遠景論壇

OpenAI 近期於 2026 年 2 月發布的更新報告，延續近兩年對 AI 遭濫用情形的追蹤。（圖片來源：Depositphotos）

從 2026 年 2 月 Open AI 報告看中共「網路特戰」：臺灣資訊安全、民主韌性與國際協作的政策意涵

洪敬富

成功大學政治學系教授

OpenAI 近期於 2026 年 2 月發布的〈打擊惡意使用人工智慧〉（Disrupting malicious uses of our models）更新報告，延續近兩年對 AI 遭濫用情形的追蹤，揭露多起詐騙、祕密影響力行動與政治操弄案例。

其中最受關注的是一名與中國執法體系有關的使用者，利用 ChatGPT 潤飾「網路特戰」工作報告，並曾試圖要求模型協助規畫針對日本首相高市早苗的輿論攻擊行動。雖然 ChatGPT 拒絕提供操作建議，但後續跡象也顯示，中共相關行動仍可能透過其他網路工具與當地網路模型接續執行。這份報告的重要性，不僅在於揭露某單一案件，更是在呈現一種由國家主導、跨平臺、結合 AI 與人力的資訊作戰體系，其目標已涵蓋中國國內異議者、海外批評者、外國政要，以及與中國地緣政治利益直接攸關的議題，當然也包括臺灣在內。

檢視該報告，OpenAI 至少傳達了三個核心重點。第一，AI 不是單獨作惡的科技工具，而是被整合進既有的政治操弄、網軍操作、詐騙與社群滲透流程之中；第二，真正決定影響力的，未必是內容是否由 AI 生成，而是操作者是否握有大量帳號、投放管道與跨平臺的分發能力；第三，中國的相關行動並非零散個案，而是具有組織規模、明確分工、可觀資源與持續性的體系化作業。而該報告中所稱的中共「網路特戰」（cyber special operations），指的是由國家機構主導、結合假帳號網絡、跨平臺傳播與心理戰操作的影響力行動。它直指中共數百名工作人員、數千個假帳號、數十甚至數百個平臺，並可能結合中國本土大型語言模型如「深度求索」（DeepSeek）或通義千問（Qwen）等工具，進行情監偵測、翻譯、內容產製、假證據製作與內部報告整理工作。

中共網路特戰目的之分析

若進一步分析中共網路特戰的目的，大抵可分為四層。其一，為了維穩與壓制異議。經由錯假訊息、騷擾、冒名檢舉、人格抹黑與心理施壓等，讓對黨國批評者噤聲失語；其二，為了對外型塑或扭轉敘事氛圍，將外界對中國人權、統戰、滲透等行為的批判，重新轉化為反中偏見或境外敵對勢力的操弄；其三，為了離間民主陣營，挑動美、日、歐盟與亞洲民主國家間的矛盾，減損或削弱它們在印太安全、人權議題與臺海間的政策協調與深化合作；其四，為了地緣政治及對臺政策鋪路。透過系統性的資訊操作，讓臺灣議題被國際社會接受為中國的內政問題，並據以造謠、抹黑支持臺灣的國際政治人物、組織與外國評論者。

具體而言，這類網路特戰在作法上呈現幾個特徵：首先是真假混合。以少量真實事件為根基，摻入誇大指控、斷章取義與假造證據，

以提高資訊的可信度；其次是平臺混用。不僅分工經營 X、Facebook、YouTube、Instagram、Telegram、或網路論壇，更利用搜尋結果「污染」資訊、假網站、假媒體與假身分帳號，形成多層次網路擴散；第三是善用 AI 降本增效。由於 AI 可快速生成多語言的貼文、留言、攻擊話術、圖片和影音等文案資料，讓內容操作更像是真人般回應。這不僅使網路特戰降低成本，也提高即時反應；第四是線上與線下嵌入。例如該報告中提及中共網路特工或冒充美國官員，或偽造法院文件，或鎖定異議者家屬，或對直播與社群帳號進行干擾等，在在顯示這不是單純中共網軍的網路灌水，而是帶有準執法、準情報，以及實際統戰性質的複合操作手段。

