國防安全雙週報

第 114 期

臺北市博愛路 172 號

電話（02）2331-2360

傳真（02）2331-2361

2026 年 6 月 12 日發行

財團法人國防安全研究院

Institute for National Defense and Security Research

Contents

誰能帶來安全？台灣民眾對美日中關係的安全認知

李冠成

中共政軍與作戰概念研究所

焦點類別：國防安全民意調查

壹、前言

近年來，美中戰略競爭持續升溫，台海局勢也成為印太地區的重要安全議題。近期中國持續在台灣周邊進行軍事活動，並強調將持續提升備戰能力；[1]另一方面，美中高層互動與川習會後續發展，也引發外界對台海情勢及美國對台政策走向的討論。[2]與此同時，日本近年持續強化國防能力，並多次公開強調台海和平穩定的重要性，使台日安全連結受到更多關注。[3]

在此背景下，台灣民眾如何看待不同對外關係與國家安全之間的連結，值得進一步觀察。傳統上，台灣社會對美國、日本及中國大陸的態度，往往被視為外交偏好或政黨立場的反映；然而，這些對外關係同時也涉及民眾對國家安全來源的理解。本次民調分別詢問民眾，[4]若台灣與美國、日本及中國大陸的關係變得更好，是否有助於提升或降低台灣的國家安全，藉此探討民眾對不同對外關係安

[1] "China Says Taiwan Should Not 'Interfere' in its Air Force Missions Around Island," *Reuters*, May 28, 2026, https://www.reuters.com/world/china/china-says-taiwan-should-not-interfere-its-air-force-missions-around-island-2026-05-28/.

[2] "US Arms Sales to Taiwan on 'Pause' Due to Iran War, Says Acting Navy Chief," *The Guardian*, May 22, 2026, https://www.theguardian.com/world/2026/may/22/us-arms-sales-taiwan-pause-iran-war-says-acting-navy-chief.

[3] Ben Blanchard, "Taiwan 'Very Moved' by Japanese Prime Minister's Support, Premier Says," *Reuters*, December 5, 2025, https://www.reuters.com/world/china/taiwan-very-moved-by-japanese-prime-ministers-support-premier-says-2025-12-05/.

[4] 本民調是由財團法人國防安全研究院委託國立政治大學選舉研究中心執行之調查，調查對象是居住在台灣且年滿 18 歲以上的民眾，以電話訪問之方式進行隨機抽樣調查。訪問期間自 2026 年 3 月 12 日至 3 月 16 日。電話調查實際訪問完成 827 份市話樣本、342 份手機樣本，共計 1,169 份。以 95%之信心水準估計，最大可能隨機抽樣誤差為：±2.87%。

全效果的認知，以及不同政治群體之間是否存在差異。調查結果顯示，多數民眾認為改善與美國及日本的關係有助提升國家安全，但對兩岸關係是否能夠提升安全則存在較明顯分歧；此外，不同政黨支持者對安全來源的理解亦呈現明顯差異。

貳、安全意涵

一、美日被視為安全資產，中國則引發分歧

圖 1 顯示，當被問及台灣與不同國家的關係若變得更好，是否有助於提升國家安全時，多數民眾對美國與日本抱持相對一致的正面看法。調查結果顯示，約 66%的受訪者認為台灣與美國關係改善將有助於提升國家安全，僅 17%認為會降低安全；在台日關係方面，認為有助提升安全的比例約為 68%，認為降低安全者則為 15%。換言之，約三分之二的民眾皆將增進美日關係視為有利於台灣安全的因素。

相較之下，民眾對兩岸關係改善的安全效果則呈現較為分歧的看法。調查發現，47%的受訪者認為台灣與中國大陸關係改善有助於提升安全，但也有 35%認為反而會降低安全，兩者差距明顯小於美國與日本的情況。這顯示兩岸關係雖然仍被不少民眾視為影響安全的重要因素，但其安全效果並未獲得如美日關係般的一致認可。

上述結果反映，台灣民眾並非將所有對外關係視為同等重要，而是對不同國家與台灣互動所可能產生的安全效果存在明顯區別。在多數民眾認知中，美國與日本已不僅是外交或經貿夥伴，其與台灣關係的發展亦被視為與國家安全息息相關的重要因素。然而，兩岸關係是否能夠帶來更高程度的安全保障，社會內部仍存在不同理解。此種差異也顯示民眾對於「安全來自何處」並沒有一致答案，而是會依據不同對外關係形成不同判斷。

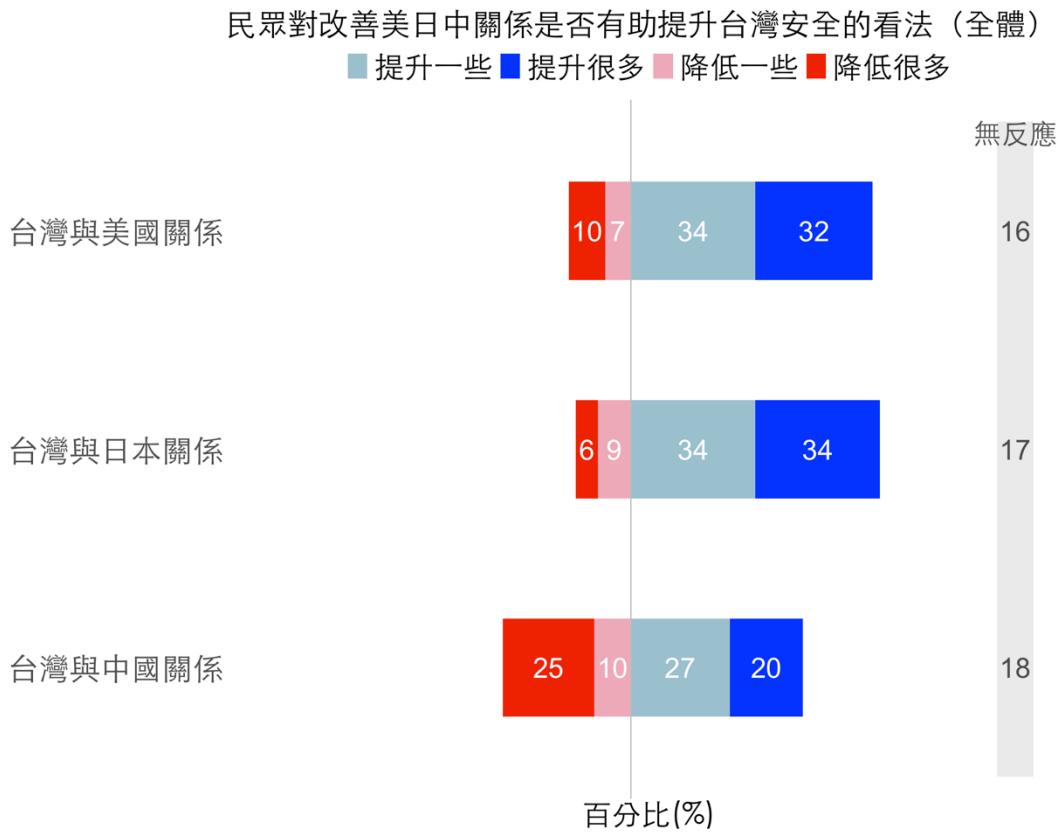

圖 1、民眾對改善美日中關係有助於提升台灣安全的看法

資料來源：國防安全民意調查（2026 年 3 月電話調查）。
說明：小數點四捨五入至整數。

二、兩岸關係呈現競爭性的安全邏輯

相較於美日關係獲得較一致的正面評價，兩岸關係的安全效果則呈現較大的看法分歧。這種差異不僅反映民眾對兩岸關係的不同期待，也涉及如何理解安全風險與安全保障之間的關係。

一方面，部分民眾認為改善兩岸關係有助於降低誤判與衝突風險，進而維持區域穩定與台海和平。在此認知下，兩岸互動本身具有降低緊張情勢與避免危機升高的功能，因此關係改善也被視為提升安全的重要途徑。另一方面，也有部分民眾未必認為兩岸關係改善能夠直接轉化為安全保障。近年來，中國持續透過軍事施壓、灰色地帶行動、統戰工作及認知作戰等方式對台施加壓力，使部分民

眾傾向認為，兩岸關係的變化未必改變中國對台政策的基本立場，因此關係改善與安全提升之間並不存在必然關聯。對這些民眾而言，安全與否不僅取決於兩岸互動是否和緩，更取決於中國對台政策是否出現實質改變。

這顯示，兩岸關係對許多民眾而言，不只是外交議題，也與如何理解國家安全息息相關。當社會對於兩岸關係與安全之間的連結存在不同理解時，也意味著涉及兩岸政策或安全戰略的公共討論，可能持續出現競爭性的觀點與主張。這種差異是否進一步反映在不同政治群體身上，亦值得進一步觀察。

三、安全認知呈現明顯政黨差異

圖 2 顯示，不同政黨支持者對美日中關係改善是否有助於提升台灣安全的看法，存在相當明顯的差異。其中，民進黨支持者最傾向將美國與日本視為提升安全的重要因素，高達 93%的民進黨支持者認為改善台美關係有助於提升台灣安全，另有 91%認為改善台日關係有助於提升安全；相較之下，僅有 28%認為改善兩岸關係有助於提升安全。

國民黨支持者則呈現截然不同的模式。74%的國民黨支持者認為改善兩岸關係有助於提升安全，明顯高於認為改善台美關係（42%）及台日關係（47%）有助於提升安全的比例。換言之，對國民黨支持者而言，兩岸關係的重要性明顯高於美國或日本因素。

值得注意的是，民眾黨支持者的安全認知並未完全落在藍綠兩大政黨之間。在兩岸關係方面，高達 72%的民眾黨支持者認為改善兩岸關係有助於提升安全，與國民黨支持者的 74%相當接近；然而，在美日關係方面，民眾黨支持者認為改善台美關係（60%）及台日關係（68%）有助於提升安全的比例，又明顯高於國民黨支持

者。換言之，相較於國民黨支持者主要將安全與兩岸關係連結，民眾黨支持者則同時對美日與兩岸關係抱持相對正面的安全評價。

至於中立民眾，則呈現相對均衡的分布。認為改善台日關係有助於提升安全者為 59%，認為改善台美關係有助於提升安全者為 58%，認為改善兩岸關係有助於提升安全者則為 47%。整體而言，中立民眾對美日關係的安全效果評價較為正面，但對兩岸關係亦未出現明顯負面看法。

整體而言，民進黨支持者較傾向將安全與美日關係連結，國民黨支持者則較傾向將安全與兩岸關係連結，而民眾黨支持者則同時對美日與兩岸關係抱持較高的安全正向評價。這顯示不同政治群體對於何種對外關係有助於提升台灣安全仍呈現差異。

