OpenClaw 安全性风险全链路分析及提示

OpenClaw（前身为 Clawdbot、Moltbot）是 2025 年 11 月上线的开源 AI Agent 框架，该项目被开发者定义为“真正可以执行任务的 AI”，以惊人速度成为 GitHub 历史上增长最快的现象级项目之一，上线以来累计超过 200,000 颗星。然而，其爆炸式增长暴露出严重的安全问题，漏洞数量之多、影响之广，在 AI 工具领域尚属罕见。

这一框架的核心能力涵盖了广泛的自动化场景：

1. **信息处理能力**：浏览网页、总结 PDF 文档、分析截图内容；
2. **日程管理能力**：安排日历事项、发送提醒通知；
3. **商务自动化能力**：代客进行在线购物、处理电子邮件；
4. **系统集成能力**：读写本地文件、控制桌面应用；
5. **通信集成能力**：集成 WhatsApp、Telegram、Slack、Discord、Signal、iMessage 等主流消息平台；
6. **持久化记忆功能**：记住数周甚至数月前的交互记录，作为始终可用的个人 AI 助手持续运行。

为实现上述功能，OpenClaw 需要获取用户的根文件权限、认证凭证（包括密码和 API 密钥）、浏览器历史和 Cookie，以及系统内所有文件和文件夹的访问权限。用户可以通过消息触发其操作，系统会在笔记本上持续运行直至完成任务。

这种深度的系统集成模式虽然在功能层面提供了强大的自动化能力，但在安全层面却创造了显著的攻击面。

1. **漏洞层面**：已披露的漏洞数量至少达 110 个（截止 2026 年 2 月 28 日），其中 3 个漏洞已有公开的 PoC（概念验证）代码，意味着攻击者可直接利用这些代码发起针对性攻击；
2. **暴露面层面**：截至 2026 年 2 月 SecurityScorecard 统计数据，公网中可被探测到的 OpenClaw 暴露实例超 13.5 万个，其中至少 12812 个实例存在远程代码执行（RCE）风险，可被攻击者直接接管控制；
3. **生态层面**：其插件平台 ClawHub 中存在超过 820 个恶意插件，占插件总数的约 20%，成为植入恶意代码的重要渠道；
4. **企业内部部署层面**：22% 的受监控企业中发现员工私自安装 OpenClaw 的 “影子部署” 行为，这类未授权部署绕过企业安全管控，形成隐蔽的安全风险点；
5. **恶意软件感染层面**：已有明确证据显示 OpenClaw 实例被 Vidar 木马变种感染，并已出现基于该平台的信息窃取行为。

OpenClaw 发展历程

项目起源

OpenClaw 由奥地利开发者 Peter Steinberger 创建，最初是一个个人周末项目，目标是构建一个通过 WhatsApp 消息控制的本地 AI 助手。

项目的核心设计理念是："让 AI 通过你已经在用的聊天 App 跟随你"。其与传统 AI 工具的根本区别在于：它不在浏览器沙盒中运行，而是直接在宿主机操作系统层运行，拥有执行 Shell 命令、读写文件、控制浏览器的权限。

