

人工智慧對國際和平的機遇與挑戰

楊惟任*

世新大學行政管理學系教授

摘要

本文旨在探討人工智慧對國際和平建構的影響，分析其作為技術工具的潛力與風險。隨著和平研究從傳統的權力平衡論演進至涵蓋「積極和平」與「混合和平」的多元觀點，人工智慧技術的介入正改變國際社會應對衝突的模式。研究指出，人工智慧透過巨量資料分析與模式辨識，能將衝突預警從被動回應轉為主動預防；利用自然語言處理工具，輔助和平談判並擴大邊緣社群的參與，促進包容性對話；透過衛星影像與地理資訊系統的整合，提升維和行動監測與人道援助資源配置的精準度與效能。另一方面，演算法偏差可能再現結構性暴力，導致弱勢群體落入演算法的識別盲區；決策過程的黑箱化造成課責困境，削弱制度正當性；資料與算力的集中引發資料殖民與權力不對稱，弱化在地主體性；資訊污染與監控濫用侵蝕社會互信。本文發現，人工智慧賦能和平的關鍵不在於技術本身，而是如何將其嵌入可課責、可協調且可檢驗的制度框架，因此本文建議架構以人機協作為核心的歸責機制，校正演算法的文化與地域偏差，並透過多邊協調緩解全球治理碎片化，使人工智慧成為建構國際和平的有效工具。

關鍵詞：人工智慧、和平研究、和平建構、演算法治理、國際安全

*英國 University of Warwick 政治學暨國際關係博士，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，研究專長包括政治學、國際事務、全球治理、人工智慧與社會科學。Email: william@mail.shu.edu.tw。

壹、前言

國際關係學（International Relations, IR）對和平的探討，可視為人類回應集體暴力與治理困境的知識演化，相應地，和平研究的理論與方法途徑，也會隨國際體系權力結構的變遷而持續調整。早期的和平概念根基於規範性哲學的問題意識，以 18 世紀康德（Immanuel Kant）為例，他聲稱和平可在制度與規範的支撐下獲得維繫，透過共和政體、國際組織與跨國貿易等制度性安排，得以使國際關係脫離弱肉強食的叢林邏輯，轉向以法治為基礎的文明秩序（Kant, 1991）。

理想主義意圖透過道德、法律和國際組織建構和平，但國際聯盟的失敗凸顯規範性力量在現實政治的侷限。Carr（1939）主張權力才是國際關係的核心，沒有實力支撐的道德或法律只是空談，缺乏實力支撐的集體安全制度必然崩潰。Morgenthau（1948）亦從權力政治的角度指出，在國際社會無政府狀態與權力擴張的作用下，和平只能以權力平衡（balance of power）為支撐，且為高度脆弱且暫時性的穩定狀態。

冷戰後期，和平研究的問題意識與研究路徑逐漸從傳統外交史與戰略研究分化出來，並在跨學科取向下形成獨立的研究領域。Galtung（1969）提出「消極和平/積極和平」的分析架構，前者指向直接暴力與武裝衝突的終止，後者以消除結構性暴力為核心，強調貧窮、不平等與壓迫對人類造成的傷害。此論點讓和平不再僅被理解為「不戰」的狀態，而被重新界定為以社會正義與制度安排的可持續狀態，為後續和平研究開啟新的議程與方向。

在 1980、1990 年代，民主和平論（Democratic Peace Theory）成為和平研究的主流議題。Doyle（1983）延續康德自由主義傳統的論證，提到成熟民主國家彼此之間發生戰爭的可能性相對較低。Keohane（1984）從自由制度主義角度指出，國際制度能透過降低交易成本、提升資訊透明度並建構互惠機制，增進無政府狀態下合作的可行性，對和平的維繫具有促進作用。進入 21 世紀，和平研究融入女性主義、後殖民與批判理論等觀點。女性主義強調性別平等是和平的必要條件，提出「女性主義和平」（feminist peace）概念，將生活的性別暴力納入分析（Tickner, 2001）。後殖民與批判和平研

究對「自由和平」(liberal peacebuilding)的普世性提出質疑，認為其經常帶來新殖民主義與精英主導的問題，因此相關研究轉而強調在地主體性，並以「混合和平」(hybrid peace)概念揭示地方實踐與外部干預交織所形成的多重秩序樣態 (Richmond, 2011; MacGinty, 2014)。

當前和平研究展現高度多元且跨領域的取向，其關注目標涵蓋和平的倒退與衝突的國際化趨勢；對正面和平與負面和平概念的再檢視；女性主義與性別論述的深化；氣候、環境變遷與永續和平的關聯；去殖民化與批判和平研究的推進；科技發展、地緣政治與新型態衝突；以及和平教育與全球公民教育等議題。整體而言，和平理論已由早期偏重道德哲學與權力平衡的討論，逐步演進為涵蓋正面/負面和平、民主與自由制度、性別、環境與去殖民等多重面向的跨學科研究領域。

在技術發展與地緣政治相互影響下，人工智慧 (Artificial Intelligence, AI) 逐步介入並改變國際和平。AI 結合巨量資料分析與模式辨識能力，可被用於辨識社會仇恨言論與動員訊號、監測區域衝突、強化衝突預警、支援人道援助與資源配置的追蹤與優化 (Zelizer, et al., 2025; Giovanardi, 2024)。其次，AI 帶來顛覆性風險，例如致命自主武器系統 (LAWS) 可能對責任歸屬與國際人道法之適用造成衝擊 (Amoroso & Tamburrini, 2019; 許智翔, 2021)，民用 AI 所引發的資訊污染與監視治理，亦可能侵蝕社會互信與公共理性，並削弱政治正當性 (SIPRI, 2025; UNIDIR, 2023; 曾怡碩, 2023)。

