中國行政評論　第 32 卷第 3 期
The Chinese Public Administration Review
Vol.32 No.3 September 2026.96-119
DOI:10.6635/cpar.202609_32(3).0005

人工智慧治理對傳統國際安全體系的衝擊與因應

楊惟任*

摘 要

生成式人工智慧與先進模型迅速擴散後，人工智慧治理已不再只是科技倫理或產業監理的附屬議題，它已逐步成為國際安全與主權秩序的核心命題。本文以「演算法權力」與「主權邊界」的關係作為分析框架，說明人工智慧治理如何牽動傳統國際安全體系的調整。本文主張，演算法權力並非單一模型或單項應用產生的工具能力，而是由資料蒐集、算力配置、模型部署、平臺分發、晶片供應、技術標準融合而成的權力形態，它經由可見性、可控性與可問責性轉移三項機制，改變國家、科技企業與國際組織的安全權威分配。本文結合結構性權力、數位主權、人工智慧治理研究，並以戰場演算法基礎設施、美中半導體和算力競爭，以及平臺推薦與選舉認知作戰三項案例作為經驗參照，發現人工智慧對國際安全的影響不限於軍事自動化，已延伸到供應鏈主權、平臺規則制定與認知安全邊界的重新劃定。基此，人工智慧治理不宜停留在科技監管層次，而須納入基礎設施韌性、跨境問責與多層次技術多邊主義，國家角色也應從單純管制者，轉向標準參與者、基礎設施投資者與跨境問責架構的推動者。

關鍵詞：演算法權力、主權邊界、數位主權、人工智慧治理、國際安全

* 楊惟任教授，英國 University of Warwick 政治學暨國際關係博士，現任職於世新大學行政管理學系，曾任世新大學國際事務長、公共事務長、共同課程委員會主任委員，研究專長包括人工智慧與社會科學、川普學、全球治理、國際氣候政治。Email: william@mail.shu.edu.tw。
本論文經兩位雙向匿名審查通過。收件：2026/5/10。同意刊登：2026/8/4。

壹、前 言

人工智慧早已超出產業競爭與科技倫理的討論範疇，進入軍事決策、供應鏈安全、平臺治理、認知作戰與國際規範競爭相互交織的領域。生成式模型快速普及，戰場感測與目標識別系統被大規模部署，晶片與算力供應鏈也在地緣政治壓力下重新排列，這些發展顯示人工智慧治理已成為牽動當代國際安全體系的關鍵制度因素（Horowitz, 2018；Scharre, 2018）。若僅以軍事自動化理解這項變化，容易忽略模型訓練、雲端服務與平臺分發在安全判斷的仲介作用，這些環節雖多屬商業或民用安排，卻會影響國家何以取得資料、判讀風險並啟動回應。

當前人工智慧治理的制度建設，以歐盟、美國、中國最為積極。歐盟《人工智慧法》（*Artificial Intelligence Act*，*AI Act*）已進入通用目的人工智慧模型義務適用與執法準備期，美國政策重心轉向以人工智慧行動計畫、基礎設施擴張與出口管制重組維持技術領先，中國透過升級人工智慧安全治理框架，並推動相關規範之法制化，強化人工智慧治理基礎。以上顯示，人工智慧治理已牽動法規競爭、供應鏈安全與國家能力建構，而非僅是倫理宣示或產業競爭問題（European Commission, 2026a, 2026b；White House, 2025a；Bureau of Industry and Security, 2025；中共中央網路安全和資訊化委員會辦公室，2025；數位發展部，2025）。

傳統國際安全體系大致建立在「國家、領土、軍力、國際法責任」等前提，並假定國家擁有疆界清楚的領土、可被追究責任的決策中樞，以及相對可量化的軍事資源。然而，人工智慧系統經常跨境部署，由私營企業營運，並依賴外部不易檢視的訓練資料、模型參數與平臺運作邏輯，一旦發生安全事件，受影響者很難判定其來源究竟是國家行為、企業決策、演算法失靈，或多方互動後產生的湧現效果。由此造成的問題在於，「看見威脅、判斷威脅、回應威脅、追究責任」四個安全治理環節，均面臨權威重新分配。

本文的問題意識是：當安全功能日益仰賴跨境技術網路與私營基礎設施時，國家主權如何在維持政治正當性的前提下，重新界定其領土、制度、基礎設施與認知邊界？

針對上述問題，本文提出三項研究問題，以釐清人工智慧治理對國際安全體系在制度與權力分配上的影響。首先，作為一種跨領域影響力的演算法權力，如何發展出有別於傳統軍事與經濟權力的獨特形態？其次，演算法權力透過何種機制重塑國家主權在領土、制度、基礎設施與認知層面的邊界？最後，現行治理機制是否足以回應人工智慧時代產生的權威分配與問責缺口？

本文主張：人工智慧對國際安全的衝擊，關鍵不在於它是否取代人類決策，更在於安全治理的若干前提條件已嵌入資料、算力、模型、平臺與技術標準之中。

換言之，國家若無法掌握或影響這些條件，即使仍保有軍事與外交資源，也難以維持自主的安全治理能力。

研究方法方面，本文採取「理論導向之質性案例分析」與「結構化、聚焦式比較」並用的研究設計。理論部分整合結構權力、數位主權與人工智慧治理文獻，案例部分則選取三個能對應本文核心命題，且彼此具有功能差異的個案，包括俄烏戰爭的戰場演算法基礎設施、美中半導體與算力競爭，以及平臺推薦與選舉認知作戰。

本文並非以這三個案例概括所有人工智慧安全事件，而是將其作為檢驗理論命題的切入點，用來觀察可見性、可控性與可問責性轉移在不同安全領域的具體樣貌。資料來源包括政府政策文件、國際組織報告、學術期刊論文、智庫研究、企業公開資訊與公開事件資料，並透過交叉比對降低單一來源偏誤。

案例的選擇標準有三：第一，具備足夠公開資料可供檢證；第二，能清楚對應本文提出的主權邊界維度；第三，能呈現國家、企業與國際制度之間的權威互動。就分析重點而言，俄烏案例著重戰場可見性與基礎設施依賴，美中半導體案例聚焦供應鏈可控性與制度外溢，平臺認知作戰案例檢視認知邊界與問責斷裂。這樣的案例配置有助於提升論證的內部一致性，但本文仍將其定位為理論生成與機制辨識，而非統計上的代表推論。

既有研究大致可區分為三個領域：「人工智慧軍事應用」研究，主要關注自主武器、決策迴圈壓縮與軍事戰略影響（Scharre, 2018；Horowitz, 2018；Bode & Huelss, 2018）；「數位主權與平臺治理」研究，焦點放在資料跨境、平臺規制與基礎設施依賴（Floridi, 2020；Bradford, 2023；Couture & Toupin, 2019）；「人工智慧倫理與全球治理」研究，討論透明、問責、公平與風險治理（Pasquale, 2015；Zuboff, 2019）。

上述文獻雖分別揭示軍事自動化、平臺規制與倫理問責的重要性，卻較少說明三者如何在國際安全結構相互強化，尤其不足之處，在於未能把資料、算力、模型、平臺與標準視為共同構成安全權威的條件。本文希冀在此缺口上，將演算法權力放入主權邊界變動的脈絡，以解釋人工智慧治理如何同時影響軍事安全、供應鏈安全與認知安全。

