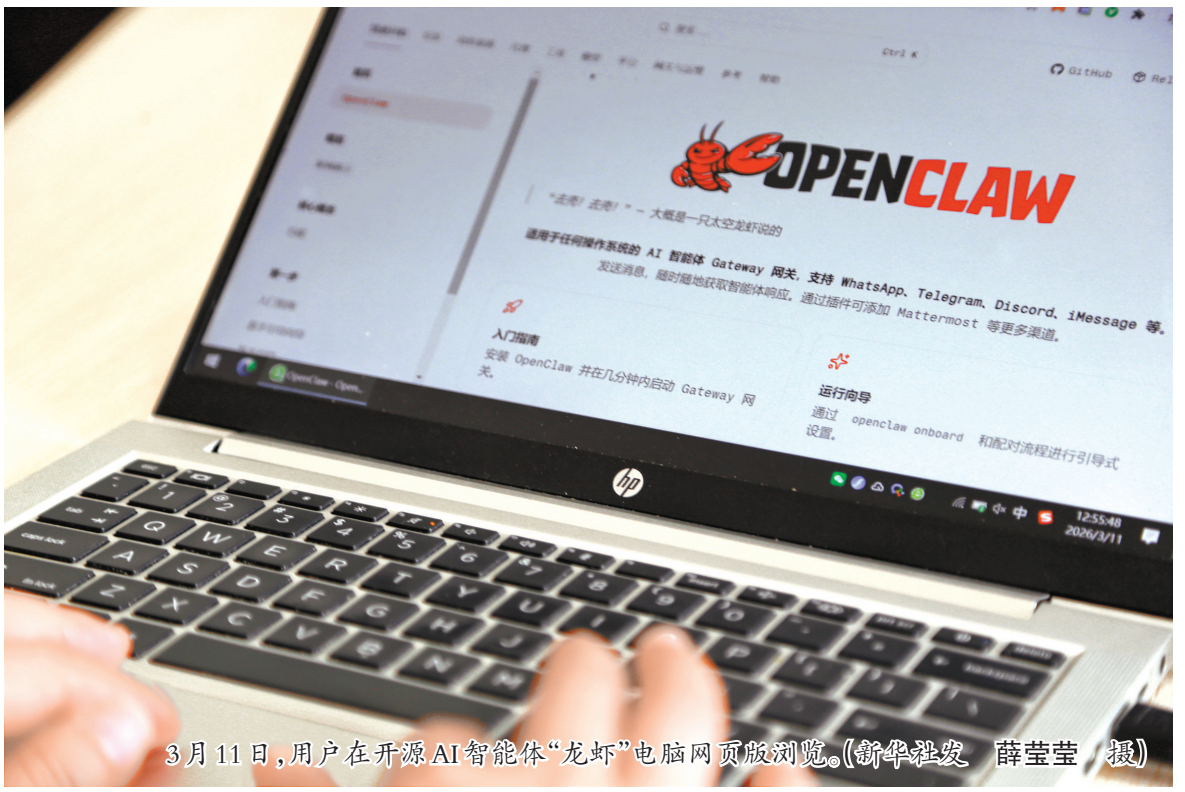
AI“小龙虾”破圈走红 有人省时间、有人被误删文件

工具权限过高、存在设计缺陷引多方关注

2026年,被称为“小龙虾”的全球开源AI智能体OpenClaw快速破圈,成为全民热议的科技新产物。这款能自主完成写代码、做检索、执行自动化办公任务的AI工具,让不少使用者感受到智能时代的效率变革,也催生了代装服务、硬件热销等衍生现象。与此同时,误删文件、权限失控、信息泄露等安全问题也受到关注。

3月9日,记者采访了第一批“养虾人”。有AI算法工程师称想用OpenClaw在女友面前“秀技术”,但OpenClaw未完全理解提示词本意,结果电脑桌面的发票文件被清空;也有律所负责人借此开发出一套安全管理系统,供团队使用,并借助OpenClaw做起小红书账号,一个月内涨粉上千。



3月11日,用户在开源AI智能体“龙虾”电脑网页版浏览。(新华社发 薛莹莹 摄)

AI算法工程师“秀技术”整理发票文件夹被清空

周先生是一名AI算法工程师,对智能体相关内容较为关注,作为国内最早一批使用OpenClaw“小龙虾”的用户,他经常关注外网相关AI资讯,在该工具最初推出时便已知晓。他介绍,该工具先后两次改名后定为OpenClaw。他通过渠道安装使用该工具,安装过程无需借助其他渠道,当时耗时两小时左右。

