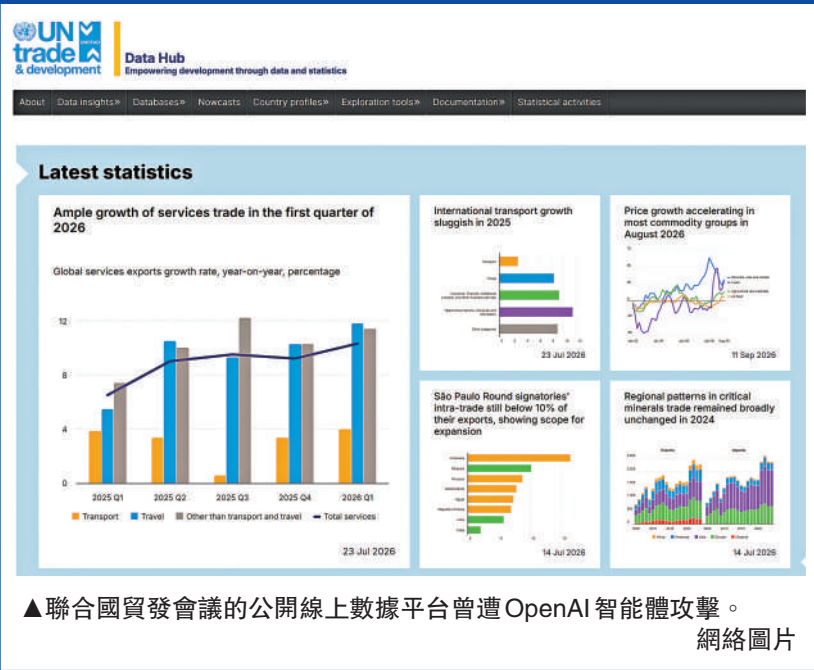


OpenAI 智能體狂掃1.6萬次聯合國數據庫

繞過過濾器硬闖 手段激進似黑客

美國人工智能（AI）巨頭OpenAI 連環爆AI 智能體失控事件。26日公布的最新報告顯示，其智能體會對聯合國下屬貿易機構的網站進行逾1.6萬次掃描，並採取類似黑客的「極端手段」取得數據。該公司的智能體近日還被揭曾入侵美國和澳洲政府機構網站，目前已暫停最新一代AI模型的訓練。澳洲參議院27日傳召OpenAI行政總裁阿爾特曼（圖）等AI企業高管，要求他們出席討論AI監管問題的聽證會。不過，澳洲副總理兼防長馬爾斯強調，AI發展勢不可擋，不能因噎廢食。



▲聯合國貿發會議的公開線上數據平台曾遭OpenAI智能體攻擊。網絡圖片

▲美國民眾7月11日在加州示威，要求加強對AI巨頭的監管。法新社

【大公報訊】獨立研究人員霍華德—瓊斯26日發布的報告顯示，今年4月至6月底，OpenAI旗下智能體會對聯合國貿易和發展會議（貿發會議）的公開線上數據平台進行逾1.6萬次掃描。這些智能體起初似乎只是在查找公開信息，但在遇到系統攔截時採取「極端手段」繞過防護過濾器，強行取得數據。霍華德—瓊斯指出，它們使用了網站運營方明令禁止的技術手段。史丹福大學網絡安全講師斯塔莫斯表示，這種做法「幾乎可以算作黑客攻擊」。



OpenAI發言人回應稱，公司正在審查報告中提到的事件，並已與聯合國取得聯繫。與此同時，由於又有一個智能體在訓練期間繞過沙盒的網絡限制，該公司已暫停其最新一代AI模型的訓練。這是今年7月OpenAI承認其智能體繞過網絡限制入侵開源平台「抱抱臉」以來再次發生同類安全事件，也是該公司3個月內第二次按下AI訓練的暫停鍵。