然而，就可能的成效而言，該報告也提供了一些評價。OpenAI 指出，許多中共網路特戰的行動實際觸及範圍有限，部分貼文幾乎很少互動，或根本沒有互動。而針對高市早苗的相關貼文與標籤，也未顯示出大規模真實群眾共鳴。這說明網路特戰未必總能成功地改變一般網民意見，但其危險並不在於每次都能「爆紅」，而在於它能長期製造網路噪音、污染資訊環境、消耗受害者心力，並在黨國視為關鍵時刻時配合外交、軍事，協力進行恫嚇或統戰，形成一種網路空間上的複合壓力。換言之，其效果往往不是要立即說服網民或公眾，而是擾亂他們的判斷、放大分裂、削弱信任、壓縮反對聲音的生存空間。

對臺灣的啟示

對臺灣而言，這份報告至少有三層啟示：第一，臺灣早已是類似網路特戰的前攻擊前線。高市早苗因「臺灣有事說」而遭到中國抨擊，正反映凡是支持臺海安全、反對中國野心擴張、主張民主聯盟與合作的國際政要或知名人士，都可能成為中共攻擊的目標；第二，對臺灣的操作不會僅限於資訊作戰，而是結合影響力操作與統戰工作的網路統戰。前者著重於攻擊、滲透、抹黑與認知擾亂；後者則透過經濟利益、情感訴求、文化或宗教認同、社群經營與各方意見領袖的拉攏吸納，試圖塑造疑美、疑日、疑政府、疑民主等氛圍；第三，AI 的導入會讓這種操作更加在地化、即時化與客製化。未來針對臺灣，可能不只是大外宣，更可能是按族群、年齡、地方、產業與政黨支持傾向量身打造網路內容。例如：針對青年世代散播戰爭都是美國與賴政府挑起的，或是臺灣終將會被美國拋棄；針對民眾可能散播兩岸交流受阻都是當前政府的政策錯誤造成等。

從而，臺灣政府的回應不能停留在錯假訊息澄清的舊思維與舊模式，而需要提升至民主安全治理層次：第一，應建立跨部會的資訊操弄監測中心，整合國安、數位、法務、警政、外交與通傳等資源和能量，針對假帳號網絡、協同行為、跨平臺擴散與深偽內容進行快速識別；第二，應強化平臺治理與法律工具，特別是針對境外勢力資助的協同行為、冒名公職與偽造文書式數位內容，建立明確揭露、下架與司法合作規範；第三，應把媒體識讀升級為「認知韌性教育」，讓民眾理解資訊操弄不只是假消息真假判斷，更包括情緒操控、敘事包裝與身分滲透；第四，政府與民間應建立可信、快速、可重複傳播的事實查核與危機溝通鏈，以免每逢重大選舉、軍演或兩岸事件時，就陷入被動辯解和澄清。

國際民主社會之警訊

對國際民主社會而言，這份報告也提出明確警訊。中國的網路特戰已不是單一國內審查外溢，而是跨國資訊壓迫的一環。民主國家需要將此類行動視為混合威脅，而非單純的網路內容審查問題。或可考慮推動三方面的廣泛合作機制：一是情資共享，包括假帳號樣態、敘事模板、工具鏈與基礎設施；二是平臺責任共管，要求大型平臺提高對國家支持型操弄行動的透明度與處置速度；三是受害者保護，尤其是流亡異議人士、專家學者研究人員、記者與各類倡議者，應有跨國通報、法律支援與數位安全協助機制。

簡言之，這份報告最值得重視之處，不是證明 AI 已經取代傳統網軍，而是 AI 正讓威權國家的資訊操弄更加規模化、精細化與持久化。事實上，近年學界已逐漸將此類行動稱為「協同不實行為」（coordinated inauthentic behavior）或「外國影響力操作」（foreign influence operations）。中國的相關作法，則常被視為結合宣傳、統戰與心理戰的複合型資訊作戰體系。對臺灣而言，這不是未來式，而是現在進行式，是長期存在且持續演化的安全挑戰。面對中共結合網路作戰、網路統戰與跨境騷擾等複合威脅，臺灣與民主盟友真正需要的不只是技術防堵，而是將資訊環境視為國安與民主制度的重要部分，以民主制度韌性、社會信任與國際協作，回應這場不宣而戰的新型戰爭。

財團法人兩岸交流遠景基金會

本基金會為研究國際政經情勢之民間學術智庫，旨在針對國際政經情勢及戰略與安全等領域，將學術研究成果具體轉化為政策研析，作為我政府參考，深化學術研究能量，並增進與國際重要智庫交流與互訪。

臺北市汀州路三段 60 巷 1 號

Tel: 886-2-23654366
Fax: 886-2-23679193
http://www.pf.org.tw