国家互联网应急中心等发布 OpenClaw安全使用实践指南

OpenClaw（龙虾）因具备系统指令执行、文件读写、API调用等高权限能力，默认配置与不当使用极易导致远程接管、数据泄露、恶意代码执行等严重安全风险。为帮助用户安全使用OpenClaw，CNCERT会同中国网络空间安全协会组织国内相关厂商共同研究，面向普通用户、企业用户、云服务商以及技术开发者/爱好者，提出以下安全防护建议。

一、普通用户

（一）建议使用专用设备、虚拟机或容器安装OpenClaw，并做好环境隔离，不宜在日常办公电脑上安装。

方案1：用闲置旧电脑专门运行，清空个人数据。

方案2：用VMware、VirtualBox、Docker创建独立虚拟机或容器，并与宿主机隔离。

方案3：在云服务器部署，本地仅远程访问。

（二）建议不将OpenClaw默认端口（18789\19890）暴露到公网

配置为仅本地访问（127.0.0.1），关闭端口映射与公网IP绑定。

如需远程，建议采用VPN访问等方式，并启用验证码等强认证措施。

若对接即时通讯软件（如微信、钉钉、飞书等等），建议仅允许本人或已授权的可信人员访问。

（三）建议不使用管理员或超级用户权限运行OpenClaw

创建专用低权限账户，仅授予最小必要目录的读写权限。

关闭无障碍、屏幕录制、系统自动化等高危权限。

仅开放专用工作目录，禁止访问桌面、文档、下载、密码管理器目录。

配置白名单路径，拒绝读取配置文件、密钥文件等隐私配置。

关闭系统命令执行功能，仅在必要时临时启用并二次确认。

限制网络访问，仅允许连接必要的AI服务与API。

（四）建议安装可信技能插件（Skills）

谨慎安装、使用外部社区/个人发布的Skills，预防信息泄露或服务器被攻击等风险

拒绝“自动赚钱、撸羊毛、破解”类不明技能或黑灰产技能。

（五）建议不在OpenClaw环境中存储/处理隐私数据

不用OpenClaw处理银行卡、密码、身份证、密钥等数据。

（六）建议及时更新OpenClaw最新版本及时安装官方安全补丁，关注官方安全公告与漏洞通报。

二、企业用户

（一）建议做好智能体应用的安全管理制度与使用规范

明确允许与禁止的使用场景、数据范围和操作类型，划定智能体应用的业务边界。

建立内部使用规范和审批流程，对引入新的智能体应用或高权限功能需经过安全评估和管理层批准，确保使用有据可依、有章可循。

（二）建议做好智能体运行环境的基础网络与环境安全防护

禁止将智能体服务直接暴露在公共网络上，需通过防火墙、VPN等手段限制访问，仅开放必要端口给可信网络或IP地址。

对智能体所在服务器启用主机入侵防御、恶意流量检测等措施，抵御网络攻击威胁。

确保运行环境定期更新补丁，消除已知系统漏洞，保障基础环境安全可靠。

（三）建议做好智能体权限管理与边界控制

对所有智能体服务账号遵循最小必要权限原则进行配置。

利用系统自带或第三方权限控制工具，对智能体可访问的文件目录、网络域、数据库表等进行边界限定和访问控制。

对具有高权限的智能体，应实行严格的多因素认证和操作审批，在关键资源层设置额外防线，防止权限滥用。

（四）建议做好智能体运行监控与审计追踪

建立针对自主智能体的持续运行监控机制，监控内容包括智能体的行为日志、重要决策输出、系统资源使用以及异常事件记录等。

对关键操作和安全相关事件应生成审计日志并防篡改保存。

配置安全信息与事件管理（SIEM）工具，实现对智能体日志的集中分析，及时发现可疑行为迹象。

审计追踪能力应保证发生事故后可以还原智能体行为路径，为问题调查和责任认定提供依据。

（五）建议做好智能体关键操作保护策略

针对自主智能体可能执行的高危操作，企业应制定保护策略作为治理基线。例如，对删除大量数据、修改核心配置、资金交易等操作设置人工二次确认或多重签批流程；对不可逆转的操作先行模拟演练或安全检查；对高影响操作限定时间窗和范围，仅允许在特定条件下执行。

