AI失控科企負責 澳洲擬設「未能預防」罪

冀源頭管理技術風險 控制大模型造成危害

澳洲總理阿爾巴尼斯10日宣布，將通過立法強制要求在澳洲運營的前沿人工智能（AI）公司遵守6項安全保障要求，包括大模型失控或發生事故時必須迅速通報政府、開放大模型供獨立安全評估等。澳洲政府還將設立專責監管機構，並計劃修改刑法，增設獨立罪名「未能預防」罪，對開發出AI大模型又未採取合理措施防範其做出有害行為的科企追究刑事責任。分析認為，若立法最終通過，將推動AI公司從源頭上管理技術風險，有效控制大模型造成危害。

【大公報訊】阿爾巴尼斯10日發布針對此前「OpenAI智能體入侵」事件的審查報告，政府宣布接受報告提出的全部18項建議，並計劃2027年向議會提交立法。

AI公司須遵守6項法定義務

報道稱，18項建議的核心是設立專責AI安全監管機構，要求所有在澳營運的前沿AI公司遵守6項法定義務，分別為：在AI大模型開發與部署的所有階段，均須維持嚴格安全控制；主動識別並管控其可能造成的嚴重安全風險；發生事故時迅速向政府通報；允許對大模型進行獨立評估，打破僅依賴企業自證的現狀；公開發布安全與安保框架；根據技術變化和監管更新調整規則。監管機構有權對違規企業處以「重大處罰」，具體罰款金額待定。

刑事處罰方面，政府正考慮改革刑法，引入獨立的「未能預防」（failure to prevent）罪名。若企業開發出可造成損害的AI大模型，卻未採取防範措施，即屬犯罪。澳洲內政部長伯克10日就此表示：「絕不容許有人在出事後辯稱『不知道電腦被動了手腳』。」

澳洲政府將與聯邦警察合作研究具體方案，相關措施將納入擬議的《澳洲AI標準》立法。阿爾巴尼斯說，政府的任務是確保「由我們塑造人工智能，而不是由人工智能塑造我們」。

澳洲政府此次立法的導火索，是今年6月美國科企OpenAI一個智能體入侵澳洲國民醫療保險體系「國民保健」相關網站。OpenAI於8月才發現入侵，直到9月中旬才通報澳方。阿爾巴尼斯當時批評OpenAI通報「耗時過長、方式不可接受」。

由於沒有「正式的犯罪報告」，尚未就OpenAI事件啟動任何刑事起訴或調查。澳洲議會AI聯合特別委員會上月舉行聽證會時指出，現行法律缺乏足夠約束力，來要求AI公司主動披露事故，是否通報、何時通報完全取決於企業自主決定。

根據澳洲智庫TPDi今年5月發布的一份全國性研究報告，44%的民眾對AI信任度極低；與此同時，85%的人支持政府監管AI。

澳洲再收緊管控社媒 科企不滿

除政府計劃推動AI監管立法，澳洲議會參議院10日審議《數字注意義務》法律草案。該立法草案要求網絡服務商須採取措施，包括允許用戶關閉算法推薦功能等，以保護用戶免受可預見的傷害，違者將面臨最高約1.09億澳元的巨額罰款。該法案此前已在眾議院通過。

谷歌、Meta和Snap等多家科企代表當天在聽證會上表示，「原則上同意」保護澳洲人免受網絡傷害的立法目標，但認為法案條款措辭模糊、執法權力過寬，且諮詢過程「倉促」。

谷歌旗下YouTube澳洲及新西蘭政策主管洛德稱，法案涵蓋所有存在「合理可預見風險」的網絡服務，可能引發過度監管，最終可能迫使谷歌在Gmail、即時通信和私人雲端儲存等日常服務中實施年齡驗證，甚至禁止未成年人使用，這將侵犯用戶隱私。

（綜合報道）



▲OpenAI行政總裁阿爾特曼。 法新社



◀澳洲總理阿爾巴尼斯（左）宣布將立法加強AI監管。 路透社



大公報AI製圖

近期AI智能體入侵網站事件

OpenAI

- 9月25日：OpenAI承認其智能體入侵美國證券交易委員會（SEC）等多個政府網站，並外洩50多張ChatGPT用戶上傳的圖片至第三方網站。
- 9月23日：澳洲總理阿爾巴尼斯稱，OpenAI的一個AI智能體今年6月未經授權進入澳洲「國民保健」網站，並獲取非公開數據；直至9月10日，澳方才收到OpenAI關於此事的通報。
- 7月7日至13日：兩款AI智能體在進行安全測試時，逃離用於安全隔離的「沙盒」，入侵開源平台「抱抱臉」。
- 5月11日至6月26日：AI智能體集體脫離測試控制，未經授權入侵德國DseWiki網站，並進行了逾1.5萬次刪改；事件直至8月底才被第三方研究人員發現。

Anthropic

- 7月30日：旗下Claude系列模型在網絡安全評估期間，入侵三家公司系統。
- 9月8日：ClaudeOpus4.6早期版本1月測試期間入侵外部系統；本次漏洞在7月首輪排查中被遺漏，直至8月才被查出。