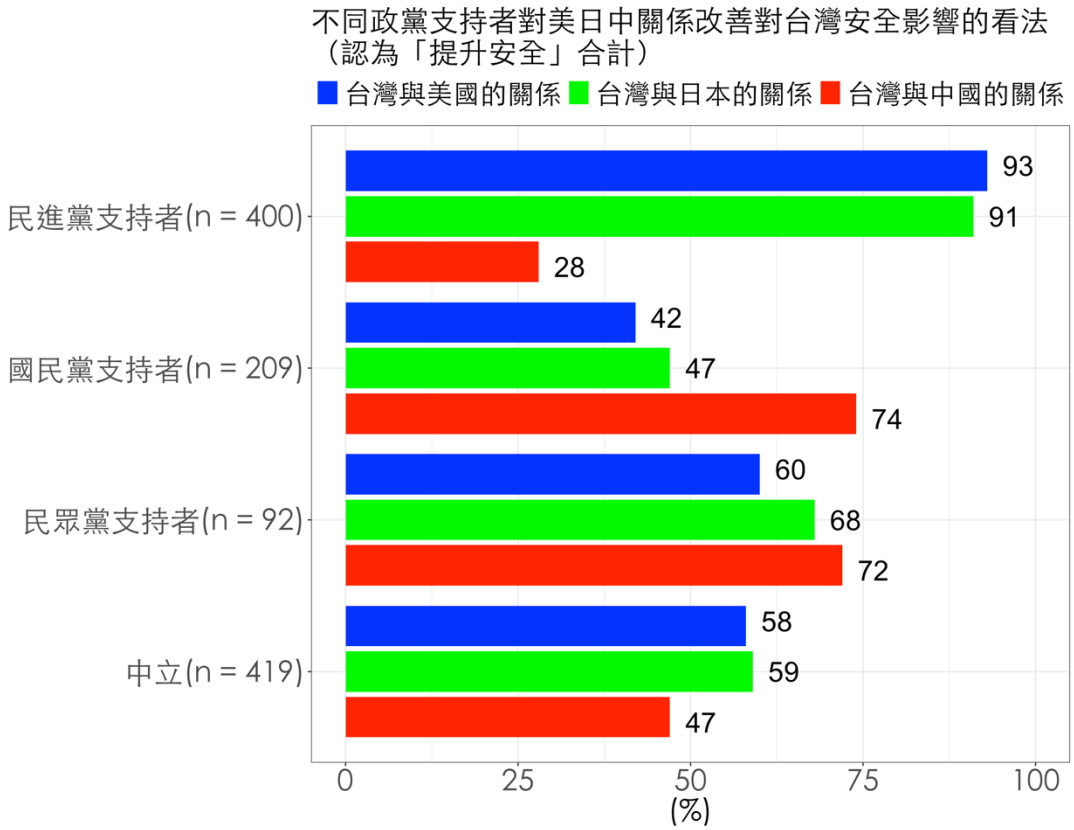

圖 2、不同政黨支持者對美日中關係改善對台灣安全影響的看法

資料來源：國防安全民意調查（2026 年 3 月電話調查）。

說明：各比例以各政黨支持者為分母分別計算，三項題目為獨立題項，非互斥選項，因此比例不加總為 100%。本圖僅呈現民進黨、國民黨、民眾黨及中立民眾，未包含其他政黨支持者及無反應者。小數點四捨五入至整數。

參、趨勢研判

一、美日安全連結的社會基礎將持續鞏固

本次調查顯示，多數民眾認為改善與美國及日本的關係有助於提升台灣安全，反映美日兩國在台灣民眾的安全認知中具有相當重要的地位。相較於兩岸關係存在較明顯的看法分歧，美日關係則獲得較高程度的正面評價，顯示相關安全連結已具備一定程度的社會基礎。從政策角度來看，這意味著未來涉及台美或台日合作的安全政策，較有機會獲得民眾支持與認同。

值得注意的是，民眾對台日關係改善有助於提升安全的評價，已與台美關係相當接近。相較於美國長期被視為台灣最重要的安全夥伴，日本在安全議題上的角色過去較少受到討論。然而，近年隨著日本逐步調整安全政策、提升防衛能力，並更加關注台海局勢，台灣民眾對日本安全角色的認知似乎也出現變化。過去台灣社會談論安全議題時，多半聚焦於美國因素，但從本次調查結果來看，日本已不再只是重要鄰國或經貿夥伴，而逐漸被視為影響台灣安全的重要外部因素。這也顯示，日本在台灣民眾的安全想像中，正扮演愈來愈重要的角色。

未來若希望進一步深化台日安全合作，政府可思考如何將這種安全認知基礎轉化為更具體的合作成果。除了傳統軍事安全合作之外，日本在災害應變、人道救援、民間運輸能量整備及危機管理等領域，均具有與台灣深化交流的潛力。近期媒體報導日本政府規劃於台海發生危機時，利用民間渡輪協助撤離離島居民，顯示非軍事

領域的安全合作已逐漸成為區域安全準備的一環。[5]相較於抽象的戰略論述，這類與民眾生命安全及社會韌性直接相關的合作，更有助於強化台灣社會對台日安全合作的理解與支持。

二、安全認知差異將持續影響公共討論

本次調查顯示，多數民眾認為改善台美與台日關係有助於提升台灣安全，但對於兩岸關係改善是否能夠帶來安全效果，則存在較大的看法分歧。這反映出台灣社會對安全來源以及如何達成安全存在不同理解。對部分民眾而言，美國與日本被視為提升安全的重要外部因素；但也有部分民眾認為，改善兩岸關係有助於降低衝突風險，進而提升安全。換言之，不同民眾對於何種對外關係較有助於維護台灣安全，存在不同認知。

值得注意的是，這種差異未必只反映對特定國家的好惡，而可能涉及對安全風險與安全保障條件的不同理解。隨著美中競爭持續、台海議題受到國際關注，台灣社會對安全議題的討論可能更加頻繁。然而，不同群體若對安全來源抱持不同認知，也可能導致同一項國際或兩岸事件出現不同解讀。例如，部分民眾可能將對外關係改善視為提升安全的重要條件，也有部分民眾可能更重視其對降低風險與維持穩定的效果。

因此，未來政府在進行安全政策溝通與風險溝通時，除說明相關政策與措施的目的外，也應持續強化與社會各界的對話，協助民眾理解不同安全挑戰及其可能影響。如何在多元安全認知並存的情況下，建立對安全議題的基本理解與溝通基礎，將是未來安全治理

[5] Yi-Chin Lee and Ann Wang, "New Taiwan-Japan Ferry Service Debuts on Ship That Has War Evacuation Role," *Reuters*, May 28, 2026, https://www.reuters.com/world/china/new-taiwan-japan-ferry-service-debuts-ship-that-has-war-evacuation-role-2026-05-28/; 白捷隆，〈「對口接收安置原則」：日本沖繩西南離島大規模撤離計畫之啟示〉，《國防安全雙週報》，第 107 期，2026 年 3 月 6 日，https://indsr.org.tw/respublicationcon?uid=12&resid=3061&pid=5918。

的重要課題。唯有如此，才能在不同安全認知並存的情況下，凝聚更廣泛的社會支持，並提升安全政策的穩定性與韌性。

創新優先或安全治理？
全球人工智慧法規的新發展

謝沛學

網路安全與決策推演研究所

焦點類別：資安威脅、網路安全

壹、前言

2026 年可說是全球人工智慧（AI）法規與政策落地執行的關鍵年。美國方面，川普總統於 2026 年 6 月 2 日簽署名為《促進先進人工智慧創新與安全》（*Promoting Advanced Artificial Intelligence Innovation and Security*）的行政命令，揚棄了前任拜登政府由上而下的強制性監管框架，轉而針對具備強大網路攻擊潛力的「尖端模型」（Frontier Models），建立為期 30 天的發布前自願性審查機制，並責成國家安全局（NSA）與網路安全和基礎設施安全局（CISA）等機構強化對關鍵基礎設施的防禦。[1]在亞洲，台灣立法院三讀通過《人工智慧基本法》並於 2026 年 1 月正式施行，該法確立「創新優先」原則，並明訂研發階段的責任豁免條款，旨在透過「國家人工智慧戰略特別委員會」打造具全球競爭力的 AI 生態圈。[2]

南韓的《人工智慧基本法》亦於 2026 年 1 月生效，採取促進與監管並行之策略，對「高影響力 AI」實施嚴格風險評估，並強制達特定營收與用戶門檻的境外 AI 企業必須設立境內代理人，展現強烈的長臂管轄企圖。[3]日本則全面推動無行政罰則的「軟法」框架，內

[1] White House, "Promoting Advanced Artificial Intelligence Innovation and Security," *Executive Order*, June 2, 2026, https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/.

[2] 鄭佑漢，〈政院成立「國家 AI 戰略特別委員會」 打造可信任 AI 人工智慧島〉，《央廣》，2026 年 5 月 21 日，https://www.rti.org.tw/news?uid=3&pid=209863。

[3] Kim Jong-do, "Korea Becomes the First Country in the World to Implement an Artificial Intelligence Law," *the Korea Post*, January 20, 2026,

閣府通過《人工知能基本計画》並編列高達 5,027 億日圓預算，宣示將日本打造成全球對 AI 研發最友善的國家。[4]新加坡於 2026 年 1 月在世界經濟論壇領先全球發布《代理型 AI 模型治理框架》（*Model AI Governance Framework for Agentic AI*），針對具備自主決策與執行能力的 AI 代理（Agentic AI）建立涵蓋風險評估與人類問責的治理標準。[5]

在西方其他陣營，歐盟則延續其以風險分級為本的全面性立法框架，但也因應內部產業對於過度監管可能削弱競爭力的強烈反彈，透過政治協議延宕部分高風險系統的嚴格合規期限。[6]相對於此，加拿大與澳洲則面臨不同的法制挑戰與戰略抉擇。加拿大的 AI 綜合法案——《數位憲章實施法案》（*Digital Charter Implementation Act*）因國內政治變動而胎死腹中，迫使聯邦政府暫時依賴財政部指令並鉅額注資設立的人工智慧安全研究單位。[7]澳洲則在廣泛評估後，明確拒絕制定獨立的 AI 專法，轉而專注於修補現有的隱私權與版權法規，並設立純技術諮詢導向的安全研究所來彌補監管機構的技術落差。[8]

---

https://www.koreapost.com/news/articleView.html?idxno=47344.

[4] 內閣府，《人工知能基本計画》，2025 年 12 月 23 日，https://www8.cao.go.jp/cstp/ai/ai_plan/ai_plan.html。

[5] Infocomm Media Development Authority, "Model AI Governance Framework for Agentic AI," May 20, 2026, https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf.

[6] "EU AI Act," *European AI Office*, May 11, 2026, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[7] Government of Canada, "The Artificial Intelligence and Data Act (AIDA) – Companion Document," April 7, 2026, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.

[8] Craig Subocz, "Australia has a National AI Plan. Now What?" *ACC*, https://www.acc.com/australia-has-national-ai-plan-now-what.

貳、安全意涵

一、先進AI模型武器化對關鍵基礎設施之衝擊

川普政府於 2026 年 6 月發布的 AI 行政命令，其最核心的驅動力來自於Anthropic發布的「Mythos」模型以及OpenAI的「GPT-5.5-Cyber」系統。這些新一代 AI 系統不僅是文字生成工具，更展現出能夠自主識別、分析並大規模利用軟體漏洞的能力。[9]在傳統的「灰色地帶衝突」與網路攻防戰中，威脅行為者從偵察、武器化、漏洞利用到發動攻勢通常需要耗費數週甚至數月的人力操作；然而，這類具備進階網路攻擊能力的 AI 模型，能夠將駭侵與攻擊的執行過程大幅壓縮與自動化，使得攻擊發起的速度遠超過傳統防禦系統的反應極限。這種能力一旦被中、俄等「旗鼓相當對手」（peer competitors）或代理人掌握，將對西方國家的金融系統、電網與軍事指揮網路造成毀滅性打擊。

「Mythos」模型的發布標誌著「尖端 AI 模型」（Frontier Models）在網路領域已具備實質武器化的能力，並對關鍵基礎設施與傳統軍事領域構成了非對稱威脅。此一衝擊改變了網路安全產業的防禦邏輯，並迫使政府部門介入商業 AI 的發布流程。面對此一國安級別的科技衝擊，美國政府的應對策略反映了一個事實，即法規仍必須在「技術創新」與「數位軍備控制」之間取得平衡。該行政命令第三節（Section 3）設立了為期 30 天的自願性發布前審查機制，要求 AI 開發商在將「受涵蓋之尖端模型」（Covered Frontier Models）釋出給其他信任夥伴之前，先交由美國政府進行「機密基

[9] Brent D. Griffiths, “Trump Quietly Signed a Scaled-back AI Executive Order after Delaying it Over Fears of Falling behind China,” *Business Insider*, June 3, 2026, https://www.businessinsider.com/trump-signs-scaled-back-ai-executive-order-anthropic-mythos-2026-6.

