

OpenAI 主管辭職警告：須嚴控 AI 風險

「試錯時代」已過 應借鑒核安全措施

近期美國人工智能（AI）安全事件頻發，引發業界對AI「失控」風險的深切擔憂。美國人工智能巨頭OpenAI負責AI安全的工程師當地時間3日撰文宣布辭職，並批評OpenAI內部風險意識嚴重不足，成為又一位因與公司安全理念不合離職的前沿研究人員。該工程師警告，AI的「試錯時代」已過，呼籲業界迅速行動，效仿核工業與航空業，嚴格設置多重安全防護措施，以避免災難發生。

【大公報訊】曾在OpenAI安全團隊領導政策規劃及透明度方面工作的戴維·魯濱遜當地時間3日在美國《大西洋》月刊發表一篇題為《我辭去OpenAI的工作，因為企業文化已崩壞》的署名文章，成為又一位因安全理念分歧而選擇離職的前沿AI研究人員。

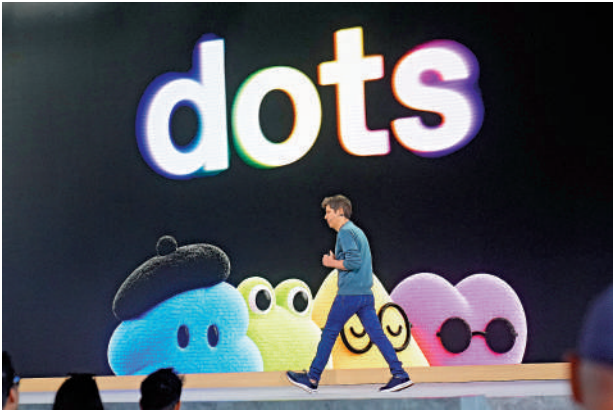
魯濱遜曾在OpenAI任職3年多，是該公司資深員工之一。他在文中透露，自己曾參與起草OpenAI的風險應對框架，並負責監督12次前沿模型發布的安全評估。他直言：「OpenAI不斷推出新產品，但在審慎管控方面，遠未達到我認為應有的水平。」

今年夏季，OpenAI和另一個AI巨頭Anthropic先後披露，其AI模型在網絡攻防能力測試中突破測試環境進入公共互聯網，並未經授權訪問其他機構的系統，引發外界對AI「失控」風險的擔憂。路透社與益普索集團9月22日發布的民調結果顯示，四分之三的美國人擔心，AI企業在防範人工智能對社會造成嚴重危害方面做得遠遠不夠。OpenAI行政總裁阿爾特曼隨後加入呼籲強化AI安全的行列，OpenAI也於9月28日以安全疑慮為由喊停新模型發布，但翌日便高調推出多款新產品，被外界詬病。

OpenAI 先發布產品再上安全護欄

魯濱遜提到，OpenAI在模型發布方面推行「迭代部署」原則：先發布產品，觀察產品是否出現問題，再補上安全護欄。但他警告，隨着AI模型能力呈指數級增強，這種試錯模式極度危險，一旦超強AI模型出現失控，人類可能連「從錯誤中學習」和「重新迭代」的機會都沒有。魯濱遜表示，面對可能造成嚴重或不可逆轉傷害的系統，業界必須做到「首次發布即接近完美」。

魯濱遜強調，「試錯時代已過」，前沿AI實驗室應該效仿核工業與航空業，設置多重安全防護措施並進行謹慎規劃，「如此一來，即便不可避免地出現人為失誤，也不至於釀成災難。」魯濱遜認為，AI公司也需要更多地依賴其他領域已有的安全專業知識。他舉例，在核電站或繁忙的



▲OpenAI行政總裁阿爾特曼9月29日在開發者大會上宣布推出新產品。美聯社

機場，即便設備損壞或操作員誤按按鈕，嚴密的縱深防禦體系也能確保不致發生堆芯熔毀或空難；相較之下，目前的AI公司在部署可能帶來毀滅性後果的技術時，嚴重缺乏基本的防護設計與嚴謹規劃。

此前從Anthropic辭職的研究人員考克森亦指出，AI巨頭在開發AI時未承擔應有責任，AI企業對爭奪技術領先地位的競爭考量正凌駕於安全考量之上，這些企業在「拿我們所有人的生命做賭注」。