當前 IR 文獻對 AI 的討論仍多集中於軍事層面的安全威脅，對 AI 在和平建構的積極作用及其伴隨風險，尚缺乏全面的檢視。隨著 AI 議題在國際關係與和平研究的能見度日增，「技術如何賦能和平」遂成為亟待回應的核心問題：如何應用 AI 以建構國際和平？在應用過程可能衍生哪些風險？鑒此，本文旨在分析 AI 介入和平建構之機制與條件，辨識其風險生成的主要路徑，並提出兼顧貢獻與風險的治理原則，以回應國際社會對穩定秩序與可持續和平的需求。

透過這項研究，得以更精確地界定 AI 介入和平建構的具體作用機制，例如強化社會信任的可能技術路徑，亦可就 AI 應用於和平進程所伴隨的風險提出可操作的預警指標。本文期能為國際社會推動「AI 促進和平」之政

策規劃與制度設計，提供嚴謹的學術論證與風險控管框架，一方面有助於維繫國際秩序的穩定與韌性，另一方面可回饋並擴充既有和平研究理論，在理論與實踐之間建立更有力的連結。

貳、文獻探討

本文旨在探討 AI 如何介入並重塑國際和平的建構，分析其作為賦能工具的技術潛力與風險。本節從國際關係理論著手，首先檢視和平研究如何從傳統權力政治與規範制度，轉向關注數位時代的「混合和平」與結構性正義，其次剖析 AI 技術如何透過預測、資源配置與協調機制介入和平建構，最後從批判角度闡述演算法可能帶來的認知偏差、權力不對稱及信任危機。透過對既有文獻的探討，瞭解 AI 在和平研究的角色，以及當前研究在「技術效能」與「政治正當性」之間的論述缺口，為這項研究建立論述框架。

一、和平研究的理論座標

早期的和平論述呈現理想主義與現實主義的辯證張力。康德在《永久和平論》提出，和平並非自然狀態，但透過理性的制度設計，人類可終結戰爭，實現永久和平（Kant, 1991）。此說法雖為後來的自由制度主義奠基，卻受到現實主義嚴厲的批評。Carr（1939）與 Morgenthau（1948）皆強調，在無政府狀態下，國際政治是一場零和遊戲，和平只是暫時的權力平衡狀態。自由主義和現實主義的辯論對 AI 和平研究的啟示在於：AI 究竟是促進康德式的合作，抑或強化現實主義的權力結構，取決於其嵌入的制度脈絡。

在 1960-70 年代，和平研究逐步從國際關係學分立出來，形成跨學科的獨立研究領域。Galtung（1969）提出「消極和平/積極和平」的分析架構，將和平的定義從單純的消除直接暴力（止戰），拓展至消除結構性暴力（貧窮、歧視與不平等），此一概念框架超越傳統現實主義的界定，並在冷戰結束後持續主導和平研究的理論議程。延續自由主義傳統的民主和平論與自由制度主義，提供另一組評估技術工具的概念。Doyle（1983）論證成熟民主國家之間較不易發生戰爭，Keohane（1984）宣稱國際制度可透過降低交

易成本、提升資訊透明與建立互惠期待以促進合作，若將此論證邏輯延伸至數位時代，AI 可成為降低不確定性、提升資訊處理能力與跨機構協作效率的制度性配套，但 AI 亦可能因平台壟斷、資料不對稱與演算法黑箱，造成不透明與不可課責，形成新的制度失靈來源（李建良，2021）。

近期和平研究的「在地轉向」與「批判視野」，為分析 AI 議題提供不可或缺的理論框架。Tickner（2001）從女性主義出發，將性別秩序與日常暴力納入和平分析，Richmond（2011）與 MacGinty（2014）則批判自上而下的自由和平建構，提出「後自由和平」、「混合和平」等概念，倡議外部干預與在地實踐的多重秩序。此一取向說明 AI 介入和平建構不但是效率問題，更涉及知識生產與治理權力的再分配，誰壟斷數據與算力、誰掌握指標的定義權，誰就擁有定義並實踐「和平」的權力。

二、技術賦能和平

Zelizer 等人（2025）的研究指出，AI 結合巨量資料分析與自然語言處理（Natural Language Processing，以下簡稱 NLP），已將和平建構帶入預測性時代。透過監測社群媒體的情緒光譜、經濟波動指標，乃至氣候變遷數據，機器學習模型能在衝突發生前，精準辨識動員訊號與仇恨言論的路徑。Giovanardi（2024）將此機制概念化為「人工智慧助力和平」（AI for Peace），主張這種技術大幅降低訊號雜訊比，使在特定制度框架下運作的多邊行為者，如聯合國政治及和平建構事務部、非洲聯盟等區域安全機制，得以從被動的危機反應，轉向主動的風險預防。此論點呼應 Keohane（1984）的核心主張：制度的關鍵功能在於改善行為者的資訊環境、降低不確定性，從而提升合作的可行性。AI 在此所扮演的角色，乃強化制度性資訊功能的技術機制，而非替代制度本身或賦予行為者強制執行的能力，從被動反應轉向主動預防的實現，仍須以有效的制度框架與政治意願作為前提條件。