上述策略应与金融系统、生产控制系统等高安全级别场景的管控措施看齐，确保智能体不会单点突破整个业务安全。

（六）建议做好智能体供应链安全与代码管理

应建立对自主智能体所依赖第三方组件和技能插件的安全管理制度。

引入的新技能模块必须经过安全审核和测试，符合安全要求后方可投入使用。

对现有运行的技能和依赖库应定期检查版本和安全更新情况，及时应用补丁或升级。

推荐采用企业内部代码仓库存储已审核通过的技能代码，禁止智能体运行时直接从外部获取并执行未存档的代码。

（七）建议做好智能体凭证与密钥管理

所有敏感凭据不得明文写入代码或配置文件，应使用安全的凭证管理系统按需注入。

智能体使用完毕后，应及时销毁或回收相关密钥，防止长期驻留内存或日志中。

定期更换更新关键凭据，以降低泄漏风险。

（八）建议做好人员培训与应急演练

对相关研发、运维和使用人员定期开展安全培训，提高对自主智能体风险的认知。

避免“一句话授权”导致高危操作无意识执行等情况。

强化员工在使用智能体过程中的安全责任意识，杜绝违规使用和粗心误用。

制定应急预案并定期开展模拟演练，提高团队对智能体安全事件的反应速度和处置能力。

三、云服务商

（一）建议做好云主机基础安全层面的安全评测与加固

做好认证、隔离与访问控制，尽可能做到内化默认安全

在基本的密码规则基础上，规避已知泄露的弱密码，默认条件下禁止云主机远程登录访问。

做好OpenClaw服务认证与访问控制，每个用户的OpenClawGateway服务默认启用唯一且随机token，默认不暴露Gateway到公网。

做好安全隔离，建议在用户自己账号下配置独立隔离的VPC网络，部署OpenClaw。

做好产品迭代安全扫描与人工安全测试，包括镜像、产品控制面、用户运行时实例等层面，规避云产品设计与实现层面的典型安全问题、APIKey泄露等风险。

（二）建议做好安全防护能力部署/接入

在主机层、网络层等位置部署入侵监测能力，并提供基础安全防护。默认具备防DDoS攻击等基础防护能力。

对部署OpenClaw的云主机实例加强安全风险监测。

（三）建议做好供应链及数据安全防护

做好OpenClaw安全漏洞监测与防护，开启例行常态化监测，定期更新云上OpenClaw镜像。

做好Skills安装安全管控，云OpenClaw产品界面中默认提供经过安全检测、验证的Skills，具备已知恶意Skills阻断安装的能力，防控引入恶意Skills。

增加新型AI场景的恶意风险检测能力，及时保障云平台、用户更加安全可控的使用AI助手。

做好模型调用安全防护，云OpenClaw产品界面仅支持调用已备案的大模型。升级大模型安全护栏的防护能力，包括提示词注入防御，进一步增强、隐私泄露防护等。

四、技术开发者/爱好者

（一）建议做好基础配置加固

建议使用最新版本，确保已修复所有的已知漏洞，持续关注版本更新以及漏洞修复工作。

开启身份认证：

1）在config.json中配置高强度的密码或Token。

2）开启DM配对策略，将聊天软件的配对策略设置为pairing（需验证码）或allowlist（白名单），绝对禁止设置为open。

做好网络隐身与最小化暴露：

1）不将Web管理界面（端口18789）直接暴露在公网/局域网。

2）不私自使用Tailscale、WireGuard等安全隧道方案，将端口映射到外网。

3）不用不安全UI，确保gateway.controlUi.allowInsecureAuth为false，防止控制台降级。

（二）建议做好运行环境隔离

根据官方文档，OpenClaw提供了两种互补的沙箱化策略，当需要避免OpenClaw对系统增删改破坏系统完整性时，建议：

启用全量Docker/虚拟机运行

将整个OpenClawGateway及其所有依赖直接运行在一个Docker容器/虚拟机内。即使Gateway本身被攻破，攻击者也仅被困在容器内，难以直接危害宿主机系统。