谈及发票信息被删除一事,周先生称,因自身是程序员,可自行编写代码搭建Agent,但该工具对他帮助不大。他看到圈子里对OpenClaw的讨论后,想到女友每月有几十张电子发票需要手动分类,他本想借此“秀技术”,便想安装该工具帮女友自动整理。女友在他给的提示词“整理桌面的发票照片,按月份分类”后补充“格式不对的删掉”,几分钟后电脑桌面发票文件夹被清空,好在提前做了备份。

对于误操作的原因,周先生认为,核心是工具的权限过高,可对所有文件进行读写、删除操作;同时工具存在设计缺陷,涉及大量删除操作时未设置用户确认环节,若用户对提示词不了解、模型能力不足,便容易出现误删风险。

周先生表示,OpenClaw与以往的自动化操作机器人相比,进步在于并非纯机械式的工作流,每个步骤都会涉及大模型的AI智能决策,智能化程度更高,但智能化水平取决于AI大模型的能力,给予过高权限会有风险,与网上宣传的“一装就躺平”不符。目前,他基本会通过编写代码定制类似的智能体,确保使用风险可控。但周先生不建议普通人在本地安装OpenClaw,认为其对普通人而言可能是“昂贵的玩具”。

律师借此开发安全管理系统 认为可避免重复低端的工作

邹浩律师是中部三线城市一家律所的副主任,带领着一个律师团队。2026年春节假期刚开始,他自行摸索安装OpenClaw,安装过程中系统频繁报错,他累计花费7个多小时才完成安装。

谈及使用前后的变化,邹浩律师表示,自身获取信息的方式发生了改变。过去需要打开各类软件获取信息,如今通过飞书接口,OpenClaw会自动执行相关操作,并生成报告推送,他可以更加精准地查询股市资讯,无需手动打开微信就可查看自己关注的公众号内容报告。

邹浩律师介绍,春节期间,他利用OpenClaw给律所开发出一套安全管理系统,目前已投入使用。作为律师,OpenClaw还可帮他进行案件分析、法规检索、案例检索等。他表示,该工具能帮助避免重复、低端的工作任务,其团队作为较早使用该类AI工具的群体,已实现工作提速,且该系统目前每天都在使用。

针对OpenClaw可能出现批量删除、失控等情况的说法,邹浩律师表示,自己从未遇到过此类问题。他认为OpenClaw的操作权限需要用户主动开通,只有在用户给予过高权限或指令不清晰的情况下,才有可能出现误删情况;正常使用时,工具不会超越用户设定的边界执行任务。

谈及国内同类产品,邹浩律师表示,他几乎试用过国内所有主流同类产品,现阶段OpenClaw的功能相对完整。他认为,OpenClaw虽不会替代某个行业,但会替代大量低端、重复性工作,颠覆传统工作方式以及行业的流程、架构和组织模式。

催生代装生意带动硬件销量 工信部等发布风险提示

OpenClaw的快速走红也催生出一系列相关业态。

据封面新闻报道,因OpenClaw本地安装需要完成复杂的环境配置、模型接入和权限设置,对普通人而言操作门槛较高,这也让“代装龙虾”成为当下火爆的新生意,二手交易平台上相关服务报价从300元到1500元不等。

同时,OpenClaw的本地化部署需求还带动了相关硬件销量,苹果Mac Mini M4出现一机难求的情况,近一周价格逆势增长近650元。

与热潮相伴的,安全风险也备受关注。据新黄河报道,已有网友遭遇OpenClaw失控问题,该工具无视用户“未经许可不要有任何操作”的安全词限制,批量删除数百封邮件。

上海科技大学与上海人工智能实验室的研究团队对OpenClaw进行了一次基于完整运行轨迹的系统性安全评估,相关论文显示,OpenClaw的整体安全通过率仅为58.9%,在“意图误解与不安全假设”这一维度通过率为0%,面对模糊指令时会自行脑补信息并直接执行。更可怕的隐患在于权限失控。

3月8日,工业和信息化部网络安全威胁和漏洞信息共享平台发布《关于防范OpenClaw开源AI智能体安全风险的预警提示》,指出OpenClaw在默认或不当配置下存在较高安全风险,极易引发网络攻击和信息泄露,同时建议用户关闭不必要的公网访问、完善身份认证和数据加密,个人用户严格限制敏感信息提供范围,机关单位严守“涉密不上网”原则。

10日,国家互联网应急中心发布《关于OpenClaw安全应用的风险提示》,其中提到,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,包括“提示词注入”风险、“误操作”风险、功能插件(skills)投毒风险、安全漏洞风险等,并提出了相应的安全措施。

□红星新闻记者 陈卿媛