白宮 AI 沙皇反對放緩發展

圍繞AI監管議題，美國國內分歧嚴重。雖然多名AI研究人員呼籲加強監管，但美國總統特朗普明確拒絕對AI設置安全「護欄」。美國總統特朗普當地時間4日確認，將任命國家情報總監克萊頓兼任白宮新成立的AI特別工作組負責人，即外界所稱的「AI沙皇」。

特朗普4日在社媒平台發文稱，在克萊頓的領導下，「超級智能部隊」（Super Intelligence Force）將確保美國在AI發展領域「持續領先全球」。

美媒3日報道，克萊頓將領導「超級智能部隊」在120天內提交一份有關人工智能風險和機遇的報告。工作組成員包括副總統萬斯、國防部長海格塞斯、財政部長貝森特和白宮辦公廳主任懷爾斯。克萊頓在報道中說，成立工作組旨在保護公眾利益，以及保持美國在人工智能領域的領先地位。他也曾表示，「超級智能是國家安全問題」，但也明確反對美國企業暫停或放慢AI模型的研發腳步。

（綜合報道）



▲示威者在OpenAI年度開發者大會場外用假電鋸切割一隻機器狗，以示抗議。美聯社

美國各界 近期就AI監管表態

反對加強監管

支持加強監管

●美國總統特朗普一直反對更嚴格的AI監管。他9月29日與美國6家科技企業代表在白宮簽署自願性質的安全協議，主張業界自我監管。

●美國財政部長貝森特3日表態稱AI「末日論」純屬危言聳聽且毫無助益，他呼籲AI行業進行自我監管並拿出解決方案。

●美國國家情報總監兼白宮AI事務顧問傑伊·克萊頓明確反對美國企業暫停或放慢AI模型研發。

●包括英偉達CEO黃仁勳等在內的科技巨頭據報在9月29日的白宮會面中，當面指出Anthropic行政總裁阿莫戴對AI危險性的警告過於極端。

▲戴維·魯濱遜近期因與OpenAI安全理念不合而辭職。網絡圖片

●OpenAI前員工魯濱遜3日加入要求加強AI安全的呼籲行列。他認為AI行業應效仿核工業與航空業，設置多重安全防護措施，以避免災難發生。

●包括Anthropic行政總裁阿莫戴、OpenAI行政總裁阿爾特曼等在內的多位科技巨頭都呼籲放緩AI開發、加強監督。

●美國共和黨參議員蒂利斯公開反駁特朗普將AI威脅稱為「騙局」的說法，堅持必須為AI設置「護欄」。

大公報AI製圖

韓國金融業個資外洩 AI疑為幫兇

【大公報訊】綜合路透社、韓聯社報道：韓國5大商業銀行及其他金融機構此前相繼遭黑客疑似利用人工智能（AI）工具輔助進行攻擊，導致信息外洩。韓國總統府4日表示，總統李在明已下令徹查，並採取應對措施。

據報道，韓國新韓銀行9月28日被黑客入侵，洩漏逾2.5萬個客戶資料。此後，KB國民銀

行、韓亞銀行、BNK釜山銀行，Yegaram儲蓄銀行和現代汽車集團旗下金融平台「現代資本」相繼遭到黑客襲擊，導致個人資料外洩，不止涉及姓名和身份證等一般個人情報，還包括年收入、信用額和受害人僱主身份等。另外，友利銀行、農協銀行亦證實遭受黑客攻擊，但未證明是否有信息外洩。

據報道，韓國當局調查發現案件涉及AI代理驅動自主滲透系統ARTEX AI，其可通過可從外部（包括境外）存取的業務系統入侵。韓國監管機構研判，這些惡意工具可能廣泛攻擊多家金融公司的安全漏洞，而非針對單一機構。韓聯社引述銀行向國會議員提交的資料報道，這些攻擊的流量來自美國、日本、新加坡、越南及英國等

多個國家的IP位址。李在明4日聽取事件報告後，指示有關官員「要以嚴肅的態度面對此事，竭盡全力進行徹底調查並制定對策」。韓國金融服務委員會（FSC）發表聲明指出，委員會主席李伍源（音）4日已與金融業協會、監管機構及受影響機構高層召開緊急會議，並警告金融業必須保持最高警戒應對。