除了衝突預警之外，AI 在人道救援與多邊外交的應用亦備受關注。在戰後重建或災難因應的混亂狀況，AI 整合衛星影像與地理資訊系統，可動態優化物資運送路徑，確保資源精準送達受創最鉅的區域（SIPRI, 2025）。此外，UNIDIR（2023）探討 AI 在「信任建立措施」的潛力，透過賽局模

擬與情境推演，AI 輔助決策系統能協助談判各方尋找核心利益的均衡點，在高度缺乏互信的無政府狀態下，降低誤判風險，強化外交談判的韌性。

然而，當 AI 嵌入衝突預警、人道分配與平台治理等公共決策流程後，風險已不再只是技術失準的工程問題，而是觸及和平建構的核心條件：信任、正當性與可課責性。相較於 LAWS 的致命性衝擊，民用 AI 更常以低可見度、長期累積的方式發生效應，它透過資料蒐集與分類邏輯重塑「風險」與「優先性」，並在黑箱決策改變責任歸屬與民主究責。尤其在權力不對稱下，資料、算力與平台控制成為新的支配資源，使弱勢國家與群體因「治理」之名被監測、被排序、甚至被排除。基此，本節將依文獻從演算法的偏差、黑箱課責、權力不對稱與監控治理四面向，分析風險路徑與制度後果（李建良，2021；許智翔，2021）。

其中普遍的問題包括演算法的偏差，源自資料不完整、標註偏差或模型遷移失靈，可能對預警、援助與治理排序造成結構性問題，使弱勢群體承擔不成比例的成本。若採用 Galtung（1969）的概念，這可被理解為「結構性暴力」在數位時代的再編碼：不平等不再以顯性的規範呈現，而是透過看似中立的分類邏輯、門檻設定與風險評分等機制被制度化。Tickner（2001）指出的日常權力不對稱，亦提示在難民審核、援助分配或線上言論治理等場域，資料缺口與偏差可能使特定群體更容易被排除。

其次、黑箱課責。縱使在非軍事領域，若政策制定過度仰賴模型輸出，在缺乏可解釋性與監管框架的情境下，一旦 AI 的和平應用引起不可回復的錯誤，責任歸屬就容易陷入模糊地帶，甚至被轉嫁為形式上的「系統建議」。此一問題與致命自主武器系統討論的責任困境在邏輯上相通，其結果是公共課責與民主究責機制被削弱，甚至動搖制度正當性（Amoroso & Tamburrini, 2019）。

再者，掌握資料、算力與平台的國家或企業，可採用技術優勢制定標準與規則，使和平建構資源與治理能力向少數行為者集中，弱勢國家與群體則淪為被監測、被抽取資料的對象。此風險呼應批判和平研究對外部干預與知識支配的警惕：AI 所承載的並非純粹技術理性，而是特定政治經濟結構的延伸（Richmond, 2011；MacGinty, 2014）。

另外，AI 驅動的監視技術與深偽、不實訊息的擴散，可能破壞公共領域的基本共識，降低社會互信與協商能力，為政治極化與衝突動員創造有利條件（SIPRI, 2025; UNIDIR, 2023；沈伯洋、王希，2022）。就民主和平論而言，一旦透明性與課責機制的基礎遭到侵蝕，制度性和平所賴以維繫的根本將隨之動搖（Doyle, 1983）。就自由制度主義來說，當資訊環境遭到污染、可信訊號不足時，行為者對互惠合作的預期將更難形成與維繫（Keohane, 1984）。

概括而言，AI 對國際和平具有雙重性，它既可提升治理效能與人道救援精準度，也能加劇結構性暴力、削弱社會互信。新興的 AI for Peace 文獻雖提出預警、協作與資源配置等技術路徑，但常以準確率與效率作為正當性基礎，對演算法偏差、權力不對稱與在地主體性等政治和倫理問題的理論化與可操作治理設計仍有缺漏。以下將就前述問題詳細說明。

參、人工智慧對國際和平的機遇

在 IR 文獻中，AI 常被框限於軍事安全與威脅評估中，但如果轉向和平建構的治理，AI 的角色則顯著表現於強化資訊處理、風險辨識與跨機構協作的的能力。AI 對國際和平的運用與貢獻，在於作為和平進程的工具，強調預防、對話、監測與信任的功能，它無法直接消弭衝突根源，而是透過數據驅動機制，提升衝突升級的可預測性、介入的即時性，以及資源分配的精準度。本節以實證案例與相關文獻，從衝突預警、和平談判、維和行動、人道援助、社會信任建構等面向，分析 AI 的機遇，並考量其所需的治理條件，確保技術應用嵌入可課責的制度框架。此分析回應 IR 理論的建構主義與自由主義論點，亦凸顯 AI 作為治理工具的轉型潛力，強化「積極和平」的制度基礎。