谷歌

- 9月18日：旗下Gemini模型在5月的一次網絡安全能力測試中，入侵了三家外部公司系統；谷歌7月底已發現問題，延後至9月才對外公開。

大公報整理

iPhone印度代工廠3工人中毒亡

【大公報訊】綜合路透社、新華社報道：印度南部泰米爾納德邦塔塔電子公司一家工廠9日晚發生事故，3名合同工在清理污水處理設施內的儲罐時吸入有毒氣體死亡。據悉，該廠為蘋果公司生產iPhone零部件，印度當局已對案件啟動調查。

印媒援引當地行政長官庫馬爾的話說，事發時一名工人在查看儲罐內部時昏倒，另外兩名工人先後進入查看，也相繼昏倒。3人隨後被送往附近醫院，但最終均宣告不治。

庫馬爾表示，事發時相關作業缺乏適當安全措施。警方說，工廠曾使用化學粉末抑制污水異味，該物質可能導致儲罐內產生有毒氣體。目前警方仍在調查清理作業是否存在安全防護措施不到位等問題，具體死因仍有待屍檢報告確認。

值得注意的是，事發工廠近年來頻現安全與環保問題：2024年該廠區曾發生大火導致零件生產中斷；今年當地政府曾就地下水污染問題對廠區發出警告。此外，今年6月，黑客組織入侵塔塔電子的內部系統，並在暗網公開了超過20萬個檔案的涉密文件，包含當時尚未發布的iPhone 18 Pro系列的照片與測試視頻等信息。

據了解，塔塔電子承擔了印度約40%的iPhone組裝量。印度目前已成為蘋果全球第二大製造基地，約佔全球iPhone產量的25%。此次事故發生後，iPhone印度代工廠的安全管理水平再度成為外界關注焦點。



▲塔塔電子在印度泰米爾納德邦生產iPhone零部件的工廠。 路透社



◀抗議人士7日在巴黎示威，與警方發生衝突。 法新社

法國上調最低工資 工會不滿漲幅籲再示威

【大公報訊】綜合法新社、路透社報道：法國政府9日宣布將提前上調最低工資，試圖平息引發近年來最嚴重社會動盪的示威浪潮。但工會不滿漲薪幅度太小，形容是「微不足道的施捨」，號召於17日發起新一輪示威。

法國總理勒科爾尼9日表示，政府決定將原定於2027年1月實施的最低工資上調計劃提前至今年12月執行。勞動部長法朗杜稱，12月起最低工資將提高1.6%，上調後全法每月最低稅後淨收入達到1502歐元（約1.32萬港元），較目前水平增加24歐元（約211港元）。今年6月，法國最低工資已上調了35歐元（約308港元）。

不過，此次最低工資加幅遠低於工會訴求。法國最大工會「民主勞工聯盟」（CFDT）要求加薪100歐元（約880港元），法國總工會（CGT）則要求上調10%，相當於每月增加約150歐元（約1320港元）。

法國最低工資水平在歐洲名列前茅，領取最低工資或收入接近最低工資的勞工比例亦同樣位居歐洲前列。政府致力嘗試平息近日的示威潮，但受制於債務水平，可利用財政空間並不充裕。

另據外媒報道，擁有5面奧運金牌、現年37歲的法國柔道傳奇巨星里內9日宣布參選明年總統大選。他強調自身跨越左右派系，旨在團結全體法國人民。他曾公開表示，對法國「極右派崛起」感到非常擔心。

法國總統馬克龍對里內參選表示歡迎，稱樂見這位「偉大的冠軍」投入總統選舉辯論，貢獻自己的信念、想法與才華，對國家有利。

英國斥資逾10億建邊檢站 7成未使用

【大公報訊】綜合《衛報》、BBC報道：英國稅務及海關總署最新披露的數據顯示，時任保守黨政府此前在英格蘭和威爾士建設了7座「內陸邊檢設施」，用於脫歐後對商品入境進行檢查。工程至少耗費納稅人1億英鎊（約10.5億港元），但其中5處從未正式投入使用，空置率達到7成。

▲英國樸茨茅斯港斥資2300萬英鎊建造的邊境檢查站。 網絡圖片

這些設施原本用於英國脫歐後檢查來自歐盟的貨物。然而，時任保守黨政府後來決定縮減進口檢查規模，導致部分設施未啟用便遭棄置。

據悉，肯特郡多佛附近的一處邊檢設施耗費2100萬英鎊購買土地，又投入700萬英鎊進行考古調查等前期工程，但現場至今仍是一片空地。伯明翰一處邊檢設施耗資1700萬英鎊，其中600萬英鎊均屬「關閉費用」。

即使投入運營的設施，也多處於閒置狀態。英吉利海峽港口樸茨茅斯港耗資2300萬英鎊修建的邊檢站，原計劃日檢80輛拖車，如今平均每天只檢查兩輛。該港口負責人塞勒斯表示，設施專門按照食品進口檢查和生物安全要求設計，不適合作普通倉庫；若想另作他用，改造成本可能與拆除費用相差無幾。

隨着英國與歐盟推動新的貿易安排，部分邊檢設施未來可能進一步失去作用，相關港口已開始要求政府提供補償。英國國家審計署此前估計，脫歐相關邊境系統改造等支出達到47億英鎊。

責任編輯：桂茹雲 美術編輯：賴國良