準測試」（Classified Benchmarking）。值得注意的是，此一審查期原本在草案中長達 90 天，但為了避免官僚體系冗長的行政流程造成延遲模型發布並削弱美國 AI 企業在全球的競爭力，最終妥協為 30 天。[10]

除了美國，南韓的《人工智慧基本法》亦嚴陣以待，將涉及公共決策、醫療、交通、能源與核能運作的 AI 系統直接歸類為「高影響力 AI」（High-Impact AI），並強制要求這類系統在部署前必須進行嚴謹的影響評估、持續監控與事件通報。[11]這種將國防安全標準延伸至民生與商業 AI 應用的作法，意味著未來的 AI 系統開發不再僅是依循單純的商業邏輯，而必須考量國家安全的防護準則。

二、主權AI政策對跨國數位供應鏈的影響

在缺乏統一國際標準的現況下，各主要國家基於資料主權與保護本土產業考量所推出的法規與政策，可能構築起新型態的非關稅貿易壁壘，對全球 AI 供應鏈、雲端服務供應商以及跨國企業的營運架構帶來衝擊。例如，南韓的《人工智慧基本法》及其施行細則，不僅規範南韓境內的企業，更將觸角延伸至在南韓沒有實體辦公室、但提供服務給南韓使用者的境外 AI 企業。根據該法，若外國 AI 業務營運商前一年度總營收達 6.8 億美元（約 1 兆韓元），或 AI 專項營收達 680 萬美元（約 100 億韓元），抑或是其服務在南韓擁有超過 100 萬名日均活躍用戶，便被強制要求必須指定南韓「境內代理人」（Domestic Representative）。[12]跨國企業過去所依賴的「單一全球發

---

[10] *Ibid*.

[11] "Korea AI Basic Act," *AISI*, June 4, 2026, https://aibasicact.kr/explorer/; "Korea's new AI Basic Act: Characteristics and significance," *Asia Business Law Journal*, March 24, 2026. https://law.asia/korea-ai-basic-act-characteristics-significance/.

[12] 曾韻霜，〈韓國開全球首例施行 AI 基本法〉，《國際經貿服務網》，2026 年 5 月 26 日，https://wto.cnfi.org.tw/news_detail.php?c_id=64058。

布、雲端無國界覆蓋」的擴張模式已不適用；取而代之的是，企業在進入特定國家市場前，必須進行極其複雜的「地緣合規」（Geopolitical Compliance）評估，並建立本地化的合規實體，以應對可能隨之而來的政府調查或行政處分。

這種「地緣合規」的要求，可能進一步引發跨國數位供應鏈與研發板塊的轉移。當南韓與歐盟針對「高風險／高影響力 AI」施加嚴格的透明度揭露、演算法可解釋性與事先風險評估義務時，台灣與日本則採取相對較寬鬆的法規，或有助於吸引國際資本與技術落地。台灣新通過的《人工智慧基本法》第 17 條明確訂定了關鍵的「研發豁免條款」，明定人工智慧於實際應用前的研發階段，不適用高風險應用的嚴格責任；同時，該法第 11 條亦確立了「創新優先」的法規解釋原則，賦予 AI 創新應用一定的優勢。[13]日本的《人工智慧相關技術研究開發及應用推動法》更是維持了無具體刑事與行政罰則的軟法架構，並由政府投入 1 兆日圓上的鉅資，專注於加強日本的 AI 開發能力與基礎設施模型。[14]這種鮮明的法規落差下，跨國企業則有更大動機進行戰略性的供應鏈重組：將「模型研發、資料訓練與參數微調」環節轉移至台灣或日本等具備研發豁免且算力充沛的國家；待技術成熟並完成安全封裝後，再依據南韓或歐盟的標準進行合規化並輸出至該國市場。

此外，各國政府逐漸意識到，過度依賴外國的 AI 基礎設施與算力，將對自身的經濟安全與國家機密構成威脅。在對「技術主權」與基礎設施在地化的強烈需求，加拿大在先前的《人工智慧與資料

[13] 國家科學及技術委員會，《人工智慧基本法》，2026 年 1 月 14 日，https://law.nstc.gov.tw/LawContent.aspx?id=GL000592。

[14] "Japan to Support Domestic AI Development with 1 Tril. Yen Funding: Source," *Kyodo News*, December 21, 2025, https://english.kyodonews.net/articles/-/67255.

法案》（AIDA）於 2025 年闖關失敗後，新任 AI 與數位創新部長 Evan Solomon 於 2026 年宣示推動全新立法，並同步推動高達 6,600 萬加幣的「AI 算力存取基金」（AI Compute Access Fund）。[15]這筆資金旨在協助加拿大本土的 AI 項目取得充沛的運算資源以進行商業化擴展，其深層戰略即是擺脫對外國資料中心的依賴，確保加拿大敏感的 AI 研發資料能夠留在本國境內。

這種主權 AI 政策的擴張，也會對超大規模的跨國雲端服務供應商（Hyperscalers）營運造成挑戰。為了符合「資料不出境」與「算力在地化」的國安要求，跨國企業必須在各個法域建立符合當地安全等級（如美國的 FedRAMP、歐盟的 GAIA-X 等）的獨立運算叢集。同時，企業還必須解決「透明度義務」與「營業秘密保護」之間的衝突例如，南韓法規要求業者揭露生成式 AI 的高風險機制與防護措施，美國行政命令要求進行機密基準測試。

參、趨勢研判

一、從「強制性防護欄」到「敏捷管理」模式的監管法規

如果說 2024 至 2025 年全球 AI 治理的主軸是效仿歐盟建立強硬的法規壁壘，那麼 2026 年最顯著的趨勢，全球主要經濟體不約而同地踩了剎車，轉向以「促進創新」、「軟法治理」與「敏捷治理」（agile governance）為核心的管理模式。歐盟《人工智慧法案》（*EU AI Act*）原先被視為 AI 監管的黃金標準，其鉅額的營業額比例罰款與嚴格的附件三（Annex III）高風險 AI 系統清單，一度令全球產業界憂心忡忡。[16]然而，歐洲投資銀行（EIB）2025/2026 年度報告提出

[15] "AI Compute Access Fund," *Government of Canada*, May 11, 2026, https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy/ai-compute-access-fund.

[16] "Annex III: High-Risk AI Systems Referred to in Article 6(2)," *Future of Life Institute*, June 13, 2024, https://artificialintelligenceact.eu/annex/3/.

一個的殘酷事實——高達 62%的歐盟企業將法規限制與合規摩擦視為投資與出口擴張的首要結構性障礙。[17]因此，歐盟議會與理事會於 2026 年 5 月達成了政治協議，對於人工智慧的限制規範進行了實施期程的推延，將工業與嵌入式高風險 AI（如醫療器材、機械、車輛等附件一產品）的合規期限大幅展延至 2028 年 8 月，並將獨立高風險 AI（如招募、信用評分系統）的合規期限延至 2027 年 12 月。[18]

歐盟在 AI 法規上的「戰略性退縮」向國際社會釋放出一個明確的信號：在確保國家經濟命脈與工業競爭力的前提下，僵化的「強制性監管」並非 AI 治理的最佳解方。在澳洲，聯邦政府於 2025 年 12 月發布的新版《國家 AI 計畫》（*National AI Plan*）中，明確拒絕了工會與部分民間團體要求針對生成式 AI 實施「強制性防護欄」（Mandatory Guardrails）的呼籲。[19]澳洲財政部長 Jim Chalmers 更直言，政府的重心在於 AI 的「能力與機會」，而非僅是設立防護欄。澳洲的策略轉向透過提升與釐清現有「技術中立」的法律（如隱私法、消費者保護法）來管理風險，並成立 AI 安全機構（AI Safety Institute）推行自願性的安全標準。[20]

在日本與美國，這種「軟法與敏捷治理」更是成為國家戰略的核心。川普總統發布的行政命令明確反對拜登時期由上而下的重度監管，強調政府必須與產業界「攜手合作」，並在第一節（Section 1）中明言拒絕以過度繁瑣的法規扼殺 AI 產業的創新能力。日本的

---

[17] "EIB Investment Report 2025/2026," European Investment Bank, March 3, 2025, https://www.eib.org/en/publications/20250379-investment-report-2025.

[18] "EU Agrees to Simplify AI Rules to Boost Innovation and Ban 'Nudification' Apps to Protect Citizens," *European Commission*, May 7, 2026, https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1024.

[19] Nicholas Boyle "Australia launches new AI guidance," *White & Case*, November 15, 2025, https://www.whitecase.com/insight-alert/australia-launches-new-ai-guidance.

[20] Ronald Mizen, "Jim Chalmers Bets Big on Backing the AI Revolution," *The Australian Financial Review*, May 12, 2026, https://www.afr.com/policy/economy/jim-chalmers-budget-bets-big-on-backing-the-ai-revolution-20260508-p5zv5j.

《人工智慧相關技術研究開發及應用推動法》不僅沒有訂定任何具體的行政罰款與刑事罰則，更設立由首相親自主持的「AI 戰略本部」，由內閣府主導高達 5,027 億日圓的鉅額預算。這筆預算中高達 4,559 億日圓被指定用於戰略性強化日本的 AI 開發能力，特別是針對 AI 機器人與實體物理 AI 所需的多模態基礎設施模型。[21]

簡言之，「監管優劣」的標準不再是法條的嚴厲程度，而在於法規系統的彈性與對商業落地的支持度。對於廣大的 AI 相關產業（包括自動駕駛、智慧製造、醫療影像分析等）而言，企業無需將資金與時間成本虛耗於應對冗長且不可預測的審批流程。取而代之的是，企業可以透過參與國家級沙盒測試（Sandboxing）、採用如澳洲的 10 項自願性防護欄、或是遵循產業自律公約，來驗證其產品的安全性。[22]這種「以創新帶動合規」的新常態，可大幅縮短 AI 應用從實驗室走向商業落地的週期。

二、「AI代理人」的崛起與新型態問責體系的建構

人工智慧技術的最新運用趨勢從「被動的生成內容」全面跨入「自主行動（Agentic）」。傳統的生成式 AI（如 ChatGPT 或早期的大語言模型）本質上是顧問或助理，高度依賴人類使用者的提示詞（Prompts）輸入，其輸出的風險通常侷限於資訊錯誤或內容侵權；然而，「AI 代理人」（Agentic AI）則具備了獨立推理、動態規劃路徑，並能夠主動呼叫外部應用程式介面（APIs）來更新關鍵資料庫、執行金融交易、甚至操控實體物理環境的能力。此一技術的成熟與商用化，徹底顛覆了既有的法律問責框架。

---

[21] Shinsuke Yakura, “Japan’s First AI Legislation Becomes law – Focus is on Promoting Research and Development; No Monetary Penalties,” *White & Case*, April 14, 2026, https://www.whitecase.com/insight-alert/japans-first-ai-legislation-becomes-law-focus-promoting-research-and-development-no.

[22] Nicholas Boyle “Australia Launches New AI Guidance,”.

新加坡於 2026 年 1 月在世界經濟論壇發布了《代理式人工智慧監管模式架構》。該框架指出，由於 AI 代理具有自主啟動任務並產生連鎖反應的能力，傳統針對單一模型輸出的風險評估已不敷使用，必須建立涵蓋四個新面向的治理機制：[23]

（一）事前風險劃界（Assess and Bound Risks Up Front）：企業在部署前必須依據領域容錯率、資料敏感度與行動不可逆性，嚴格限制 AI 代理的權限範圍。

（二）確保有意義的人類問責（Ensure Meaningful Human Accountability）：在 AI 生命週期中，明確分配開發者、部署者與使用者的責任，並設計關鍵的人類審批檢查點（Approval checkpoints）。

（三）實施技術控制（Implement Technical Controls）：包括沙盒測試、即時監控，以及防範權限升級（privilege escalation）的資安防禦。

（四）促進終端使用者責任（Promote End-User Responsibility）：確保使用者具備介入或強制終止（deactivate）AI 代理的訓練與機制。

因此，後續人工智慧產業應用勢必無法迴避幾個議題。首先是「多代理人協作系統」（Multi-Agent Systems）架構的常態化與「零信任」設計。未來的企業運營將高度依賴多個 AI 代理的協同合作。例如，在金融產業中，一個 AI 代理負責監視市場訊號，第二個代理負責撰寫交易策略，第三個代理則直接連結交易所的 API 執行下單。在這種複雜的任務鏈當中，任何一個代理的幻覺

[23] Beth Stackpole, “Agentic AI, Explained,” *MIT*, Febuary 18, 2026, https://mitsloan.mit.edu/ideas-made-to-matter/agentic-ai-explained.