一、衝突預警

傳統衝突預警系統的侷限在於依賴已發生狀況的數據或主觀判斷，難

以捕捉零散的訊號，如社群媒體情緒波動、地方暴力事件或經濟指標異動，AI 透過機器學習整合多元數據，包括衛星影像、地理空間資訊與社群媒體分析，顯著提升預測準確度與可操作性，將和平行動從被動回應轉向主動介入，此轉變不但縮短政策反應時間落差，還制度化風險管理流程，降低資訊不對稱與協調成本（蘇紫雲，2022）。例如聯合國政治及和平建構事務部（United Nations Department of Political and Peacebuilding Affairs，簡稱 UNDPPA）在葉門、利比亞與伊拉克部署 AI 輔助工具，運用 Remesh 平台蒐集公民意見，分析歷史模式與社群情緒，提供即時意見，協助地方和平建構者制定針對性策略，避免暴力擴散，經由此類技術應用支援「積極和平」的實現（UNDPPA, 2023）。

此外，文獻指出以 AI 為基礎的衝突警示系統，可辨識隱藏資訊如仇恨言論激增或經濟不穩，進而降低人道危機的成本（Armed Conflict Location and Event Data Project, 2022; Violence and Impacts Early-Warning System, 2023）。在非洲脆弱國家，此技術強化國家能力，透過大數據預測內戰觸發因素，支持非洲聯盟等區域組織的預防外交。

相較傳統方法，AI 的連續監測更具彈性，特別適用於跨國衝突情境，但其效用取決於預測輸出能轉譯為可執行的行動，並搭配地方知識與質性判讀，避免資料偏差導致誤判（Schrodt & Gerner, 2000）。因此，AI 在衝突預警的功能取決於規範化的治理流程，而非僅追求技術精確，除了提升預防外交的效能，也能回應 IR 理論預防治理的需求。

二、和平談判

在和平談判中，AI 藉由 NLP 技術與模擬分析扮演中介角色，有效促進包容性對話，並化解溝通障礙，此機制實踐建構主義提到「互信生成」的核心觀點，展現語言與意義在形塑國際和平的功能（Wendt, 1992; 張登及，2021）。AI 平台如 Remesh 或 Talk to the City 能擴大參與規模，匿名蒐集邊緣社群意見，進行情緒分析與意見分群，輔助談判者識別共識點。

聯合國的數位對話案例顯示，AI 工具能提升弱勢群體的能見度，避免菁英主導模式，轉而採用「人力-AI 混合架構」，落實因地制宜。此外，AI

基於歷史和平協議數據，模擬談判情境，預測成功模式如資源分配或停火條款，協助外交官調整策略。在非洲和平進程，AI 用於蒐集公民輸入，整理大型諮詢，強化談判合法性。Höne (2019) 指出，AI 分析溝通模式可化解意義衝突，深化語言理解，促成協議，此應用不僅加速談判，還透過即時翻譯跨越語言障礙，支持跨文化對話，如後衝突地區的族群調解。

儘管 AI 將談判轉化為更包容的參與過程，提升協議的持久性，但其貢獻限於輔助分析，而非取代政治判斷，調停核心仍是信任建構與利益交換 (Conciliation Resources, 2021)。吾人需建立人機協作規程，保留人類在價值判斷與政治責任的角色，並界定模型輸出的適用範圍，以強化談判的效能，促進持久和平的制度韌性 (蘇紫雲, 2022)。

三、維和行動與人道救援

維和行動中，AI 結合無人機、衛星影像與地理資訊系統，監測停火違規、裁軍進度及人權侵害，降低人員風險並提高效率，從靜態部署轉向動態監控。在人道援助方面，AI 優化資源分配，預測需求熱點，確保援助精準抵達。譬如聯合國開發計劃署 (United Nations Development Program, 以下簡稱 UNDP) 的 RAPIDA (AI-powered Early Recovery Assessment) 以高解析度衛星影像快速辨識受損情形，協助危機後 72 小時內的應對，紅十字國際委員會 (International Committee of the Red Cross, 以下簡稱 ICRC) 亦應用 AI 提升人道救援的效率和可信度 (Mol, 2002)。

再者，在不穩定區域，AI 能透過社群媒體的大數據分析，精準預測疾病爆發或市場動盪等衍生風險，從而優化資源分配機制，從根本上遏止因資源匱乏而加劇的社會衝突。相較於人力侷限，AI 全天候運作確保連續監測，提升維和與援助效能，並強化 IR 文獻有關人道治理的技術整合路徑，提高危機復原的能力 (Floridi, et al., 2018)。

四、社會信任建構

從社會層面來說，AI 可透過 NLP 工具匿名翻譯輸入，促成中立交流，

降低不同群體的對立。後衝突地區案例顯示，AI 監測仇恨言論與市場變化，早期預警系統幫助志工介入調解、重建互信。另外在外交領域，AI 分析媒體與網路訊息，預測危機並輔助公共外交，可以提升透明度。此外，AI 支援教育與文化交流，線上平台打破地理障礙、傳播和平敘事，呼應自由主義 IR 理論，說明技術進步深化制度間的依賴程度（Keohane, 1984; 張登及，2021）。

然而，AI 應用需注重在地知識整合，避免文化偏差。聯合國全球脈動（United Nations Global Pulse, 以下簡稱 UNGP）成立於 2009 年，經由推動大數據、AI 與數位創新，以應對全球永續發展、人道主義行動及和平挑戰，是為推動 AI 永續與和平的創新實驗室，這類平台能縮減跨組織的資訊落差，使政策建立在共同事實基礎上；但其成效取決於是否具備完善的權限治理機制與資訊安全架構。聯合國政治及和平建構事務部導入 AI、遙測與多語資料分析，以強化早期預警、支援預防外交，並為調停提供資訊支持；這些工作除了強化社會信任，也提升了制度韌性，回應 IR 理論關於技術如何重構全球治理的論述，具體展現 AI 作為建構國際和平的潛力（UNDP, 2023）。

