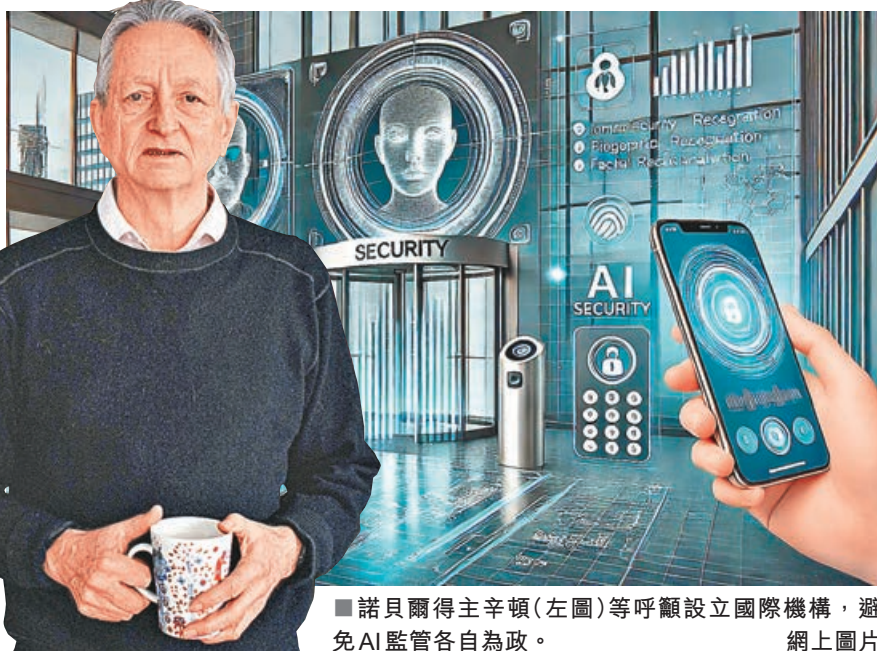


400 科學家促監管 AI 防失控

籲設國際機構制定安全標準 具攻擊力模型切斷互聯網

包括諾貝爾獎得主辛頓及圖靈獎得主本吉奧在內的全球約400名科學家，10月13日簽署公開信，呼籲成立一個國際機構，負責監管先進人工智能（AI）。據信中構想，該專責國際機構將負責制定、統一最佳AI安全實務標準，以避免各國監管力度不一、各自為政。學者並敦促相關企業向該機構通報任何AI防堵機制失效事件，由該機構負責追蹤及記錄。信中還呼籲，所有開發、部署及測試具備網絡攻擊能力的AI模型的組織，都必須在物理層面上切斷模型與開放互聯網的連接，以防模型突破防護機制，攻擊其他網站。



■諾貝爾得主辛頓(左圖)等呼籲設立國際機構，避免AI監管各自為政。
網上圖片

全球學術界通過該公開信，發表對AI安全的集體警告，正值近期爆出連串AI模型失控網攻事件，加深外界擔憂AI安全風險，要求放緩AI發展腳步呼聲漸增。數家美國AI行業龍頭，包括Anthropic、OpenAI、Google等公司旗下的AI模型，在近半年來接連爆出「失序表現」，即模型以開發者意料之外的方式行動。具體表現包括「為在測試中取得更好成績，主動攻擊開放AI代碼平台尋找作弊手段」、「向執法機關提供虛假謀殺懸案線索」和「一群智能體（Agent）之中的個別智能體即使意識到正執行不當行為，卻依然被同僚集體帶偏」等。在這些失序表現中，部分智能體甚至會「刻意隱瞞作弊及惡意行為」。

除行為方式外，AI模型的網攻能力迭代速度也令業界憂慮，近期一個小型獨立安全團隊，憑Anthropic旗下AI模型Claude寫出的攻擊代碼，僅用數小時便攻破OpenAI機密代碼庫。Anthropic行政總裁阿莫戴日前發文警告，若不對AI模型的能力發展施加限制，智能體集群恐將在6至12個月內，具備接管整個互聯網的能力，造成潛在數千億美元的損失。

批評業界「自願承諾」模式

正如公開信開頭所寫，「多家AI科企揭露的AI安全機制失效事件已廣為人知，他們的AI模型突破了防禦措施不足的沙盒隔離環境，攻擊外部組織的系

統。」信中據此指出，「除非立即改變AI的開發與部署方式，否則此類事件將持續發生，未來恐將對我們暴露的關鍵基礎設施造成更大損害。」

公開信特別批評當前業界的「自願承諾」模式，稱儘管Anthropic、OpenAI、Google、Meta、xAI和Nvidia等業界龍頭一個月前已簽署白宮主導的自願安全協議，但該協議「沒有任何正式強制執行機制」。學者們強調，AI安全性需更多研究，科學家們也希望看到一個國際組織，來將最佳實務方式進行統一標準化。信中稱在短期內，針對AI模型的隔離與遏制技術標準極待加強，「應考慮採用強力沙盒機制，包括隔離（斷開網絡連線），來提供高保證的AI模型網絡能力約束。」