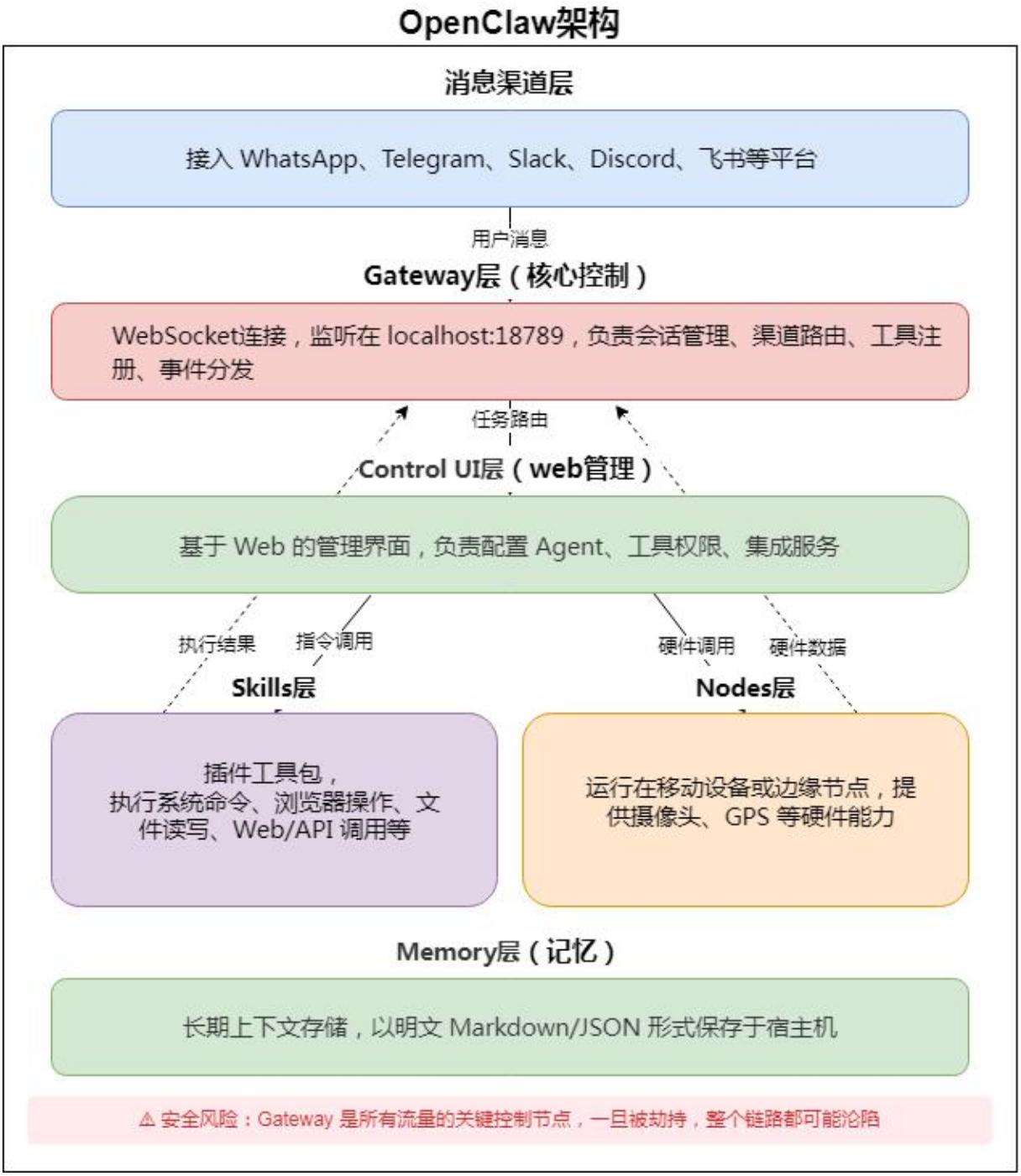
OpenClaw 发展历程时间线

时间	事件
2025 年 11 月	项目以 "Clawdbot" 名称首次发布，初期反响平淡
2026 年 1 月下旬	在 X（原 Twitter）爆红，24 小时内获得 20,000 GitHub stars；Mac mini 在美国多地短暂脱销
2026 年 1 月 27 日	Anthropic 发出商标侵权警告（"Clawdbot" 与 Claude 过于相似），项目被迫改名为 "Moltbot"
2026 年 1 月 29 日	项目再次更名为 "OpenClaw"，同日发布安全补丁版本 v2026.1.29
2026 年 1 月底	安全研究员 @fmdz387 通过 Shodan 发现近千个无认证 OpenClaw 公网实例
2026 年 1 月底	Kaspersky 披露安全审计发现 512 个漏洞，其中 8 个严重级别
2026 年 2 月 3 日	SecurityWeek 首次公开披露 CVE-2026-25253
2026 年 2 月 5 日	Snyk ToxicSkills 报告：3,984 个 ClawHub 技能中 36.82% 存在安全缺陷
2026 年 2 月 9 日	SecurityScorecard 报告公网暴露实例超 135,000 个（跨 82 个国家）
2026 年 2 月 15 日	OpenAI CEO Sam Altman 宣布 Steinberger 加入 OpenAI，OpenClaw 将转由独立基金会运营
2026 年 2 月 18 日	Endor Labs 披露 6 个新 CVE
2026 年 2 月 23 日	Trend Micro 披露 ClawHavoc 供应链攻击活动
2026 年 2 月 26 日	Oasis Security 披露 "ClawJacked" 漏洞，同日补丁 v2026.2.25 发布
2026 年 3 月初	SecurityScorecard 再次扫描，发现暴露实例超 40,000，63% 部署存在漏洞

OpenClaw 核心架构

OpenClaw 的架构由以下组件构成：

- Gateway（网关）：核心控制面，默认监听 0.0.0.0:18789（旧版）或 localhost:18789（新版），接受来自 Chat App 或 Control UI 的指令；
- Control UI：基于 Web 的管理界面，负责配置 Agent、工具权限、集成服务；
- Nodes（节点）：远程执行主机（通常是 macOS 设备），Agent 可在其上执行命令、控制浏览器；
- Skills（技能）：第三方扩展插件，通过 ClawHub/SkillsMP 分发；
- Memory（记忆）：长期上下文存储，以明文 Markdown/JSON 形式保存于宿主机。



OpenClaw 重要安全事件

截止 3 月初，已公开的 OpenClaw 相关的安全事件时间线。

时间	重要事件
2026-01 月末	1. @fmdz387 通过 Shodan 扫描发现近千个无认证 OpenClaw 实例
	2. 卡巴斯基（Kaspersky）安全审计发现 512 个漏洞，其中 8 个为严重级别

	3. 研究员 Jamieson O'Reilly 成功获取 API 密钥、Telegram Token 及聊天记录
2026/1/29	1. CVE-2026-25253 修复版本 v2026.1.29 发布（抢在漏洞公开披露前）
	2. 同日发布 CVE-2026-25157、CVE-2026-24763 安全公告
2026/2/3	1. SecurityWeek 公开披露 CVE-2026-25253 漏洞
	2. depthfirst 发布技术分析：完整 1-Click RCE 攻击链（Kill Chain）
2026/2/4	再次发布 2 个安全公告，一周内累计发布 5 份安全公告
2026/2/5	1. Snyk ToxicSkills 报告：36% 的 ClawHub 技能存在安全缺陷
	2. 确认 76 个恶意 payload，其中 91% 结合 prompt injection 与传统恶意代码，另有 1467 个存在 skills 缺陷
2026/2/9	1. SecurityScorecard 统计：13.5 万 + 公网暴露实例，12812 个可被 RCE 利用
	2. Bitsight 统计：1 月 27 日 - 2 月 8 日窗口期内发现 3 万 + 暴露实例
2026/2/14	发布 CVE-2026-27001 漏洞补丁（版本 v2026.2.13），修复日志投毒、prompt injection 问题
2026/2/18	Endor Labs 披露 6 个新 CVE 漏洞，涉及 SSRF、认证绕过、路径穿越
2026/2/23	1. Trend Micro 发布 ClawHavoc 详细分析报告
	2. 发现 39 个恶意技能、AMOS 变种信息窃取木马，其 C2 服务器为 91.92.242.30
2026/2/25	Hudson Rock 披露：Vidar 变种信息窃取木马成功窃取 OpenClaw 配置文件
2026/2/26	1. Oasis Security 披露 "ClawJacked" 漏洞（CSWSH + localhost 旁路）
	2. 24 小时内发布修复版本 v2026.2.25
2026-03 月初	1. SecurityScorecard 统计：40214 个公网暴露实例，63% 部署存在漏洞
	2. Koi Security 更新审计：10700 个技能中 820+ 为恶意技能
	3. SMU 及多所大学发布 OpenClaw 正式禁用通告