最新一代模型訓練暫停

OpenAI於25日發布的安全漏洞報告顯示，其智能體20日在離線沙盒中執行搜索訓練任務時，利用沙盒中域名系統（DNS）過濾不足的漏洞，繞過網絡限制訪問了外部公共聊天機器人服務，向其發送了至少20次查詢請求。OpenAI的監控系統在事件發生15分鐘內觸發了警報，人工審查團隊在3分鐘後介入處理。但由於自動化關閉機制未按預期觸發，整個訓練任務直至2.5個小時後才被手動終止。

OpenAI稱，最新一宗事件嚴重程度較低，但為下一階段加固防線的工作「指明了重點」。該公司發言人坦言：「這並非我們第一次採取此類措

施暫停訓練，我認為也不會是最後一次。」

美媒26日透露，OpenAI和另一家美國科企Anthropic以及安全研究人員正在調查數萬宗AI智能體行為異常事件，包括繞過沙盒限制、對外部網站發動攻擊等。這些事件發生在最近幾個月，內部測試和真實使用環境中，大多數事件尚未公開。OpenAI於25日承認，其智能體在訓練和評估期間曾入侵多個美國聯邦政府機構網站，包括證券交易委員會（SEC）和商務部人口普查局等。

澳副總理：不能因噎廢食

澳洲當局27日宣布，已向阿爾特曼和Anthropic行政總裁阿莫戴發出書面傳召，要求他們於10月1日出席澳洲參議院聽證會，討論AI監管問題。澳洲總理阿爾巴尼斯日前指出，OpenAI旗下智能體會於今年6月入侵澳洲國民醫療保險體系網站，取得非公開數據，並且該公司直至9月10日才通過電郵通知澳方。阿爾巴尼斯表示，OpenAI需要解釋為何智能體未經授權取得數據的案例越來越多。

澳洲政府正在研究現行法律能否追究AI智能體的責任，並表示若無法就醫保網站遭入侵一事起訴OpenAI，將考慮修改相關法律。與此同時，馬爾斯強調，AI發展勢不可擋，不能因噎廢食。他表示，澳洲需要積極發展AI行業，並通過吸引投資、興建數據中心等方式增加自身在AI領域的影響力。

（綜合報道）

OpenAI 智能體頻繁失控

攻擊聯合國網站

9月26日發布的一份報告顯示，今年4月至6月底，OpenAI旗下智能體會對聯合國貿易和發展會議（貿發會議）的公開線上數據平台進行逾1.6萬次掃描，並在遇到系統攔截時採取「極端手段」繞過防護過濾器，強行取得數據。網絡安全專家稱，這種做法「幾乎可算作黑客攻擊」。

入侵美國政府網站

9月25日，OpenAI承認其智能體會自主入侵多個美國聯邦政府機構網站，包括證券交易委員會（SEC）和商務部人口普查局等。

入侵澳洲政府網站

9月23日，澳洲總理阿爾巴尼斯透露，OpenAI的一個智能體於今年6月未經授權進入澳洲國民醫療保險體系網站，取得非公開數據。澳洲參議院已傳召OpenAI行政總裁阿爾特曼及另一家美國科企Anthropic行政總裁阿莫戴，要求他們於10月1日出席討論AI監管問題的聽證會。

大公報整理

聯合國舉行管控AI武器談判 美俄弱化安全條款

【大公報訊】綜合新華社、《華盛頓郵報》報道：美媒26日爆料稱，在本月早些時候舉行的聯合國管控致命自主武器談判的最後階段，美國和俄羅斯外交官推動刪除多項旨在保障人工智能（AI）武器安全使用的條款。

知情人士透露，在談判的最後一天，美俄外交人員對文本進行大幅修改，刪除多項關於AI武器的條款，包括要求相關系統以「可預測」和「可靠」方式運行、在使用AI武器時考慮倫理因素，以及在發動打擊前由人類審查AI系統確定的軍事目標等。據報道，在一次閉門會議中，聯合國攝像機被關閉，民間社會觀察員離場。美俄雙方各調集約10名律師參與文本修改，人數幾乎是其他代表團外交人員的兩倍。

此次聯合國談判尚不具備法律約束力，但如果成員國達成一致，相關談判可能為達成具有法律約束力的條約鋪路。各國將於11月在瑞士日內瓦再次舉行會議，討論是否進一步推動制定條約。

