AI 自主性武器系統之國際規範秩序：省思與前瞻 *

林昕璇

（成功大學政治學系助理教授）

摘　要

巨量數據所驅動之預測型演算法勃興，AI 自主性武器系統相關爭議浮上檯面，成爲國際法研究之新課題。爲填補國際法研究針對 AI 新武器崛起涉及之國際規範秩序形塑過程中的空白，本文旨在以國際人道法及特定常規武器公約作爲論述憑藉，檢視此新型武器系統適用於現行國際法秩序之應然與實然。本文以文獻研究爲本，以「自主性程度之高低」、「智能決策迴路」及「特徵式列舉」作爲類型化之區別要素，比較美國、中國、德國、紅十字國際委員會及人權觀察組織對於自主性武器系統的異同之辨，接續考察武器控管之國際法主要淵源《特定常規武器公約》之談判歷程及重要里程碑。本文發現，《特定常規武器公約》締約方所僵持不下的若干倡議，已然無法周延因應自主性武器系統的潛在風險，故本文試圖以貫徹有意義之人爲控制和課責機制爲探詢思維，提出若干監管軍事演算法之規範性建議。

關鍵詞：自主性武器系統、人工智慧、特定常規武器公約、日內瓦公約第一附加議定書、武器演算法之審查

* 本文原始初稿曾發表於 2020 年 6 月 3-4 日由中央研究院歐美研究所主辦之「AI 與民主學術研討會」，誠摯感謝諸位審查人與季刊編輯委員會之寶貴賜教與悉心專業地往復修改建議，實令本文獲益匪淺且臻至完善。惟本文文責概由作者自負。

壹、前言

一、研究緣起

仰賴演算法之自主性武器系統於國際法規範體系之適用爭議，儼然爲當前科技治理之一大問題。鑑於資料驅動之人工智慧（Artificial Intelligence，以下簡稱 AI）系統具有「自我學習」各項資料特徵及模式之能力，得協助公私部門進行風險預測；[1] 相關公私部門仰賴 AI 做成自動化決策時，亦獲益於該系統得於不同選項間，就其潛在風險和發展路徑發揮精準預測之優勢，進而協助人類擬定最具優勢的競爭策略。[2]

根據學者研究，武裝衝突情境中，預測型演算法 (Predictive Algorithm) 爲主責國家安全之行政機關——如國防部、國安局——所廣泛採用，眾多行政機關莫不深刻意識到巨量資料和演算法之重要性。以美國國安部門爲例，AI 對於軍事武器系統發展之影響力主要體現於「軍事拘押」(military detention)、「涉外數位監控」(foreign intelligence surveillance) 及「自主性武器」(autonomous weapon) 等三大面向之實踐。該作爲不僅牽動國際安全之敏感神經，更引起學界廣泛重視，並以國際人道法爲基礎回應 AI 相關科技所衍生之疑慮。[3]

爲因應此牽涉廣泛且於國際社會上引發激烈討論之新興武器對

---

1. Andrew Guthrie Ferguson, “Big Data and Predictive Reasonable Suspicion,” *University of Pennsylvania Law Review*, Vol. 163, No. 2, January 2015, pp. 353-376.
2. Emily Berman, “A Government of Laws and Not of Machine,” *Boston University Law Review*, Vol. 98, January 2018, pp. 1284-1290; David Lehr & Paul Ohm, “Playing with the Data: What Legal Scholars Should Learn About Machine Learning,” *UC Davis Law Review*, Vol. 51, December 2017, pp. 670-672.
3. Ashley Deeks, “Predicting Enemies,” *Virginia Law Review*, Vol. 104, No. 8, March 2018, pp. 1547-1562.

傳統國際法規範秩序所衍生之衝擊，美國國會研究處 (Congressional Research Service) 於 2020 年發布名爲《人工智慧與國家安全》(*Artificial Intelligence and National Security*) 之研究報告，將「情報、監控與偵查」(intelligence, surveillance and reconnaissance, ISR) 及「致命自主性武器系統」(Lethal Autonomous Weapon Systems, LAWS) 並列爲美國國防部發展 AI 之主要應用領域。[4] 復根據國際安全戰略的研究成果，AI 在軍事領域得以發揮包括：（一）促進實時分析和改進戰場態勢感知；（二）爲地面部隊提供可行的情報和增強決策能力；（三）促進力量的分解或快速集中和應用致命力量，從而增強任務的準確性；（四）作爲後勤助手，提供預測性維護和軍事裝備，增加操作設備的安全性，降低作戰成本，從而提高部隊的作戰能力和部隊能力等戰略布局。[5]