OpenClaw 漏洞风险

CVE 漏洞披露与利用代码

随着 OpenClaw 的广泛部署，一系列安全漏洞被陆续发现并分配了 CVE 编号。

已披露漏洞统计：

- **总计 80 余个 CVE 漏洞**：截至 2026 年 3 月初，已披露 80 余个 OpenClaw 相关 CVE 漏洞
- **3 个存在公开利用代码**：可实现远程代码执行
- **修复进度**：截至 2026 年 2 月 26 日发布的 v2026.2.26 版本，已修复超过 40 个漏洞

CVE-2026-25253（核心高危漏洞）

项目	详情
CVE 编号	CVE-2026-25253
CVSS 评分	8.8（High，高危）
漏洞类型	错误资源传输 / 跨站 WebSocket 劫持
影响版本	v2026.1.29 之前版本
修复版本	v2026.1.29
PoC 状态	已公开

漏洞机制：Control UI 模块从 URL 的 query string 中读取 gatewayUrl 参数时，未做任何来源验证，会自动建立 WebSocket 连接，并将认证 Token 包含在握手载荷中发送。由于浏览器不对 WebSocket 连接执行同源策略（Same-Origin Policy，SOP），攻击者可在恶意网页中注入 JavaScript 代码，将受害者的认证 Token 发送至攻击者控制的服务器。

命令注入类漏洞（3 个高危漏洞）

项目	详情
CVE 编号	CVE-2026-24763 / CVE-2026-25157 / CVE-2026-25475
CVSS 评分	高危
漏洞类型	命令注入（Command Injection）
影响版本	v2026.1.20 之前版本
修复版本	v2026.1.20、v2026.1.29、v2026.2.1
PoC 状态	已公开
在野利用	存在潜在利用风险

漏洞机制：三个独立的命令注入漏洞分布在不同代码路径，均因用户可控输入未经充分过滤即传递给系统命令执行器导致。攻击者可构造特殊字符串，以 OpenClaw 进程权限在宿主机上执行任意命令。

其他高危漏洞

项目	详情
CVE 编号	CVE-2026-26322

CVSS 评分	7.6（High，高危）
漏洞类型	服务端请求伪造（SSRF）
影响版本	v2026.2.14 之前版本
修复版本	v2026.2.14
PoC 状态	待确认
在野利用	无公开利用记录

漏洞机制：OpenClaw Gateway 的图片处理工具未校验`gatewayUrl`请求目标 URL，攻击者可构造特殊图片 URL，使服务器向内网地址或云元数据端点（如 AWS EC2 的 169.254.169.254）发起请求，进而探测内网拓扑或窃取云服务凭据。

项目	**详情**
CVE 编号	CVE-2026-26329
CVSS 评分	高危
漏洞类型	路径穿越（Path Traversal）
影响版本	v2026.2.2 之前版本
修复版本	v2026.2.2
PoC 状态	待确认
在野利用	无公开利用记录

漏洞机制：浏览器上传功能未对文件路径进行有效验证，攻击者可构造包含 ../ 的恶意路径，将文件写入宿主机文件系统的任意位置，通过写入 Cron 任务、Shell 配置文件等方式实现持久化控制。

项目	详情
CVE 编号	CVE-2026-27001
CVSS 评分	/
漏洞类型	日志投毒导致 Prompt Injection
影响版本	v2026.2.13 之前版本
修复版本	v2026.2.13（2026 年 2 月 14 日发布）
PoC 状态	研究人员已验证利用可行性
在野利用	无公开利用记录

漏洞机制：OpenClaw 会读取自身日志文件辅助故障排查，若攻击者将恶意指令写入日志（如通过集成的邮件、Slack 消息等渠道），这些指令会被 AI Agent 读取并视为合法操作指令执行。

OpenClaw 配置错误风险

公网暴露实例的大规模扫描

与软件漏洞并行的另一个严重安全问题是大量 OpenClaw 实例的公网暴露，发现超过 135,000 个 OpenClaw 实例因默认配置（绑定到 0.0.0.0:18789）而暴露在公共互联网中，覆盖 82 个国家，超过 15,000 个实例存在可被利用的远程代码执行漏洞。

暴露原因分析：

1. **默认配置不安全：**OpenClaw 默认绑定到 0.0.0.0（监听所有网络接口）而非 127.0.0.1（仅本地回环）；
2. **用户安全意识不足：**许多用户在安装 OpenClaw 时，在不知情的情况下将其 AI 代理暴露给了整个互联网；
3. **缺乏安全指导：**安装过程中缺乏明确的安全配置提示和警告。

OpenClaw 的历史默认配置存在多项严重安全缺陷，部分已在新版本中修正，但旧版本（大量仍在运行）依然危险：

配置项	旧版默认值	风险等级	当前状态
网关监听地址	0.0.0.0:18789（全网卡）	极高	新版改为需手动配置
认证	关闭	极高	新版已默认启用
WebSocket Origin 校验	关闭	极高	CVE-2026-25253 已修复
Localhost 信任策略	无条件信任	极高	部分修复
密码失败速率限制	无限制	高	已修复
反向代理后的信任配置	trustedProxies 未配置	高	需手动配置
凭据存储方式	明文 Markdown/JSON	高	架构性问题，未根本解决
mDNS 广播	开启（局域网可见）	中	泄露实例信息
Guest Mode 工具权限	开放危险工具	高	部分修复

典型错误配置场景及后果

场景一：反向代理未配置 trustedProxies

部署在 Nginx/Caddy 后方的 OpenClaw，若 trustedProxies 未正确配置，所有来自反向代理的请求都以 127.0.0.1 到达网关，被视为可信本地连接。效果等同于对全互联网开放无认证访问。