總結來說，AI 在和平建構的潛力可概括為三項貢獻：提升風險可見性並縮短預防時間窗（如衝突預警）；提高危機行動的精準度與協作效率（如損害評估與資料整合）；強化調停與政策設計的分析支援（如文本彙整與情境推演），這些機遇的實現取決於治理條件的完備，包括可課責的決策流程、資料倫理與透明機制。聯合國裁軍研究所（United Nations Institute for Disarmament Research, UNIDIR）的報告將 AI 置於國際和平與安全風險情境，指出建立信任措施與透明討論的重要性，表示民用 AI 可能構成風險，需以「負責任創新」原則處理演算法偏見（Stockholm International Peace Research Institute, 2022）。

換言之，AI 賦能和平的關鍵不在技術本質，而是將其融入可預測、可協調、可檢驗的治理實作，這不止是辨識 AI 的具體作用機制，還界定其適用條件，為後續風險生成路徑與治理原則的討論奠定分析基礎，確保應用符合學術與倫理規範。透過此思考邏輯，AI 可轉化為國際和平的資產，強化全球治理效能（Bostrom, 2014; 戴豪君，2021）。

肆、人工智慧對國際和平的挑戰

AI 雖被視為和平建構的潛在助力，但其應用也有某些風險，從演算法偏差到資訊污染，這些風險放大既有不平等，也可能侵蝕和平進程的制度基礎（曾怡碩，2023；蘇紫雲，2022）。此分析回應 Galtung（1969）對結構性暴力的概念，將 AI 視為數位時代的權力再編碼機制，它也連結 Keohane（1984）對制度互依的看法，凸顯 AI 如何弱化多邊合作的訊號基礎。

一、演算法偏差與誤判

AI 模型仰賴歷史資料進行訓練，容易將既有偏差內嵌於模型之中，導致衝突預警與資源分配的誤判，此誤判可能擴大全球南北差距、加劇族群不平等，使得和平建構重製既有的全球權力結構。聯合國報告指出，以色列在衝突區使用「Gospel」與「Lavender」等 AI 系統，且在低度人力監督的情況下運作，致使平民傷亡與基礎設施毀損，凸顯資料不完整與偏誤直接影響決策結果（United Nations, 2023）。

其次，在非洲維和行動中，訓練資料主要源自西方社會，未能捕捉在地文化與社會互動模式，因此造成預警機制失靈。再者，在外交談判或跨語言協商平台，NLP 系統經常無法等量處理不同語言與語用差異，致使特定群體的敘事被降低權重、弱勢聲音遭邊緣化，背離包容性和平的規範要求。對此，Floridi 等人（2018）與李建良（2021）都強調，偏差並非單純的技術瑕疵，而是與權力結構的問題，加上黑箱效應削弱模型的可解釋性，讓外交決策難以對預測結果進行有效驗證，放大全球不平等並侵蝕行動者之間的信任。Tickner（2001）亦指出，演算法偏差可被理解為結構性暴力的數位再現，不平等透過分類與標籤的邏輯被制度化，並轉化為例行性的排除機制。

當治理看似客觀且中立，實則可能持續排除弱勢群體，其後果包括援助延宕、資源錯置與程序不公，甚至使原本旨在促進和平的工具反而成為衝突的催化劑，此一風險在資源分配過程尤其顯著，弱勢群體的需求被低估，長期下來將侵蝕和平制度的韌性與正當性。

二、黑箱化與課責困境

當 AI 作為決策支援工具卻欠缺足夠可解釋性時，責任歸屬將被稀釋，形成課責問題。在高風險的和平情境中，這種困境會放大誤判的後果，甚至推動局勢升級。就維和與人道援助而言，若 AI 模型的輸出不可追溯、推理路徑不可重建，即使出現錯誤，也難以將責任明確歸屬於特定決策者，最終常被合理化為系統錯誤，而非可被追究的治理失當（蘇紫雲，2022）。聯合國裁軍研究所將此置於人機互動的風險脈絡，主張透過信任建立措施提升透明度與可監督性，以降低誤用與濫用的可能（UNIDIR, 2021）。此一問題亦與 LAWS 的責任困境相互呼應，當人類控制被弱化，歸責機制與規範執行便更容易受阻（Amoroso & Tamburrini, 2019; 許智翔，2021；歐錫富，2022）。

在和平建構的條件下，AI 黑箱是技術挑戰，將侵蝕民主究責的制度基礎。當公眾無從理解決策依據，便難以有效監督、質疑或提出申訴，致使救濟與課責機制失靈，和平政策的正當性也難以累積。此一說法回應民主和平論的核心命題：透明與課責是制度性和平得以成立的重要條件，一旦這些條件被削弱，制度性和平的根基將因此動搖（李建良，2021）。