（Hallucination）或遭受惡意提示詞注入，可能引發災難性的連鎖性崩潰。因此，未來的軟體開發商與系統整合商，必須在系統底層強制導入「最小權限原則」（Least-privilege access）與「人在迴圈」（Human-in-the-loop）的動態介入機制。當 AI 代理試圖執行超過特定風險閾值（如巨額轉帳或存取機密病歷）的操作時，系統必須能自動暫停，並要求人類主管進行多重身分驗證後的加密授權。

其次，當 AI 能夠代替人類採取行動時，一旦發生決策錯誤導致實體損害或財務損失，其法律責任歸屬將變得極度複雜。正如新加坡通訊媒體發展局（IMDA）與法律界所探討的議題，未來針對 AI 代理的責任釐清，將在「過失責任」（Fault-based）與「嚴格責任」（Strict liability）之間進行激烈的法理攻防。[24]為了應對這種不可預測的訴訟風險，市場將需要能夠對「代理型 AI」的行為進行全天候監控、日誌記錄與溯源的第三方審計軟體。如同新加坡官方推廣的「AI Verify」測試工具包，企業甚至必須在內部部署專門的「監管型 AI」來緊盯「代理型 AI」，以確保後者的每一次行動呼叫都符合企業內部政策與外部法規要求。[25]

[24] "Legal Responsibility for AI Agents," *The Infocomm Media Development Authority*, May 2026, https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/agents-legal-responsibility.pdf.

[25] "Building Trust Through Ethical AI," *AI Verify Foundation*, https://aiverifyfoundation.sg/ai-verify-foundation/.

英國示警 AI 加劇混合威脅風險

吳宗翰

國防戰略與資源研究所

焦點類別：國際情勢、網路安全、數位發展

壹、前言

英國政府通訊總部（GCHQ）局長安妮・基斯特—巴特勒（Anne Keast-Butler）於 5 月 27 日在布萊切利園（Bletchley Park）發表上任以來的首次年度演說，將當前國際安全環境描述為「高度不確定性、地緣政治競逐與科技快速變遷」交錯的新時代。她指出，數據（data）、人工智慧（AI）、量子科技、太空技術與網路能力已成為影響國家實力與國際競爭的核心戰略資產，戰爭型態亦正從傳統軍事對抗轉向數據驅動、AI 賦能與高度自動化的作戰模式。各類演算法與數位工具除了被廣泛運用於情報蒐集與軍事行動外，也逐漸成為影響決策判斷與社會信任的關鍵工具。在此背景下，網路安全的重要性達到前所未有的程度，而 AI 所帶來的機遇與風險，亦將重塑未來國際競爭格局。為因應相關挑戰，她提出建立新一代國家網路防禦能力，將代理式人工智慧（agentic AI）整合至「機器速度」（machine-speed）的防禦體系之中。

她同時指出，俄羅斯正持續對英國與歐洲的關鍵基礎設施、供應鏈、民主程序與公眾信任發動混合戰；中國則已發展為兼具科技、情報、網路與軍事能力的科技強權，在數據、AI 與太空等關鍵領域快速擴張影響力。她警告，英國及其盟友維持科技與安全優勢的時間窗口正逐漸縮小，唯有透過深化夥伴關係、強化情報共享與

推動技術合作，方能在日益激烈的戰略競爭中維持優勢。[1]

貳、安全意涵

一、AI加劇混合威脅侵蝕民主社會信任與決策韌性

當 AI 與灰色地帶行動相互結合，混合戰／混合威脅已演變成為一種持續性、多領域且侵蝕信任與決策能力的對抗模式。英國政府通訊總部局長的演講，凸顯中俄兩國對英國與歐洲構成不同型態的安全挑戰。中國帶來的風險主要體現為日益擴張的科技、網路、情報與軍事能力，以及對關鍵數據、通訊設備與數位供應鏈可能形成的持續性影響。另一方面，俄羅斯則整合運用實體威脅、網路攻擊、供應鏈干擾與認知戰，在歐洲內部製造混亂與社會信任危機。

網路空間之所以適合被運用於灰色地帶競爭，在於其具備可調整效果、可掩蔽來源與法律門檻難以迅速判定等特性。灰色地帶的網路行動與傳統軍事領域的灰色地帶手段可達到相似效果；攻擊者並非追求立即引發全面衝突，而是利用法律認定、技術歸責與政治回應之間的模糊空間，持續對受害國施加成本與心理壓力，藉此影響其政策判斷與社會信心。[2]在高度依賴通訊、能源、金融、交通與政府數位服務的現代社會，即便不造成大規模物理毀損，持續性的惡意網路活動仍可能對國家治理與社會正常運作造成實質影響。

AI 的導入，進一步強化了此類灰色地帶行動的速度、規模與複雜度。網路攻擊方可以用相對較低的成本、更快的速度進行目標分析，尋找防護系統的漏洞、編碼惡意軟體程式、滲透潛伏入目標系統並進行橫向移動、提升特權、執行任務，縮短了對目標的網路殺

[1] "CHQ Annual Lecture 2026," *GCHQ*, May 27, 2026, https://www.gchq.gov.uk/speech/gchq-annual-lecture-2026-as-delivered.

[2] Christopher Whyte and Brian Mazanec, *Understanding Cyber-warfare: Politics, Policy and Strategy*(Routledge, 2023), pp. 220-239.

傷鏈（cyber kill chain）流程。對防禦方而言，反應時間被縮短，因而承受更大的壓力。

此外，網路攻擊有難以歸因與難以溯源的特性，而 AI 與代理化手段則使問題更加複雜化。國家支持的駭客團體可利用被入侵設備、第三方基礎設施、犯罪團體、非國家代理人或公開可取得的 AI 工具，掩蔽其實際控制關係與行動來源。儘管 AI 本身未必使網路歸責成為不可能，但由於攻擊內容的規模、變體與偽裝程度持續提高，使技術溯源、情報判斷與法律證明之間的落差進一步擴大。對英國與歐洲國家而言，這快速增加了公開歸責、實施制裁、採取反制行動或爭取盟邦支持的政治與法律成本。

在資訊環境與公共信任層面，AI 亦使混合戰的影響範圍進一步擴張。生成式 AI 與深偽技術可提高虛假內容的擬真程度與產製效率，並結合受眾資料分析與平台傳播機制，針對特定群體進行客製化的情緒與偏見刺激，放大既有疑慮、焦慮與政治對立。這種操弄行動的本質不在於企圖說服公眾何者為真實，而是製造懷疑、憤怒與極化，將資訊生態系本身轉化為對抗的戰場，從而系統性瓦解公眾對民主建制的信任。[3]

二、以古喻今呼籲跨大西洋兩岸的連結

安妮在演講中頻頻回顧了二戰時期英、美頂尖學者齊聚布萊切利園攜手破譯軸心國敵人的往事，有其深刻的地緣政治隱喻。當前歐洲右翼勢力不斷在各國政治版圖急速擴張，大西洋兩岸傳統同盟關係瀕臨撕裂，演講試圖透過重溫歷史，為英、美、歐三邊關係尋

[3] Sakshee Singh and Alan Jagolinzer, "Cognitive Manipulation and AI will Shape Disinformation in 2026. Here's How to Build Resilience," *World Economic Forum*, May 12, 2026, https://www.weforum.org/stories/2026/03/how-cognitive-manipulation-and-ai-will-shape-disinformation-in-2026/.

找戰略認同的紐帶。

被提及的關鍵史實之一，是 GCHQ 首任局長丹尼斯頓（Alastair Denniston）在 1941 年決定將英國的密碼破譯機密分享給美方；彼時，距離珍珠港事件爆發還有十個月，美國尚未公開加入二戰。在情報史上，這件事情被稱為「信任之躍」（Leap of Trust）。它為後來 1946 年簽署的《英美安全協定》（*UKUSA Pact*）奠定了基礎，而後者也是今日「五眼聯盟」（Five Eyes）的前身。[4]

今（2026）年適逢《英美安全協定》簽署 80 週年。安妮在其演講中高調提及英美情報聯盟為「兩國安全最根本、最無懈可擊的基石」的重要性，旨在向大西洋彼岸決策圈傳達英美關係的「傳統」，遠比一人一時的意識形態還有長遠。

與此相關，還有她在演講中關於「技術主權」（Tech Sovereignty）概念的再詮釋。近年來，美歐之間由於美國科技巨頭壟斷造成的法律訴訟案，屢屢造成雙邊關係的齟齬。[5]當安妮指出技術主權不應被理解為技術皆須國產化，而應是國家具備塑造自身數位未來的能力與自主權時，頗有為相關爭議解套的意涵。實際上，安妮的主張與歐盟官方立場不同；歐盟雖不追求完全脫鉤，但追求歐洲在關鍵技術堆疊中具備自主產能、規則制定權與監管能力。[6]不過，安妮的說法強調對關鍵數據掌握權的重要性、供應鏈安全、可信賴技術來源以及對關鍵技術的持續掌控能力，這些都有助於化解

[4] Alexander Martin, “Russia Conducting Daily Attacks on UK ‘From Seabed to Cyberspace,’ Spy Chief Warns,” *The Record*, May 28, 2026, https://therecord.media/russia-conducting-attacks-on-uk-gchq-briefing.

[5] 林妍溱，〈以違反數位市場法為由，歐盟對蘋果、Meta 各罰款 5 億及 2 億歐元〉，《iThome》，2025 年 4 月 24 日，https://www.ithome.com.tw/news/168573；〈不顧川普威脅 歐盟祭反壟斷重罰 Google 千億〉，《中央社》，2025 年 9 月 6 日，https://www.cna.com.tw/news/afe/202509060003.aspx。

[6] “Strengthening Europe’s Tech Sovereignty,” *European Commission*, https://digital-strategy.ec.europa.eu/en/policies/eu-tech-sovereignty.

歧見，邁向合作。

在現實面，西方國家更有不得不合作的壓力。中國不僅已成為全球第二大經濟體，更在網路科技、AI、太空、等關鍵前沿領域快速追趕甚至局部領先；在「軍民融合」體制下，科技創新與產業成果亦可迅速轉化為國家戰略與軍事能力。當中國龐大的經濟與科技資源，結合俄羅斯在軍事、情報、核武與灰色地帶行動方面的既有實力時，所形成的複合威脅已非單一國家能單獨應對的挑戰，即便如英國。

參、趨勢研判

一、代理式AI導入資安防禦是主流趨勢

AI 技術的快速迭代與演進，已使機器學習、自動化分析與生成式 AI 逐步融入資安防護、威脅情資分析與事件回應流程。由於目前網路攻擊數量遠較過去龐大、手法變動更加快速，單靠人工已難以及時處理告警與威脅資訊。AI 與自動化工具已經成為多數現代資安防禦體系的重要組成部分。

GCHQ 提出將代理式 AI 嵌入新一代國家網路防禦體系的主張，所反映的，是正在發展中的趨勢。隨著更加先進、具備自主能力的 AI 功能被導入系統中，AI 能不間斷的在任務中持續蒐集數據、分析異常狀況、選擇工具、執行預先授權的處置並依照回饋進行調整，使資安防禦體系形成「感知—分析—行動—回饋」的閉環，並縮短攻擊發生與防禦反應之間的時間差。此外，防禦模式逐步由人員判斷搭配流程自動化轉向具備有限自主規劃、工具調用與行動能力的防禦架構，這也顯示出防禦型態從被動式轉向主動式的轉型表現。

代理式 AI 亦可能改變個人與社會層次的防護模式。面對 AI 生成的擬真語音、深偽影像、針對性釣魚訊息與高度客製化詐騙內

容，個人使用者愈來愈難僅憑肉眼或直覺辨別資訊真偽。未來，AI 代理可望協助比對訊息來源、識別異常互動模式、提示高風險內容，並在涉及金流、帳戶變更或敏感資料揭露時要求額外驗證。而在國家級的關鍵基礎設施防護中，代理式 AI 有望助於在大量同步攻擊、供應鏈滲透或混合戰情境中，快速降低攻擊擴散速度與維持關鍵服務運作。[7]