本文首先建立理論框架，界定演算法權力並提出主權邊界的四個維度，其次分析人工智慧對戰略穩定、認知安全與軍備競賽邏輯的衝擊，再者討論科技企業、平臺、供應鏈與基礎設施如何改寫主權邊界，此外評估現行治理機制的侷限並提出制度方向，最後整理研究發現、學術貢獻與政策啟示。

貳、理論框架與文獻探討

一、演算法權力的定義

本文將演算法權力（algorithmic power）定義為：「行為體透過資料蒐集、模型訓練、算力配置、平臺分發、晶片供應與技術標準，持續影響他者感知、選擇與行動範圍的能力」，這項定義可從三個面向加以解釋。

第一，演算法權力具有分散式特徵，並不集中於單一權力中心，它散佈在資料管線、模型權重、運算基礎設施、應用程式介面與內容分發平臺，並由跨國企業、政府機構、研究實驗室與終端使用者共同生產與承載（Pasquale, 2015）。因此，分析重點不宜停留於模型輸出是否正確，而須追問資料來源、介面設計與部署情境如何限制使用者的判斷。

第二，演算法權力不只是工具性權力，更是一種能事先安排選擇條件的力量，它不直接命令行為者，而是透過設定可選項目以影響決策。搜尋引擎將特定資訊排在前列、雲端平臺調整服務條款，或晶片廠商限制特定客戶取得最新製程，這些作法未必帶有明確的政治指令，卻已實質改變他者的行動空間。

第三，演算法權力具有跨領域滲透性，遍及經濟、軍事、文化與認知領域，傳統按部門分工的治理架構難以完全因應。因此，單一國家機構通常無法獨立監管，國際組織也難以找到對等且穩定的治理對手，這也表示，安全風險常以民用服務、商業契約或平臺規則的形式出現，若仍依循傳統部門邊界分別處理，便容易低估其政治後果。

換言之，這裡所稱之權力並非指程式碼本身具有政治意志，而是指技術架構被嵌入制度與市場安排後，會改變行為體可取得的資訊、可使用的資源與可承擔的責任。這樣的界定有助於區分演算法作為技術物件與演算法作為治理條件的不同層次。

二、與既有權力理論的對話

演算法權力的概念可與既有權力理論對照理解。Barnett 與 Duvall（2005）將國際關係的權力分為強制（compulsory）、制度（institutional）、結構（structural）與生產（productive）四類，演算法權力雖涉及制度、結構與生產三個面向，卻無法完全歸入其中任何一類。

Strange（1996）的結構權力理論指出，國際政治經濟的結構性權力奠基於安全、生產、金融與知識四大結構。此邏輯推論，演算法權力可視為「知識結構」在數位轉型後衍生出的次級形態，凡掌握資料、模型與算力者，即能左右其他結構的運作條件。不過，若僅把它理解為知識結構的延伸，仍會低估算力與雲端部

署造成的物質限制，模型能否訓練、更新與跨境提供服務，取決於能源、晶片與資料中心等條件。

Lukes（2005）的三面向權力觀提供另一條分析線索：第一面向是公開決策權力，第二面向是議程設定，第三面向是偏好塑造。演算法權力同樣作用於這三個層次：自動化決策對應第一面向，排序與推薦對應第二面向，大規模個人化內容分發則可能介入第三面向。

三、作用機制

演算法權力對國際安全體系的影響，主要經由三項機制呈現。

（一） 可見性機制決定哪些威脅得以被察覺，以及它們以何種形式進入決策視野。情報蒐集、目標識別、社群監控與內容審核，均涉及演算法對特定關注對象的篩選。一旦可見性被外包給少數模型或平臺，國家對「威脅圖景」的掌握便可能受到削弱（Brundage et al., 2018）。這種篩選不一定來自刻意操控，也可能源於訓練資料偏誤、標註規則或感測器部署位置，正因其原因分散，決策者更難確認被忽略的威脅為何。

（二） 可控性機制透過排序、預測與自動化手段改變行動選項。當人工智慧系統介入危機決策、平臺流量分配或關鍵基礎設施運作，行動選項的範圍與先後順序已被事先安排，雖然最終決定仍由人類作成，可供選擇的空間已受到演算法條件的約束（Scharre, 2018）。因此，可控性問題並非只涉及演算法能否替人作決定，也涉及排序結果如何改寫人類決策者對急迫性、可行性與成本的理解與認知。

（三） 可問責性轉移機制使責任分散於國家、企業、平臺、開發者與模型之間。當深偽影像在多個平臺間擴散、自主武器系統造成附帶傷害，或推薦演算法加劇社會分裂時，傳統國際法預設的「誰決策、誰負責」的責任歸屬架構，便不容易被重新拼接（Chesney & Citron, 2019；Bode & Huelss, 2018）。責任被切割後，受害者只能面對平臺規範、服務契約或國家安全例外，卻難以取得足以說明因果關係的技術資料。

四、主權邊界的維度

為提高分析的操作性，本文將主權邊界區分為四個維度，這四個維度彼此相關，但仍須分開分析，才能辨識哪些主權能力正在被削弱，哪些仍可由國家透過制度安排補強。

（一） 領土邊界關乎資料跨境流動、雲端服務管轄與境外平臺規制，當資料、模型與運算分散於多重司法管轄區，傳統領土主權便面臨「資料即領土」的挑戰（DeNardis, 2014）。

（二） 制度邊界攸關規則與標準制定權的歸屬。技術標準、模型評測規範、資料保護法制與內容治理規則，正成為新的制度競爭領域(Bradford, 2023)。

（三） 基礎設施邊界牽涉晶片、算力、雲端與平臺依賴的政治意涵，當一國的關鍵服務、軍事決策與政府運作高度依賴外部基礎設施，其主權的「物質基礎」便會受到外部條件牽制（Miller, 2022 ）。

（四） 認知邊界涉及社會事實的建構、政治信任與威脅感知的形成，當公民的資訊環境受到外部演算法塑造，主權應保護的「公共理性」便面臨外部滲透風險（Singer & Brooking, 2018 ）。

五、理論命題

依據上述討論，本文提出三項可檢驗的理論命題，這些命題並非預設單向因果，而是說明當資料、算力與平臺愈趨集中時，主權能力如何以不同速度移轉，並可在後續案例檢驗其適用範圍與限制。

命題一：演算法權力越集中於少數平臺與基礎設施供應者，國家安全能力被功能性外包的程度越高，主權自主性越低。

命題二：安全領域越依賴即時資料與自動化判斷，傳統主權的責任推定架構越容易被稀釋，跨域問責成本也越高。

命題三：人工智慧治理規範越分裂，國際安全體系越可能走向技術陣營化與規範碎片化，全球公共財供給能力也越容易削弱。

參、安全衝擊：戰略穩定、認知戰與演算法軍備競賽

一、戰略穩定與決策迴圈壓縮

人工智慧已深度嵌入情報、監視、偵察、目標識別與指揮控制等環節，並改變戰略穩定的基本條件。Horowitz（2018）提及，人工智慧對權力平衡的影響速度與幅度，可能堪比蒸汽機與核能等通用技術。關鍵不在單一武器系統的性能高低，而在整體決策節奏被大幅壓縮。因此，戰略穩定不僅取決於武器毀傷力，也取決於資料回傳速度、模型更新頻率，以及指揮鏈是否能辨識錯誤訊號，若這些環節仰賴不同承包商或跨境平臺支撐，危機溝通的可預期性就會降低。