三、權力不對稱與資料殖民

AI 高度依賴資料、算力與平台生態，當資源分配長期失衡，標準與規則必然由大國或大型企業主導，弱勢國家淪為資料被抽取與再利用的對象，形成「資料殖民」（蔡政修，2024）。在和平治理背景下，這種結構性不對稱將「規制定義權」集中於少數行動者，削弱在地主體性與政策自主。由於許多發展中國家在資料主權與監管能力上相對薄弱，容易遭遇不對等的資料取得、技術鎖定與利益外溢不足，放大地緣政治分歧。因此，AI 治理必須正視能力差距與制度落差，避免和平建構在實務上被外包化為由外部技術供應方主導的流程（SIPRI, 2022; 戴豪君，2021）。

Richmond（2011）與 MacGinty（2014）闡明，和平的可持續性有賴在地自主與社會－政治能動性，若 AI 加深對外部平台、模型與專業的依賴，

將弱化政策主權與在地治理能力，和平方案更趨技術依附而非共同建構。另一方面，在 IR 的賽局結構中，AI 治理逐漸成為新的競爭場域，中美關係緊張提高規範分裂與陣營化風險，開發中國家對 AI 軍事化尤為警惕，所以更強調人文導向、發展優先與規範保障，反映多邊治理在規範整合與制度協商的困境。

值得一提的是，權力不對稱本身並非 AI 所獨創的現象，現實主義早已確認強國與弱國之間的結構性差距是國際體系的常態 (Morgenthau, 1948)。本文的論點並非主張 AI 憑空製造了不對稱，而是指出 AI 在三個面向使既有的不對稱產生質性的加深：其一，AI 賦予資料與算力優勢者以指數級的情報處理速度，使傳統意義上的知識落差在規模與即時性上出現非線性擴大；其二，AI 所形成的不對稱在相當程度上是結構性不可見的，弱勢行為者難以知悉自身被分析、被分類、被預測的程度，使其無從採取傳統的外交或戰略因應；其三，構成 AI 能力的核心要素具有強烈的規模報酬遞增特性，使先行者的技術優勢趨向自我強化，後行者的追趕成本持續攀升，進而賦予此種不對稱更高的結構固化傾向。

根據前述，AI 所反映的雖是真實世界的權力格局，但它也以新的機制深化並鞏固了既有的不對稱，而非僅作為靜態的映射工具

四、隱私侵害、監控濫用與資訊污染

AI 以大規模數據監測易引發隱私危機，可能被威權政府用來作為鎮壓工具，危害人權與和平正當性。紅十字國際委員會聲稱，非洲案例顯示在「安全」名義下的監控可轉為政治壓迫，在維和場域，AI 無人機與臉部辨識暴露平民資料，恐觸發報復循環，在人道援助，生物識別追蹤難民，提高外洩與追索風險 (ICRC, 2020)。聯合國安理會警告，缺乏監督的 AI 難以符合國際人道法，AI 監控缺乏全球規範，將助長權力集中並弱化民主 (United Nations Security Council, 2024)。

此外，生成式 AI 的不實訊息可操縱輿論、破壞談判與社會凝聚 (沈伯洋、王希, 2022)。在烏克蘭或中東案例，深偽引起辨識誤判、升級與擴散風險，引發外交危機，危及和平行動並提高人員風險 (United Nations, 2023)。

聯合國裁軍研究所關切資訊誤傳會放大風險，使和平建構陷入「誤判—升級—擴散」的路徑風險（UNIDIR, 2021）。Radsch（2022）提及，AI 驅動的不實訊息威脅民主基礎，社群平台放大迴聲室，阻礙跨群體對話，使安全困境再現。

五、雙重用途與治理缺口

AI 的雙重用途特性使和平技術易被轉化為軍事能力，在缺乏透明度的情況下，決策責任容易落入課責真空（蘇紫雲，2022）。聯合國安全理事會警告，AI 的發展速度正超越既有治理能力，尤以與軍事體系整合的風險最為敏感（United Nations Security Council, 2024）。在維和任務中，自主或半自主系統可能因辨識誤判、互動失靈而引發非預期升級，而相關規範仍缺乏先例與可操作標準。

當前全球治理呈現碎片化趨勢，且包容性框架不足，規則制定往往由大國主導，資料庫訓練資料、算力與技術門檻亦加劇強國和弱國間的不平等，所以聯合國呼籲國際社會對於 AI 的運用建立跨領域規範，涵蓋倫理原則、風險評估與驗證機制，不過能否落實終究取決於先進國家的政治意願與制度承諾（United Nations, 2023）。

伍、結論

本文依據 IR 與和平研究理論，針對 AI 在和平建構的機會與挑戰進行分析，核心問題包括「AI 如何參與和平建構過程？」；「其應用可能引發哪些風險？」以及「在確保技術效能與政治正當性的前提下，應當建立何種治理原則？」研究發現，AI 於國際和平呈現顯著的二元性，一方面，它既是維護安全的賦能工具，提升和平建構的預測性、精準度與包容性，另一方面，它亦是破壞穩定的潛在誘因，可能以隱蔽、累積的方式，削弱和平過程賴以維繫的信任和正當性。

本文發現：AI 能否成為建構國際和平的有效工具，取決於它是否被嵌入可課責、可協調、可檢驗的制度框架，而非取決於技術本身的精確度或效

率。這一論點貫穿本文對衝突預警、談判輔助、維和行動與社會信任建構的分析，也是對演算法偏差、黑箱課責、資料殖民與監控濫用等風險的回應。

從理論層面而言，本文的貢獻在於將 AI 議題嵌入和平研究的既有理論光譜。Galtung 的積極和平框架提示，技術應用若以中立之名複製結構性不平等，即構成數位時代的結構性暴力；Keohane 的自由制度主義有助於雙向評估 AI 對制度合作條件的強化或侵蝕；Richmond 與 MacGinty 的批判視角揭示，AI 治理的正當性危機。這些理論說明，AI 並非和平研究的外來議題，而是既有 IR 理論在數位時代的延伸場域。