然而，在武裝衝突中使用自主性武器實則非爲晚近的發展趨勢，自動化武器系統之先驅最早可溯至美國於二次世界大戰中使用無人戰鬥飛行載具 (Unmanned Combat Aerial Vehicle) 供軍事偵察用途。2009 年上任後之歐巴馬政權更進而擴增其使用頻率，基於打擊恐怖主義於該年 8 月 8 日單日內，針對蓋達組織 (Al-Qaeda) 嫌疑犯展開軍事行動，狙殺了 12 名嫌疑犯，從而面臨來自國際社會和人權組織的監督和關注。[6] 相較於小布希政權任內僅 57 次的紀錄，歐巴馬

4. Kelley M. Sayler, *Artificial Intelligence and National Security* (Washington, D.C.: Congressional Research Service, 2020), pp. 1-43.
5. J. Burton & S. R. Soare, "Understanding the Strategic Implications of the Weaponization of Artificial Intelligence," paper presented at 11th International Conference on Cyber Conflict (CyCon) (Tallinn: NATO Cooperative Cyber Defence Centre of Excellence, May 28-31, 2019), p. 12.
6. Elias Groll, "The Sudden and Unexpected Return of the Drone War," August 8, 2013, *Foreign Policy*, <https://foreignpolicy.com/2013/08/08/the-sudden-and-unexpected-return-of-the-drone-war/>.

政府的八年任期內，美國政府採取了 563 次無人機定點行動，旨在打擊位於葉門、索馬利亞、巴基斯坦等地的蓋達組織，估計造成約 384 名至 807 名無辜平民傷亡之人道危機。[7] 此外，由美國主導的反恐戰爭，亦於阿富汗及巴基斯坦等軍事干預上，使用將攻擊恐怖分子及平民視爲標的的遙控無人飛行載具。此等戰鬥方式因涉及對生命權之嚴重侵犯，遂引發學理競相討論其適法性，而其性能隨著可持續性監控 (persistent surveillance) 擴張至遠端精準制導武器 (precision-guided weapon) 之複合式運用，更觸發學界與政府專家對此種新武器型態是否牴觸國際法對於戰爭行爲規範之激烈論辯。[8]

二、本文之範疇界定

武器系統能夠於電腦預設程式下產生行動方案，並因應戰場情況盱衡情勢，據以發動帶有敵意之軍事行動。[9] 斯德哥爾摩國際和平研究所 (Stockholm International Peace Research Institute, SIPRI) 發布之《建構武器系統自主性的發展》(*Mapping the Development of Autonomy in Weapon Systems*) 指出，當今致命自主性武器已然具備完

7. Jessica Purkiss & Jack Serle, "Obama's Covert Drone War in Numbers: Ten Times More Strikes Than Bush," January 17, 2017, *The Bureau of Investigative Journalism*, <https://www.thebureauinvestigates.com/stories/2017-01-17/obamas-covert-drone-war-in-numbers-ten-times-more-strikes-than-bush>.
8. Michael Carl Haas & Sophie-Charlotte Fischer, "The Evolution of Targeted Killing Practices: Autonomous Weapons, Future Conflict, and the International Order," *Contemporary Security Policy*, Vol. 38, No. 2, August 2017, pp. 281-306; Michael Mayer, "The New Killer Drones: Understanding the Strategic Implications of Next-generation Unmanned Combat Aerial Vehicles," *International Affairs*, Vol. 91, No. 4, July 2015, pp. 765-780.
9. Jürgen Altmann & Frank Sauer, "Autonomous Weapon Systems and Strategic Stability," *Survival*, Vol. 59, No. 5, September 2017, pp. 123-124.

整且不受人類參與做成決策之自主目標辨識能力。[10] 因此，自主性武器其實際運用與規範評價，結合 AI 執行任務之精準性，不僅涉及作戰方法手段之改變，亦牽動人類對武裝衝突之認知變革；其與現行國際人道法規範與監管機制之接軌建構，遂成爲亟待探詢之課題。

軍事武器系統乃係國家權力競逐和具體操作之實踐場域，與國際安全、衝突預防及維和、衝突後之穩定維護、衍生之人道危機間具高度關聯性，國際政治及國際法律規範亦須隨之做出相應調整。AI 自主性武器之持續發展，牽動科技發展、國家安全與國際人道法三者間之動態平衡。在武裝衝突中，國家法益和個人法益恆常處於須取捨其一之扞格情境，亦即讓渡國際人道法對生命權之高密度保障，以換取國家安全及軍武科技所擔負之「開疆闢土」前導性任務，似屬不可迴避的對抗分立。倘若兩者有所衝突，攸關確保國家安全領域之 AI 研發確有其必要性、正當性及合法性，且取決於貫徹有效監督控制之制度設計。學理上饒富爭議的問題在於：AI 自主性武器系統於運作過程所衍生之系統性監管難題爲何？現行國際法於何種程度內適用於具備 AI 技術優勢之自主性武器系統？AI 自主性武器系統應體現何者規範調適要素？

