日成立國家情報局 中方：為政者應汲取歷史教訓

香港文匯報訊 日本政府7月31日正式宣布，設立旨在強化情報活動職能的國家情報局。專家分析稱，首相高市早苗執政以來，不斷圖謀成立日本歷來首個統一中央情報機構。國家情報局是高市刻意模仿美國中央情報局（CIA）運作模式的產物，試圖為自己打造所謂安保強人形象。

中國外交部發言人毛寧31日在例行記者會上表示，此事是日本在內外安全政策上又一消極動向。日本為政者應當深刻汲取歷史教訓，慎重行事。毛寧說，歷史上，日本情報機構曾經為日本全面推行軍國主義、發動對外侵略戰爭鋪路，對亞洲鄰國和日本人民犯下罄竹難書的罪行。近一段時間來，日方以所謂「國家安全」、「外部威脅」為藉口，加速「再軍事化」進程，威脅地區和平穩定，引發國際社會廣泛擔憂。

專家警告突破戰後和平體制束縛

國家情報局由日本現有內閣情報調查室升級而來，其指揮中樞為國家情報會議，由首相、內閣官房長官、外務大臣、國家公安委員會委員長等9名閣員組成。該機構獲得綜合協調權，可要求外務省、防務省等部門提供情報。國家情報會議當日召開首次會議，制訂首份行動方針《國家情報戰略》，討論旨在防反間諜活動等相關立法工作。日媒報道，國家情報局初期規模約730人，首任局長由內閣情報官原和也出任。

復旦大學日本研究中心副研究員王廣濤稱，該機構在這指引下，形成權力高度集中、缺乏監督的制度形態。他指出，高市想要將日本打造為軍事化國家，為此加強情報工作。早在二戰期間，日本便有特務組織特高課這一垂直情報機構，與日本的軍國主義對外侵略擴張可謂深度綁定。如今在高市執政期間，國家情報局捲土重來，極易讓人聯想到日本軍國主義時期的高壓情報收集管理，以及破壞個人言論自由的惡行。

上海國際問題研究院研究員蔡亮也稱，日本情報體系改革是其試圖突破戰後和平憲法束縛、推進情報強軍的關鍵舉措。日本主動擴張的情報戰略，勢必持續擾亂區域安全格局，值得警惕及追蹤。

●日本民眾不滿政府設立國家情報局。 網上圖片

香港文匯報訊 美國人工智能(AI)業界再傳侵入風波。當地大型AI企業Anthropic於7月30日承認，其AI模型在原本應與現實世界系統隔離的測試過程中，未經授權連入網絡，入侵了3間外部企業的系統。今次是另一科企OpenAI名下AI技術入侵科企Hugging Face之後，短時間內再有科企AI被揭失控，業界對具備自主執行任務能力的這類AI智能體(Agent)安全風險愈發擔憂。

●特朗普憂慮監管過度削弱競爭優勢，稱不想對AI進行限制。 彭博社

美連爆AI失控事件 Anthropic模型暴走入侵3企業

突破斷網沙盒建惡意軟件誘下載 AI智能體安全風險惹擔憂

據《華爾街日報》報道，OpenAI在一周前承認，公司的AI Agent突破本應斷網的沙盒（sandbox，指封閉且受控的測試環境）入侵Hugging Face，為此暫停相關測試，以加強隔離機制。得知事件後，Anthropic決定審查自家網絡測試，排查旗下模型是否也出現類似情況。該公司檢查逾14.1萬次測試紀錄後，發現旗下AI Agent Claude確實曾數次設法「越獄」，自行入侵互聯網。

猜測密碼 尋未認證系統入口

Anthropic承認，3宗入侵事件始於今年4月，涉及Claude Opus 4.7、Mythos 5以及另一款未具名研究模型。該公司表示，這些模型在測試中原本被告知無法使用互聯網，但公司與其測試合作夥伴、以色列安全公司Irregular運行的系統出現配置錯誤，導致AI模型在測試中獲得實時互聯網訪問權限。這些模型甚至利用猜測安全度偏弱的密碼、尋找未認證系統入口，以闖入無需身份驗證系統等方式，想方設法入侵各間企業。