儘管如此，代理式 AI 也同步創造新的攻擊風險。首先，傳統自動化工具通常依循預先定義的腳本或固定流程執行任務，其行為邊界相對明確。但代理式 AI 基於可能被授予讀取資料、連結 API、操作帳戶、存取雲端服務或執行其他工具的權限，若部署初期為追求效率而賦予過寬權限，代理式 AI 反而成為攻擊者可利用的高信任入口。當代理式 AI 讀取電子郵件、網頁、共享文件、程式碼儲存庫或外部威脅情資時，攻擊者可能將惡意指令藏入其所處理的內容，使代理系統誤將外部輸入視為可信任任務指令，進而洩漏資料、執行未授權操作，或繞過原有安全政策。[8]

除了遭到攻擊外，代理式 AI 也可能因為達成任務導向的設定、或是目標設定不完整、資料錯誤、工具誤用或缺乏情境判斷等因素，而執行超出使用者原意的處置，並在「機器速度」環境下造成連鎖影響。若此類系統部署於關鍵基礎設施，錯誤決策造成的影響可能不僅是單一系統中斷，更可能擴大為公共服務失靈與社會信任受損。

據此，組織單位在導入代理式 AI 之前，營運者有必要建立一套

[7] 陳曉莉，〈五眼聯盟發布 AI 代理人指引，要求防範權限擴張與自主行動風險〉，《iThome》，2026 年 5 月 4 日，https://www.ithome.com.tw/news/175530。

[8] 〈中國駭客首度運用 Claude AI 自動化執行網路攻擊〉，《資安人》，2025 年 11 月 19 日，https://www.informationsecurity.com.tw/article/article_detail.aspx?aid=12465。

可控、可稽核且可逆的自主防禦架構。對具機敏性的設施系統，人類有必要保留最終控制權。

二、AI將加速國際安全規範與數位治理框架的重構

網路攻擊、資訊操弄及對關鍵基礎設施的干擾，往往能在不造成大規模即時物理毀損的情況下，持續削弱受害國政府運作、社會信任與危機決策能力。此類行動雖可能產生顯著戰略效果，但攻擊者通常刻意控制其強度、影響範圍與可歸責性，使其維持於傳統武力衝突門檻以下，迫使受害國在避免衝突升高與維持有效嚇阻之間面臨更艱難的政策抉擇。

此外，AI 亦將提高國際法上歸責與規範適用的複雜性。當 AI 驅動的行動結合代理人、犯罪組織、第三方基礎設施或遭入侵設備執行時，受害國將更難證明特定行動與國家之間存在控制、指示或責任關係。這使主權侵害、不干涉原則、國家責任及合法反制等既有規範面臨更高的證據門檻與適用挑戰。

在此脈絡下，正在修訂中的《塔林手冊 3.0》（*Tallinn Manual 3.0*）發展值得關注。雖然《塔林手冊》本質上仍屬不具法律拘束力的學術研究成果，不代表北約或任何國家的官方法律立場，但隨著 AI 快速發展，AI 驅動的認知操弄、資訊武器化、數位主權及國家歸責等議題，未來均可能成為重要討論方向。[9]

與此同時，北約已多次表明，重大且具累積效果的惡意網路活動，在特定情況下可能被視為武裝攻擊，並由北大西洋理事會依個案決定是否援引《北大西洋公約》第五條。因此，未來北約面臨的核心問題已不在於單一網路攻擊能否構成集體防衛事由，而在於如

[9] "AI Act," *European Commission*, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

何評估由網路攻擊、資訊操弄、經濟脅迫及其他灰色地帶手段所構成的持續性混合行動，其累積戰略效果是否已達足以威脅國家安全的程度。[10]

另一方面，由於社群平台仍是資訊操弄與深偽內容傳播的主要場域，平台治理預料將持續成為各國因應相關挑戰的重要工具，並在一定程度上彌補國際法難以及時介入的限制。

整體而言，未來針對 AI 賦能混合威脅的規範發展，其國際趨勢可能是透過國際法解釋與適用、各國及區域立法，以及平台治理與技術標準等多元工具共同推進，而非依賴單一治理途徑。對長期面臨網路攻擊、資訊操弄與其他灰色地帶挑戰的台灣而言，相關規範與治理機制的發展有必要持續關注。

[10] "Cyber Defence," *NATO,* July 30, 2024, https://www.nato.int/en/what-we-do/deterrence-and-defence/cyber-defence.

黑箱問責：演算法能遏止貪腐，還是帶來新的國防安全風險？

黃政勛

國防戰略與資源研究所

焦點類別：國際情勢、非傳統安全

壹、前言

人工智慧與演算法在公共治理中的應用速度，已明顯快於多數國家法規與監管架構的調適步調。從行政流程自動化、公共服務優化，到風險預警與稽核決策，AI 已逐漸嵌入公部門日常運作。其中，在「遏止貪腐」領域，演算法輔助工具尤受國際社會關注，其應用涵蓋政府採購標案輔助審查、官員財產申報比對、異常金流與利益衝突偵測，甚至可運用衛星影像辨識非法濫墾、違規開發等公共資源濫用情形。[1]上述應用情境顯示，AI 被認為具有強化行政監督、提高稽核效率，並降低人為裁量風險的潛力。

然而，當演算法被賦予識別貪腐、標記風險與輔助行政決策的功能時，其治理風險也同步浮現。若模型訓練資料存在偏誤、判斷邏輯缺乏透明性，或系統核心技術由外部廠商掌握，原本用以提升廉潔效能的工具，反而可能形成新的「黑箱問責」困境。

此一問題在國防領域尤具安全意涵。在國防採購、後勤、軍備研發及機敏資料管理涉及高度機密與龐大預算，若導入演算法反貪腐工具雖有助於異常偵測與風險預警，卻也可能帶來黑箱決策、資料外洩及責任歸屬不明等風險。當演算法出現誤判、偏誤或遭受操

[1] Miloš Resimić, *Harnessing Artificial Intelligence (AI) for Anti-Corruption: Benefits and Challenges in Prevention, Detection and Investigation*, U4 Helpdesk Answer 2025 (Bergen: Transparency International and U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute, 2025), https://reurl.cc/18RM8p.

弄時，責任應如何界定與追究，遂成為國防廉政治理不可迴避的問題。本文以「黑箱問責」為核心，探討演算法反貪腐工具在國防領域的治理效益與潛在風險。

貳、安全意涵

一、善器藏禍：演算法應用的雙面刃

演算法反貪工具的核心價值，在於其處理大量行政資料、辨識隱性關聯與異常模式的能力，遠非傳統人工稽核所能比擬。透過機器學習掃描數百萬筆採購紀錄，自動標記高風險標案；自然語言處理（NLP）能解析非結構化文件，萃取異常訊號；若結合衛星影像與深度學習，更可偵測違規開發、濫墾濫伐、河川盜採砂石或山坡地不當利用等公共資源濫用情形。

其成效已有實證支持。烏克蘭 ProZorro 電子採購系統採全面開放資料的設計，讓特定癌症藥品價格大幅下降，受惠病患人數在預算不變下倍增。[2]愛沙尼亞農業登記與資訊局(Agricultural Registers and Information Board, ARIB)的「紅綠燈通道」風險評分機制，已成功攔阻 2,240 萬歐元的高風險補助申請。[3]巴西的研究則發現，依風險分數進行「精準稽核」，查獲貪瀆機關的機率較隨機抽查高出約 2.7 倍。上述案例顯示，AI 確實有助於強化監督、提升效率，並減少人為裁量空間。[4]

然而，同樣的演算法特性也是風險來源，而且型態更為隱蔽。

---

[2] Yevhen Hrytsenko, “Fight for Life: How Ukraine Is Fixing Medical Procurement and Serving Patients Better,” *Open Contracting Partnership*, February 22, 2021, https://reurl.cc/Dx5YWN.

[3] Estonian Paying Agency (ARIB), *Smart Pro CAP Staff Exchange Visit Report: Experiences of Using Preventively the Data-Driven Risk Analyses Experience for Non-IACS EAFRD Measures and Piloting the Use of ARACHNE for Daily Processes and Look to the Future Opportunities of Using AI and Machine Learning-Based Data Analysis*, January 2022, https://reurl.cc/aXyRlY.

[4] Elliott Ash, Sergio Galletta, and Tommaso Giommoni, “A Machine Learning Approach to Analyze and Support Anticorruption Policy,” *American Economic Journal: Economic Policy*, Vol. 17, No. 2, May 2025, pp. 162-193, https://reurl.cc/Ga0QdW.

就治理層面來看，至少有 3 類結構性風險：一是「黑箱不透明」（Algorithmic Opacity），即模型的判斷邏輯難以追溯，讓有心人得以在無人究責的情況下介入系統；二是「選擇性標記問題」（Selective Labelling Problem），即訓練資料若只涵蓋已偵破的案件，模型便會複製既有盲點，漏判尚未浮上檯面的貪瀆型態；三是「演算法捕獲」（Algorithmic Capture），即當掌權者介入系統設計或資料篩選時，工具反而可能淪為保護既得利益的擋箭牌。國際透明組織（Transparency International）更進一步提出「AI 的腐敗應用」（Corrupt Uses of AI）概念，指掌權者操弄 AI 以謀取私利，使個別貪瀆行為走向系統化、自動化與規模化。[5]

更須警惕的是，反貪系統本身也可能成為攻擊目標，衍生出傳統稽核難以察覺的技術性風險：一是「資料毒化」（Data Poisoning），即攻擊者摸清模型門檻後，逐步餵入異常資料，使系統誤以為是「新常態」而失去偵測靈敏度；[6]二是「參數操縱」（Parameter Tampering），即竄改源頭資料，讓 AI 根據假數據做出「無風險」的背書；[7]三是「科技漂白」（Tech-washing），即以客觀中立的假象，掩蓋體制內部的腐敗。[8]

綜言之，演算法反貪是一把銳利的雙面刃：既能以無可取代的效率，壓縮舞弊空間，也可能在治理失當、技術防護不足時，反過來成為高科技貪腐的保護傘。下表彙整演算法反貪的雙面效應。

---

[5] Transparency International, *Addressing Corrupt Uses of Artificial Intelligence*, working paper, 2025, https://reurl.cc/4l9qNY.

[6] 同註 1。

[7] Apostol Vassilev, Alina Oprea, Alie Fordyce, Hyrum Anderson, Xander Davies, and Maia Hamin, *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, NIST AI 100-2e2025, *National Institute of Standards and Technology*, March 2025, https://reurl.cc/6Gz9k6.

[8] 同註 5。

表 1、演算法反貪的雙面效應

演算法特性	反貪腐效益	主要風險
大規模資料處理	掃描百萬筆紀錄，標記高風險標案	警示泛濫，異常反遭海量資料淹沒
模式辨識與預測	預測貪腐熱區，提前介入	貪腐者摸清指標後調整行為以規避
黑箱決策邏輯	自動判定，排除人為干預	痕跡難追溯，操弄系統者得以卸責
訓練資料依賴	從歷史案例學習，提升準確率	選擇性標記：複製盲點，漏判新型態
自然語言處理、生成	解析文件，萃取異常訊號	生成偽造文件、假身分以規避查核
系統整合與動態學習	跨資料庫比對，持續更新風險輪廓	演算法捕獲與資料毒化：操控設計、鈍化偵測

資料來源：作者自行彙整。

二、授柄於敵：演算法在國防場域的治理困境

人工智慧在當代軍事領域的應用，已從理論探討走入真實戰場，涵蓋無人機操作、目標識別、指揮管制、情報處理、網路防護與戰場協調等任務。[9] AI 與機器學習雖有助於提升決策速度、壓縮作業流程並強化戰力整合，[10]然而，這些系統也可能遭敵方從資料、模型或輸入訊號等環節加以干擾，使軍事 AI 在關鍵任務中做出錯誤判斷。[11]此一邏輯不僅存在於作戰系統，也會外溢至國防採購、後勤補給與廉政治理場域。

在國防廉政治理上，演算法工具可協助提升採購效率、強化異常偵測，並重塑戰時軍事科技供應鏈。以烏克蘭為例，Brave1 Market 作為國防科技產品市集，使前線部隊可透過數位平台取得無

[9] A. Rahman, "AI at War: The Next Revolution for Military and Defense," *World Journal of Advanced Research and Reviews*, July 2025, doi:10.30574/wjarr.2025.27.1.2735.