本文行文安排如次：於文獻回顧後，以國際人道法與《禁止或限制使用某些可被認爲具有過分殺傷力或濫殺濫傷作用的常規武器公約》（*Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapon Which May be Deemed to be Excessively Injurious or to have Indiscriminate Effects*，以下視行文需要穿插使用特定常規武器公約或 UNCCW）之規範文本爲研究途徑，並以「自主性程度之高低」、「智能決策迴路」及「特徵式列舉」作爲類型化要素，分析美國、中國、德國、紅十字國際委員會 (International Committee

10. V. Boulanin & M. Verbruggen, *Mapping the Development of Autonomy in Weapon Systems* (Solna: SIPRI, 2017), p. 26.

of the Red Cross) 及人權觀察組織 (Human Rights Watch) 對於自主性武器系統在定義與概念內涵上的異同之辨。繼以依循《特定常規武器公約》之發展演替及其對自主性武器之規範觀點爲主軸加以著墨；接續檢證當今武裝衝突情境中，日益仰賴 AI 之自主性武器系統如何參酌既存之國際人道法秩序遵循調適。最後以貫徹有意義之人爲控制 (meaningful human control)、人機協作之軍事決策及課責機制作爲探詢思維，試圖提出若干監管軍事演算法之規範性反思與建議。

三、文獻回顧與探討

冷戰後國際政治體系之結構變遷及軍武科技蓬勃發展，諸多國家莫不重視推動軍武科技及軍備競賽；惟對於國際條約和國際習慣法之創設和遵循，仍眾說紛紜、莫衷一是。職此，2016 年特定常規武器公約之締約方組建籌設政府專家小組（Group of Governmental Experts，以下簡稱 GGE）以應對此種挑戰。[11] 然而，各國及非政府組織對於現有之國際人道法框架是否可以解決自主性武器系統所帶來之法律問題認定一節，仍難以取得共識。俄羅斯認爲現有法律規定適用於任何一種型態之武器，英國則認爲國際人道法的監管強度已足，惟若干國家支持以不具拘束力之方式訂定相關規定。綜言之，大多數國家與紅十字國際委員會依然認爲必須爲自主性武器系統建立一套周延且具拘束力之國際法規範。[12]

---

11. Thompson Chengeta, "Is the Convention on Conventional Weapons the Appropriate Framework to Produce a New Law on Autonomous Weapon Systems?" in Frans Viljoen, Charles Fombad, Dire Tladi, Ann Skelton, & Magnus Killander, eds., *A Life Interrupted: Essays in Honour of the Lives and Legacies of Christof Heyns* (South Africa: Pretoria University Law Press, 2022), pp. 379-397.
12. Thompson Chengeta, "Is the Convention on Conventional Weapons the Appropriate Framework to Produce a New Law on Autonomous Weapon Systems?" pp. 379-397.

學者克羅多夫 (Rebecca Crootof) 研究指出，以俄羅斯和美國爲首之傳統強權，傾向將國際人道法視爲一發展成熟且周延全面的規範框架，故採取所謂「適用與遵循」(apply-and-comply) 或「觀望」(wait-and-see) 模式立場，有條件承認國際人道法應與自主性武器系統之發展接軌。[13] 中國則立於上述主張之對立面，認爲該事務領域欠缺與既存國際組織間之制度性連結，且制定相關法規仍難以根本性解決問題。[14] 另巴西與墨西哥等其他國家亦各有立場，希冀以談判促成具法律約束力之規範文本，樹立若干指導方針或制度性指引。[15]

依據筆者之歸納，迄今諸多研究者對於自主性武器系統之詮釋觀點，大抵略分爲「從嚴否定說」、「風險評估說」及「加強課責說」等三種類型。「從嚴否定說」係由道德倫理及人性尊嚴之角度著眼，對無人機定點清除及自主性武器系統抱持從嚴解釋之態度，認爲軍武科技將傳統武器所難以企及之關鍵目標打擊能力予以擴張，毋寧爲一種透過「減輕人類士兵心理負擔」(relieve[s] human soldiers of the

13. Rebecca Crootof, "The Killer Robots Are Here: Legal and Policy Implications," *Cardozo Law Review*, Vol. 36, August 2015, pp. 1837-1915.