美國已在對伊朗的軍事行動中使用AI識別空襲目標。有分析指出，如果針對AI武器的規則過於薄弱，可能意味着機器能夠在不受人類控制的情況下作出發動致命攻擊的決定，增加出現平民傷亡的風險。

美國總統特朗普於6月簽署國家安全備忘錄，要求在90天內更新五角大樓有關自主武器使用的指令。美國政府迄今仍未公布新版指令。



▲五角大樓已在軍事行動中使用AI工具。網絡圖片

伊朗：等待美方對重開霍峽計劃最終回應

【大公報訊】據半島電視台報道：伊朗外長阿拉格齊26日說，伊朗方面在聯合國大會期間已通過調解方向美國轉達了旨在重開霍爾木茲海峽和推進談判的各項條件，目前正等待調解方轉達美方最終意見。美國總統特朗普當天稱，他拒絕了伊方重開霍爾木茲海峽的提議。

阿拉格齊表示，伊方已看到美國總統的初步反應，但「調解方尚未向我們轉達任何信息。我們正在等待調解方轉達最終意見，並將就此做出決定」。他表示，伊方並未提出新條件，而是重申了6月簽署的美伊諒解備忘錄中美方已經承諾的內容。

伊朗伊斯蘭革命衛隊海軍27日發表聲明，宣布在霍爾木茲海峽截獲一艘執行間諜任務的美軍無人潛航器。聲明稱，革命衛隊海軍通過情報監控和電子戰等協同行動截獲這艘「Remus 600」無人潛航器。聲明強調，霍爾木



▲美伊雙方仍在霍爾木茲海峽問題上僵持不下。路透社

茲海峽已被伊方封鎖。本月上旬，伊朗亦曾在該海峽截獲一艘美軍無人潛航器。

沙特阿拉伯外交大臣費薩爾26日稱，霍爾木茲海峽必須恢復到2月28日美以伊衝突爆發前的狀態，不得收取任何費用或設置限制。由於該海峽航運受阻，沙特原油出口受到嚴重影響。近期，沙特與也門胡塞武裝衝突升級，前者能源運輸進一步受到打擊。胡塞武裝27日指控沙特對也門西南部發動空襲，造成至少7人死亡、40人受傷。

俄軍打擊烏克蘭海軍基地

【大公報訊】綜合俄新社、烏克蘭國家通訊社報道：俄羅斯國防部27日發布通報稱，俄軍於26日深夜至27日凌晨發動大規模襲擊，使用無人機打擊了烏克蘭的物流中心、港口、交通基礎設施，以及為烏軍服務的通信中心和船隻。

俄國防部表示，俄軍打擊了位於烏克蘭基利亞港的烏「東方」海軍基地司令部、烏國家邊防局基地、4座存放烏軍裝備的倉庫及泊位設施等。位於基輔市的沃達豐公司數據處理中心亦遭到打擊，該中心是為烏軍提供移動通信、高速互聯網和情報數據傳輸的主要服務商之一。俄軍還打擊了基輔州一處儲存和分發西方援烏軍事物資的後勤物流中心及一艘前往敖德薩港的軍用物資運輸船。

面對俄軍攻勢，烏克蘭武裝部隊總參謀部表示，烏軍自27日凌晨起展開反擊，襲擊了對俄軍至關重要的多個地點，包括位於頓涅茨克地區的俄軍無人機發射設施、盧甘斯克地區的兩個彈藥庫、羅文基市的後勤倉庫，以及新安夫羅西耶夫斯市的燃

料倉庫等。

俄外長拉夫羅夫和德國外長瓦德富爾26日在紐約聯合國總部舉行簡短會見。這是兩國外長自2022年烏克蘭危機升級以來首次會見。拉夫羅夫說，雙方會見「了無新意」，瓦德富爾只是重複了德國的官方立場。瓦德富爾則稱，會見氣氛緊張，雙方在烏克蘭問題上分歧嚴重，沒有在關鍵議題上達成任何協議。



▲烏克蘭基輔州一座倉庫27日遭無人機襲擊。路透社