AI智能体“龙虾”为何引发广泛警惕

今年年初以来，一款俗称“龙虾”的人工智能（AI）智能体工具“开放之爪”（OpenClaw）凭借其自主执行复杂任务、可扩展技能包等强大能力，在开源社区迅速崛起。但爆火之后，“开放之爪”接连被曝出存在多重安全隐患。

目前，多国监管机构和科技企业已陆续发布针对“开放之爪”的使用指南和规范。4月1日，中国国家知识产权局发布风险提示说，“开放之爪”等智能体工具被曝光默认安全配置脆弱，易引发严重安全风险。与此同时，使用此类智能体撰写专利申请文件，也可能诱发多重风险。

安全漏洞频发

“开放之爪”由奥地利软件工程师彼得·施泰因贝格尔开发，是一款开源AI智能体软件。该智能体采用层级化架构，将社交即时通讯软件与自动化智能体深度耦合，同时借助插件系统扩展各种工具能力。这种分层架构虽赋予了“开放之爪”灵活性与可扩展性，但也带来了多维度的安全风险。

1月下旬，开源平台GitHub上发布的一项安全审计报告显示，“开放之爪”存在512项安全漏洞，其中有8项被归类为“严重”，涵盖了身份验证、机密管理等领域。

2月下旬，国际网络安全机构“绿洲安全”研究人员发布报告说，“开放之爪”核心系统中存在一个名为“ClawJacked”的重大安全漏洞，攻击者可能通过恶意网页接管该智能体，从而获取设备权限和访问系统数据。“开放之爪”团队将漏洞定级为“高度危险”，并在24小时内发布了修复版本。

3月30日，中国360数字安全集团在官方微信公众号上发文说，在“开放之爪”平台中发现一处高危漏洞，影响范围覆盖全球50多个国家和地区。

攻击风险广泛

美国微软公司安全团队发布的风险报告显示，使用“开放之爪”可能面临两类攻击风险：恶意技能插件和间接提示词注入。

“开放之爪”的执行能力依赖于社区平台提供的技能插件。绿盟科技公司近期发布的安全报告指出，如果缺乏严格的代码审计和签名校验，攻击者可通过发布包含恶意提示词和代码的恶意技能插件实现“代码投毒”。用户可能只因一次点击就加载了此类插件，攻击者可在受害者系统中获得持久驻留能力。而攻击者上传自定义技能插件的门槛非常低，只需要注册一个非实名的GitHub账号即可。

据美国派拓网络公司2月发布的数据，研究人员已在相关平台上发现超过800个针对“开放之爪”的恶意技能插件。

提示词注入是一种针对大语言模型的攻击技术，分为直接注入（攻击者直接输入恶意指令）和间接注入（通过网页、文档等外部数据源实现攻击）两种方式。

美国“众击”网络安全服务公司近期在官网发文说，提示词注入的首要威胁是敏感数据泄露，考虑到“开放之爪”对敏感文件与系统的高访问权限，这一风险尤为严重。间接注入则会进一步放大风险，因为攻击者无需直接与“开放之爪”交互，只需污染其读取的数据，恶意指令即可悄悄进入软件决策流程。

多国发布使用规范

对于“开放之爪”是否适合在企业中部署应用，“众击”公司的文章指出，若员工在企业设备上部署“开放之爪”或将其接入企业系统，且配置不当、缺乏安全保护，它就可能成为系统“后门”，执行攻击者的指令。

业内人士建议，个人或企业用户不要在常规办公与涉密设备上运行“开放之爪”，如需部署须采取权限治理、沙箱机制、持续监控与全周期安全防护等严格管控措施。

据媒体报道，出于风险管控的考虑，美国元宇宙平台公司、韩国多音通讯公司等多国科技企业已禁止员工在办公设备上使用“开放之爪”。与此同时，多国监管机构也发布了关于使用“开放之爪”的安全指南。

荷兰数据保护局2月发布公报，建议用户和组织不要在存有敏感或机密数据（如访问码、财务行政资料、员工数据、私人文档或身份证明文件）的系统上使用“开放之爪”及类似AI智能体；建议谨慎对待外部插件，实施严格的访问控制，在存在泄露风险时及时更新登录信息。该监管机构还呼吁将“开放之爪”等AI智能体纳入欧盟《人工智能法》的管辖范围。