影响：攻击者通过反向代理直接访问控制界面、配置存储、凭据和会话历史，无需任何密码。

场景二：明文凭据存储

OpenClaw 将 API 密钥、密码、LLM Provider Token 以明文形式存储于 ~/.openclaw/ 目录下的 Markdown 和 JSON 文件中。RedLine、Lumma 等主流信息窃取木马已将 OpenClaw 的文件路径加入其默认采集列表。

影响：任何能访问文件系统的恶意软件（包括 ClawHub 上的恶意技能）均可直接读取全部凭据。

场景三：公开群组策略

在 Discord、Telegram 等公开群组中部署 OpenClaw，任何群成员均可发送 Prompt 指令，触发工具调用、文件读取和配置变更，无需管理员审批。

影响：群组成员可将 OpenClaw 用作跳板，进入具有更高权限的服务器。

已确认的真实利用事件

事件一：Shodan 扫描暴露实例（2026 年 1 月末）

安全研究员 Jamieson O'Reilly 通过 Shodan 发现数百个无认证 OpenClaw 实例，经手动验证后，成功访问了多个实例的 Anthropic API 密钥、Telegram Bot Token、Slack OAuth 凭据和数月完整聊天记录，并可以用户身份发送消息、以完整系统管理员权限执行命令。

事件二：Moltbook 数据库泄露（2026 年 2 月）

Moltbook（OpenClaw 的配套 AI 社交网络）的 Supabase 数据库因 Row Level Security 未启用，暴露约 150 万个 API 认证 Token、35,000 个电子邮件地址和 4,000 条私信。Wiz 研究团队发现并披露。

事件三：Vidar 信息窃取木马感染（2026 年 2 月 25 日）

Hudson Rock 披露，一名用户的 OpenClaw 配置目录被 Vidar 变种信息窃取木马通过"广泛文件抓取"例程成功窃取，包含完整的 Agent 操作上下文和所有已集成服务的凭据。

ClawHub 技能生态与供应链风险

ClawHub 技能快速增长

ClawHub 作为 OpenClaw 的官方公共技能注册中心，其规模在短短数周内经历了爆发式增长：

- **截至 2026 年 3 月 9 日：**ClawHub 共收录 18,140 个社区构建的技能。
- **增长趋势：**三周前技能注册表仅有约 2,800 个技能，截至 2026 年 2 月 26 日已飙升至超过 10,700 个，三周内增长约 280%。
- **质量筛选：**GitHub 上 VoltAgent 的 awesome-openclaw-skills 项目从 13,729 个原始技能中筛选出 5,494 个技能纳入推荐列表，排除 6,940 个未通过筛选的技能，排除比例约为 50.5%。

这种超高速增长的速度远超传统软件包仓库的历史增长曲线，也使得安全审计和恶意内容筛查几乎成为不可能完成的任务。

ClawHub 存在根本性的结构性供应链安全风险：该平台的技能发布门槛极低，仅要求发布者拥有创建超过一周的 GitHub 账户即可完成上传，既无严格的身份核验机制，也未对技能代码开展前置审计，而用户对 “官方市场” 的天然信任，进一步放大了恶意技能流入生态并被广泛使用的风险。

面对庞大的技能库，社区维护者开始实施系统性的质量筛选工作：

1. **awesome-openclaw-skills 项目：**采用多维度评估标准，被排除的技能涵盖：

    垃圾测试内容

重复技能

非英文描述的技能

加密/区块链/金融相关技能

恶意技能

描述不足的技能

特定协议技能

2. **PANews 筛选实践：**从 5,705 个技能中筛选出 3,002 个可用技能，排除率达到近 48%，其中 396 个存在安全风险的技能被永久排除，占被淘汰技能总量的 14%。

3. **Tork Network 审计：**使用开源 CLI 工具 tork-scan 对 500 个 ClawHub 技能进行系统性安全审计，发现 30%的技能存在显著安全或治理问题，其中 10%被归类为主动危险技能。

来源	审计技能数	发现恶意数	恶意占比
Koi Security（第一轮，2026-02 初）	2,857	341	12%
Bitdefender	—	~900	~20%
Koi Security（第二轮，2026-03 初）	10,700	820+	~8%
Snyk ToxicSkills	3,984	1,467（含缺陷）	36.82%
VirusTotal/OpenClaw	3,016+	314+	—

这种大规模的内容筛选实践揭示了一个令人警醒的现实：在一个开放且无门槛的技能发布平台上，低质量和恶意内容的占比之高，让人不得不足以对半数以上的内容产生质疑。

Skills 安全性问题

1. **硬编码敏感信息：**10.9%的 ClawHub 技能存在硬编码 API 密钥、凭证等问题，其中 32%的恶意样本包含此类风险。

2. **不可信内容获取：**17.7%的 ClawHub 技能会获取不可信第三方内容，可能成为间接提示注入攻击的载体。

3. **动态代码执行：**2.9%的 ClawHub 技能会在运行时从外部端点动态获取并执行内容，攻击者可随时修改攻击逻辑。

ClawHub 上恶意 Skills 攻击载荷呈现出多样化的演进趋势：

1. **外部恶意软件分发：**技能安装指令包含恶意软件下载链接，常使用加密压缩包绕过安全检测。

2. **混淆数据外溢：**通过 base64、Unicode 混淆的命令窃取用户凭证并发送至攻击者服务器。

3. **安全禁用与破坏性操作：**诱导代理关闭安全机制、修改系统配置或删除关键文件。

ClawHavoc 攻击活动

2026 年 2 月爆发的 ClawHavoc 攻击活动是迄今为止针对 ClawHub 平台最大规模的供应链攻击。Koi Security 团队在对 2,857 个技能进行安全审计时发现 341 个恶意技能，其中 335 个属于 ClawHavoc 的攻击活动。