當預警、評估與反應時間由小時縮短至分鐘，甚至進入秒級時，危機決策的「人機協作」將承受更大壓力，決策者若不信任演算法輸出，可能錯失反應時機，若全然依賴演算法輸出，則可能出現誤判或受到樣本誤導（Scharre, 2018）。Cave 與 ÓhÉigeartaigh（2018）指出，人工智慧戰略競爭的論述可能加劇此一困境，各方在系統可靠性尚未充分驗證前，便因恐懼而加速部署，從而提高誤判風險。

更棘手的是，演算法系統的失效模式經常難以預測。傳統武器系統的失效多屬機械性，通常較易診斷，深度學習系統的失效則可能呈現統計性、情境性或對

抗性特徵，且不易解釋。這意味著，傳統嚇阻依賴的「可信承諾」邏輯正受到挑戰，當行為體本身都無法確定系統將如何反應，敵方也難以準確評估其能力與決心（蘇紫雲、曾怡碩，2018）。尤其在早期預警與目標分類階段，錯誤未必表現為明顯故障，反而可能以高信賴度分數呈現，使決策者表面上仍保有裁量權，實際判斷卻受到模型輸出影響。

二、自主武器與人類控制爭議

自主武器系統（Lethal Autonomous Weapons Systems, LAWS）的規範爭議，重點在於如何界定具實質意義的人類控制（meaningful human control）。Bode 與 Huelss（2018）指出，國際人道法雖未明文禁止自主武器，但「區分原則」與「比例原則」都預設存在可問責的人類決策者，一旦模型在毫秒內自行完成決策，這些原則的適用便會出現難以迴避的制度裂縫。

在聯合國《特定常規武器公約》（*Convention on Certain Conventional Weapons, CCW*）架構下，政府專家小組自 2014 年起討論自主武器規範，但因為主要技術強權立場分歧，所以迄今仍未形成具拘束力的條約。最新進展是，2026 年 3 月第一階段會議依 2023 年授權，以「不預判法律性質」方式研擬可能的實質要件，並以 2025 年動態草案為基礎，討論致命自主武器系統的定義、人類判斷與控制、國際人道法適用及風險降低措施（United Nations Office for Disarmament Affairs, 2026），這顯示規範討論已具較明確的文本基礎，但也反映人工智慧軍事應用的擴散速度仍快於國際規範形成。

事實上，人類控制的爭議不能簡化為「在迴圈內」或「在迴圈外」。在實務上，「人在迴圈上」（human-on-the-loop）與「人在迴圈外」（human-out-of-the-loop）之間存在不少灰色地帶，人類雖名義上保有決策權，卻可能因資訊過載、時間壓力或自動化偏誤而難以實質介入。傳統軍備管制工具對這些灰色地帶同樣缺乏有效驗證手段。因此，判斷人類控制是否充分，不能只看最後是否有人按下授權鍵，還必須檢視其是否理解系統限制、能否取得反證資訊，以及是否有足夠時間否決建議。

三、認知戰與社會韌性

平臺推薦演算法、生成式內容、深偽影像與微定向傳播，使得認知安全成為主權秩序不可分割的一環。Singer 與 Brooking（2018）認為，社群媒體已成為可被「武器化」的場域，無論是國家或非國家行為體，都能以低成本、快速擴散的內容操作，影響他國公民的認知與判斷。

Chesney 與 Citron（2019）對深偽（deepfake）技術影像的研究顯示，倘若合成媒體達到真假難辨的程度，社會判斷真實的基準便可能被侵蝕，受損的不但是選舉公正，也包括民主治理所依賴的「公共理性」與「事實共識」。

認知戰的特殊性在於其屬於灰色地帶行動，這類作為未必達到傳統武力使用門檻，難以歸因於特定國家行為體，損害也不易量化。因此，當前國際法工具，無論是禁止干涉內政原則、國家責任法則，或武力使用規範，都難以直接適用於此類行動（Schmitt, 2017）。

從輔助案例可見，多次重要選舉期間都曾出現由境外行為體協同推動的不實資訊行動，部分甚至結合生成式人工智慧大量產製內容，這些行動未必只為促成特定候選人當選，更常見的目的是擴大社會分裂並削弱民眾對民主制度的信任，若認知環境長期遭到污染，主權邊界的認知維度便會弱化。

四、演算法軍備競賽的新邏輯

傳統軍備競賽主要以平臺數量、兵力規模與武器性能衡量，人工智慧競爭的核心則轉向算力、資料、模型參數、晶片、雲端基礎設施與人才，其競爭邏輯可歸納為幾項特徵。

其一，「軍民兩用性更強」。前緣模型多由商業實驗室訓練，既可用於民用服務，也可投入軍事任務，使傳統「軍用－民用」二分的出口管制框架難以有效切割；其二，「能力擴散更快」。模型權重在釋出後幾乎無法收回，開源模型的全面普及也降低能力取得門檻，使非國家行為體取得高階人工智慧能力的成本隨之下降（Brundage et al., 2018）；其三，「競爭範圍更廣」。人才、資料、算力、晶片設備、製程、雲端服務與標準制定都成為關鍵節點，任何一環的優勢都可能透過系統連動被放大；其四，「量化評估更困難」。傳統軍力尚可透過衛星偵察、條約透明條款與情報蒐集相對可靠地估算，模型能力、訓練資料品質與算力部署卻難以由外部完整觀測，使相互威懾依賴的「能力可知性」假設受到挑戰。

五、輔助案例：俄烏戰爭的戰場演算法

（一）案例脈絡與實務發展。俄烏戰爭是觀察人工智慧如何重塑戰場條件與國家主權互動的指標性場域。商業衛星影像、低軌道通訊、開源情報與資料融合平臺同步運作，顯著擴大了戰場透明度，並將指揮決策的反應迴圈壓縮至前所未見的尺度，私營企業提供的衛星服務、雲端運算與通訊網路，已構成戰場感測與指揮體系不可或缺的環節（Buchanan, 2020）。截至 2026 年 5 月，烏克蘭政府與 Palantir 持續深化合作，藉由 Brave1 Dataroom 平臺訓練並部署戰場資料分析工具，另有多家民間企業投入空中威脅偵測模型的部署（Reuters, 2026）。此一發展並不意味戰場已全面

交由自主系統決斷，而是凸顯戰場資料、模型訓練與私營平臺的協作，正逐步內化為國家軍事韌性的構成要件。

（二）**理論命題呼應與機制檢證。**本案例具體檢證本文第貳節所提出的理論命題。就命題一（能力外包與主權自主性流失）而言，烏克蘭的戰場可見性與感測能力高度仰賴跨境私營企業所提供的低軌衛星（如 Starlink）與資料融合服務，當關鍵安全職能被「功能性外包」予私營技術網絡時，國家的安全自主性即受制於企業的營運裁量與合規契約，恰與可見性機制相互呼應。就命題二（即時資料依賴與責任稀釋）而言，戰場演算法對前線目標的即時辨識與自動化排序，大幅壓縮人為判斷的時間餘裕，一旦自動化流程引發附帶損害或演算法失靈，責任便分散於資料供應商、模型開發者與前線指揮官之間，使傳統國際人道法的責任推定架構遭到稀釋，跨域問責成本隨之攀升，正對應可控性與可問責性轉移機制。

