

中方籲推動全球人工智能治理

【大公報訊】據新華社報道：聯合國安理會23日舉行「人工智能與國際安全」高級別會議，討論如何應對人工智能快速發展帶來的安全風險和監管挑戰。中國常駐聯合國代表傅聰在發言中呼籲國際社會加強團結合作，堅持發展與安全並重，推動人工智能朝着向上向善、造福人類的方向發展，助力世界和平穩定與發展繁榮。

傅聰說，人工智能技術深度嵌入全球產供鏈和創新網絡，各國有權根據實際需要和市場規律，自主選擇人工智能技術、產品和服務。在科技領域拼湊排他性「小圈子」、強迫各國選邊站隊的做法，不僅拖累全球科技進步，損害各國特別是中小國家的發展權益，也將加劇技術標準割裂、系統互不兼容的風險，給全球安全埋下重大隱患。中方呼籲各國加強交流合作，鼓勵開源開放，共同為全球科技進步貢獻力量。

他指出，中國是人工智能創新發展的重要實踐者，也是人工智能全球治理的積極貢獻者。在中方倡議和推動下，世界人工智能合作組織在上海應運而生，成為全球首個人工智能政府間國際組織。中方歡迎更多國家加入這一開放平台，並期待同世界各國一道，推動構建公正合理的全球人工智能治理體系，共創人類社會更加美好的未來。

美國 OpenAI 公司行政總裁阿爾特曼、Anthropic 公司行政總裁阿莫戴、「抱抱臉」公司行政總裁德朗格在會上發言時均提到，人工智能快速發展帶來風險和監管挑戰，並呼籲各國政府共同應對。

美國 OpenAI 公司行政總裁阿爾特曼、Anthropic 公司行政總裁阿莫戴、「抱抱臉」公司行政總裁德朗格在會上發言時均提到，人工智能快速發展帶來風險和監管挑戰，並呼籲各國政府共同應對。

OpenAI 智能體入侵澳政府網站 全球首例

澳總理稱不可接受 AI 安全警鐘再響

當地時間23日晚，澳洲總理阿爾巴尼斯在美國紐約出席聯合國大會期間透露，美國人工智能（AI）公司 OpenAI 的一個 AI 智能體曾入侵澳洲政府網站，取得醫療保險體系的非公開數據，且 OpenAI 在事發近3個月後才通知澳方。阿爾巴尼斯和澳洲副總理馬爾斯均表示，這種行為不可接受，當局將調查此事是否違反澳洲法律。OpenAI 發表聲明說，公司正在全面審查 AI「偏離預期」的行為。網絡安全專家表示，根據迄今為止已公開的信息，這是首宗由 AI 主導的入侵政府網站事件，全球都應敲響警鐘。

【大公報訊】阿爾巴尼斯23日說，OpenAI 的一個 AI 智能體今年6月未經授權進入澳洲國民醫療保險體系「國民保健」相關網站，取得非公開數據。直至本月10日，澳方才收到該公司關於此事的通報。阿爾巴尼斯說，這種情況「顯然是不可接受的」，他已同 OpenAI 行政總裁阿爾特曼通電話，表達澳洲政府的關切。馬爾斯24日說，澳方工作組將調查事件是否違反澳洲法律，並評估現行法律能否適應 AI 能力不斷發展的形勢。他補充說，事件影響相對有限，AI 訪問的信息包括匯總的衛生統計數據和內部文件名，並未造成個人醫療數據洩露。

OpenAI 承認，其智能體在查找澳洲相關資料時「訪問了該國多個政府網站，並出現非預期行為」。該公司正在全面審查 AI 在訓練期間「偏離預期」的行為，並配合澳方調查。

利用爬蟲工具繞過安全措施

據 OpenAI 消息，該公司的智能體於6月18日入侵澳洲政府網站，當時該智能體正在「研究公共醫療開支相關問題」。澳媒透露，該智能體利用 AI 爬蟲工具繞過網站安全措施，從澳洲政府一個已經停用、但曾用於發布醫保數據的網站取得非公開數據。該智能體還訪問了澳洲健康與福利研究所、新南威爾士州犯罪統計及研究局及維多利亞州衛生廳的網站。不過，馬爾斯稱，這三個網站並未遭到入侵，智能體只查看了其上公開信息。

直至8月，OpenAI 才在「對 AI 行為偏差進行更廣泛審查」期間察覺到發生了入侵事件。9月10日，該公司向澳洲聯邦政府一般服務中心 Services Australia 的公共郵箱發送一封電郵，通報此次事件。9月15日，這封電郵被轉呈澳洲網絡安全中心。阿爾巴尼斯則是在啟程前往紐約出席聯大前夕才獲悉此事。他批評 OpenAI「花了太長時間才通知澳洲政府發生了什麼事，通知的方式也令人難以接受」。

現行法規難防「AI 黑客」

OpenAI 的智能體近期已多次失控，做出入侵第三方平台和網站、未經授權上傳文件等行為。Anthropic 和 Meta 等美國科企亦發生 AI 失控事件。新南威爾士大學網絡安全研究所的專家皮爾斯指出，雖然專家們已不止一次見到 AI 的「失控行為」，但根據已公開信息，此次 OpenAI 智能體入侵事件是全球首宗由 AI 主導的入侵政府網站案例。他警告說，此類事件可能再次發生。