在政策層面，本文提出三項可操作性的原則：其一，以可課責性作為 AI 和平應用的准入門檻，建立決策追溯機制並保留人類最終判斷權；其二，透過在地參與機制校正演算法的文化與地域偏差，防止技術中立論述掩蓋結構性排除；其三，透過多邊協調建立最低共同規範標準，遏制資料殖民與治理碎片化趨勢。

特別要說明的是，「在地知識」偏差校正並非 AI 本身或是任何 AI 設計者可解決的問題，許多脆弱國家與衝突後社會的在地知識高度依賴口述傳統、非正式網絡與社群記憶，既未被系統性記錄，也難以轉化為 AI 訓練所需的結構化資料。此意味著「在地知識」，本身即預設一定程度的制度能力與資料基礎建設，而這剛好是需要國際社會優先投資的前置條件，而非 AI 偏差校正的即時解方。

再者，部分在地知識包括族群分布、資源配置、社會衝突的歷史脈絡，對執政者而言具有高度政治敏感性，甚至涉及國家安全的核心利益，這類知識在性質上呈現雙重屬性，一來是治理主體的主權資產，另則是屬於人類共同知識遺產的組成部分，兩者之間的界線，無法由 AI 設計者或國際組織單方面劃定，必須通過主權國家與國際社會的協商機制加以確立。

面對上述困境，本文建議依知識的敏感程度與用途性質，建立分層知識釋出機制；對於無法直接釋出的敏感在地知識，可由具有公信力的在地研究機構、公民社會組織或獨立第三方擔任知識代理人，以匿名化、去識別化的方式提供資料，既避免直接暴露敏感資訊；將知識釋出的談判納入多邊 AI 治理議程，與資料跨境流動、演算法透明度等議題並列，形成完整的知

識治理框架。

總之，和平的實現從來不是純粹的技術問題，而是技術、制度與人文價值三者之間的動態平衡。本文期望為「AI 促進和平」的政策規劃提供另一個學術論證，並為後續結合氣候正義、去殖民化與永續治理等交叉議題的研究，開啟對話空間。

參考文獻

- 李建良，2021。〈人工智慧與法律：法學研究的挑戰與契機〉《臺大法學論叢》50期，頁 1165-1216。
- 沈伯洋、王 希，2022。〈數位威權主義下的資訊操弄：以台灣為例〉《問題與研究》61 卷 1 期，頁 1-40。
- 張登及，2021。〈大國競爭下的全球治理危機與轉型〉《政治科學論叢》88 期，頁 1-36。
- 許智翔，2021。〈人工智慧與未來戰爭：技術、戰術與倫理的交織〉《國防情勢分析》。台北：國防安全研究院。
- 曾怡碩，2023。〈人工智慧與網路安全：大國競爭的新場域〉《國防安全趨勢報告》。台北：國防安全研究院。
- 歐錫富，2022。〈AI 武器化：軍事人工智慧發展與挑戰〉《國防安全短評》。台北：國防安全研究院。
- 蔡政修，2024。〈美中人工智慧競爭及其對印太安全秩序之影響〉《國際關係學報》55 期，頁 35-72。
- 戴豪君，2021。〈人工智慧治理原則與國際發展趨勢：台灣的觀點與回應〉《資訊法務透析》33 卷 1 期，頁 34-45。
- 蘇紫雲，2022。〈數位時代的國家安全與科技：人工智慧的戰略意涵〉《戰略與評估》13 卷 2 期，頁 1-22。
- Amoroso, Daniele and Guglielmo Tamburrini. 2019. "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues." *Current Robotics Reports*, Vol. 1, No. 4, pp. 187-94.
- Armed Conflict Location and Event Data Project. 2022. "CAST: Conflict Alert System Toolkit." (<https://acleddata.com/cast/>) (2026/1/17)
- Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Carr, Edward. 1939. *The Twenty Years' Crisis, 1919-1939: An Introduction to the Study of International Relations*. London: Macmillan.
- Doyle, Michael. 1983. "Kant, Liberal Legacies, and Foreign Affairs." *Philosophy and Public Affairs*, Vol. 12, No. 3, pp. 205-35.
- Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi,

- Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. "AI4People-An Ethical Framework for A Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines*, Vol. 28, No. 4, pp. 689-707.
- Galtung, Johan. 1969. "Violence, Peace, and Peace Research." *Journal of Peace Research*, Vol. 6, No. 3, pp. 167-91.
- Giovanardi, Michele. 2024. "AI for Peace: Mitigating the Risks and Enhancing Opportunities." *Data and Policy*, No. 6, pp. 1-15.
- Höne, Katharina. 2019. "Artificial Intelligence and International Conflict Resolution: A Constructivist Perspective." *International Studies Review*, Vol. 21, No. 3, pp. 456-78.
- International Committee of the Red Cross (ICRC). 2020. "Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centered Approach." (<https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach>) (2026/1/20)
- Kant, Immanuel. 1991. "Perpetual Peace: A Philosophical Sketch," trans. by M. Campbell Smith, in Hans Reiss, ed. *Kant: Political Writings*, 2nd ed., pp. 93-130. Cambridge: Cambridge University Press.
- Keohane, Robert. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.
- MacGinty, Roger. 2014. "Everyday Peace: Bottom-Up and Local Agency in Conflict-Affected Societies." *Security Dialogue*, Vol. 45, No. 6, pp. 548-64.
- Mol, Arthur. 2002. "Ecological Modernization and the Global Economy." *Global Environmental Politics*, Vol. 2, No. 2, pp. 92-115.
- Morgenthau, Hans. 1948. *Politics among Nations: The Struggle for Power and Peace*. New York: Alfred A. Knopf.
- Radsch, Courtney. 2022. *AI and the Future of News: The Promise and Perils of Artificial Intelligence in Journalism and Media*. Oxford: Oxford University Press.
- Richmond, Oliver. 2011. *A Post-Liberal Peace*. London: Routledge.
- Schrodt, Philip and Deborah Gerner. 2000. "Cluster-Based Early Warning Indicators for Political Violence in the Contemporary Levant." *American Political Science Review*, Vol. 94, No. 4, pp. 803-18.
- Stockholm International Peace Research Institute (SIPRI). 2025. "Addressing the Risks that Civilian AI Poses to International Peace and Security: The Role of

- Responsible Innovation.” (https://www.sipri.org/sites/default/files/2025-11/1125_civilian_ai.pdf) (2026/1/23)
- Stockholm International Peace Research Institute (SIPRI). 2022. “Artificial Intelligence: A Strategic Opportunity for Global Arms Control?” (<https://www.sipri.org/publications/2022/other-publications/artificial-intelligence-strategic-opportunity-global-arms-control>) (2026/1/22)
- Tickner, Ann. 2001. *Gendering World Politics: Issues and Approaches in the Post-Cold War Era*. New York: Columbia University Press.
- United Nations (UN). 2023. “AI-Driven Misinformation and Its Impact on Peace and Humanitarian Operations.” (<https://www.un.org/en/ai-misinformation-peace>) (2026/1/20)
- United Nations Department of Political and Peacebuilding Affairs (UNDPPA). 2023. “Innovation in Peacebuilding: AI Tools and Applications.” (<https://dppa.un.org/en/innovation>) (2026/1/24)
- United Nations Institute for Disarmament Research (UNIDIR). 2023. “AI and International Security: Understanding the Risks and Paving the Path for Confidence-building Measures.” (<https://unidir.org/publication/ai-and-international-security-understanding-the-risks-and-paving-the-path-for-confidence-building-measures>) (2026/1/19)
- United Nations Security Council. 2024. “Debate on AI and International Security. *United Nations*.” (<https://www.un.org/securitycouncil/debate-ai-2024>) (2026/1/23)
- Violence and Impacts Early-Warning System. 2023. “VIEWS: Predicting Political Violence.” (<https://views.pcr.uu.se/>) (2026/1/23)
- Wendt, Alexander. 1992. “Anarchy Is What States Make of It: The Social Construction of Power Politics.” *International Organization*, Vol. 46, No. 2, pp. 391-425.
- Zelizer, Craig, Joshua Miller, and Stephen Thompson. 2025. “Leveraging Artificial Intelligence for Conflict Early Warning: A New Era of Predictive Peacebuilding.” *Journal of Peace and Conflict Studies*, Vol. 32, No. 1, pp. 45-68.

Opportunities and Challenges of Artificial Intelligence for International Peace

William Yang*

Professor, Department of Public Administration and Management, Shih Hsin University, Taipei, TAIWAN

Abstract

This study explores the impact of Artificial Intelligence (AI) on international peacebuilding, analyzing its potential as a technological tool alongside its associated risks. As peace studies evolve from traditional balance-of-power theories to diverse perspectives encompassing “positive peace” and “hybrid peace,” the intervention of AI technology is transforming how the international community addresses conflict. The research indicates that AI, through big data analysis and pattern recognition, can shift conflict early warning from passive response to proactive prevention. Using Natural Language Processing (NLP) tools, AI assists in peace negotiations and expands the participation of marginalized communities, fostering inclusive dialogue. Furthermore, the integration of satellite imagery and Geographic Information Systems (GIS) enhances the precision and effectiveness of monitoring peacekeeping operations and allocating humanitarian aid resources. On the other hand, algorithmic bias may reproduce structural violence, leading to the systemic exclusion of vulnerable groups. The “black-boxing” of decision-making processes creates accountability dilemmas, weakening institutional legitimacy. The concentration of data and computing power triggers data colonialism and power asymmetries, undermining local agency. Moreover, information pollution and the abuse of surveillance erode social mutual trust. This paper finds that the key to AI-empowered peace lies not in the technology itself, but in how it is embedded

* Ph.D. in Political Science and International Relations, University of Warwick. Previously Dean of International Affairs, Dean of Public Affairs, and Chair of the General Education Committee. Email: william@mail.shu.edu.tw.

within an accountable, coordinated, and verifiable institutional framework. Therefore, this study suggests establishing an accountability mechanism centered on human-AI collaboration, correcting the cultural and local biases of algorithms, and mitigating global governance fragmentation through multilateral coordination, making AI an effective tool for building international peace.

Keywords: artificial intelligence (AI), Peace Studies, peacebuilding, algorithmic governance, international security