（三）**主權邊界重劃效應。**俄烏經驗顯示，當戰時關鍵能力分散於跨境企業網絡，主權的「基礎設施邊界」已被重新劃定：國家安全能力與私營企業的物質基礎形成深度綁定，能否於危機中合法動員、徵用、保護並規範這些節點，遂成為當代安全治理無可迴避的核心課題。

簡言之，演算法權力帶來的，不只是新威脅，更是安全體系在時間尺度、資訊結構、能力基礎與規範適用條件上的明顯位移，理解這種位移，是建立新治理框架的前提。

肆、主權邊界重劃：科技企業、供應鏈與數位領地

一、科技企業作為準主權行為體

由於大型人工智慧企業、雲端服務商、晶片業者、衛星通訊公司與社群平臺，掌握模型能力、資料通道、算力配置與資訊分發，足以影響公共秩序與安全決策。Bradford（2023）將歐洲、美國、中國圍繞全球技術治理的競爭，描述為「數位帝國」的規範博弈，大型科技企業在其中既是被規範的對象，也是可被動員的工具。

在安全治理上，關鍵不在於企業是否擁有正式主權，而在於其技術決策會事先影響風險如何呈現、服務何時開放，以及資料能否保存，一旦國家必須依賴外部平臺進行威脅辨識或危機通報，安全權威便會被分散到契約條款、技術標準與介面設計等細部判斷。

不過，將這些企業稱為「準主權行為體」，並非它們已取代國家，而是指出它們在特定領域承擔過去由國家掌握的功能：界定可接受言論、決定關鍵服務是

否可用、管理跨境資料流動，並在某些情況下行使武力以外的強制力（如平臺封鎖、雲端中止服務、應用程式介面限制。

Zuboff（2019）提出「監控資本主義」概念，指出部分企業已建立超越單一國家的資料蒐集與行為預測能力，這種強制力通常不以命令形式出現，而是透過登入權限、資料可攜性、雲端區域選擇或模型存取額度展現，其基礎雖是私法契約，實際效果卻可能接近公共權力。

準主權行為體的形成，並非企業單方面擴權的結果，而是國家委託、市場競爭與技術路徑交互作用的產物。國家在情報、國防與認知安全等領域，已部分依賴這些企業提供的服務與基礎設施，因此「準主權」概念的分析重點，在於說明國家安全能力與企業營運決策相互綁定，而非將企業直接視為國家。

二、平臺主權與制度邊界

平臺的服務條款、應用程式介面政策、模型釋出規則、內容審核標準與雲端可用性，實際上決定了國家、企業與個人的行動空間。Floridi（2020）指出，數位主權的競逐已不再限於國家，也延伸到國家與企業之間，以及不同治理模式的角力。當平臺以風險控管、濫用防治或合規為由調整模型存取條件時，外界往往難以判斷其決策依據，制度邊界也被推向企業內部治理文件與技術審核程序。

歐盟透過《一般資料保護規則》（*General Data Protection Regulation, GDPR*）、《數位服務法》（*Digital Services Act, DSA*）、《數位市場法》（*Digital Markets Act, DMA*）與《人工智慧法》，建立以風險分級為核心的規範路徑，並把部分制度邊界轉化為全球平臺的合規要求，美國透過行政命令、出口管制與安全標準，推動兼顧安全與創新的治理模式，中國以《中華人民共和國網路安全法》、《中華人民共和國資料安全法》與《生成式人工智慧服務管理暫行辦法》等規範，形成主權安全與內容治理並行的體系。其治理模式的差異，除了是法律選項不同，更反映出制度層次上對主權邊界的不同劃定方式（Bradford, 2023；Couture & Toupin, 2019）。

這種治理模式反映出威權體制在演算法治理由主權邏輯所主導的特徵。相對於歐盟捍衛公民權利、美國聚焦市場創新與國防需求，以中國為代表的威權體制，其人工智慧治理更接近「國家中心的演算法全面管控」。

在制度邊界劃定，威權國家並非被動地對跨國平臺進行事後規制，而是透過演算法備案、內容安全審查與社會信用體系連動等事前與過程工具，將政權存續的需求直接編碼進生成式模型的訓練文本與模型權重之中。由此形成一種「制度邊界內置化」機制，平臺不再是與國家競逐主權的外部行為體，而被吸納為國家主權能力的延伸載體。

威權國家將政治紅線轉譯為技術標準，強制平臺承擔高密度的認知過濾義務，把傳統的邊境資訊審查升級為動態且即時的「演算法數位疆界」。此舉雖大幅壓縮企業的自主裁量空間，卻在形式上以極高的強度鞏固國家控制力，恰與民主國家在公私邊界模糊化過程中所面臨的權威流失形成鮮明對照（Bradford, 2023）。

截至 2026 年 5 月，歐盟《人工智慧法》已進入分階段適用期，通用目的人工智慧模型義務自 2025 年 8 月生效，歐盟執委會對相關義務的執法權限則將自 2026 年 8 月啟動（European Commission, 2026a, 2026b）。美國在 2025 年後，轉向以人工智慧行動計畫、基礎設施擴張、聯邦採購治理與晶片出口管制調整，推動結合創新、基礎設施與國際競爭的治理敘事（White House, 2025a, 2025b；Bureau of Industry and Security, 2025）。中國於 2025 年發布《人工智慧安全治理框架》2.0 版，強調風險分類、分級與安全可控（中共中央網路安全和資訊化委員會辦公室，2025）。台灣於 2025 年通過《人工智慧基本法》，為產業發展、人權保障、數位平等與安全應用建立初步法制基礎（數位發展部，2025）。

對中型國家而言，關鍵難題是在三方規範競爭下維持制度自主，避免被動承擔「規範採購」（regulatory borrowing）的成本。因此，制度自主不應被理解為排斥外部規範，而應包含本國進行風險評測、法規轉譯與執法優先排序的能力，否則合規成本就會轉化為安全政策的被動依賴。

三、供應鏈主權

先進晶片、半導體製造設備、電子設計自動化工具、雲端算力與關鍵礦物，已成為國際安全競爭的核心資源。Miller（2022）指出，全球半導體供應鏈高度集中於少數國家、企業甚至工廠的運作狀況，就足以牽動全球安全結構的穩定。從這個角度而言，半導體不單是一般生產投入，而是人工智慧模型訓練、推論與大規模部署的前提，若缺乏穩定的晶片與能源供應，演算法能力便難以轉化為可持續的安全能力。

Farrell 與 Newman（2019）以「武器化的相互依賴」（weaponized interdependence）說明，全球網路的咽喉節點如何被掌握者用於資訊蒐集與排除對手。半導體供應鏈正是典型的咽喉節點，先進製程集中於少數企業，設備仰賴特定供應商，設計工具受少數平臺主導，任何節點都可能成為地緣政治競逐的籌碼。換言之，供應鏈風險的政治性，來自節點稀少與替代成本過高，一旦某國或企業掌握咽喉節點，其他行為體在政策選擇就必須預先評估被排除、延遲交貨或授權中止的風險。

美中圍繞先進晶片、製程設備與算力出口的競爭，已不再只是貿易議題，而是主權邊界重劃的標誌性範疇。其影響也不限於中美兩強，而是透過協力廠商合規要求、出口管制延伸與供應鏈重組，對所有占據關鍵節點的中型國家造成衝擊。