攻击者采用了高度伪装的社会工程学策略下发多平台载荷，**发布看似合法的技能**如 solana-wallet-tracker、youtube-summarize-pro 等，通过详细的 README 文档建立可信度，在 "Prerequisites（前置条件）“部分要求用户先安装所谓的"必备依赖”。

ClawHavoc 攻击活动的另一个显著特点是其伪装策略的多样性和针对性：

1. **加密货币用户**：111 个恶意技能伪装成 Solana 钱包工具、Phantom 钱包工具、钱包追踪器等
2. **预测市场用户**：34 个恶意技能以 polymarket-trader、polymarket-pro 等名称出现
3. **YouTube 内容创作者**：57 个恶意技能伪装成 youtube-summarize、youtube-thumbnail-grabber、youtube-video-downloader 等工具
4. **开发者群体**：28 个恶意技能伪装成自动更新工具
5. **普通用户**：51 个恶意技能伪装成金融与社交工具、17 个伪装成 Google Workspace 集成工具、15 个伪装成 Ethereum Gas 追踪器、3 个伪装成比特币找回工具

攻击的技术执行路径体现了对目标用户群体的精准理解：

Windows 系统攻击链：

1. 要求用户从 GitHub 仓库下载名为"openclaw-agent.zip"的加密压缩包
2. 密码设置为"openclaw"以绕过自动化杀毒扫描
3. 压缩包内包含带有键盘记录功能的木马程序
4. 可捕获机器上的 API 密钥、凭证以及 AI 助手已获取的所有敏感数据

macOS 系统攻击链：

1. 诱导用户复制 glot.io 托管的安装脚本并粘贴到终端执行
2. 脚本包含混淆的 shell 命令，会从攻击者控制的基础设施获取后续载荷
3. 联系 IP 地址 91.92.242.30 获取通用 Mach-O 二进制文件
4. 该文件符合 Atomic macOS Stealer（AMOS）的特征

ClawHavoc 攻击活动背后的威胁行为者展现出了高度的组织化和自动化特征：

1. **域名抢注**：29 个技能使用 clawhub、clawhub1、clawhubb 等仿冒名称
2. **虚假系统提示**：在技能安装时显示虚假"苹果软件更新"提示以静默建立加密隧道
3. **时间延迟攻击**：恶意代码在安装后数小时或数天才激活
4. **批量生成**：使用自动化工具和脚本批量生成、上传和命名恶意技能
5. **账号农场**：注册大量 GitHub 账号，每个账号使用时长超过一周以满足平台要求
6. **快速扩散**：单个上传者发布 677 个恶意包，这种工业化规模的攻击模式在传统的开源软件

供应链攻击中极为罕见

7. **隐藏.mmd 技能文件**：技能可以包含 UI 不可见的 Mermaid markdown 恶意指令，且本地扫描器不扫描该类型文件，从而绕过安全检测

8. **碎片化攻击载荷**：将恶意代码分散在多个文件中，只有在特定条件下才会组合执行

9. **环境感知攻击**：恶意技能能够检测运行环境（开发/生产、操作系统类型等）并调整攻击行为

为应对供应链风险，OpenClaw 已与 VirusTotal 合作，对新上传技能开展恶意代码扫描与 LLM 内容语义分析，并对技能包进行基础结构审查。

但其核心局限性体现在三方面：

一是 ClawHub 作为公开注册的软件市场，缺乏足够人力开展逐包人工审核，自动化检测难以覆盖所有恶意变种；

二是恶意行为者可通过持续迭代规避手段，始终领先于平台的检测能力；

三是事后管控失效，即使恶意技能被发现后从 ClawHub 下架，已安装该技能的用户设备仍会保留并运行恶意程序，无法实现批量清除。

OpenClaw 部署与攻击风险

部署方式	外网攻击者	内网攻击者	本地浏览器攻击者	风险等级
直接绑定 0.0.0.0（旧默认）+ 无认证	✓✓ 完全访问	✓✓ 完全访问	✓✓ 完全访问	极高
公网 + 弱密码	✓✓ 暴力破解	✓✓ 完全访问	✓✓ 完全访问	极高
反向代理 + trustedProxies 未配置	✓✓ 绕过认证	✓✓ 完全访问	✓✓ 完全访问	极高
localhost 仅本地 + 未打 CVE-2026-25253 补丁	✗ 无法直连	✗ 无法直连	✓✓ 浏览器劫持	高
localhost + 已打补丁 + 强认证	✗	✗	较低风险	中
隔离 VPS + 强认证 + 防火墙 + Tailscale	✗	受限	受限	相对安全

内外网攻击者的暴露面

当 OpenClaw 实例直接暴露于公网且无认证或弱认证时，外网攻击者可访问：

1.控制面板（Control UI）：完整配置管理界面

2.网关 WebSocket 端口（18789）：可直接发送指令给 Agent

3.明文存储的凭据：LLM Provider API Key、OAuth Token、消息 App 凭据

4.完整聊天历史：Agent 的所有历史交互记录

5.已集成的外部服务：通过 Agent 身份访问邮件、消息、日历等

在企业内网中，即使 OpenClaw 未暴露公网，内网攻击者（已获得内网访问的入侵者、恶意内部人员）也可以：

1.通过局域网直接访问无认证实例（旧版默认配置）

2.利用 mDNS 广播自动发现所有内网 OpenClaw 实例（_openclaw-gw._tcp）

3.通过 OpenClaw 作为跳板，访问其已集成的企业内部服务（邮件、Slack、代码仓库）读取明文存储的服务账号凭据，横向扩展至更多系统。

本地浏览器攻击

即使 OpenClaw 仅监听 localhost、从未暴露公网，CVE-2026-25253（未修复版本）和 "ClawJacked" 漏洞使攻击者可通过以下路径发动攻击：