3月22日，中国国家互联网应急中心等发布了“开放之爪”安全使用实践指南。此前，工业和信息化部网络安全威胁和漏洞信息共享平台组织相关机构研提了“六要六不要”建议，以防范“开放之爪”开源智能体安全风险。 据新华社电

OpenAI融资超1200亿美元

据新华社电 美国开放人工智能研究中心（OpenAI）3月31日宣布结束最新一轮融资，共筹集资金1220亿美元，当前公司估值达8520亿美元。按英国媒体说法，这让OpenAI成为全球市值最高的私有企业之一。OpenAI正积极筹备今年内在美国首次公开募股（IPO）。

陪孩子看屏幕比限制时长更利于培养儿童亲社会行为

据新华社电 新加坡一项最新研究发现，相比单纯限制孩子使用电子屏幕的时间，如果父母在孩子看屏幕时陪伴观看、交流讨论或一起玩耍，更能帮助孩子培养“亲社会行为”。

研究团队在2018至2019年追踪了2449名3至6岁的新加坡儿童。研究显示，父母与孩子共同使用屏幕的时间，与孩子的“亲社会行为”呈正相关。这一研究表明，当父母与孩子一起观看、提出问题并讨论所见内容时，屏幕使用时间也可以促进社交学习。

3月全球AI观察

从对话到执行 智能体重塑AI应用格局

继聊天机器人之后，能“自主干活”的智能体正快速进入人们生活。今年以来，一款名为“开放之爪”（OpenClaw）的开源智能体在全球科技圈迅速走红。数据显示，它在开源平台GitHub上线仅两个多月便获得超过30万颗“星标”，显示出开发者社区对其高度关注。在中国，在设备上部署该工具被称为“养龙虾”，“养虾”热潮以超乎想象的速度火爆出圈。

支持智能体的AI模型、技术方案等也不断演进。美国开放人工智能研究中心（OpenAI）3月5日宣布推出最新升级版大模型GPT-5.4，称该模型是其首个能直接操作计算机的通用模型，不仅擅长编写代码，还能根据屏幕截图发出鼠标和键盘操作指令，使智能体能够操作计算机，并在不同应用程序之间执行复杂的工作流程。

英伟达公司创始人兼首席执行官黄仁勋3月16日在该公司年度GTC大会（GPU技术大会）上发布名为“NemoClaw”的软件栈，用于支持“开放之爪”智能体的开发和部署。

智能体等AI应用的兴起，使得“词元”（token）概念越来越多地进入公众视野。作为AI模型处理和生成信息的基本单位，词元逐渐成为衡量智能服务成本与价值的重要标尺。全球主要AI模型聚合平台“开放路由器”数据显示，用户通过该平台调用AI模型的词元总量从今年年初的一周约6万亿大幅上涨到截至3月22日一周的20.4万亿。

为抓住AI发展机遇，多国政府和企业纷纷加大政策扶持与产业投资，在智能经济布局、国际合作与规则完善、算力基建扩容等方面持续发力。

中国在今年的政府工作报告中首次提出“打造智能经济新形态”，包括深化拓展“人工智能+”，促进新一代智能终端和智能体加快推广，推动重点行业领域人工智能商业化规模化应用，培育智能原生新业态新模式等。

韩国政府3月2日表示，韩国将与新加坡加强在AI领域的合作，韩方计划在新加坡设立3亿美元的全球投资基金，吸引对初创企业和AI领域的投资。

欧盟理事会3月13日就一项简化部分AI监管规则的提案达成一致，主要涉及精简欧盟AI监管框架，减轻企业合规负担等内容。

德国政府3月17日公布的一项数据中心扩容战略规划显示，到2030年，德国通用数据中心的算力将在2025年基础上至少翻一番，其中专门用于AI的算力将至少增至2025年的4倍。

美国知名企业家埃隆·马斯克3月21日宣布，一个名为Terafab的大型芯片制造项目将落户得克萨斯州奥斯汀。根据他的构想，该项目将建设一座集设计、制造、封装和测试于一体的先进晶圆厂，目标支撑每年最高达1太瓦级算力需求，并服务自动驾驶出租车、人形机器人及太空数据中心等应用场景。 据新华社