對中型國家而言，關鍵不只是是否在大國競爭中選邊，而是能否將自身的節點優勢轉化為穩定合作的籌碼，並避免單一市場或單一技術規格成為政策防禦弱點。

2025 年美國商務部撤回原定上路的人工智慧擴散規則，宣布將以替代規則與執法指引強化全球半導體出口管制，顯示出口管制工具已從限制特定企業取得晶片，轉向管理算力、雲端服務與盟友合規秩序的制度安排（Bureau of Industry and Security, 2025）。該變化支持本文「供應鏈即主權」的命題：當算力取得、設備授權與雲端部署被納入安全政策，主權邊界便會沿著供應鏈節點向外延伸。

四、公私邊界模糊

國防創新單位、情報機構、軍事承包商、人工智慧企業與雲端平臺，正共同形成新的「軍工－演算法複合體」（military-industrial-algorithmic complex）。國家安全能力日益依賴私營技術網路，企業的戰略決策也愈來愈受國家安全政策牽動，此情況引起傳統軍工複合體對國防採購與產業利益連結的關注，但更強調資料中心、模型供應商與平臺服務在危機中的運作能力。

這種現象具有雙重治理意涵。一方面，國家對核心安全能力的掌握出現「功能性外包」，必須透過契約、規制與安全審查重建可控性，另方面，企業若是參與國家安全任務，也須承擔相應的合規、透明與責任。如何在創新效率與民主問責取得平衡，已成為人工智慧時代主權邊界調整的重要挑戰之一。若契約只重視服務可用率或商業保密，便不足以處理戰時服務撤離、資料保存、模型誤判與跨境救濟等問題。基此，安全採購應將可稽核性、退出機制與事故通報納入契約條件。

五、國家的再領土化

面對數位領地擴張，國家並非只能被動承受主權侵蝕，而是可透過多種策略推動「再領土化」（re-territorialization）。再領土化並非回到封閉的網路邊界，而是在開放連結中劃定最低控制線，確保資料、模型與基礎設施在危機時仍能由國家合法調度，具體作法包括：

（一） 推動資料與算力在地化，包括關鍵資料境內儲存、政府雲端優先採用本地服務商，以及管制敏感模型訓練資料。

（二） 發展主權雲（sovereign cloud）與本國基礎模型計畫，藉此建立最低限度的基礎能力自主。

（三） 強化外資與投資審查，對關鍵技術併購、技術授權與研發合作進行安全把關。

（四） 實施出口管制與技術管控，涵蓋先進晶片、製程設備、模型權重、雲端算力與雙重用途人工智慧能力。

（五） 加強平臺規制與演算法備案，要求大型平臺在地註冊、提交演算法說明、接受透明度檢視，並承擔內容治理義務。

（六） 擴大關鍵基礎設施安全審查，將雲端、海底電纜、衛星通訊與大型資料中心納入國安資產管理框架。

這些措施應依風險高低排序，否則容易把低敏感資料納入過度管制，反而壓縮研究合作與產業創新的空間，較可行的作法，是先界定國防、選舉、能源、金融、醫療等領域的最低安全需求，再配置在地化與備援要求。

值得補充的是，上述「再領土化」策略在不同政體下的推動強度與制度依賴路徑存在本質差異。民主國家在推行資料在地化或供應鏈韌性時，經常受制於私法契約、跨國自由貿易安排與公民隱私權保障，其再領土化過程常伴隨公共權力與私營企業之間的法律攻防。相對地，威權體制的再領土化展現「總體國家安全觀」下的強制整合特徵：算力基礎設施、主權雲與資料管線不僅被視為經濟投入，更被界定為直接攸關政權安全的物質基礎。威權國家藉由國家資本大規模注資關鍵技術核心，並以國安法規將私營科技巨頭「擬制」為國家機關的技術延伸，在實務上模糊了公私界線。此種模式雖能在短期內展現高度的動員效率與控制強度，卻高度依賴行政命令與政治忠誠的運作邏輯，長期而言可能使技術創新生態趨於僵化，並在跨境問責與多邊標準協商承擔更高的國際信任赤字與制度排斥成本。

六、輔助案例：美中半導體算力競爭與威權體制治理之對比

（一） **案例脈絡與實務發展。**美中圍繞先進晶片、半導體製造設備、雲端算力與雙重用途模型的競逐，是主權邊界沿供應鏈節點向外延伸的標誌性範疇。2025 年美國商務部撤回原定實施的人工智慧擴散規則，轉以替代規則與執法指引管理雲端服務與盟友合規秩序（Bureau of Industry and Security, 2025），確認了全球供應鏈關鍵樞紐日益走向政治化與武器化的趨勢。在此脈絡下，民主同盟與威權體制展現出截然不同的主權邏輯，本案例即藉由雙政體對照檢證本文第貳節的理論命題。

（二） **理論命題呼應與機制檢證。**就命題一（能力外包與主權自主性）而言，先進晶片與算力配置是人工智慧模型部署的物質前提。在民主邏輯下，國家安全能力可能因依賴跨境私營企業而出現「功能性外包」，因此需透過私法契約與安全審查重建可控性。相較之下，威權邏輯則依循「總體國家安全觀」，將私營科技巨頭視為國家機關的技術延伸，並透過國家資本投入關鍵節點以維持主權自主，此模式雖具高度動員效率，卻可能因過度倚賴政治忠誠而抑制創新活力。就命題三（規範分裂與技術陣營化）而言，美中雙方藉由出口管制清單、實體清單與雙重用途規範，迫使全球協力廠商與盟友在合規秩序上「選邊」。美國《美國人工智慧行動計畫》、

中國《人工智慧安全治理框架》2.0 版與歐盟《人工智慧法》在執法上的實質分歧，加上威權體制特有的「制度邊界內置化」與演算法數位疆界，共同推動全球人工智慧治理走向技術陣營化與規範碎片化，並削弱跨陣營共同供給全球公共財（如跨境安全問責）的能力（Miller, 2022；Farrell & Newman, 2019）。

（三） **主權邊界重劃效應。**美中半導體競爭印證了主權並非遭人工智慧單向侵蝕，而是國家透過法規限制與供應鏈政策重建控制力的「再領土化」政治過程。領土、制度、基礎設施與認知四個維度在本案例相互影響，亦即出口管制改變晶片供給，晶片供給約束算力部署，算力部署牽動模型研發節奏，而技術主權論述又回過頭形塑社會對風險與盟友承諾的判讀。此案例的雙政體對照亦顯示，主權重劃並非主權衰退的單線敘事，而是民主與威權體制在外部依賴與內部能力建置之間持續校準的政治過程。

伍、邁向人工智慧時代的國際安全治理

一、現行治理機制的侷限

就當前局勢而言，人工智慧治理大致形成三條並行路徑：歐盟以風險分級為核心，美國強調安全與創新並進，中國著重主權安全與內容治理（Bradford, 2023）。歐盟以《人工智慧法》為代表，美國自 2025 年起，明顯轉向以《美國人工智慧行動計畫》（*America's AI Action Plan*，*AI Action Plan*）與《排除美國人工智慧領導障礙行政命令》（*Executive Order 14179, Removing Barriers to American Leadership in Artificial Intelligence, EO14179*）為政策主軸，把創新、基礎設施與國際競爭提升至國家安全戰略高度，中國以《生成式人工智慧服務管理暫行辦法》作為管理生成式模型服務的重要依據。