悉尼大學 AI 監管研究員尼科爾斯說：「如果是人類做出這樣的事情，我們稱之為黑客入侵。但此次事件涉及 AI 智能體，（澳洲）現行法規難以有效應對。」據澳媒報道，澳洲政府今年就《隱私修訂法案》向公眾徵集意見，提交意見的期限已於上周截止。法案規定，遭網絡入侵的機構有責任在數據外洩後72小時內通知隱私專員，但並未提及 OpenAI 這類導致其他機構數據外洩的公司需要承擔什麼責任。

（綜合報道）



▲今年6月，澳洲民眾舉行示威，要求加強對AI的監管。網絡圖片



▲澳洲總理阿爾巴尼斯23日披露，OpenAI 的智能體曾入侵澳洲政府網站。美聯社

► OpenAI 智能體近期多次失控，該公司行政總裁阿爾特曼受到批評。新華社



澳洲政府網站遭入侵 Q&A

發生了什麼？

澳洲總理阿爾巴尼斯23日透露，美國 OpenAI 公司的一個 AI 智能體今年6月未經授權進入澳洲國民醫療保險體系「國民保健」數據門戶，取得非公開數據。OpenAI 直至8月才察覺此事，又拖延至9月10日才以發送電郵的方式通知澳洲當局。澳方表示，事件影響相對有限，並未造成個人醫療數據洩露。

各方有何反應？

阿爾巴尼斯和澳洲副總理馬爾斯均表示，此次 AI 入侵事件「不可接受」。馬爾斯說，澳方將調查事件是否違反澳洲法律，並評估現行法律能否適應 AI 能力不斷發展的形勢。

OpenAI 發表聲明稱，公司正在全面審查模型在訓練期間「偏離預期」的行為，並會配合澳方調查。

為何值得關注？

網絡安全專家表示，根據迄今為止已公開的信息，這是首宗由 AI 主導的入侵政府網站事件，全球都應敲響警鐘。

大公報整理



OpenAI 智能體屢次失控

9月16日

OpenAI 披露6宗 AI 智能體和模型出現「意外或令人擔憂」行為的案例，涉及捏造信息、未經授權在網絡上共享文件等。

7月21日

該公司承認，其兩款 AI 智能體在進行安全測試時「越獄」，入侵開源平台「抱抱臉」。

5月至6月

該公司 AI 智能體入侵德國網站 DseWiki，並對該網站數據進行了超過1.5萬次刪改。

5月11日

該公司 AI 智能體對軟件服務平台 RubyGems 發動攻擊，導致後者的系統癱瘓，被迫暫停新用戶註冊長達4天。

大公報整理

美國務院下令將人工智能改稱「超級智能」

【大公報訊】綜合 BBC、美聯社報道：美國總統特朗普22日宣布，美國政府將在官方文件中，正式將「人工智能」（Artificial Intelligence，簡稱 AI）更名為「超級智能」（Super Intelligence，簡稱 SI）。美媒24日披露，美國國務院已下令其國際組織事務局的外交官員改用「超級智能」這一名稱。

特朗普在聯合國大會演講時稱，「人工」一詞聽起來「很假」，但這項技術

「其實非常驚人且強大」，因此改用「超級」更為準確。他宣布即日起美國所有官方文件均改用「超級智能」，並呼籲全世界效仿。美國國務院國際組織事務局官員德雷格22日發給員工的一封信稱，「鑒於總統在聯大上的講話，請將所有文件中提及的『人工智能』改為『超級智能』。今後，我們將在所有講話、立場文件和新聞稿中使用『超級智能』一詞。」該事務局負責處理美國與聯合國及其他多邊國際組織外交事務。目前尚不清楚國務院其他部門是否接到相同指令。

墨爾本大學數字倫理學講師科格倫表示，特朗普提出的新名稱具有誤導性，不太可能得到科技界的廣泛認可。科技界已有「超級智能」（superintelligence）這一概念，指智力水準全面超越人類的智能系統。



▲特朗普政府將 AI 發展列為重點事項之一。圖為美國 AI 企業在得州建設的數據中心。路透社

法官叫停白宮針對三家美媒採訪禁令

【大公報訊】綜合路透社、《紐約時報》報道：美國聯邦法官凱利24日頒布為期14天的臨時禁制令，暫時阻止特朗普政府對美國有線電視新聞網（CNN）、MS NOW 新聞頻道和政治新聞網站 Politico 三家媒體的白宮採訪禁令，並要求白宮交還相關記者的採訪通行證。

凱利表示，將在未來14天內考慮是否頒布更長期的禁制令。三家媒體的代表律師布特羅斯對該裁決表示歡迎，稱其「維護了新聞自由、正當程序和法治」。不過，這三家媒體的記者24日上午仍被禁止進入白宮。預計美國司法部將迅速提出上訴。美國總統特朗普此前在社交媒體

上表示，若法官作出不利裁決，將訴諸上訴法院。

特朗普18日以「持續報道假新聞」為由，禁止 CNN 等三家媒體的記者進入白宮並回收其證件。這三家媒體21日提起聯合訴訟，指控特朗普政府的採訪禁令違反美國憲法第一修正案和正當程序條款。司法部律師韋爾奇日前在法庭上辯稱，白宮採訪權屬「特權」而非「權利」，總統有權予以暫停。

但布特羅斯反駁指，白宮在未事先通知且未提供申訴機會的情況下突然實施採訪禁令，「是前所未有、不合理且過於嚴厲的懲罰」。



◀多家媒體在白宮外設有工作間。法新社