Hacktron藉助Claude發現OpenAI漏洞 獲6500美元賞金

【看中國記者田淨心綜合報導】美國人工智能公司OpenAI近日發現，一家網絡安全初創公司Hacktron AI的研究人員，利用Anthropic的人工智慧模型Claude，發現並利用OpenAI系統中的安全漏洞，進入了OpenAI員工的ChatGPT帳戶，並進一步觸及公司的內部軟件代碼系統。研究人員隨後主動向OpenAI報告漏洞，OpenAI修復問題，並支付6500美元漏洞賞金。

這起事件發生在今年7月，但直到9月18日才由《華爾街日報》（The Wall Street Journal）首先披露。需要說明的是，這屬於漏洞賞金計畫下的安全測試，而不是研究人員未經授權發動的普通網絡攻擊。



2026年9月16日，Anthropic公司的AI應用「Claude」的圖標透過火焰映照在數字屏幕上。拍攝地點為英國倫敦。（Leon Neal/Getty Images）

Hacktron是一家甚麼公司？

Hacktron AI是一家總部位於舊金山的網絡安全初創公司，主要研究如何利用人工智能尋找軟件漏洞。

它的方向與傳統網絡安全公司有所不同：Hacktron希望讓AI代理參與漏洞發現、代碼分析和安全測試，幫助企業在軟件上線前發現可能被攻擊者利用的問題。

Hacktron公開的測試報告顯示，該公司的系統可以讓AI代理自動分析代碼、尋找漏洞，並由人類安全研究人員進行驗證。公司此前曾對Gumroad等軟件進行安全測試，並報告發現多個安全漏洞。

Hacktron成立時間不長，目前仍是一家規模較小的創業公司。Business Insider報導，該公司員工不足10人，由Zayne Zhang擔任聯合創始人兼CEO。

也正因為如此，這次事件受到關注：一家人數不到10人的網絡安全初創公司，利用另一家AI公司的模型，對OpenAI這樣的AI巨頭進行了安全測試。

兩個漏洞被串聯起來

這次事件並不是Claude直接「攻破」了OpenAI。

研究人員首先發現，OpenAI社區論壇使用的Discourse軟件，在處理某些圖片文件時存在安全問題。

簡單來說，用戶把特殊格式的圖片上傳到論壇時，伺服器會調用圖像處理程序處理文件。研究人員發現，其中使用的底層軟件存在漏洞。

接下來，研究人員又發現了OpenAI登錄系統中的另一個問題。

兩個問題被連接起來後，研究人員可以取得身分驗證信息，也就是系統用來確認「你是誰」的登錄令牌。

這些令牌中有一些可以用於進入ChatGPT帳戶，其中包括OpenAI員工的帳戶。

SecurityWeek對技術過程的報導指出，攻擊鏈涉及圖片處理軟件中的漏洞，以及OpenAI登錄系統的問題。

Claude在其中起了甚麼作用？

Hacktron研究人員使用Anthropic的Claude幫助編寫和修改漏洞利用代碼。

研究人員最初嘗試使用一個版本的Claude，但沒有成功。隨後，Anthropic發布了新的Claude Opus 5，研究人員再次嘗試後，成功完成了漏洞利用。

整個過程的核心仍然是研究人員，而不是Claude自行決定攻擊OpenAI。

換句話說，可以把Claude理解成一個能夠幫助安全人員寫代碼、分析問題的工具。

這也是這起事件受到關注的原因之一：過去一些複雜的漏洞利用需要具有豐富經驗的網絡安全專家花費大量時間編寫代碼，而AI模型可能幫助研究人員更快完成其中一部分工作。

進入內部系統 但沒有證據顯示源代碼被盜

獲得OpenAI員工帳戶後，研究人員進一步發現這些帳戶能夠連接到OpenAI內部使用的GitHub代碼系統。

這意味著他們獲得了進入OpenAI部分內部軟件環境的許可權。

但這裡需要特別區分「能夠訪問」和「偷走源代碼」。

Hacktron表示，研究人員沒有繼續查看內部源代碼，而是通過提交一個代碼修改請求，證明他們已經獲得了相應許可權，然後停止測試，並向OpenAI報告。

OpenAI隨後進行內部檢查。SecurityWeek報導，OpenAI發現的是對私有代碼庫元數據和代碼提交紀錄的有限讀取，以及研究人員提交的一次代碼修改請求；沒有公開證據顯示OpenAI的模型權重或完整源代碼被下載。

OpenAI如何處理？

Hacktron通過OpenAI的漏洞報告渠道提交了相關問題。

OpenAI隨後撤銷受到影響的登錄令牌和會話，並縮小社區登錄令牌的許可權。

OpenAI還向Hacktron支付了6500美元漏洞賞金。公司表示感謝研究人員報告問題，並確認已經採取措施修復漏洞。

需要注意的是，Hacktron測試OpenAI社區論壇所使用的第三方Discourse軟件，本身並不完全屬於OpenAI漏洞賞金計畫的範圍。研究人員後來也向Discourse報告了相關軟件漏洞，Discourse進行了修復。