在多邊層面，相關安排包括聯合國體系下的討論、七大工業國集團廣島人工智慧進程（G7 Hiroshima Process on Generative Artificial Intelligence）、《經濟合作暨發展組織人工智慧原則》（*OECD AI Principles*），以及由英國發起並延續的「人工智慧峰會」系列。前兩屆分別於 2023 年在英國舉行，2024 年在韓國首爾舉行，稱為「人工智慧安全峰會」（AI Safety Summit），2025 年在法國巴黎舉行，名稱調整為「人工智慧行動峰會」（AI Action Summit），反映議題重心從安全承諾轉向行動落實與利益分配。2026 年印度召開人工智慧影響力峰會（AI Impact Summit）則凸顯全球南方與公共服務導向在治理議程的能見度提升（Élysée, 2025；IndiaAI, 2026）。

然而，上述發展雖提高國際對話密度，仍未彌平實質法律、自願性規範、企業承諾與國家安全例外的制度落差。

現行治理機制在回應人工智慧時代的國際安全問題時至少面臨四項限制。第一，具法律強制力的規定和非強制性原則之間仍缺乏有效銜接，企業自律承諾在民主正當性與可驗證性明顯不足。第二，國家安全例外條款在實務上可能成為規避透明與問責的理由，而傳統軍備管制工具也難以涵蓋算力、模型與資料流動等新型態競爭。第三，跨境平臺與前緣模型的權責歸屬網絡仍受限於「平臺所在地、資料所在地、損害所在地」的管轄競合。第四，標準制定機構雖具技術專業，但其決策程序的代表性與透明度仍待加強。這些限制說明，若人工智慧治理僅停留在產業合規或倫理宣示層次，便難以處理安全能力外包、危害難以歸因，以及制度責任被切割等問題。

二、治理缺口與需求

前述分析顯示，現行治理機制至少有三個關鍵問題值得注意。

（一） 權威缺口：當安全功能依賴跨境技術網路，主權國家難以單獨建立有效規範，既有國際組織也未必具備足夠的尖端技術理解與執行能力。問題不在於國家權威消失，而在於權威必須透過平臺規則、雲端契約與技術標準才能發揮，若國家缺乏判讀模型能力與基礎設施風險的專業，形式上的主權雖仍存在，安全判斷卻可能受外部條件牽制。

（二） 問責缺口：當損害由多方行為體共同造成，且因果鏈難以重建時，受害者多數時候難以獲得有效救濟，行為體也缺乏內化外部成本的誘因，困難不只在於單一責任人不願承擔責任，更在於資料供應者、模型開發者、平臺部署者與使用者各自只掌握部分資訊，導致法律責任與技術因果難以順利銜接。

（三） 韌性缺口：當關鍵基礎設施高度集中於少數節點，整體系統便會出現深層脆弱性，但現行治理機制較少把韌性視為需要共同投資與協調的國際公共財。晶片、雲端、海底電纜、資料中心與大型模型服務的集中，使單點中斷可能擴大為跨國安全事件，因此韌性不僅是各國內政，也應成為國際安全協調的共同議題。

三、制度方向

針對上述缺口，本文提出六項彼此連動的制度建議。

其一，建立高風險人工智慧系統的透明、測試、審計與事故通報機制。透明不單是指「公開原始碼」，還應涵蓋用途說明、訓練資料來源類別、評測結果、已知限制、紅隊測試發現與重大事故揭露。歐盟《人工智慧法》對高風險系統的合規要求可作為借鏡，但也須避免將透明義務過度加諸資源有限的中小行為體。

較可行的作法是把透明區分為「公開透明」、「監理透明」與「機密審計」三個層次，讓商業秘密與國家安全資料無須完全公開，但仍可接受合格第三方檢視。

其二，將基礎設施韌性納入國際安全治理。相關措施包括晶片供應鏈多元化、雲端服務跨境互通、關鍵模型部署備援，以及跨境資料依賴的風險評估。Farrell 與 Newman（2019）主張將「去武器化的相互依賴」概念落實到人工智慧治理，在保留合作效益之際，降低單一節點被工具化的風險。對中型國家而言，這意味著安全政策不能只追求國產化，還應建立替代供應、相互援助與中斷情境演練。

其三，強化跨境認知安全合作。可行作法包括深偽內容標示與可追溯性標準、平臺間協同共享不實行為資料、選舉期間快速通報機制，以及對公民社會韌性的長期投入，此類合作高度仰賴公私部門與跨國網路協作，必須在保障言論自由與遏止惡意操縱維持精細平衡（Singer & Brooking, 2018；Chesney & Citron, 2019）。尤其在民主社會，若治理工具過度依賴下架與封鎖，反而可能削弱公共信任，更穩健的作法是提升內容來源辨識、平臺廣告透明與公民媒體素養。

其四，發展可驗證的人工智慧軍備管制工具。傳統軍備管制依賴可量化的硬體特徵，在人工智慧時代則需發展模型卡（model card）、運算用途報告、訓練算力閾值通報、評測基準與信賴建立措施（confidence-building measures）等工具。雖然短期內難以達成全面禁令，但局部、漸進的透明與通報機制仍可作為過渡安排。此處不宜直接套用《不擴散核武器條約》（*Treaty on the Non-Proliferation of Nuclear Weapons，NPT*）模式，因為模型能力可複製、可微調，也可能包裹於民用服務，較可行的方向是優先處理高風險軍事用途、訓練算力門檻與自動化攻擊程序。

在此基礎上，國際社會更應正視自動化決策所衍生的問責真空，並發展兩項具批判性的法律工具。一為「自動化偏誤的法律推定責任」，當致命或高風險決策係在人為監督形同虛設的自動化流程中作成時，舉證責任宜由受害方轉向系統的部署者與開發者，推定其對可預見的演算法失靈負有注意義務，藉以扭轉「人在迴圈」流於形式，導致責任被稀釋的困境。另為「跨境救濟的證據移轉程序」，針對模型日誌、訓練資料類別與決策軌跡等關鍵技術證據，建立跨司法管轄的強制揭露與保全機制，使受害者得以在損害發生後重建因果鏈，避免商業機密或國家安全例外淪為規避救濟的制度屏障。

其五，建立多層次技術多邊主義。在全球層面維持基本原則對話，在區域層面深化具信任基礎的合作（如同盟內部的演算法情報共享），並在功能層面建立議題導向的治理聯盟（如深偽治理或關鍵礦物供應）。這種多層次設計有助於緩解「全有或全無」的治理困境，也能降低治理安排完全陣營化的風險，其制度重

點不在追求一套立即普遍適用的單一條約，而在於逐步累積可互認的測試、通報、審計與救濟程序，形成穩定慣例。

其六，建立因應陣營對抗的技術互認與分級分流機制，以緩解規範碎片化。隨著美中等大國將人工智慧納入地緣戰略核心，全球治理日益呈現民主陣營強調人權、信任與安全，威權陣營強調政權可控與內容審查的分裂格局。2025 年巴黎「人工智慧行動峰會」，大國圍繞基礎設施控制與合規秩序的對抗，顯示跨陣營形成單一強制性條約的可能性已相當有限（Élysée, 2025）。

面對此一「技術鐵幕」，較務實的作法是採取「分級分流與技術互認」，在低敏感度且具全球公共財性質的領域，如氣候模型、跨國醫療影像標準與重大事故通報，可透過聯合國或經濟合作暨發展組織等技術多邊平臺，推動去政治化的最低技術共識與互認機制，在高敏感且涉及地緣安全的領域，如核心技術參數外流、戰場演算法供應鏈與認知作戰防禦，民主陣營則應深化內部「高信任技術同盟」，並以七大工業國集團「廣島人工智慧進程的志願性報告框架」（*Voluntary Reporting Framework*）為參照，將紅隊測試標準、運算用途報告與供應鏈關鍵環節管制轉化為可互相承認的規範，以降低重複合規成本。

