小心AI代理變資安破口 資安署提醒導入OpenClaw應落實五項資安防護

強化資安韌性

發稿日期：115年3月25日
新聞聯絡人：李昱緯主任 0920-072628

具備自主執行能力的 AI 代理工具（AI Agent，如近期廣受關注的開源專案 OpenClaw，俗稱龍蝦）已被廣泛應用於日常自動化任務。數位發展部資安署表示，此類工具在提升作業效率的同時，因具備極高的系統權限與24小時自主運作特性，若未妥善設定防護機制，極易成為駭客入侵個人主機與企業網路的破口，可能導致使用者個資、機敏資訊如帳號密碼及金融資料等外洩，引發身分冒用與財產損失風險。提醒使用者導入OpenClaw時，應落實資安防護與環境隔離。

資安署提醒導入相關工具的單位與民眾，AI 代理的資安風險並非單一漏洞問題，而是涉及架構層面的系統性風險。例如近期Oasis Security的研究人員揭露的ClawJacked漏洞（CVE-2026-25253），攻擊者僅需誘導使用者瀏覽惡意網頁，就能在不觸發瀏覽器安全警報的情況下，對AI代理的管理員權限進行暴力破解（該漏洞已於2026年1月29號修補）。在評估AI代理工具的風險時，應特別注意以下幾類資安威脅情境：

1. 惡意指令可能出現在外部網頁： AI代理在瀏覽外部網頁或讀取真實世界的社群留言時，若內容中暗藏攻擊者預埋的惡意指令， AI代理有可能執行刪除檔案、竄改系統設定等危險操作。

2. 第三方技能包暗藏惡意程式： AI代理可以安裝名為「技能（Skill）」的擴充來執行訂票、製作影片等複雜任務。網路上已有開放平台供使用者分享自製的擴充包，攻擊者可將惡意行為指令寫入其中並偽裝成正常的Skill上架，一旦安裝即可能被植入後門或惡意程式。

3. 長時間運作導致安全守則遺失： AI 代理能處理的資訊量有限，長時間運作後會自動壓縮早期內容以騰出空間。在此過程中，原本設定好的安全規則與權限設置可能被刪減，導致AI代理逐漸「忘記」哪些事不該做，產生失控行為。

資安署建議，各界在評估與導入此類新型的AI代理工具時，應提高警覺並落實下列實務措施：

1.落實環境隔離： 避免將AI代理安裝於存放機密資料或日常作業的環境。應將其部署於全新、已格式化的獨立電腦，或是專屬的虛擬機（Virtual Machine）或容器（Container）中，以此進行有效的風險管控。

2.外部帳號權限最小化： 為AI代理註冊專屬的獨立帳號（包含專用電子郵件及社群平台帳號），避免將個人日常使用的帳號與密碼直接提供給AI代理。若AI代理必須登入外部服務，建議設定具有時效性的臨時授權憑證，時間一到權限即自動失效，避免日後因疏於管理而導致帳號遭竊。

3.設置人類審核機制： 針對高風險操作（如存取憑證、發送郵件或執行系統指令），應於系統設定中強制啟用人工審核，要求每次執行前必須經由人員手動確認方可放行。

4.親自審查Skill擴充套件： 在安裝任何第三方技能擴充套件前，應先對其內容說明與程式碼進行完整的安全掃描。若發現內容中有要求下載不明檔案、連線至不明網站等可疑行為，應立即停止安裝並向平台檢舉，同時企業用戶應於組織內部進行資安通報。

5.將安全守則寫入「長期記憶」： 定期審閱且備份AI的長期記憶檔。務必將重要的安全限制（例如：刪除郵件前必須經過人員同意）直接寫入「核心記憶檔案」（如：OpenClaw的MEMORY.md）中，確保每次運作時都會強制載入安全守則，避免因記憶壓縮而遺忘設定好的防護設定。

數位發展部資安署強調，AI代理技術能帶來顯著的創新效益，但須在「環境隔離」、「人工審核」的前提下進行測試與應用，方能兼顧數位發展與資訊安全。

小心AI代理變資安破口

資安署提醒

導入OpenClaw前
資安署提醒您落實下列防護

1. 落實環境隔離
2. 外部帳號權限最小化
3. 設置人類審核機制
4. 親自審查Skill擴充套件
5. 將安全守則寫入「長期記憶」

moda 數位發展部 Ministry of Digital Affairs

資安三大風險

1 惡意指令可能出現在外部網頁

2 第三方技能包暗藏惡意程式

3 長時間運作導致安全守則遺失

moda
數位發展部
Ministry of Digital Affairs

防護措施一

1 落實環境隔離

避免將AI代理安裝於存放機密資料或日常作業的環境。應將其部署於全新、已格式化的獨立電腦，或是專屬的虛擬機(Virtual Machine)或容器(Container)中，以此進行有效的風險管控。

moda
數位發展部
Ministry of Digital Affairs

防護措施二

2 外部帳號權限最小化

為AI代理註冊專屬的獨立帳號(包含專用電子郵件及社群平台帳號)，避免將個人日常使用的帳號與密碼直接提供給AI代理。若AI代理必須登入外部服務，建議設定具有時效性的臨時授權憑證，時間一到權限即自動失效，避免日後因疏於管理而導致帳號遭竊。

moda 數位發展部 Ministry of Digital Affairs

防護措施三

3 設置人類審核機制

針對高風險操作(如存取憑證、發送郵件或執行系統指令)，應於系統設定中強制啟用人工審核，要求每次執行前必須經由人員手動確認方可放行。

moda
數位發展部
Ministry of Digital Affairs

防護措施四

4 親自審查Skill擴充套件

在安裝任何第三方技能擴充套件前，應先對其內容說明與程式碼進行完整的安全掃描。若發現內容中有要求下載不明檔案、連線至不明網站等可疑行為，應立即停止安裝並向平台檢舉，同時企業用戶應於組織內部進行資安通報。

moda
數位發展部
Ministry of Digital Affairs

發布單位：資通安全署　　建立日期：2026-03-25　　更新日期：2026-03-25