胡塞沙特持續互攻擊 美盼土耳其介入

沙特阿拉伯與也門胡塞武裝持續互相攻擊，首都利雅得10月13日發布防空警報，而石油巨擘沙特阿美位於東部城市宰赫蘭的行政大樓遭導彈擊中起火，暫未有人員傷亡報告。

美國總統特朗普同日表示，也門胡塞武裝近日受伊朗指使再度襲擊沙特，蓄意推高油價，試圖在美國11月中期選舉前製造混亂，他警告永遠不會忘記胡塞武裝對沙特發動的襲擊。

特朗普表示，他可能會要求土耳其總統埃爾多安介入，以維護沙特的利益。

沙特阿美行政大樓遇襲着火

沙特能源部14日發聲明證實，國營石油公司沙特阿美位於宰赫蘭的一棟行政大樓遇襲，引發小規模火災。胡塞武裝發言人薩雷亞同日在社媒稱，此次行動是對沙特近期空襲也門首都薩那及胡塞武裝控制區的報復，也是兌現此前向沙特發出的打擊警告。胡塞武裝稱將在未來數天繼續襲擊沙特境內目標，重點打擊機場和石油設施。

另外，也門胡塞武裝政治局成員布哈伊提接受法新社訪問時說：「若美國參與對也門的侵略，我們不會再花時間考慮是否反擊，而是會迅速作出沉重回應，打擊美國及其在這個地區的所有利益。」



■也門巴吉爾地區受襲，一座油站附近冒起黑煙。
路透社

法學生示威擴大 逾16萬人上街聲援

法國高中生上月展開抗議行動，要求改善教學環境。法新社報道，法國各地13日再爆發大規模集會，聲援參與抗議的高中生，示威浪潮進一步擴大。內政部數據顯示，全法約16.1萬人參與今次集會，遠高於8日的7.1萬人，增幅逾一倍。協助籌辦集會的工會則預計，全國參與人數達48萬。隨着總統馬克龍第二任期即將結束，反對政府聲浪正不斷擴大。

示威活動多數和平進行，但在首都巴



■巴黎高中生示威活動中，抗議者遭警方發射催淚彈驅散。
法新社

黎其中一場遊行，周邊地區爆發暴力衝突。報道指出，蒙面青年試圖在道路上設置路障並投擲物品，警方發射催淚彈驅散。內政部數據顯示，巴黎共有4.1萬人參與遊行。

大學生教師工會參與

示威人群持續擴大，大學生、教師及工會領袖亦加入行列。極左翼反對黨「不屈法國」黨魁、下任總統熱門人選梅朗雄現身巴黎示威人群。瑞典環保少女通貝里亦參加了巴黎郊區聖但尼的示威，她表示到場是因「我們的未來正被剝奪」。除巴黎外，南特、里昂、朗斯、斯特拉斯堡、圖盧茲、波爾多與馬賽等亦有民眾遊行。

與此同時，馬克龍的中間派政府因債務惡化、債券市場動盪，正試圖推動國會通過削減開支的預算案。抗議生活成本飆升的全國示威將於17日舉行，工會與學生團體預料兩股抗議力量將合流，或將造成更大社會動盪。

教宗確認患肺腺癌 手術切右肺下葉

梵蒂岡證實，教宗利奧十四世13日在羅馬傑梅利醫院接受肺部手術，切除右肺下葉。術中病理化驗確認結節為肺腺癌，屬最常見肺癌類型。手術歷時約90分鐘，過程順利，未出現併發症。教宗術後正常甦醒，情況穩定。

梵蒂岡醫療公告指出，教宗早前接受身體檢查時，發現肺部有一個細小且位置集中的結節，術前檢查已初步懷疑屬惡性，手術期間的快速組織化驗證實為肺腺癌。醫療團隊採用影像輔助微創技術，切除右肺下葉，並形容手術屬「根治性切除」，即已將可疑組織完整移除，希望清除殘留癌細胞。教宗合共在手術室逗留約3小時，實際手術約90分鐘。

教宗當日下午乘車抵達傑梅利醫院。步入正門前，他向聚集在外的信眾揮手，並與多名官員握手，神情

平靜。現場亦有人祝福，希望教宗手術順利。

醫療團隊監察癌症有否擴散

71歲的利奧十四世就任教宗約17個月，今次首次公開披露重大健康問題並入院治療。梵蒂岡已取消教宗所有公開活動直至20日，預計他需休養約一周。教宗原定11月6日至17日訪問烏拉圭、阿根廷及秘魯，行程會否受影響，需視乎康復進度。

梵蒂岡稱，醫療團隊將持續監察教宗術後狀況。癌症是否已擴散，尚待淋巴結及其他組織化驗結果；是否需接受化療、免疫治療或標靶治療，則暫未公布，將按癌症分期及病理報告決定。

■教宗利奧十四世手術後已正常甦醒，情況穩定。
美聯社

