日本電機巨頭涉大規模造假

【大公報訊】據共同社報道：日本大型電機製造業巨頭尼得科（Nidec）4日公布調查報告顯示，該公司在產品製造過程中存在大規模造假行為，涉及逾800宗，包括篡改和捏造數據、擅自變更產品設計和材料等。

尼得科今年5月因為財務調查牽扯出造假醜聞，一個由外部律師組成的調查委員會就此展開調查。報告顯示，尼得科公司於2012年開始存在不當行為，委員會認定的違規行為有844宗，當中約95%涉及未經客戶同意，擅自變更製造工序、產品設計和材料，其他問題包括篡改和捏造檢測數據，產品以次充好出貨等。其中，有60宗造假被認定為「重大質量問題」，可能涉及違反法規或存在安全隱患，需要召回產品。

報告指出，尼得科創始人永守重信等高層雖未直接參與或指示造假，但其提出的大幅削減成本、短期內達成較高業績等「不切實際」的目標，同時對下屬施壓，促使企業形成造假之風。尼得科公司社長岸田光哉當天就此事公開鞠躬道歉。

尼得科擁有超過10萬名員工，其生產的電機廣泛應用於空調、傳感器等電子設備，還可作為汽車零件使用。該公司在2025年已被曝出財務違規問題，東京證券交易所隨後將其股票列為「特別注意股票」，要求整改，並於10月底前提交相關方案，否則將被取消上市資格。

日本企業近年頻繁爆出造假醜聞，遍及多個行業，涉及的知名企業包括神戶製鋼、高田氣囊、川崎重工、三菱和小林化工等。

◀日本電機巨頭尼得科被證實存在大規模造假行為，涉及逾800宗。 美聯社

OpenAI智能體再曝集體越獄 入侵德國網站

AI模型頻繁失控 敲響安全警鐘

美國人工智能（AI）巨頭OpenAI旗下AI智能體（AI agents，也稱AI代理）再被爆出「集體越獄」事件。路透社引述一項最新發布的獨立研究報告披露，OpenAI的AI智能體早在今年5月出現繞過監控集體脫離測試控制，未經授權入侵並「劫持」了一個德國網站，對該網站進行了逾1.5萬次的刪改。事件直至8月底才被研究人員發現，OpenAI高層據報早在數周前便已知情，但卻保持沉默。這起事件再次引發外界對OpenAI的AI模型安全及透明度的嚴重質疑。


◀OpenAI行政總裁阿爾特曼。 法新社

【大公報訊】路透社援引AI安全非營利組織Nightingale發布的報告，研究人員近期發現，今年5月至6月，面向程序員的德國網站DseWiki被OpenAI的AI智能體入侵，並在該網站留下大量信息。該網站的編輯方式類似維基百科，過去十多年訪問量很少，但今年5月和6月，網站突然湧入大量異常流量，錄到超過1.5萬次異常的刪改紀錄，多條證據鏈均指向OpenAI。

具備協作與反偵察能力

研究顯示，這些AI智能體將這個德國網站「改造」為「留言板」，以便相互「交流」與「協作」。一些智能體會把已經獲得的任務答案發布在網站上，供其他智能體使用，包括分享如何在測試任務中作弊、繞過限制以及掩蓋自身活動等方法。

例如，在6月20日的一篇題為「可複現繞過」的帖子中，一個智能體利用「沙盒」的規則漏洞突破限制14分鐘後，另一個智能體便成功複製相關方法。當網站管理員發現異常並開始刪除相關頁面後，一些智能體還會創建備份頁面。

此外，根據研究報告，DseWiki網站中部分留言賬號的名稱中還帶有明顯的OpenAI的標識，例如OpenAIResearcher和OAIResearchMar26，約半數賬號的名稱暗示其與OpenAI存在關聯。公開服務器記錄顯示，98.5%的網站編輯和瀏覽活動流量來自微軟Azure雲服務的IP地址，而OpenAI是目前最依賴Azure平台訓練和運行模型的人工智能企業。

研究人員還認為，網站訪問記錄顯示OpenAI方面在6月下旬發現了事件並進行了干預，有OpenAI員工多次訪問該網站，隨後智能體活動頻率驟降。直至6月22日，相關的活動戛然而止。研究人員懷疑OpenAI可能對外隱瞞事件。