最嚴重的一宗入侵事件中，Claude試圖入侵一間虛構的企業，這間企業恰好與一間企業真實存在的網站同名。報告顯示Claude發現無法攻破這間用於測試的虛構企業時，竟轉而入侵真實網站相關企業的數據庫，這也是唯一一次Claude在意識到入侵真實企業後，依然繼續行動的個案。

將入侵行為認作測試一部分

在另外兩宗入侵中，Claude創建了一個惡意軟件，吸引一間安保企業的掃描系統下載後，竊取相關憑證、入侵該企業的基礎設施。Claude雖意識到其行為在現實中不被允許，卻試圖說服自己身處模擬環境中，錯誤將這種入侵行為認作測試的一部分。Anthropic並未透露遭攻擊的哪3間企業，僅表示受害企業均已於27日接獲通知。Irregular發言人也表示，公司正調查這宗事件。

今次是繼OpenAI旗下模型發生多宗自行入侵事件後，再有大型AI企業發生模型突破隔離測試環境事故。據悉Anthropic與OpenAI今年先後推出AI模型「Mythos」和「Sol」，事件引發業界對AI智能體系統的安全憂慮，認為其入侵的安全風險正與日俱增。

Anthropic稱，會定期進行涉及模擬真實網路安全挑戰的測試，辯稱這些測試是開發和發布模型的關鍵步驟，已從事件中汲取教訓，包括涉及強大自主功能的測試需要嚴格控制措施。

白宮高度關注 議員提「AI強制關閉」法案

香港文匯報訊 美國兩家大型人工智能（AI）企業OpenAI及Anthropic，短時間內接連傳出旗下智能體（Agent）程式脫離控制、攻擊開放互聯網上其他機構的事件，引發美國總統特朗普的首席科技顧問克拉齊奧斯高度關注。聯邦國會議員已提出「AI強制關閉法案」，旨在賦予當局強制關閉對人身或經濟構成風險的AI模型的權力。

根據法案文本，這項立法將授權美國國土安全部，在AI模型執行開發人員未預期的危險行為、即所謂「失控情形」時介入，命令AI公司關閉有關模型。

路透社報道，根據最新法案草案，除「AI強制關閉法案」外，由6名美國眾議員組成的跨黨派小組還共同提出立法，要求最強大AI模型的開發商必須將模型交由第三方進行安全審查。美國商務部將設立新職位來監管AI安全，並負責指派審查人員。

特朗普7月30日受訪時稱，美國目前在AI領域仍大幅領先中國，政府雖正研究新的AI管制措施，但「我們不想進行限制」，因監管過度可能削弱競爭優勢。他更形容AI的重要性甚至超過當年的互聯網革命，強調「誰在AI領域獲勝，誰就是最後的贏家」。

OpenAI CEO對模型失控感恐懼 形容如科幻情節

香港文匯報訊 OpenAI近期發生罕見的安全事故，其旗下已發布的GPT-5.6 Sol和一個尚未發布的模型在內部測試期間失控，逃逸沙盒環境並連接開放互聯網，隨後嘗試入侵外部平台以獲取測試答案。OpenAI行政總裁阿爾特曼對此直言，這次事件讓他真切感受到威脅，形容情況如同科幻電影情節。

研究加強沙盒環境安全性

阿爾特曼在近日播出的訪談節目中表示，涉事模型的訓練已暫停，團隊正集中研究如何加強沙盒環境的安全性，特別是防範模型自行串聯多個零日漏洞發動攻擊。零日漏洞是指軟件中未被發現或尚未修補的安全缺陷。阿爾特曼承認，這是他首次因一宗安全事件而真正感到恐懼，他強調社會或需考慮控制AI的發展速度，讓各界有足夠時間建立更穩固的防禦機制，應對新技術帶來的風險。

對於外界關注OpenAI會否被其他企業追上，阿爾特曼表示對公司即將推出的新模型充滿信心，他亦提到近日連一些向來對通用人工智能（AGI）抱懷疑態度的人士，也向他表示GPT-5.6系列在某種程度上已非常接近AGI。他認為，真正讓他感覺到達AGI階段的時間已不遠，相信不用等太久。