14. United Nations, "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects," April 11, 2018, *United Nations*, <https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/CCW_GGE.1_2018_WP.7.pdf>.

15. 奧地利、巴西及墨西哥等國主張透過談判促成一份具有法律約束力之文書；此外，支持或至少「願意進一步討論」具政治約束力文書者，則包含澳洲、比利時、德國、芬蘭、法國、愛爾蘭、義大利、挪威、波蘭、西班牙、瑞典及瑞士等國。請見 Ray Acheson, "New Law Needed Now," *CCW Report*, Vol. 6, No. 9, August 30, 2018, pp. 1-7。

psychological burden)途徑剝奪人類生命之軍事系統，[16]應遵循聯合國 GGE 所述「由武器系統之全生命週期以觀，有鑑於責任無從轉移至機器，使用自主性武器系統之際，其責任應從嚴歸屬於人類」[17]作爲自主性武器規範指導原則之一。

另一方面，「風險評估說」一派學者蓋斯特 (Edward Geist) 認爲學理爭辯之焦點，應由積極面或消極面評估致命自主性武器系統之容許性，轉而正視軍武科技之擴增已然不可逆，故應就現實主義觀點，實質評價其被廣泛應用後所致潛在風險之影響。[18]另學者修斯 (Marcus Schulzke) 則認爲，技術擴散之不可控性來自於 AI 技術應用之軍、民兩用性質，此意謂無論是軍用或民用之外溢效應，皆會強化擴散的潛在風險。[19]

相較於此，「加強課責說」則著意建構監管機制與課責機制，力主對自主性武器所致問責差距 (accountability gap) 或問責眞空 (accountability vacuum) 等癥結——亦即對自主性武器系統之決策、行動和效果等——強化制度性監管力道。[20]此論點亦與紅十字國際委

16. Srđan T. Korać, "Depersonalisation of Killing: Towards a 21st Century Use of Force 'Beyond Good And Evil'?" *Philosophy and Society*, Vol. 29, No. 1, January 2018, p. 62; Aaron M. Johnson & Sidney Axinn, "The Morality of Autonomous Robots," *Journal of Military Ethics*, Vol. 12, No. 2, August 2013, pp. 129-141.
17. United Nations Group of Governmental Experts, *Emerging Commonalities, Conclusions and Recommendations* (Geneva: United Nations, 2018), pp. 1-5.
18. Edward Moore Geist, "It's Already Too Late to Stop the AI Arms Race—We Must Manage It Instead," *Bulletin of The Atomic Scientists*, Vol. 72, No. 5, August 2016, pp. 318-321.
19. Marcus Schulzke, "Autonomous Weapons and Distributed Responsibility," *Philosophy & Technology*, Vol. 26, No. 2, June 2013, pp. 203-219.
20. Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology*,

員會之一貫立場互有呼應，其基本邏輯認爲，針對自主性武器系統使用武力之決策流程應加諸道德責任箝制，確保有意義、有效且適當之人爲控制應屬制度設計上不可或缺之樞紐。[21] 學者艾肯霍夫 (Merel Ekelhof) 指出，當今圍繞自主性武器系統的辯論，不乏強調將「有意義之人爲控制」概念作爲一項規範性要求，然實際上尚缺乏對該概念之實踐意涵與通用定義。[22] 承襲前述邏輯，近年學界開始自不同面向廓清「有意義之人爲控制」之內涵要素。舉例而言，學者霍洛維茲 (Michael C. Horowitz) 和夏爾 (Paul Scharre) 即闡述對所謂「有意義之人爲控制」所應具備之要素，渠等認爲有效能之監管至少應涵蓋問責制、道德責任及可控性等三項重要內涵；[23] 而洛夫 (Heather M.

Vol. 6, No. 3, September 2004, pp. 175-183; Rebecca Crootof, "War Torts: Accountability for Autonomous Weapons," *University of Pennsylvania Law Review*, Vol. 164, No. 6, May 2016, p. 1347; Robert Sparrow, "Robots and Respect: Assessing the Case Against Autonomous Weapon Systems," *Ethics & International Affairs*, Vol. 30, No. 1, March 2016, pp. 93-116; Esther Chavannes, Klaudia Klonowska, & Tim Sweijs, "Governing Autonomous Weapon Systems," February 3, 2020, *The Hague Centre for Strategic Studies*, <https://hcss.nl/wp-content/uploads/2021/01/HCSS-Governing-AWS-final.pdf>.

21. "New SIPRI and ICRC Report Identifies Necessary Controls on Autonomous Weapons," June 2, 2020, *SIPRI*, <https://www.sipri.org/media/press-release/2020/new-sipri-and-icrc-report-identifies-necessary-controls-autonomous-weapons>.