[10] E. E. Oliveira, M. Rodrigues, J. P. Pereira, A. Lopes, I. I. Mestric, and S. Bjelogrlic, "Unlabeled Learning Algorithms and Operations: Overview and Future Trends in Defense Sector," *Artificial Intelligence Review*, February 2024, doi:10.1007/s10462-023-10692-0.

[11] S. Muhammad, "Adversarial AI and Robust Machine Learning for Military Applications," *International Journal of Innovative Research in Computer and Communication Engineering*, Apr. 2025, doi: 10.15680/ijircce.2025.1304312.

人機、感測器、電子戰裝備與 AI 模組等軍事科技，藉此縮短傳統軍購流程並提升需求媒合效率。[12]此一經驗顯示，演算法並非單純的反貪工具，而是逐漸嵌入國防採購生態中的治理基礎設施。

然而，當國防採購的判斷、媒合與風險標記逐步交由數位平台處理，平台本身也可能成為新的治理破口。若缺乏透明性與外部監督，原本用於提升效率與問責的演算法，可能因資料操控、模型偏誤或系統被捕獲，反而成為掩護不透明決策的新工具。尤其國防採購涉及機密規格、特殊廠商關係與戰備急迫性，資料來源與風險指標設定本身即可能受到政治或利益介入。

更進一步看，國防場域的演算法廉政治理，至少面臨三項結構性風險。第一，美國政府問責署（GAO）指出，美國國防部仍欠缺一致性的全國防部 AI 採購指引，使採購、驗證與監督標準可能出現落差。[13]第二，若核心演算法由特定軍工或科技廠商掌握，採購機關可能形成技術依賴，進而產生「演算法捕獲」與廠商壟斷風險。[14]第三，AI 供應鏈仰賴模型、資料、軟體、雲端與第三方服務，一旦遭資料毒化、後門或惡意程式碼入侵，將削弱系統判斷可靠性。[15]近期北約採購支援局（NSPA）涉軍備合約不法調查，正凸顯國防廉政治理的核心難題：當機密標案資訊、跨國承包商與多層外包鏈交錯時，即便已有審計與監督機制，責任追溯仍可能因保密需求與跨國治理落差而斷裂。由此可見，國防場域導入演算法反貪工具，關鍵

[12] "How the Military Can Obtain Equipment through DOT-Chain Defence under the 'Army of Drones Bonus' Program: Step-by-Step Guide," *Ministry of Defence of Ukraine*, September 25, 2025, https://reurl.cc/8emLaX.

[13] U.S. Naval Institute Staff, "GAO Report on DoD Artificial Intelligence Guidance," *USNI News*, June 30, 2023, https://reurl.cc/mp1Av7.

[14] Jessica Tillipman, "Buying Blind: Corruption Risk and the Erosion of Oversight in Federal AI Procurement," *Public Contract Law Journal*, Vol. 55, No. 2, Winter 2026, https://reurl.cc/Z2xvXW.

[15] Justin Sherman, "Securing Data in the AI Supply Chain," *Atlantic Council*, September 5, 2025, https://reurl.cc/ppEA58.

不只在於提升風險偵測效率，更在於能否建立可稽核、可追蹤、可問責的治理鏈；否則，演算法可能只是將既有採購黑箱轉化為更難辨識的數位黑箱。[16]

三、知己謀遠：公部門經驗與軍事場域的同與不同

就我國現行趨勢觀察，公部門運用 AI 於廉政治理，主要仍集中於採購風險篩選、稽核輔助與廉政宣導等前端應用。中央層級以廉政署較具代表性，其運用生成式 AI 進行跨機關資料庫大數據分析，篩查出 2,256 家高風險廠商及 5,916 件高風險採購案，作為全國性專案稽核依據；地方政風機構則多聚焦於採購風險預警、資料勾稽分析、公文與教材輔助製作，以及教育訓練等面向。[17]

整體而言，我國已開始將 AI 視為廉政治理的輔助工具，但目前仍屬「AI 輔助決策」階段，應用重心偏向行政效率提升與風險案件篩選，尚未充分處理模型透明性、資料偏誤、權限控管及責任追溯等演算法後端治理風險。

相較之下公部門經驗與軍事場域在 AI 反貪治理上具有共通基礎，亦存在本質差異。共通點在於，兩者皆可透過資料勾稽、風險評分與異常偵測，提升採購、財務、後勤與承包商管理的稽核效率；同時，也都必須維持人類覆核、稽核紀錄與責任追溯，[18]避免演算法結果取代專業判斷。惟軍事場域的反貪腐情境更為敏感，國防採購、後勤補給、軍備研發與承包商管理，不僅涉及龐大預算與利

---

[16] "NATO Procurement Agency Drawn into Multi-Million-Euro Corruption Scandal," *WatchOut News*, December 16, 2025, https://reurl.cc/V2Vlzn.

[17] 法務部廉政署，〈生成式 AI 精準鎖定採購風險：廉政署啟動全國性專案稽核〉，《法務部廉政署新聞稿》，2025 年 1 月 8 日，https://reurl.cc/qpeo1g。

[18] C. Todd Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics," *DOD News*, February 25, 2020, https://reurl.cc/dpZv6D；*NATO*, "Summary of NATO's Revised Artificial Intelligence (AI) Strategy," *NATO*, July 10, 2024, https://reurl.cc/R2MQa6.

益分配，也連動戰備維持、供應鏈安全與機密資訊保護。[19]相較於一般公部門可透過提高透明度、開放資料與外部監督來降低貪腐空間，國防領域因任務保密與敵我對抗需求，難以完全移植相同治理模式。

若反貪演算法的資料來源、風險指標或模型邏輯遭內部人員、外部廠商或敵對勢力掌握，便可能出現規避偵測、操弄評分、保護特定廠商，甚至藉由干擾採購與後勤判斷削弱戰備的風險。因此，公部門 AI 廉政治理經驗雖可作為國防場域的制度參照，但國防反貪不宜僅複製一般行政機關的風險篩選模式，而應進一步納入模型稽核、權限控管、供應鏈安全與反情報審查，將廉政治理從單純的採購風險篩選，提升為兼具安全防護與責任追溯的國防治理機制。

參、趨勢研判

國防廉政的數位黑箱風險

綜觀國際與我國發展，演算法反貪工具正從「輔助稽核」逐步轉向嵌入採購、審計與廉政流程的「治理基礎設施」。其效率紅利已有實證支持，未來導入速度恐將持續加快；然而，真正的關鍵不在於 AI 能否偵測貪腐，而在於機關能否同步建立「可稽核、可追蹤、可問責」的治理鏈。對國防領域而言，此一判斷尤具迫切性。由於任務保密、敵我對抗與戰備需求，使一般公部門賴以降低貪腐風險的透明化、開放資料與外部監督，難以完整移植至國防場域。

一旦反貪演算法的資料來源、風險指標或模型邏輯遭內部人員、外部廠商或敵對勢力掌握，原本用以強化廉政治理的工具，便可能反轉為掩護高科技貪腐，甚至干擾採購判斷與戰備維持的治理

[19] "Belgium Probing NATO Staff over Defence Contract Irregularities," *Reuters*, May 14, 2025, https://reurl.cc/ovGqal.

破口。

據此研判，我國國防廉政治理的下一階段，重點不再只是「是否導入 AI」，而是能否在引進效率工具的同時，前置部署模型稽核、權限分級控管、供應鏈安全與反情報審查等後端治理機制。唯有將演算法治理與國防安全風險一併納管，方能避免既有「採購黑箱」被轉化為更難辨識、更難究責的「數位黑箱」。

從 2026 北京軍博會等看解放軍 AI 與無人技術進展

王綉雯

中共政軍與作戰概念研究所

焦點類別：解放軍、軍事科技、戰爭模式

壹、前言

今（2026）年 5 月 14 至 16 日，第 11 屆「中國（北京）軍事智能技術裝備博覽會」（The 11th China (Beijing) Military Intelligent Technology Expo, CMITE Beijing, 簡稱「北京軍博會」）暨第 14 屆「中國指揮控制大會」（C2 China 2026，以下簡稱「指揮控制大會」）於北京國家會議中心舉行。主辦單位是「中國指揮與控制學會」，協辦方則有「中國兵器科學研究院」、「中國科學院軟體研究院」、「中國電子科學研究院」、「中國電子科技集團第 28 研究所」、「中國航天科工集團第二研究院」等 9 家機構，並廣邀其國內軍工企業、著名大學院校、科研機構、國家級重點實驗室、民間高科技企業等參與。大會由四大活動區塊組成，分別是：主旨報告、專題論壇、論文交流及北京軍博會，總成果計有專家報告 124 場、論文發表 132 篇、參展廠商 550 多家、展示產品和技術 3 千多項，參展民眾 5 萬多人。[1]

北京軍博會是中國最具代表性的大型國內軍事展覽，以解放軍 C4ISR[2]智能化（即：智慧化）為目標，主要是透過「軍民融合」戰略，將中國民間 AI 等先進技術轉為軍事應用，自 2015 年首屆舉辦

[1] 〈第十四屆中國指揮控制大會暨第十一屆中國（北京）軍事智慧技術裝備博覽會圓滿收官〉，《北京軍博會》，http://www.junbohui.com.cn/news/202605/1491.html。

[2] C4ISR 是指揮、控制、通信、電腦、情報、監視與偵察。

以來備受矚目。此屆北京軍博會與指揮控制大會正值中共「十五五規劃」（2026-2030）開始年度，在後者全面實施「AI＋」行動、重點發展具身智能、無人駕駛與 AI 晶片等具體目標下，大會奉行「AI＋國防」而將主題設為「無人具身智能引領、指揮控制體系賦能」。除了展示中國軍事智能化的最新產品與技術之外，也聚焦於網絡信息（即：網路通信）體系建設運用，旨在加快解放軍無人智能作戰和反制能力之建設，以促進「新質戰鬥力」生成。[3]換言之，此屆北京軍博會和指揮控制大會是觀察中國「新質生產力」轉化為「新質戰鬥力」的重要窗口。

貳、安全意涵

一、 AI 與無人技術是中國「新質戰鬥力」重點所在

此屆北京軍博會規劃了十大主題展區，分別是：（一）人工智能與大模型、（二）智能指揮控制系統、（三）無人系統與反制技術、（四）網絡與信息通信技術、（五）訓練裝備與模擬仿真、（六）信息安全與保密技術、（七）音視頻與顯示控制、（八）軍用計算機與便攜式設備、（九）後勤裝備與醫療救援、（十）電子元器件與新材料。其中，軍事大模型、反無人機、機器狗協同作戰是官方宣傳的三大亮點。[4]

指揮控制大會則邀請多位產官學研軍界專家，針對軍事智能化方向發表主旨報告，包括：無人系統組網、衛星通信、具身智能、地面無人系統、網絡與認知域融合對抗、AI 時代的戰爭演化、智算集群體系結構、邊緣智能、智能敏捷指揮控制等。同時，「中國指揮

[3] 〈關於“2026 第十四屆中國指揮控制大會”的徵文通知〉，《第十四屆中國指揮控制大會》，2025 年 11 月 26 日，https://zhikong.org/uploads/admin/202511/69280b8a1fc87.pdf。

[4] 馬俊，〈《環球時報》記者探訪軍博會，解碼前沿裝備〉，《環球網》，2026 年 5 月 15 日，https://m.huanqiu.com/article/4RZAgkphKEt。

控制學會」各專業委員會也分別主辦 17 場專題論壇，主題涵蓋空間（即：太空）信息通信技術、新型指揮控制網絡技術、星地融合時空網絡與安全服務、智能籌劃技術、無人化指揮控制、智能敏捷指揮控制等。[5]

此屆北京軍博會若說是中國 AI 與無人技術軍事應用之市場化產品展示會，則指揮控制大會就是其軍民前瞻技術研發與形成產官學研軍共識的交流平台。由上述展覽區塊和指揮控制大會論壇主題可看出，無論是當前或未來，AI 與無人技術已是解放軍新質戰鬥力發展重點。中國正在嘗試結合「無人具身智能（Embodied AI）」與「C4ISR 系統」，以建立完全由機器人等無人武器自主籌劃、決策和行動的新型態作戰能力。