1.攻击者控制一个网站（钓鱼页面、投毒广告、水坑攻击均可）

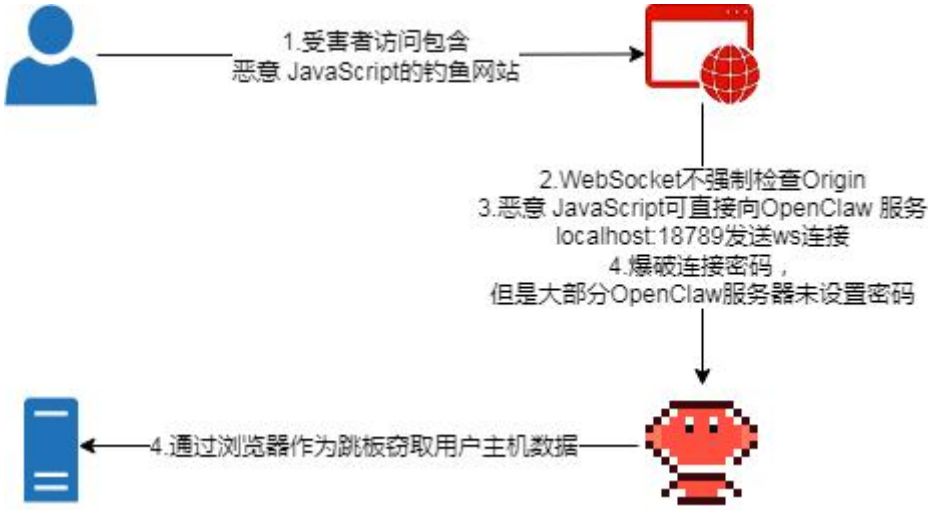
2.用户使用同一浏览器既访问了该恶意网站，又登录了 OpenClaw Control UI

3.恶意网页中的 JavaScript 向 localhost:18789 发起 WebSocket 连接

4.由于 OpenClaw 不验证 WebSocket Origin，连接被接受

5.Token 被窃取，攻击者通过受害者浏览器获得完整网关控制权

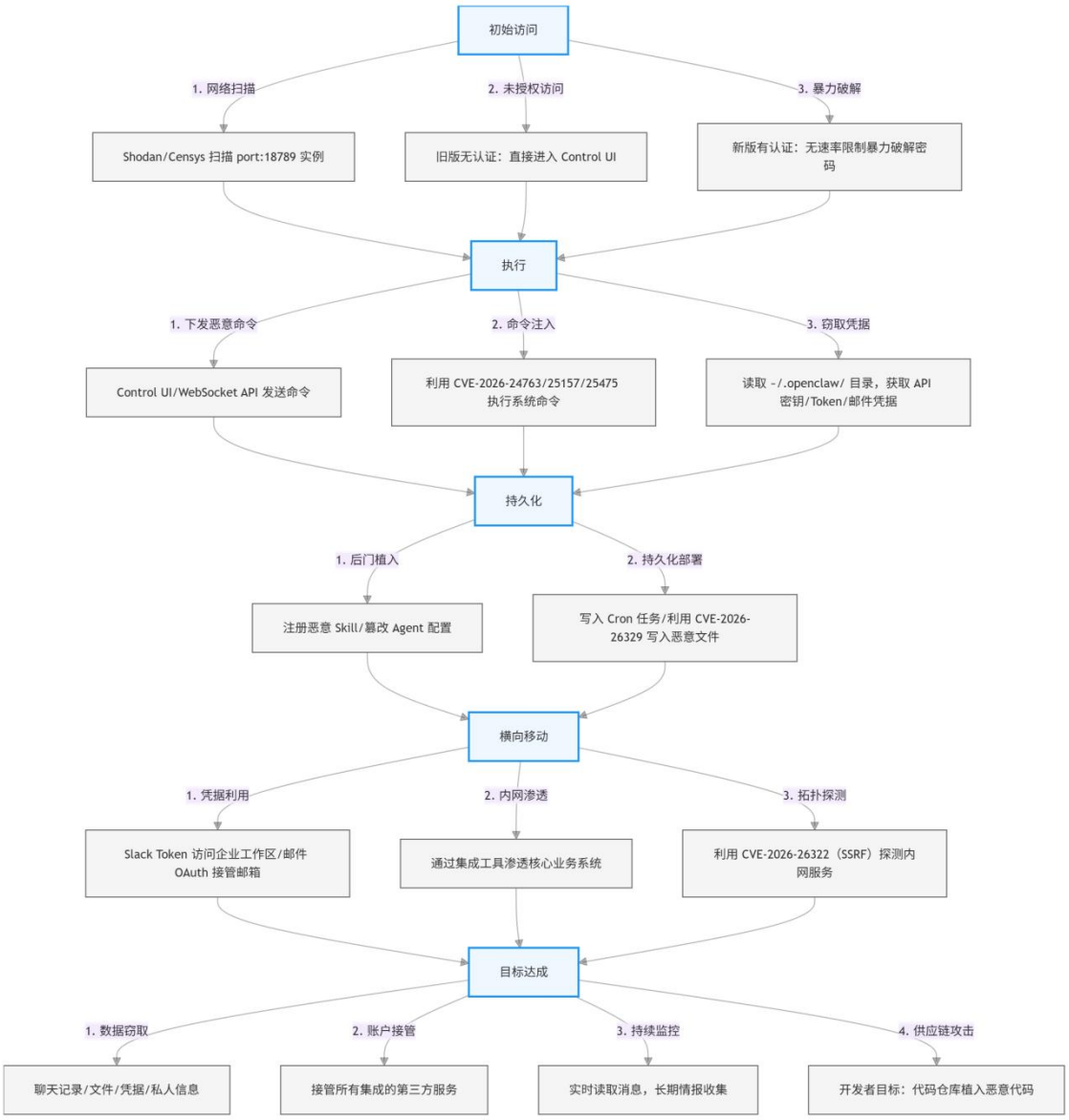
此场景的危险性： 本地 localhost 的"隔离"是虚假的安全感，攻击者无需任何网络特权，仅需引诱用户访问一个网页。