這件事真正值得關注的地方

這起事件涉及三個不同的AI安全問題。

第一，AI公司自身的網路系統仍可能存在傳統軟件漏洞。

第二，AI模型可以幫助網路安全人員更快分析漏洞和編寫利用代碼。

第三，如果企業把AI代理連接到代碼庫、聊天軟體和其他內部系統，一旦賬戶許可權管理出現問題，AI輔助的攻擊可能擴大傳統漏洞的影響範圍。

但從目前公開資料看，這起事件屬於授權安全研究發現漏洞，OpenAI已經修複相關問題，也沒有公開證據顯示核心模型被盜。

警告期結束！ MTA施工區自動測速9月22日起開具罰單

（上接B2）據MTA公布的統計，目前測速設備已部署在約200個活躍的施工區域，期間共記錄到超過3.3萬輛汽車以至少高於限速10英里/小時的速度行駛，這部分違規車輛的平均超速幅度達到14.5英里/小時。更為怵目驚心的是，MTA記錄到了極端超速案例：有車輛在限速45英里/小時的施工區域內，車速飆升至111英里/小時；另有車輛在限速35英里/小時的區域內，以109英里/小時的危險速度飛馳。

累犯最高罰款100美元 違規不計入駕駛紀錄

根據MTA實施的罰款分級標準：

首次違規：罰款50美元。

第二次違規：若在首次違規後的18個月內再次違規，罰款增至75美元。

第三次及以上違規：在同一18個月的週期內，第三次及隨後的每次違規均罰款100美元。

MTA強調，逾期未繳納罰款可能產生額外的滯納費用。值得注意的是，此類由自動攝像設備捕捉的罰單屬於「非行駛違規」（non-moving infraction），即等同於停車罰單性質。它不會對車主的駕駛紀錄產生計分，也不會被通報至汽車保險公司從而導致保費上漲。所有被抓拍的違規圖像，在正式寄出罰單前，均須經過認證技術人員的二次審核確認。

在執法方式上，測速設備並非固定安裝，而是搭載於外觀無明顯標誌的車輛上，隱蔽停放於正在進行施工作業的工作區域內。測速僅在施工人員實際作業期間激活。為了符合法律規定並提醒駕駛者，MTA會在施工區域的前方設置明顯的提示標誌，告知駕駛者即將進入自動測速監控範圍。

目前，該執法計畫覆蓋了MTA在紐約市五大行政區內運營的9座關鍵橋梁及隧道設施，具體包括：

布朗士-白石大橋（Bronx-Whitestone Bridge）

跨灣退伍軍人紀念大橋（Cross Bay Veterans Memorial Bridge）

亨利·哈德遜大橋（Henry Hudson Bridge）

休·L·凱裡隧道（Hugh L.Carey Tunnel，原布碌侖-炮臺公園隧道）

海洋公園大道—吉爾·霍奇斯紀念大橋（Marine Parkway-Gil Hodges Memorial Bridge）

皇后區-中城隧道（Queens Midtown Tunnel）

羅伯特·F·肯尼迪大橋（Robert F.Kennedy Bridge，原三區大橋）

窄頸大橋（Throgs Neck Bridge）

維拉札諾海峽大橋（Verrazzano-Narrows Bridge）

自動施工區超速執法計畫最初於2021年獲得紐約州立法機構的授權，當時僅限於紐約州級公路（如紐約州高速公路管理局及州交通部管轄的路段）。2025年，州議會通過修正案擴大了該法律的適用範圍，將MTA運營的橋梁和隧道也納入其中，賦予了MTA在橋隧施工區進行自動測速執法的權力。

MTA明確指出，推行該計畫的核心目的是為了保護施工現場工人的生命安全，同時降低駕駛者在狹窄或臨時改變路況的施工區內發生碰撞的風險。MTA援引全國數據警示，2023年全美施工區域內共發生了近900起致命交通事故，並導致約4萬人受傷。隨著9月22日正式罰款階段的到來，MTA呼籲廣大紐約司機在途經橋隧施工區時，務必減速慢行，遵守臨時限速標誌，以避免面臨經濟處罰，並確保道路安全。

********** *****

大紐約
數字營銷
929-235-2332
• 專注各類型網站設計
• 視頻製作和推廣
• 網絡廣告投放和推廣
• 臉書和微信廣告投放和推廣

請支持
看中國
敬愛的讀者：
感謝您閱讀《看中國》。如果您認為獨立正直的媒體有存在的價值，請您考慮在這個過節的時候捐款支持《看中國》的運作。
《看中國》是非營利機構，多年來，工作人員無私奉獻，為讀者帶來豐富精彩、內涵深邃的文章。
您的愛心鼓勵就是《看中國》工作人員最大的動力。
款額不限，涓涓細流匯江河，點滴的呵護能鼓舞正的能量，您的愛心將支撐起健康運作的媒體。
《看中國》是在美國註冊的501(c)(3)非營利機構，因此，美國的捐款者可獲得減稅優惠。捐款時請務必註明捐款人姓名、地址、電郵信箱，以便給您寄送捐款證明。
支票：
支票抬頭請寫：KanZhongGuo Association Inc.
郵寄地址：KanZhongGuo Association Inc.
12 E 41st St #1302, New York, NY 10017

網上辦理安全可靠・服務體系專業權威 ・請認準官網：TUIDANG.ORG
辦理 退黨證明 要盡快
訪問官方網站 tuidang.org 自助式辦理
全球退党服务中心成立于2005年，是美国纽约501(C)(3)非营利组织 。自成立以來一直致力於幫助中國大陸以及國際社會的民眾認清中共的邪惡本質。至今，已有超過四亿中国人退出中共黨團隊組織。全球退黨服務中心頒發的【退黨證明】是您脫離中共的有效見證，普遍被國際社會認可，是您出國、移民必備的證明材料。
脫離中共
擁抱自由
出國移民
暢通無阻
全球退黨服務中心
GLOBAL SERVICE CENTER FOR QUITTING THE CHINESE COMMUNIST PARTY
20年來堅持不懈
幫助中国人擺脫中共