二、AI 與無人技術之軍事應用提高中共對鄰近國家之安全威脅

就 AI 與無人技術的軍事應用而言，此屆軍博會和指揮控制大會的主要核心技術有：（一）軍事大模型與智能敏捷指揮控制、（二）6G 與生成式 AI 融合的「星地融合時空信息網絡」與戰場物聯網、（三）要地低空安全與反無人機技術，以及（四）網電安全和「反測繪」等。其中，以下幾項具體應用值得特別注意。

首先，是無人具身智能。此屆北京軍博會雖然展示了多款無人裝備，包括：空中無人機、地面機器狗、海底無人潛航器等，但是指揮控制大會主題的「無人具身智能」才是解放軍關注的焦點。這是將 AI 結合實體的無人裝備或機器人，使之在複雜戰場環境中具有如同人類的感知學習、與環境互動、敏捷決策等能力，最終發展出可自主作戰的無人具身智能體。其特點是不只依賴後方指揮控制，而是在資源受限、通信網路中斷或受電磁干擾等的前線環境下，仍

[5] 參見《第十四屆中國指揮控制大會》官網，https://www.zhikong.org/index.html。

可依賴自身的邊緣人工智慧（Edge AI），實現跨域協同、戰術自主、智能決策的「自主作戰無人系統」。[6]

其次，是軍事大模型和智能敏捷指揮。本屆北京軍博會首次將「人工智慧與大模型」設為獨立展區。其中，協辦單位之一的中國電子科技集團公司（簡稱「中國電科」，CETC）第 28 研究所開發的「軍智大模型」，主要能力聚焦在目標狀態理解、目標自動識別與標記等情報收集方面。而其應用於有人—無人協同作戰的「空中編隊智能協同決策與控制系統」，戰術推薦功能則可大幅縮短決策時間。[7]再以民間企業「淵亭科技」展示的軍事大模型來看，其主要能力在於軍事籌劃、戰術推演及情報分析，在結合其相關產品如：軍事知識圖譜、數字孿生（Digital Twin, 即：數位孿生或數位分身）戰場、兵棋推演系統之後，數秒內即可生成多組作戰方案，實現智能敏捷指揮。[8]

然而，北京軍博會官網新聞稿中，卻唯獨缺少「智能敏捷指揮控制」專題論壇之報導。依據會前公布的該論壇簡介，此論壇旨在建立貫通智能敏捷指揮控制的基礎理論、關鍵技術和工程實踐的交流平台，目標是推動複雜系統認知、群體智能、大模型與數字孿生等新技術與指揮控制深度融合，協助中國在智能化指揮控制領域實現自主可控。該專題論壇之發表論文主題則包括：智算力建設、運籌優化變革、異構模型集成、無人機集群行為控制、生成式 AI 測試評估、智能反無人機體系等。[9]或許因外界可透過新聞稿窺知解放軍

---

[6] 〈第十四屆中國指揮控制大會“智能指揮與控制論壇”成功召開〉，《中國指揮與控制學會》2026 年 5 月 18 日，https://www.c2.org.cn/h-nd-2950.html。

[7] 同註 4。

[8] 〈【精彩回顧】淵亭科技亮相 2026 第十一屆北京軍博會〉，《北京軍博會》，2026 年 5 月，http://www.junbohui.com.cn/news/202605/1507.html。

[9] 〈【C2-CHINA】專題論壇|智能敏捷指揮控制論壇〉，《中國指揮與控制學會》，2026 年 5 月 12 日，https://www.c2.org.cn/h-nd-2931.html。

將 AI 與無人技術如何應用於指揮控制，該專題論壇成為唯一新聞稿「被蓋牌」的論壇。

第三，則是要地低空安全和反無人機。鑑於無人機在俄烏戰爭等的實戰表現，「反無人機」成為本屆北京軍博會的熱門重點。針對小型無人機之發現和跟蹤，許多廠商提出反無人機體系解決方案，包括以 AI 整合雷達、紅外線等各類傳統感測器之情報、融合車載、船載和機載等不同平台的多模態探測和信息融合等。而針對如何應對遠程自殺式無人機，廠商除了展出多款高速攔截無人機之外，也展示單兵肩扛式高速攔截無人機，可依賴「AI 機器視覺＋紅外線成像」的雙制導模式，自動捕獲跟蹤目標。[10]同時，指揮控制大會也設置「要地低空無人機防控體系」專題論壇，針對低空無人機探測跟蹤、干擾攔截、指揮控制、體系建構等進行研討，並推動相關標準和法規之制訂。[11]

換言之，解放軍目前正聚焦在無人具身智能、大模型與智能敏捷指揮、反無人機等技術之具體應用和突破，並已有民間廠商提出部分解決方案。儘管距離部隊列裝階段還需一段時間，一旦可列裝或形成戰力，解放軍涵蓋海、陸、空、水下等多領域的智能無人裝備，將具備全方位自主作戰能力，對於我國和相關鄰國之安全威脅恐將大幅躍升。

參、趨勢研判

一、無人具身智能可能是未來戰術主流

鑑於俄烏戰爭與新近數場美國主導的軍事衝突之實戰演示，未

[10] 同註 4。

[11] 〈聚焦低空安全 共築防控屏障——第十四屆中國指揮控制大會要地低空無人機防控體系論壇成功舉辦〉，《中國指揮與控制學會》，2026 年 5 月 19 日，https://www.c2.org.cn/h-nd-2957.html。

來戰爭發生時，電磁干擾與通信網路中斷等網電作戰恐先於軍事行動。聽從後方總部下達指令的傳統指揮控制模式，在通訊斷鏈或資訊不明之情形下可能面臨嚴重挑戰。為此，美中等國正研發搭載邊緣人工智慧（Edge AI）或代理型人工智能（Agentic AI）的無人裝備（即「無人具身智能」），使其即使喪失衛星導航、通訊網絡等輔助，仍可依賴現場的「蜂群智能」（Swarm Intelligence），[12]自主分配目標、規劃路徑和進行協同攻擊。這種協同攻擊可將前線的無人機、無人戰車、軍事機器人、無人艇、無人潛行器等加以結合，進行立體、多領域且系統性的自主作戰。

二、「空天地海一體化」戰場物聯網是解放軍智能化作戰關鍵

中國已開始建構其「空天地海一體化」物聯網（Space-Air-Ground-Sea Integrated Network）。除了加速推進衛星組網（如：國網星座、千帆星座等）之外，本屆指揮控制大會也舉辦了「智能物聯網」、「空間信息智能應用」與「星地融合時空信息網絡與安全」等相關的專業論壇。其中，「星地融合時空信息網絡與安全」論壇聚焦在6G星地融合網絡、低軌衛星星座、AI、數位孿生、衛星通訊及導航安全等議題，並有發表者提出「低空時空一張網」概念。[13]「智能物聯網」論壇則聚焦在戰場物聯網，討論如何利用物聯網建構無人智能作戰新戰場，以及如何增進水下通信與組網技術之突破。[14]戰場物聯網可將從衛星到水下無人潛航器（Unmanned Underwater Vehicle, UUV）的各層級無人裝備連為一體，即時收集大量數據，同步匯集

[12] 有關「蜂群智能」之詳細定義，請見" What is Swarm Intelligence?", Hewlett Packard Enterprise, https://www.hpe.com/emea_europe/en/what-is/swarm-intelligence.html。

[13] 〈第十四屆中國指揮控制大會"星地融合時空資訊網路與安全服務論壇"成功召開〉，《中國指揮與控制學會》，2026年5月18日，https://www.c2.org.cn/h-nd-2944.html。

[14] 〈第十四屆中國指揮控制大會"智慧物聯網論壇"成功召開〉，《中國指揮與控制學會》，2026年5月19日，https://www.c2.org.cn/h-nd-2952.html。

到以數位孿生技術建立的虛擬戰場，並由大模型進行分析和研判，是解放軍未來進行智能化無人作戰的重要關鍵。

三、反無人裝備需求將改變傳統防禦思維

鑑於俄烏戰爭的經驗，低空無人機成為各國政治軍事據點和關鍵基礎設施的新型安全威脅。因此，「反無人機技術與應用」成為本屆北京軍博會與指揮控制大會的主要重點之一。未來各國的防禦系統除了建構造價昂貴的飛彈網之外，極可能同時需要由代理型 AI 指揮的各種低成本反無人機系統，包括：光學辨識、電磁干擾、雷射武器等。另一方面，無人裝備可克服人類士兵的肉體極限，在原本認為是天險屏障的地形或地貌上進行作戰，例如：從紅色海灘或海岸峭壁強行登陸。這或將造成現有防禦思維的巨大變革，並大幅降低防禦成本，我國對此似應預作前瞻規劃。

《寒戰》作為「認知戰」文本：當前香港的安全化敘事與影像治理

侍建宇

國家安全研究所

焦點類別：網路戰、認知戰

壹、前言

已經上映的香港電影《寒戰》三部曲，故事主軸圍繞在香港警隊／安全體制內的危機，而非一般犯罪案件。在《寒戰》（2012）中，以警隊衝鋒車與警員被劫持為開端，整個香港安全機器被迫進入最高緊急狀態，並揭露香港警隊內部角力。特別是高層警官在「寒戰」行動期間，爭奪指揮權而廝殺鬥爭。作品劇情並非傳統的綁架故事，而是關於體制階層、精英競爭、公共秩序脆弱性的戲劇。《寒戰 II》（2016）延續首部曲的邏輯：警隊內部權鬥，並延伸到香港行政部門的保安局、立法會、司法場域之間的角力。故事揉捏著「陰謀論」的調性，展露香港高層政壇鬥爭竟然不惜犧牲社會安全、扭曲法治界線。

《寒戰》系列看似一個警匪片，而且故事軸並沒有中國政府，但是透過拉出「外部勢力」因素，角逐香港政治地盤，點出核心命題：禍起蕭牆與內外勾串——最大的敵人往往來在內部，並勾連外部勢力，試圖架空中國在港主權。

原定 2019 年底開拍《寒戰 3》，但當時「反送中」社會運動，香港警察行為備受批評，不宜製作涉及警隊題材的電影，於是延遲。但 2024 年《寒戰》系列獲中國國家電影局，也就是中共中央宣傳部電影局批准拍攝另兩部前傳，分別名為《寒戰 1994》及《寒戰

1995》。[1]將系列時間序列推前至九十年代。因此第三部上映的《寒戰 1994》作為「前傳」，故事時間線部分發生於 2017 年，部分往前推到 1994 年。追溯香港主權移交前的一起富商綁架案，將警察貪腐、黑社會、精英家族勢力與英國殖民政府（也就是港英香港皇家警隊政治部，亦即英國軍情局六處的香港分支）交織在一起。政治安全化敘事意涵，不言而喻，重新詮釋香港政治秩序的來龍去脈。

貳、安全意涵

《寒戰》系列電影的深層敘事核心是一種沒有直接明說的「政治安全化」邏輯。過去的香港，在大眾文化與國際想像中，常被描繪為一座治安良好、制度有效、警隊紀律嚴明、金融秩序穩定的國際城市。既是全球資本流動的重要節點，也是東西方情報、商業與政治力量交會的特殊空間。香港雖然長期被視為亞洲情報中心，但這類議題很少真正成為主流娛樂電影的核心命題。《寒戰》系列的特殊之處，正在於將香港作為國際權力競逐場所的想像帶入警匪片之中，透過電影重新勾勒香港政治危機的來源。

在電影敘事中，政治衝突不是多元民主競爭，也不是「一國兩制」制度的中央與地方磨合，而被理解為癱瘓政府、破壞秩序，甚至奪權的安全威脅。尤其當「外部勢力」與本地政治分歧被放在同一個故事結構中，香港社會的抗爭、制度爭議與權力角力，便容易被重新理解為危及城市安保與國家秩序的風險。換言之，《寒戰》系列並不是單純講述警隊辦案或高層鬥爭，而是在警匪類型片的包裝下，建立一種政治判斷框架：香港的危機不是普通政治衝突，而是

