可能被篡改，沦为生成虚假信息、实施网络诈骗的工具。

而在技术方面，“龙虾”缺乏专业维护和漏洞修复机制，攻击者可能通过恶意插件等方式，诱导其突破权限，主动窃取核心信息，其隐蔽性远超传统木马。

如何做个安全的“养虾人”，手册指出，用户给自己的“龙虾”全面体检，检查控制界面是否暴露、权限配置是否过高、插件来源是否可信，发现严重风险立即隔离处置。同时，严格遵循最小权限原则，对敏感数据加密存储，建立完整审计日志，尽量在隔离环境中运行。要让“龙虾”听话好用，需要清晰认知“龙虾”并非供人娱乐的数字宠物，而是能够自主执行任务、承担流程操作、持续学习成长的“数字员工”，要理性看待、规范使用，确保在合规、安全、可控的前提下服务于生产生活。

OpenClaw引发AI焦虑?

“龙虾”不是人人都要养

“想到即做到”是人类对AI的最高期望。OpenClaw这只半岁不到的“龙虾”，似乎正让幻想照进现实，动动嘴，它就能替你干活。随之而来的，就是AI焦虑，还带火了一个心理学名词“FOMO(Fear of Missing Out的缩写)”，意思是“错失恐惧症”。

别人都养上了，我不养，是不是要被时代甩下？华南理工大学未来技术学院教授许言午表示，“技术热潮里最容易出现一种误区，就是把‘能用’误认为‘都需要用’。”对于OpenClaw这个新生事物，使用前还需要官方认证和引导，普通人现阶段还无法驾驭这个“超能力工具”。

许言午说：“OpenClaw这类智能体确实代表了很有前景的AI方向，但并不意味着每个人、每个岗位、每个场景都必须马上配一个。”

许言午还提到成本问题，“这类智能体在执行任务时，往往不是一次简单对话就结束，它需要不断理解指令、拆解任务、调用工具、持续修正，背后会产生大量模型调用，成本可能远超人工。”

许言午认为，从现实需求看，除了需要体验前沿产品的AI研究者，真正适合使用智能体的人，往往是那些重复性任务多、数字化流程重、愿意折腾工具链、对效率回报很敏感的人，比如运营人员、研究人员、内容处理岗。“如果一个人的日常工作本来就高度非标准化，或者对隐私、合规、稳定性要求很高，那他未必适合第一时间‘养龙虾’。”许言午说。