對處於規範碎片化夾縫中的中型國家，如台灣，較可行的方向是推動「技術安全雙軌制」。一方面，在核心國防與政府資產建立排除威權演算法基礎設施的「高信任數位邊界」，另方面，運用自身在半導體製造與供應鏈關鍵環節的不可替代性，將「節點實力」轉化為參與風險審查與演算法可稽核標準制定的實質發言權，藉由技術韌性鞏固主權自主。

四、國家角色的轉型

人工智慧治理不會取代主權，但會促使主權功能轉型，國家角色應重新定位。首先，由單純管制者轉為標準參與者。技術標準既是競爭場域，也是規範擴散的重要管道，與其被動接受既成規則，不如主動參與制定。其次，由消極監督者轉為基礎設施投資者，在算力、資料、人才與評測能力維持基本自主，是取得國際治理發言權的前提。

再者，由危機回應者轉為風險協調者。國家應建立跨部門、跨產業、跨國界的長期協調機制，將個案處理轉化為制度學習。此外，由國內法執行者轉為跨境問責架構的形成者。透過雙邊、區域與多邊合作，補足單一管轄區的能力限制。

換言之，國家不僅是事後裁罰者，也應在事故發生前建立資料保存、證據移轉、共同調查與救濟互認程序，讓主權在跨境技術網路具備可操作性。

五、對中型國家的策略意涵

對台灣等中型國家而言，重點不在追求全面技術自主，而在於作出正確的策略選擇，包括建立最低關鍵能力，例如基礎模型評測、紅隊測試與事故調查；掌握全球供應鏈的關鍵節點優勢，將節點價值轉化為治理發言權；積極參與國際標準與規範形成，避免被動承擔他國規範外溢成本；提升選舉與社會韌性，將認知安全納入國家安全戰略主體（國家科學及技術委員會，2023）。

就此而言，台灣的優勢不僅在於半導體製造，更在於可信賴的民主制度、供應鏈管理經驗與高科技產業網路的結合，若能把節點優勢轉化為評測、通報與標準合作能力，便可在大國競爭下維持穩定的制度發言位置。

2025 年《人工智慧基本法》通過後，台灣已具備制定風險分級、資料治理、產業創新與安全應用配套規範的法制基礎，後續關鍵不在於抽象宣示「可信賴人工智慧」（Trustworthy AI），而在於能否建立可稽核的模型評測、政府採購準則、關鍵基礎設施 AI 風險管理與跨境事故通報制度（數位發展部，2025）。

陸、結 論

本文以演算法權力與主權邊界的關係為主軸，探討人工智慧治理如何衝擊傳統國際安全體系，以及各國可採取的回應方向。具體而言，本文並非要求各國回到封閉式技術民族主義，而是在兼顧「開放合作」與「安全自主」的前提下，設計可被檢驗的制度界線，若缺乏這種界線，人工智慧治理容易在效率、合規與安全三種目標之間反覆擺盪。

近年來，人工智慧治理制度的演進與戰場實務的發展，印證本文論點。人工智慧治理已從倫理原則與產業監理，轉向涵蓋算力、模型、平臺、資料、能源與供應鏈的安全治理。歐盟分階段推動執法、美國強化基礎設施與出口管制、中國更新安全治理框架、台灣通過《人工智慧基本法》，以及《特定常規武器公約》對自主武器的討論逐步條文化，皆顯示主權邊界不再只是法律界線，而是由技術、基礎設施與跨境問責共同塑造的制度邊界。

本文研究發現如下：

一、 **演算法權力透過「功能性外包」侵蝕主權自主性。**人工智慧對國際安全的衝擊，核心在於戰場感測、目標識別與供應鏈關鍵環節等重要功能，已深度嵌入跨境技術網路與私營基礎設施。俄烏戰爭與美中半導體競逐的經驗證據顯示，當國家安全高度依賴外部平臺與晶片設備時，其對威脅圖景的「可見性」與行動選項的「可控性」即面臨實質外包，從而在維護國家自主權上面臨體制性制約，印證了命題一與可見性、可控性機制的聯動。

二、 **自動化決策與跨域分散網路重劃主權邊界，並稀釋責任歸屬。**主權並非遭人工智慧單向侵蝕，而是在領土、制度、基礎設施與認知四個維度經歷動態重劃。然而，正如俄烏戰場演算法與威權體制「制度邊界內置化」呈現的灰色地帶，安全領域越依賴即時資料與自動化判斷，傳統國際法預設的「誰決策、誰負責」責任架構就越容易被稀釋，可問責性轉移使損害的因果鏈難以重建，導致跨域問責與法規轉譯的政治成本大幅攀升，驗證了命題二下的問責稀釋與高昂成本。

三、 **治理規範分裂加劇技術陣營化，削弱全球公共財供給。**歐盟風險分級、美國創新防禦與中國威權式演算法控制的三方規範競爭，已在實務上造成全球人工智慧治理體系的實質割裂。美中算力競爭與出口管制的外溢效應，印證人工智慧治理規範越分裂，國際安全體系越走向技術陣營化與規範碎片化。此一地緣政治鐵幕切斷跨陣營建立單一強制性條約的可能，使跨境安全救濟與重大事故通報等全球公共財的供給能力嚴重削弱，並迫使中型國家轉而採取「技術安全雙軌制」以求自保，驗證了命題三對全球公共財供給能力的削弱。

本文有三項學術貢獻。第一，將演算法權力界定為一種結構性權力，並以可見性、可控性與可問責性轉移三項機制加以操作化，使其能與既有結構權力理論展開實質對話，而非被簡化納入。第二，提出主權邊界的四維分析框架，將「主權重劃」轉化為可比較的研究路徑。第三，將戰場演算法基礎設施、半導體與算力競爭，以及平臺與認知作戰三類看似不同的現象，置於同一理論架構中分析，呈現人工智慧治理對國際安全的整體影響。

這些貢獻的意義在於，本文並未把人工智慧視為外部衝擊，而是將其置於國際安全權威分配的長期變遷之中，因而能解釋技術能力、企業角色與國家主權如何相互牽動。

在政策啟示上，本文建議各國建立最低關鍵能力，包括模型評測、紅隊測試、事故調查與供應鏈風險監測，掌握關鍵供應鏈節點優勢，並將其轉化為國際治理發言權與安全合作籌碼。此外，各國也應積極參與國際標準制定，避免被動承擔規範外溢成本，且將認知安全納入國家安全戰略，以提升選舉與社會韌性。

儘管本文具有前述的學術與實務價值，仍有三項研究限制。第一，公開資料不足以完整呈現模型部署、情報合作與平臺內部決策的實況，因此部分論證仍須依賴間接證據。第二，本文選用的案例多為高能見度事件，可能低估演算法權力在日常治理中的潛在作用。第三，本文主要採取民主治理視角，對非民主治理模型的內部邏輯著墨較少。

因此，本文的推論範圍應界定為「機制導向」而非「總體效果估計」，本文可說明演算法權力如何重組安全權威，但尚無法量化人工智慧治理對戰略穩定、民主信任或供應鏈韌性的整體效果，這亦顯示，後續研究仍需補充可比較資料與跨國測量指標。