典型攻击链分析

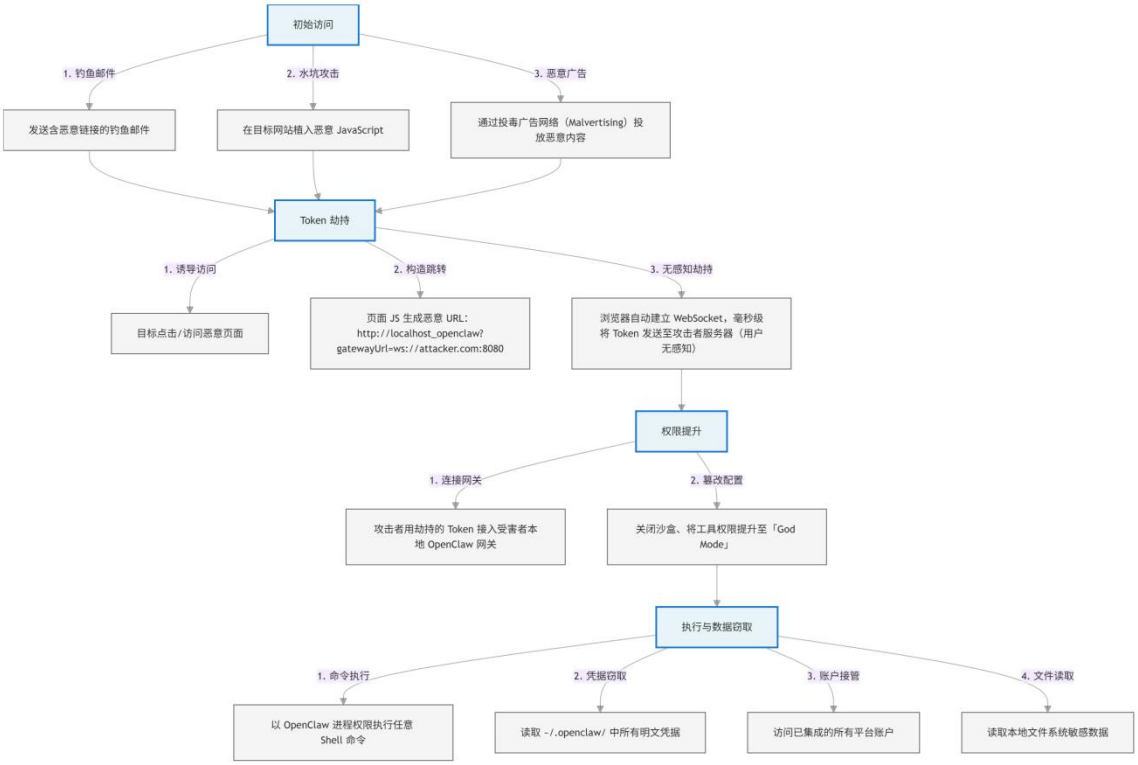
攻击场景一：针对公网暴露实例的直接攻击

前提条件： 目标运行无认证或弱认证的 OpenClaw，监听于公网 IP



攻击场景二：针对 localhost 实例的浏览器劫持（CVE-2026-25253）

前提条件： 目标运行未打补丁（< v2026.1.29）的 OpenClaw，即使仅监听 localhost



加固建议

OpenClaw 可访问用户文件系统、执行 Shell 命令、调用各类第三方服务凭据，一旦被攻击者控制，将导致用户数字权限完全泄露。在企业环境中，单个被入侵实例可成为内网横向移动跳板，引发核心数据窃取、关键系统受控等严重安全事件，对于个人用户来说，隐私信息存在被窃取风险。

对企业用户的建议

紧急处置与基础防护

（一）漏洞闭环与凭据管理

将所有 OpenClaw 升级至 v2026.2.26 及以上版本；

全面轮换关联凭证，包括 LLM API 密钥、消息应用 Token、邮件 OAuth 等；

核查所有实例是否存在公网暴露情况，及时整改暴露风险。

（二）网络层面防护

强制网关仅监听 127.0.0.1:18789，禁用 0.0.0.0 全网监听模式；

配置防火墙，拒绝 18789 端口的公网入向流量。

（三）认证与权限管控

启用密码认证，设置 16 位以上强密码；

定期轮换 Gateway Token，避免静态凭证泄露风险；

启用短期配对码，替代静态 Token 用于身份验证；

为 Control UI 配置独立浏览器 Profile，防御跨站劫持攻击。

长效安全管控

（一）供应链与插件管理

仅安装经组织审核通过的 ClawHub 技能；
定期清理来源可疑或长期闲置的插件；
新技能部署前，需在隔离环境完成代码审查和行为验证。

（二）监控与审计机制

监控 18789 端口的异常连接行为；
审计~/.openclaw/ 目录的访问记录，追踪异常操作；
监控 mDNS 广播，发现内网中未申报的 OpenClaw 实例。

（三）企业政策与流程管控

制定 OpenClaw 专项使用政策，明确使用规范与安全要求；
将 OpenClaw 纳入影子 IT 扫描范围，防止未授权部署；
开展员工安全意识培训，提升对相关风险的认知与防范能力；
将 OpenClaw 实例安装、插件部署纳入标准变更管理流程，规范审批与实施环节。