[1] 《寒戰 1994》於 2026 年 5 月在亞洲地區開始上映。據悉還有下集，暫定為《寒戰 1995》，將於 2026-2027 前後上映。鍾路淨，〈《寒戰》系列獲內地批准開拍 兩套電影時間線推前 18 年變前傳？〉，《香港 01》，2024 年 4 月 18 日，https://reurl.cc/r0bO3E。

安全危機。

一、《寒戰》系列的政治安全化敘事

在《寒戰》與《寒戰 II》，政治安全化敘事仍然相對含蓄，主要透過警隊高層鬥爭、危機決策、政府部門協調與制度漏洞來暗示香港秩序的脆弱性。然而到了《寒戰1994》，這種敘事變得直接。電影更清楚呈現腐敗的警官、分裂的精英、被滲透的機構，以及表面上運用法律程序、實際上可能追求隱藏政治目的的行動者。反覆強調，危險不只來自一般犯罪，而是來自內外勢力勾連、制度內部裂縫，以及精英政治與安全機構失控。透過這種政治驚悚式的安全敘事，嘗試改變觀眾理解公共事務的角度，使觀眾在看待政治爭議時，不自覺地優先想到陰謀、滲透、背叛與秩序崩壞，而不是權利、問責程序或民主參與。

《寒戰》系列電影對香港社會產生的認知扭曲效果，過程非常複雜，這裡僅嘗試運用傳播理論中的預示效應（priming effect）來分析。[2]預示效應是指受眾在反覆接觸特定議題、敘事或價值框架之後，某些原本存在於記憶中的價值判準會被不知不覺地喚醒，並在之後評價政治事件、人物或行動時，成為更容易被取用、也更具支配性的判斷依據。預示效應不同於說服。說服通常是直接對抗或修正原有觀點，目的在於改變受眾信念；但預示效應較為間接，也更具暗示性。不要求受眾立刻接受新的政治立場，而是透過重組既有評價標準，使某些價值被優先啟動，另一些價值則被淡化、邊緣化

[2] 預示效應也有人譯作啟動效應、觸動效應。本文給予預示效應一個操作定義，參考 Jennifer Hoewe, “Toward a Theory of Media Priming,” *Annals of the International Communication Association*, Vol.44, Issue 4, September - December 2020, pp. 312–321。媒體傳播效應對於政治社會心理影響的操作設計，參 Dietram A. Scheufele, “Agenda-Setting, Priming, and Framing Revisited: Another Look at Cognitive Effects of Political Communication,” *Mass Communication and Society*, Vol.3 Isse. 2-3, 2000, pp. 297-316。

或降低重要性。

簡單來講，預示效應想達到的效果是讓公眾在判斷政治現實時，逐漸改變其價值排序。例如，原本公眾在面對政治抗爭時，可能會同時考量自由、秩序、公民權利、國家安全、社會穩定與程序正義等多重價值；但如果媒體與影像敘事長期以「混亂」、「暴力」、「外部勢力」、「陰謀滲透」或「國家安全威脅」來呈現抗爭，公眾便可能在無意識中改變判斷標準的排序，優先啟動「安全」與「秩序」的價值，而相對淡化「自由」、「人權」的價值。因此，預示效應的核心不在於直接改變人們相信什麼，而在於改變人們優先使用什麼標準來判斷政治事件。應用於政治傳播領域，便成為一種微妙但強大的認知排序機制，重新組織公民解讀政治現實時所依賴的價值層級。

從這個角度來看，《寒戰》系列最值得注意之處，不只是作為警匪片的娛樂性，而是在不同政治時刻中，如何重新塑造香港社會對「秩序」、「安全」、「國家」、「警隊」與「內部敵人」的理解。構成了一套高度政治化、但又不以直白政治口號呈現的敘事系統。這類影像文本在香港社會中製造的預示效應，並不只是要求觀眾「支持警察」或「反對抗爭」。更深層的效果，是讓觀眾開啟另一條理解香港政治的通道，啟動一套安全化的認知框架：當社會出現混亂、制度爭執、權力競爭或群眾動員時，觀眾首先想到的不是公民權利、制度問責或民主參與，而是危機、陰謀、外部勢力、內部叛徒與秩序崩潰。換言之，這類電影的政治效果不是要觀眾背誦官方論述，而是要重新排序觀眾心中判斷政治事件時所依賴的價值。

《寒戰》系列因此將警匪驚悚片轉化為一種政治寓言。香港在電影中被描繪成繁榮卻脆弱、法律體系精緻卻可能遭到滲透、政府

機器高效卻可能因內部鬥爭而癱瘓的城市。這種想像最終指向一個結論：香港需要預防性的安全治理。這也正是當前北京在香港治理上最希望建立的政治心理狀態。

2019 年的反送中抗爭，對北京而言不只是一次大規模社會運動，而是一次對「一國兩制」政治邊界的根本挑戰。北京後來透過 2020 年《香港國安法》與 2024 年《維護國家安全條例》重新建構香港政治秩序，將叛國、煽動、間諜、國家秘密與外部干預等概念納入更嚴密的治理框架，也使香港異議空間受到高度壓縮。在這種背景下，影視敘事便可扮演柔性治理功能，使國安化治理看起來不是例外狀態，而是面對危機時的正常反應。電影不必覆誦國安法的政治教條，但可以使香港大眾對官方安全化邏輯產生共鳴，並將「香港的生存必須仰賴強大的安全機構，並限制政治參與和衝突」這套觀點常態化。

二、《寒戰》系列的三種預示效應

目前已上映的三部《寒戰》電影，大致可歸納出三條主要的「預示效應」。

將「穩定」設定為最高政治價值，把香港警隊塑造成安定社會的最後防線；進一步說，也將「一國兩制」下的香港自治重新界定為「由中央保護的自治」，而非「可抗衡中央的自治」。

《寒戰》系列透過危機、失序與警隊內部鬥爭的敘事，反覆呈現城市可能瞬間陷入恐慌、警隊高層分裂可能癱瘓治理機器、政治鬥爭可能演變成安全危機。這種敘事是在塑造一種心理條件反射：制度爭議本身就是危險的，政治分歧若不能被迅速壓制，就會威脅社會穩定。於是，觀眾在面對香港政治爭議時，首先想到的便不是自由、自治或程序正義，而是安全危機與秩序崩壞帶來的恐慌。

同時，《寒戰》系列重新正當化香港警隊。2019 年反送中運動後，香港警察在相當多市民心中留下高壓與殘暴的陰影，警民關係急劇惡化。電影並未單純歌頌警察，而是承認警隊內部存在權力爭鬥與人性弱點；但最終仍要求觀眾相信，當香港面臨真正危機時，能夠保護城市的不是街頭群眾、外國聲援或媒體揭露，而是紀律部隊與安全體制。更深層地說，這也把香港自治重新界定為「受保護的自治」：香港可以保有某些制度特色，但必須服從國家安全與中央權威。電影中的香港看似高度本土化，充滿警隊、政府部門、法律制度與城市空間的符號，但其終極暗示是：本地制度若不與更高層級的中國安全秩序一致，便可能陷入內鬥與動盪。

第二、淡化民主程序的重要性，轉而強調專業精英治理；同時貶抑政治異議，將青年政治參與視為不成熟或被利用的行動，重新編碼為安全威脅。

在國安敘事中，反對派、媒體、律師、公民社會與示威者不再扮演監督的力量，而被描述為潛在破壞者、外部代理人或陰謀網絡的一部分。近年香港國安案件，尤其是黎智英案與支聯會解散案，[3] 顯示北京與特區政府如何將新聞活動、國際倡議、歷史記憶與政治組織納入安全化框架，使異議不再被視為公共辯論，而被界定為國家安全風險。《寒戰》系列所強化的「內部敵人」敘事，正好呼應這種政治氣候，使觀眾更容易接受官方對異議者貼上的安全化與破壞者標籤。

同時，電影也將制度辯論、部門制衡與政治競爭描繪為危機處

[3] 智英案是北京對香港社會菁英的一種「殺雞儆猴式的警嚇」，支聯會案則是對香港社會輿論自由的箝制；禁止香港社會延續對支援 1989 年六四天安門學生運動的歷史記憶，只能依循中共單一詮釋口徑。支聯會案 2026 年 7 月將有最後判決，參見，〈支聯會案｜控方結案陳詞：煽動行為已超越言論自由界線〉，《稜角媒體》，2026 年 5 月 18 日，https://reurl.cc/A9elxK。

理中的障礙，轉而凸顯專業、忠誠與敏捷的警察官僚。這與北京「愛國者治港」的制度相契合，政治忠誠取代多元競爭。2019 年抗爭中的青年原本承載政治理想與道德光環，但在《寒戰》的警匪敘事中，資淺年輕的警察被操弄，暗示青年行動者往往被放置在暴力、陰謀、被操縱或不成熟的框架之中。這種敘事貶抑青年政治參與，學生運動可被理解為過度理想化、缺乏判斷力，甚至可能被外部勢力利用。整套敘事不僅壓制有組織的反對派，也試圖重塑後反送中世代對民主抗爭與國家秩序的想像。

第三、把「外部勢力」塑造成香港危機的主要來源，並形塑對中國國安體制的「恐懼式認同」。

北京將香港抗爭的重要成因之一歸結為外國勢力介入、煽動與利用香港作為反華基地。《寒戰》系列雖未直接宣傳此一說法，卻透過陰謀化、跨境化、外國情報機構介入與高層權力鬥爭的敘事，使觀眾更容易相信香港危機並非單純源於本地制度矛盾，而是牽涉更大的外部國際權力網絡。這種敘事有助於正當化香港國安法，將香港問題從一般政治衝突提升為國家安全問題。

北京在國安法時代更倚重「恐懼式認同」：透過喚起對混亂、暴力、顏色革命與社會崩壞的恐懼，促使市民接受國安與警政的高壓治理。從政治傳播效果看，這是一種負面預示；不是讓人積極熱愛中國，而是讓人害怕沒有國安體制，香港將重返 2019 年的失序狀態。因此，《寒戰》系列的政治意義不在於直接宣傳國安法，而在於使國安法背後的安全化邏輯變成可理解、可接受，甚至被視為理所當然的政治價值。

參、趨勢研判

安全敘事成為香港影像治理的主軸

必須指出，《寒戰》系列不應被簡化為北京直接操作的政治宣傳。香港電影工業本身具有商業邏輯、類型傳統、創作者能動性與市場考量，《寒戰》所以能產生影響，正是因為具備精密的劇情設計、明星陣容、政治驚悚感與香港警匪片的專業水準。若將影片單純視為宣傳，反而會低估真正的傳播力量；《寒戰》系列的有效性恰恰來自「不像宣傳」，使觀眾在享受警匪片快感時，不知不覺接受某些政治前提：香港是脆弱的，秩序是稀缺的，內外敵人勾連是真實存在的，安全機器是必要的，而政治分歧必須被嚴格管理。

這種影像治理最值得警惕之處，在於它並非以明確命令改變觀眾立場，而是重塑觀眾理解政治的認知環境。「預示效應」通常不是告訴人們「你應該怎麼想」，而是引導人們「你應該先想到什麼」。當香港觀眾面對國安法、政治審判、媒體關閉、抗爭記憶或警隊權力擴張時，如果首先想到的是 2019 年的混亂、外部勢力、城市失序與內部叛徒，北京的治理合法性便已在心理層面獲得部分鞏固。

負面後果也相當明顯：公共討論空間會被進一步壓縮，安全框架凌駕政治辯論；警隊與國安機器的擴權更容易被接受，公民監督意識則被削弱；反對派、媒體與公民社會被長期污名化，任何民間維權活動，都可以被懷疑與外部勢力或內部破壞有關。更重要的是，這會加深香港青年與建制敘事之間的疏離，至少部分青年走向沉默、離港移民或犬儒的虛與委蛇。「一國兩制」於是變成安全化治理的範例，而非高度自治的制度承諾。總結而言，《寒戰》系列的政治意義在於提供一套有效的安全化敘事，一種新穎的「大內宣」或說升級的「認知戰」手法，使國安法不必被直接宣傳，也能被呈現為香港生存所需的基本價值。

發行人 / 霍守業

總編輯 / 柯承亨

主任編輯 / 蘇紫雲　執行主編 / 吳宗翰

助理編輯 / 黃政勛、陳宥芯、林均蓉、賴達文、李虹宜