後續研究可朝三個方向延伸：第一，建立可比較的演算法權力衡量指標，為跨國量化比較奠定基礎；第二，比較中型國家在人工智慧安全領域的戰略選擇，以辨識不同國家規模與制度條件下的可行路徑；第三，擴大研究視野，檢視人工智慧與量子科技、生物科技、太空基礎設施交互作用形成的複合安全效應，以掌握下一波技術匯聚對國際安全體系的可能衝擊。未來若能取得平臺治理與模型部署資料，將有助於檢驗本文命題，並補足資料的侷限。

參考文獻

中共中央網絡安全和信息化委員會辦公室（2025）。**《人工智能安全治理框架》2.0 版發布**。9 月 15 日。https://www.cac.gov.cn/2025-09/15/c_1759653448369123.htm [Office of the Central Cyberspace affairs Commission. (2025). *Release of the Artificial Intelligence Safety Governance Framework (Version 2.0)*. September 15. https://www.cac.gov.cn/2025-09/15/c_1759653448369123.htm]

國家科學及技術委員會（2023）。**台灣 AI 行動計畫 2.0（2023-2026 年）**。[National Science and Technology Council. (2023). *Taiwan Artificial Intelligence Action Plan 2.0 (2023-2026)*.]

數位發展部（2025）。**立法院三讀通過《人工智慧基本法》 構築我國 AI 創新與安全治理基石**。12 月 24 日。https://moda.gov.tw/press/press-releases/18316 [Ministry of Digital Affairs. (2025). *The Legislative Yuan passed the third reading of the "Artificial Intelligence Basic Act," building the foundation for Taiwan's AI innovation and safety governance*. December 24. https://moda.gov.tw/press/press-releases/18316]

蘇紫雲、曾怡碩（編）（2018）。**2018 國防科技趨勢評估報告**。國防安全研究院。[Su, T. Y., & Tzeng, Y. S. (Eds.). (2018). *2018 defense technology trend assessment report*. Institute for National Defense and Security Research.]

Barnett, M., & Duvall, R. (2005). Power in international politics. *International Organization, 59*(1), 39-75. https://doi.org/10.1017/S0020818305050010

Bode, I., & Huelss, H. (2018). Autonomous weapons systems and changing norms in

international relations. *Review of International Studies, 44*(3), 393-413. https://doi.org/10.1017/S0260210517000614

Bradford, A. (2023). *Digital empires: The global battle to regulate technology*. Oxford University Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., ÓhÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., ... Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute, University of Oxford. https://arxiv.org/abs/1802.07228

Buchanan, B. (2020). *The hacker and the state: Cyber attacks and the new normal of geopolitics*. Harvard University Press.

Bureau of Industry and Security. (2025). Department of Commerce announces rescission of Biden-era Artificial Intelligence Diffusion Rule, strengthens chip-related export controls. U.S. Department of Commerce, May 13. https://www.bis.gov/press-release/department-commerce-announces-rescission-biden-era-artificial-intelligence-diffusion-rule-strengthens

Cave, S., & ÓhÉigeartaigh, S. S. (2018). An AI race for strategic advantage: Rhetoric and risks. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 36-40). Association for Computing Machinery. https://doi.org/10.1145/3278721.3278780

Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs, 98*(1), 147-155.

Couture, S., & Toupin, S. (2019). What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society, 21*(10), 2305-2322. https://doi.org/10.1177/1461444819865984

DeNardis, L. (2014). *The global war for internet governance*. Yale University Press.

Élysée. (2025). Statement on inclusive and sustainable artificial intelligence for people and the planet. February 11. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

European Commission. (2026a). AI Act. Shaping Europe’s Digital Future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission. (2026b). Guidelines for providers of general-purpose AI

models. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers

Farrell, H., & Newman, A. L. (2019). Weaponized interdependence: How global economic networks shape state coercion. *International Security, 44*(1), 42-79. https://doi.org/10.1162/isec_a_00351

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology, 33*(3), 369-378. https://doi.org/10.1007/s13347-020-00423-6

Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. *Texas National Security Review, 1*(3), 36-57. https://doi.org/10.15781/T2639KP49

IndiaAI. (2026). India AI Impact Summit 2026. https://impact.indiaai.gov.in/

Lukes, S. (2005). *Power: A radical view* (2nd ed.). Palgrave Macmillan.

Miller, C. (2022). *Chip war: The fight for the world's most critical technology*. Scribner.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Reuters. (2026). Zelenskiy meets Palantir CEO as Ukraine expands use of AI in war. Reuters.

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. W. W. Norton.

Schmitt, M. N. (Ed.). (2017). *Tallinn manual 2.0 on the international law applicable to cyber operations*. Cambridge University Press.

Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The weaponization of social media*. Houghton Mifflin Harcourt.

Strange, S. (1996). *The retreat of the state: The diffusion of power in the world economy*. Cambridge University Press.

The White House. (2025a). White House unveils America's AI Action Plan. July 23. https://www.whitehouse.gov/releases/2025/07/white-house-unveils-americas-ai-action-plan/

The White House. (2025b). Fact sheet: President Donald J. Trump takes action to enhance America's AI leadership. January 23. https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/

United Nations Office for Disarmament Affairs. (2026). Chair's summary – first 2026 session of the GGE on LAWS (CCW/GGE.1/2026/WP.2). https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_%282026%29/CCW-GGE.1-2026-WP.2.pdf

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

The Impact of AI Governance on the Traditional International Security System: Challenges and Responses

William Yang*

Abstract

With the rapid proliferation of generative AI and advanced models, AI governance is no longer merely a subordinate issue within technological ethics or industrial regulation. It has progressively emerged as a core proposition of international security and the sovereign order. This article uses the relationship between "algorithmic power" and "sovereign boundaries" as an analytical framework to elucidate how AI governance necessitates adjustments to the traditional international security system. This paper argues that algorithmic power is not a tool-based capability arising from a single model or application, but a composite form of power synthesized from data collection, computational resource allocation, model deployment, platform distribution, chip supply, and technical standards. Through three mechanisms, i.e., the transfer of visibility, controllability, and accountability, it reshapes the distribution of security authority among nation-states, technology enterprises, and international organizations. Integrating theories of structural power, digital sovereignty, and AI governance, and referencing empirical cases such as battlefield algorithmic infrastructure, U.S.-China semiconductor and computing competition, and platform recommendation-driven cognitive warfare, this study finds that the impact of AI on international security is not confined to military automation. Rather, it extends to supply chain sovereignty, the establishment of platform rules, and the redrawing of cognitive security boundaries. Consequently, AI governance should transcend technical regulation and instead incorporate infrastructure resilience, cross-border accountability, and multi-level technological multilateralism. The role of the state must transition from a mere regulator to a participant in standards, an investor in infrastructure, and a facilitator of cross-border accountability frameworks.

Keywords: Algorithmic Power, Sovereignty Boundaries, Digital Sovereignty, Artificial Intelligence Governance, International Security

---

* William Yang received his Ph.D. in Political Science and International Relations from University of Warwick (U.K.) and is currently Professor in the Department of Public Administration and Management at Shih Hsin University. He has previously served as Dean of International Affairs, Dean of Public Affairs, and Chair of the General Education Committee. His research expertise includes AI in Social Science, Trumpism, Global Governance, International Climate Politics. Email: william@mail.shu.edu.tw.
The paper was published under two double-blind reviews. Received: May 10, 2026. Accepted: August 4, 2026.