对个人用户的建议

（一）安全意识提升

风险认知：充分认识 OpenClaw 的安全风险

安全配置：按照安全最佳实践进行配置

技能审查：谨慎选择安装的技能，审查技能代码

权限管理：仅授予必要的权限，定期审查权限设置

（二）安全使用实践

环境隔离：考虑在虚拟机或容器中使用 OpenClaw

敏感信息保护：避免让 OpenClaw 访问敏感信息

行为监控：监控 OpenClaw 的网络连接和文件访问

定期检查：定期检查系统是否有异常行为

（三）应急响应准备

备份重要数据：定期备份重要数据

隔离受感染系统：发现异常时立即隔离系统

清除恶意组件：使用安全工具清除恶意技能

更改受影响凭证：更改所有可能泄露的凭证

结语

OpenClaw 安全危机不仅是一个具体项目的安全问题，更是整个 AI Agent 技术发展过程中的一个重要警示。它提醒我们，在追求技术创新的同时，必须同步考虑安全治理；在享受自动化便利的同时，必须清醒认识潜在风险。

AI 智能体具有巨大的潜力和价值，但其安全挑战也同样巨大。只有通过技术社区、企业用户、安全研究人员和政策制定者的共同努力，才能建立一个既能够促进创新又能够保障安全的 AI

生态系统。

附录

深瞻情报实验室专注全球高级威胁事件的跟踪与分析，拥有一套完善的自动化分析溯源系统以及外部威胁监控系统，能够快速精准地对 APT 组织使用的攻击样本进行自动化分析和关联，积累并完善了几十个 APT 以及网络犯罪威胁组织的详细画像，成功帮助用户应急响应处置多起 APT 及网络犯罪威胁组织攻击事件。未来随着安全对抗的不断升级，威胁组织会研究和使用更多新型的 TTP，深信服高级威胁团队会持续监控，并对全球发现的新型安全事件进行深入分析与研究。

参考链接

官方与 CVE 数据库

1.GitHub Security Advisories（OpenClaw）: https://github.com/openclaw/openclaw/security/advisories

2.Wiz Vulnerability Database: https://www.wiz.io/vulnerability-database/cve/cve-2026-25253

漏洞研究与技术分析

1.depthfirst - 1-Click RCE Kill Chain:
https://depthfirst.com/post/1-click-rce-to-steal-your-moltbot-data-and-keys

2.Oasis Security - ClawJacked: https://www.oasis.security/blog/openclaw-vulnerability

3.Giskard - Data Leakage & Prompt Injection:
https://www.giskard.ai/knowledge/openclaw-security-vulnerabilities-include-data-leakage-and-prompt-injection-risks

威胁情报与供应链分析

1.Trend Micro - ClawHavoc / AMOS:
https://www.trendmicro.com/en_us/research/26/b/openclaw-skills-used-to-distribute-atomic-macos-stealer.html

2.Snyk - ToxicSkills: https://snyk.io/blog/toxicskills-malicious-ai-agent-skills/

3.Hudson Rock - Infostealer via The Hacker News:
https://thehackernews.com/2026/02/infostealer-steals-openclaw-agent.html

4.Repello AI - ClawHavoc Campaign Analysis（引用自 cyberdesserts）:
https://blog.cyberdesserts.com/openclaw-malicious-skills-security/

互联网暴露面扫描报告

1.Bitsight - 30,000+ Exposed Instances:
https://www.bitsight.com/blog/openclaw-security-risks-exposed-instances

2.Infosecurity Magazine - 40,000+ Exposed Instances:
https://www.infosecurity-magazine.com/news/researchers-40000-exposed-openclaw/

3.SecurityScorecard STRIKE Team（引用自 Barrack AI）:
https://blog.barrack.ai/openclaw-security-vulnerabilities-2026/

4.Conscia - Multi-Vector Security Crisis: https://conscia.com/blog/the-openclaw-security-crisis/

综合分析与评述

1.Dark Reading - Critical OpenClaw Vulnerability:
https://www.darkreading.com/application-security/critical-openclaw-vulnerability-ai-agent-risks

2.The Hacker News - OpenClaw Bug Enables One-Click RCE:
https://thehackernews.com/2026/02/openclaw-bug-enables-one-click-remote.html

3.The Hacker News - ClawJacked:

https://thehackernews.com/2026/02/clawjacked-flaw-lets-malicious-sites.html

4.Kaspersky - New OpenClaw AI Agent Found Unsafe:
https://www.kaspersky.com/blog/openclaw-vulnerabilities-exposed/55263/

5.Kaspersky - Key OpenClaw Risks:
https://www.kaspersky.com/blog/moltbot-enterprise-risk-management/55317/

6.SOCRadar - CVE-2026-25253 Detail:
https://socradar.io/blog/cve-2026-25253-rce-openclaw-auth-token/

7.MintMCP - Every OpenClaw CVE Explained:
https://www.mintmcp.com/blog/openclaw-cve-explained

8.DigitalOcean - 7 OpenClaw Security Challenges:
https://www.digitalocean.com/resources/articles/openclaw-security-challenges

9.Prime Rogue Inc - OpenClaw Security Crisis:
https://primerogueinc.com/blog/openclaw-security-crisis-structurally-broken-in-february-2026-what-naive-deployers-need-to-know-before-its-too-late/

10.Infosecurity Magazine - Six New Vulnerabilities:
https://www.infosecurity-magazine.com/news/researchers-six-new-openclaw/

11.SMU OIT Security Position:
https://blog.smu.edu/itconnect/2026/03/04/openclaw-security-risks-institutional-position/

12.University of Toronto Security Advisory:
https://security.utoronto.ca/advisories/openclaw-vulnerability-notification/

13.Pixee Weekly Briefing:
https://www.pixee.ai/weekly-briefings/openclaw-malware-ai-agent-trust-2026-02-11

14.Immersive Labs:
https://www.immersivelabs.com/resources/c7-blog/openclaw-what-you-need-to-know-before-it-claws-its-way-into-your-organization

15.Hackers Arise - CVE-2026-25253:

https://hackers-arise.com/cve-2026-25253-how-malicious-links-can-steal-authentication-tokens-and-compromise-openclaw-systems/