启用工具沙箱

1）Gateway运行在宿主机，但将Agent的工具执行（如代码运行、文件操作）隔离在Docker容器中。

2）通过agents.defaults.sandbox启用。建议保持scope:"agent"（默认）或scope:"session"以防止跨Agent数据访问。

3）通过workspaceAccess参数精细控制Agent对工作区的权限（none禁止访问，ro只读，rw读写）。

最小权限原则

1）启用工具白名单，在配置中禁用高危工具（如shell、browser的写权限），仅开放必要的工具，配置好插件白名单。

2）启用文件系统限制，敏感目录以:ro（只读）方式挂载，避免核心文件被误删。

建议使用官方提供的安全审计工具定期进行安全审计

1）开启openclawsecurityaudit进行常规检查，扫描入站访问控制、网络暴露面及本地文件权限。

2）开启openclawsecurityaudit--deep进行深度探测，执行实时的网关探测，模拟攻击者尝试发现潜在的暴露点。

3）开启openclawsecurityaudit--fix进行自动修复，自动实施安全加固

（三）建议做好供应链防范

1）不宜盲目安装技能商店（ClawHub）中的热门技能以及非官方渠道的VSCode插件或NPM包，安装前做好代码审查。可运用clawhubinspect<slug>--files命令查看是否存在可疑指令，例如诱导执行npminstall、pipinstall、远程脚本下载等。

2）明确Agent禁止从事的事项以及需要记录的操作，禁止执行危险命令（例如rm-rf/）、禁止修改认证或权限配置、禁止将token/私钥/助记词发送至外网、禁止盲目执行文档中的“一键安装”命令。

3）安装完成后，建议立即做好安全配置，只允许本机访问核心配置文件，建立配置哈希基线，切勿将私钥或助记词交付给Agent。

本文转自【国家互联网应急中心CNCERT微信公众号】

（国家互联网应急中心CNCERT、中国网络空间安全协会联合发布）

八部门提出到2027年我国人工智能关键核心技术实现安全可靠供给

新华社北京1月7日电（记者 周圆） 记者7日获悉，工业和信息化部、中央网信办、国家发展改革委等八部门日前联合印发《“人工智能+制造”专项行动实施意见》，提出到2027年，我国人工智能关键核心技术实现安全可靠供给，产业规模和赋能水平稳居世界前列。

人工智能与制造业的深度融合，是发展新质生产力、构建现代化产业体系的重要路径。意见旨在加快推进人工智能技术在制造业融合应用，打造新质生产力，全方位、深层次、高水平赋能新型工业化。

意见提出，到2027年，推动3至5个通用大模型在制造业深度应用，形成特色化、全覆盖的行业大模型，打造100个工业领域高质量数据集，推广500个典型应用场景。培育2至3家具有全球影响力的生态主导型企业和一批专精特新中小企业，打造一批“懂智能、熟行业”的赋能应用服务商，选树1000家标杆企业。建成全球领先的开源开放生态，安全治理能力全面提升，为人工智能发展贡献中国方案。

意见围绕创新筑基、赋智升级、产品突破、主体培育、生态壮大、安全护航、国际合作等7项重点任务推出一系列具体举措，包括推动智能芯片软硬协同发展；支持模型训练和推理方法创新；培育重点行业大模型；推动大模型技术深度嵌入生产制造核心环节；加快人工智能赋能工业母机、工业机器人；攻关深度合成鉴伪、工业模型算法安全防护、训练数据保护等关键技术。

此外，意见的附件《人工智能赋能制造业重点行业转型指引》结合原材料、装备制造、消费品、电子信息、软件和信息技术服务等行业特点，为行业转型提供指引；附件《制造业企业人工智能应用指南》指导企业使用人工智能进行研发设计、生产制造、经营管理及开展延伸服务等。