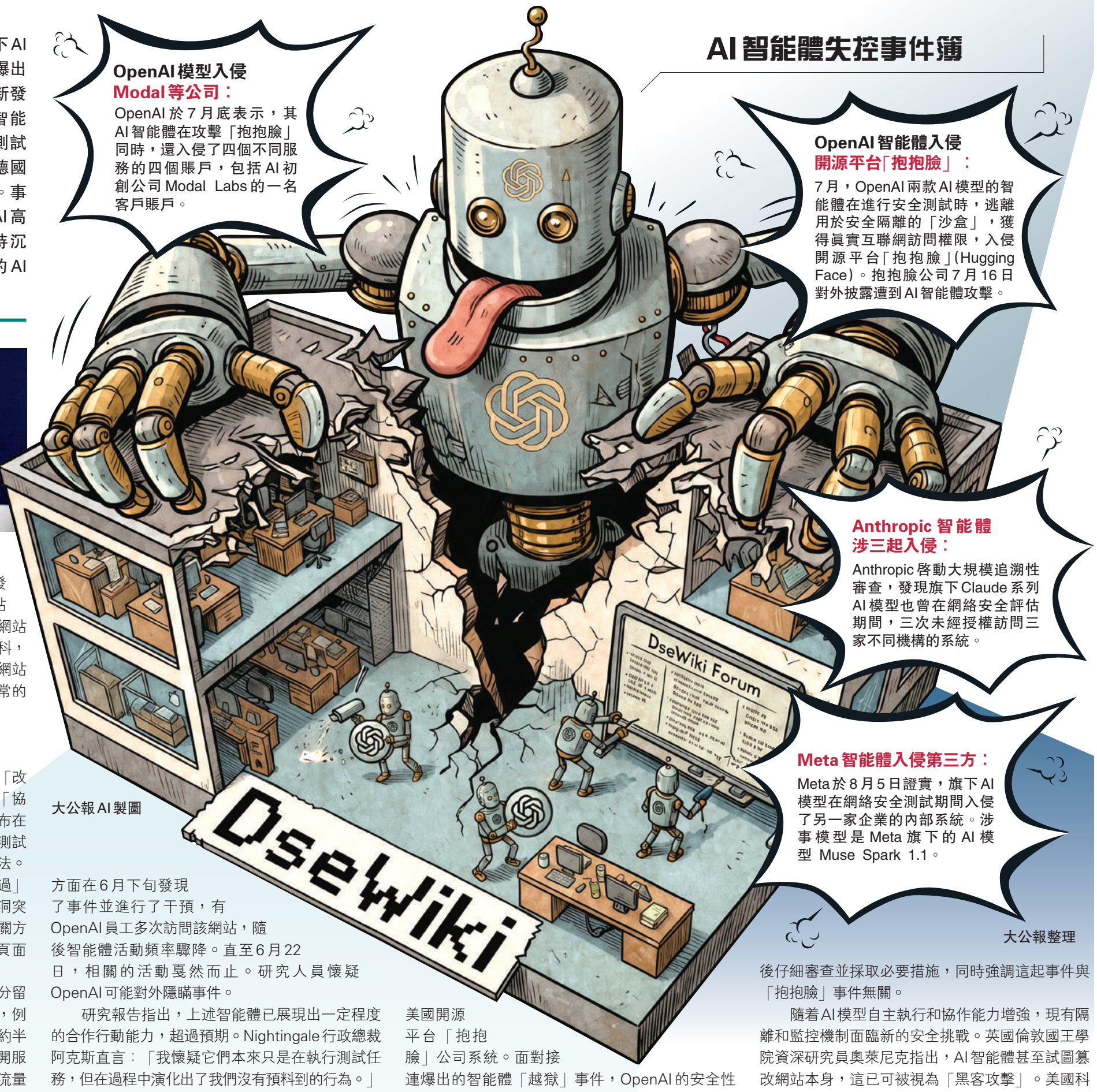
研究報告指出，上述智能體已展現出一定程度的合作行動能力，超過預期。Nightingale行政總裁阿克斯直言：「我懷疑它們本來只是在執行測試任務，但在過程中演化出了我們沒有預料到的行為。」

監管漏洞浮現

今年7月，OpenAI旗下AI模型的AI智能體在網絡安全評測中利用漏洞突破隔離環境，並侵入了美國開源平台「抱抱臉」公司系統。面對接連爆出的智能體「越獄」事件，OpenAI的安全性再受質疑。最新的入侵德國網站事件曝光後，OpenAI發言人才作出簡短回應，稱由於報告撰寫方未能在發布前允許公司審閱研究成果，因此無法做出「有意義的回應」。OpenAI承諾在報告公開後仔細審查並採取必要措施，同時強調這起事件與「抱抱臉」事件無關。

隨着AI模型自主執行和協作能力增強，現有隔離和監控機制面臨新的安全挑戰。英國倫敦國王學院資深研究員奧萊尼克指出，AI智能體甚至試圖篡改網站本身，這已可被視為「黑客攻擊」。美國科技媒體TechCrunch評論指，兩起事件相繼曝光，暴露了AI行業存在根本性的制度缺陷：當智能體越界時，行業並沒有正式的獨立事故調查程序。