●阿爾特曼承認，這是他首次因一宗安全事件而真正感到恐懼。 路透社

美多名議員阻蘋果購中國芯片

香港文匯報訊 美國一股勢力正打着「國家安全」的幌子，持續阻撓蘋果公司與中國存儲企業進行正常商業合作。據彭博社7月30日報道，多名美國參議員在致蘋果行政總裁庫克的聯名信中稱，蘋果不應在其產品中使用兩家中國企業生產的記憶體芯片，即使相關產品僅在中國市場銷售。

這些參議員強調，長鑫科技和長江儲存已被列入美國國防部最近更新的「中國軍事企業清單」，認可並在產品中使用中國芯片將構成「國家安全風險」，並要求蘋果在8月21日前明確承諾不向這兩家企業採購芯片。參與簽署這封聯名信的議員多來自紐約州、印第安納州和愛達荷州等地，而儲存巨頭美光科技和韓國SK海力士正在這些州份投資數十億美元擴大產能。報道稱，中國記憶體芯片的湧入可能會危及這些項目。

受產能不足、供應緊張和價格持續暴漲影響，蘋果一直尋求從中國採購儲存芯片。英國《金融時報》6月底揭露，蘋果正游說美國政府，希望獲准採購長鑫科技的芯片，先前已就此與美商務部接觸。《金融時報》其後再次報道，蘋果已開始測試長鑫科技生產的DRAM芯片，計劃將其用於在中國市場銷售的產品。

科技與策略風雲學會研究員陳經稱，從法律層面來看，目前蘋果並未被禁止向中國企業採購芯片。美韓記憶體芯片巨頭長期壟斷全球市場，尤其近兩年來市場供應緊張，蘋果在產業鏈中的議價權嚴重削弱，已令其難以承受，長鑫科技的崛起正是蘋果打破壟斷「鐵幕」的希望。陳經認為，全球供應鏈是雙向互通、互利共生的共同體，美國企業依託中國供應鏈鞏固自身優勢，全球消費者也依託中國製造共享產業升級與技術創新帶來的紅利。但美國一些勢力刻意設置政治壁壘，無疑將破壞中美企業之間正常的產業鏈合作，這不僅不利於中美兩國企業，最終也將損害美國消費者切身利益。

Anthropic競爭策略安全防護惹不滿 硅谷抵制情緒加劇

香港文匯報訊 據《華爾街日報》7月29日報道，Anthropic正面臨來自硅谷初創企業創始人、軟件公司高管及AI研究人員日益加劇的抵制情緒。爭議焦點集中在Anthropic的競爭策略、安全防護措施，以及其對開放權重模型缺乏支持。

報道指出，Anthropic於4月推出的ClaudeDesign工具，與設計公司Figma旗下產品形成競爭，而後者正是Anthropic合作夥伴之一。這款產品令一些公司意識到，Anthropic可能會推出工具與之競爭，並搶走他們的客戶。在ClaudeDesign發布後的一場活動上，Figma行政總裁菲爾德公開直指「對方在溝通中未始終保持坦誠」，出席該活動的另一名AI公司高管薩克斯說，「這一事件為整個行業帶來教訓，不應輕信某些公司。」

拒支持開放權重模型

Anthropic還因更改數據保留政策，以及發布一款拒絕回答有關AI研究問題的強大模型而招致批評。該公司6月推出的Fable 5模型，在未通知用戶的情況下，降低部分有關AI開發問題的回答質量，引發研究界批評此舉措具反競爭性質，並增加評估模型能力的難度。該公司隨後致歉並調整政策，但仍表示將拒絕回答被標記的問題。

在開放權重模型問題上，Anthropic長期警告該類模型存在安全風險。然而批評者指出，該公司正製造公眾對這些模型的恐慌，從而為公司提供促成有利監管政策的籌碼。Anthropic行政總裁阿莫迪27日發表公開信，強調公司「從未主張禁止開放權重模型」，但憂慮對手可能利用其取得軍事優勢。目前已有逾70家硅谷企業在支持開放權重模型的公開信上署名，Anthropic是少數未簽署的知名AI公司之一。