諾貝爾醫學獎今揭曉 GLP-1研究成大熱

【大公報訊】綜合《紐約時報》、路透社報道：2026年諾貝爾獎各獎項將在5至12日陸續揭曉。其中，生理學或醫學獎結果將在香港時間5日17：30率先出爐。今年無論是人類科學家或是生成式聊天機器人，在預測生理學或醫學獎時，都押寶開創GLP-1腸道激素研究的學者。

GLP-1是人體分泌的腸道激素的簡稱，它負責向人體發出「我們吃飽了」的信號。通過對該領域的研究，研究人員開發出了治療糖尿病和肥胖症的革命性藥物，例如風靡全球的針劑Ozempic與Wegovy。

2026年引文桂冠獎將生理學或醫學領域的榮譽授予了多倫多大學的德魯克（Daniel J. Drucker），哥本哈根大學的霍斯特（Jens Juul Holst）以及紐約洛克菲勒大學的莫伊索夫（Svetlana Mojsov），使GLP-1成為本年度最大熱門。

據報道，GLP-1的故事始於上世紀80年代：喬爾·哈貝納（Joel Habener）等人發現胰高血糖素基因可以編碼GLP-1；莫伊索夫鑑定出具有生物活性的GLP-1（7-36）酰胺；霍斯特證明它能夠以葡萄

糖依賴的方式促進胰島素分泌；德魯克則系統闡明了GLP-1的生理學功能，並推動其向藥物轉化。

根據諾貝爾基金會公布的信息，2026年每個完整獎項的獎金金額為1200萬瑞典克朗（約938萬港元），較2025年的1100萬瑞典克朗增加100萬瑞典克朗，漲幅約9.1%。如果獎項由多位獲獎者共同獲得，該獎金仍將按照諾貝爾獎的相關規則在獲獎者之間分配。頒獎典禮定於諾貝爾逝世紀念日12月10日，在斯德哥爾摩和奧斯陸舉行，並通過官方線上頻道直播。



▲2026年諾貝爾獎各獎項將在5至12日陸續揭曉。美聯社

巴西大選首輪投票 盧拉有信心獲勝

【大公報訊】綜合路透社、美聯社報道：巴西2026年總統選舉首輪投票於當地時間4日上午8時（本港時間晚7時）拉開帷幕。選前民調顯示，現任總統盧拉微幅領先主要對手、前總統博爾索納羅之子弗拉維奧，但預計兩人均無法獲得過半數的有效選票，極可能進入10月25日的第二輪投票決一勝負。

巴西現任總統、勞工黨候選人盧拉正尋求其第4個非連續總統任期。他4日現身聖保羅郊外的一所學校投票，並表示有信心在首輪選舉中獲勝。民調機構Datafolha當地時間3日發布的最新民調顯示，盧拉的支持率為42%，領先弗拉維奧約2個百分點。分析指出，兩人可能均難以在4日取得過半選票，



▲巴西現任總統盧拉4日現身投票站投票。路透社

10月25日舉行第二輪角逐或成定局，且民調預測兩人在次輪投票中也将勢均力敵。除總統選舉外，巴西選民亦將在4日投票選出27名州長與州議員、眾議院全部513席以及參議院81席中的54席。

當地時間上午8時起，巴西選民陸續前往各投票站投票。69歲的全職家庭主婦羅桑格拉表示，支持

盧拉扶助弱勢群體的社會政策，並讚賞其應對美國對巴西政策的堅定立場。她直言，現在的美國政府不斷「插手巴西內部事務」，「若盧拉勝選，這些企圖就不會得逞。」據當地媒體3日報道，盧拉的競選團隊已請求巴西最高法院，將美國駐巴西使領館2日暫停線下領事服務一事，納入正在進行的外部勢力干預巴西選舉相關調查。

外界普遍認為，此次選舉結果將影響拉美未來政治格局，因此受到美國干涉。弗拉維奧此前已明確表示，將讓巴西成為特朗普的親密盟友。專家認為，一旦弗拉維奧當選，就意味着巴西這一拉美大國倒向美國，拉美「右翼回潮」趨勢會進一步加強。