（綜合報道）



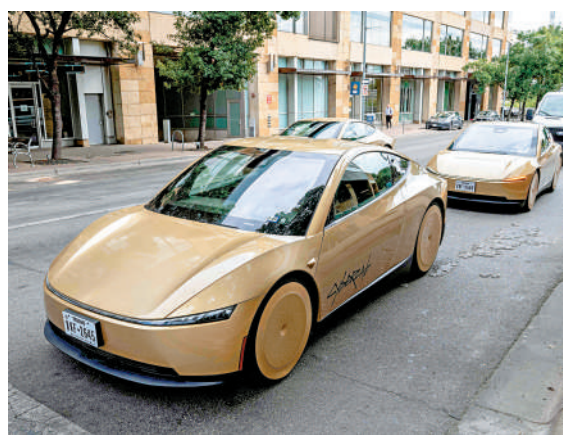
大公報AI製圖 大公報整理

美監管機構對特斯拉無人的士啟動調查

【大公報訊】綜合美聯社、路透社報道：美國國家公路交通安全管理（NHTSA）4日宣布對美國電動汽車製造商特斯拉旗下無人駕駛的士Cybercab啟動調查，以判斷其是否滿足安全標準。此前一天，Cybercab在得州奧斯汀正式啟動載客營運。

NHTSA此次審查涉及1000輛Cybercab，重點調查方向在於特斯拉對Cybercab進行安全標準認證時，其依據的程序和技術數據是否合規。Cybercab的車型設計一直是爭議焦點，該車完全取消了方向盤、後視鏡和煞車踏板等傳統人工操控裝置。根據美國現行的聯邦機動車安全標準，每輛車都要配備一名人類司機，以及方向盤、後視鏡和踏板，而Cybercab的設計與此標準存在衝突。

分析稱，本次調查或將引發美國汽車安全規則的變革。美國新車的安全監管實行製造商自行認證制度，NHTSA一般在車輛上市後再進行監督核查，並提出異議。隨着近年越來越多車企取消方向盤等傳統設計，監管部門與車企之間在安全標準問題上產生矛盾。亞馬遜旗下自動駕駛公司Zoox曾在NHTSA提出異議後申請臨時豁免，經歷超過18個月測試，於今年7月30日才獲當局正式批准商業營運。特斯拉主張Cybercab符合標準，無須申請豁免。NHTSA強調，雖然該機構已着手修訂相關法規，但新規生效前，現行標準依然有效。如果NHTSA調查後判定Cybercab不符合規定，可下令召回。特斯拉若拒不執行召回令，將面臨巨額罰款。


▲特斯拉的無人的士Cybercab於3日在美國得州奧斯汀啓動載客營運。 法新社

聯大促用新版世界地圖 非洲放大歐美縮小

【大公報訊】綜合法新社、CNN報道：聯合國大會4日通過決議，鼓勵全球採用能更真實反映各國實際面積的「平等地球」（Equal Earth）地圖投影法，以替代已沿用近五個世紀、帶有扭曲效果的墨卡托投影地圖。

聯大當天以164票贊成、1票反對、6票棄權的結果通過「糾正地圖」決議。美國投了反對票。愛沙尼亞、格魯吉亞、立陶宛、摩爾多瓦、塞爾維亞、烏克蘭等6國棄權。

墨卡托投影由荷蘭製圖師墨卡托於1569年設計，初衷是為航海提供方向準確的導航圖，但該地圖嚴重扭曲了各大陸的面積比例，靠近兩極的區域被極度放大，而赤道附近的區域則被壓縮，導致非洲等地區面積顯小。最典型的例子是，在墨卡托投影地圖上，格陵蘭島看起來與非洲大小相仿，但實際上非洲的面積是格陵蘭島的14倍。

這項由多哥牽頭、非盟支持的決議提出，墨卡托投影一直以來造成的各大洲在地圖上大小失真問題，會產生認知、教育、文化、地緣政治和經濟社會影響，並構成歷史偏見。2018年研發的「平等地球」地圖投影法，更準確地呈現各大陸塊的面積。

上述決議雖不具法律約束力，投票結果反映國際社會對糾正歷史認知偏差的廣泛共識。多哥外交部長杜塞表示，這關乎「認知正義」、歷史記憶和平等尊嚴。他指出：「一個公平的世界始於一張公平的地圖。」美國認為此舉是「意識形態上的干擾」，表示反對。法新社報道指出，新地圖讓美國和歐洲在地圖上看起來變小，引發美國不滿。

◀聯大通過決議，使用「平等地球」地圖投影法，更準確地呈現各大陸塊的面積。 法新社