22. Merel Ekelhof, "Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation," *Global Policy*, Vol. 10, No. 3, March 2019, pp. 343-348.

23. Michael C. Horowitz & Paul Scharre, "Meaningful Human Control in Weapon Systems: A Primer," March 13, 2015, *Center for a New America Security*, <https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf>.

Roff) 和莫耶斯 (Richard Moyes) 則自三個時間序列——在敵對行動開始之前（戰前）、在敵對行動期間（交戰中）及在敵對行動之後（戰後）——切入，檢證何謂「有意義之人爲控制」。[24] 在國際人道法適用於自主性武器系統的前提下，學說不乏認爲向來被忽略之必要性原則 (necessity) 及最小武力原則 (the requirement of minimal force) 當屬自主性武器規範指導原則之一環。[25]

國內學者亦紛紛嘗試提出各項解釋，然迄今未獲共識。如由《特定常規武器公約》對於致命自主性武器發展之觀點，探討 AI 運用於武裝衝突場合之際所致之規範衝突，主張進一步增訂《特定常規武器公約》第六議定書之途徑，將軍武科技管理予以規範化；[26] 另有論者以國家責任角度，主張各國對於 AI 自主性武器之研發與擴散確屬不可逆，故承襲國際軍事法學者克羅多夫立論，指出應提高國家責任之歸責門檻爲限制無過失責任 (limited strict liability)，希冀緩解軍武科技外溢後所致之監管難題。[27]

---

24. Heather M. Roff & Richard Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons," *Article 36*, April 11-15, 2016, <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>.

25. Alexander Blanchard & Mariarosaria Taddeo, "Jus in Bello Necessity, The Requirement of Minimal Force, and Autonomous Weapons Systems," *Journal of Military Ethics*, Vol. 21, No. 3-4, May 2022, pp. 295-298。國內論及軍事必要性構成適用國際人道法指導原則中不可或缺之變項的論著，請見郭雪眞，〈人道軍事干預與國際人道法：美國反恐戰正中關達那摩灣 (Guantanamo Bay) 被拘禁者釋憲案例分析〉，《復興崗學報》，第 101 期，2011 年 12 月，頁 17-20、34；趙國材，〈論國際人道法適用於內戰之發展〉，《軍法專刊》，第 56 卷第 4 期，2010 年 8 月，頁 122-153。

26. 林韋仲、廖宗聖，〈致命自主武器發展之國際法管制〉，《台灣國際法學刊》，第 15 卷第 2 期，2019 年 6 月，頁 9。

27. 林昕璇，〈AI 自主性武器系統於國際法適用上之研析〉，《軍法專

貳、自主性武器系統之體系發展與現況

一、發展體系與衍生爭議

（一）「自主性」意涵界定之浮動性

1. 美國國防部《3000.09 指令》

自主性武器系統之定義及監管所致之爭議，可由各國與國際組織對該系統具體意涵與射程範圍之見解不一，略見端倪。然大體而言，於此些分歧觀點中，仍可發現自主性武器系統之「類型化」、「定義內涵」與「自主性程度之高低」三者確爲息息相關。

美國國防部於 2012 年所頒布《3000.09 指令：自主性武器系統》(*DoD Directive 3000.09 Autonomy in Weapon Systems*) 乃官方發展自主性武器系統之規範依據，旨在界定何謂美國官方定義之半自主及自主性武器系統，同時賦予該系統選擇攻擊目標時於命令位階上之正當化根據。[28] 申言之，該指令係以階層化方式劃設出三種不同層級之自主性武器之概念操作型定義，分別爲：「半自主性武器系統」(semi-autonomous weapon system)、「人類監控自主性武器系統」(human-supervised autonomous weapon system) 及「完全自主性武器系統」(fully autonomous weapon system)。[29] 具體而言，第一種「半自主性武

刊》，第 67 卷第 4 期，2021 年 8 月，頁 21。

28. Justin Haner & Denise Garcia, "The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development," *Global Policy*, Vol. 10, No. 3, September 2019, pp. 331-337。國內介紹美國國防部《3000.09 指令》之文獻，請見林韋仲、廖宗聖，〈致命自主武器發展之國際法管制〉，頁 12-14；林昕璇，〈AI 自主性武器系統於國際法適用上之研析〉，頁 25-27；陳建佑，〈人工智慧監管法律—獨漏自主性武器之規範？〉，《全國律師》，第 27 卷第 6 期，2023 年 6 月，頁 43-44。

29. U.S. Department of Defense, *Department of Defense Directive 3000.09: Autonomy in Weapon Systems (2012)*, November 21, 2012, pp. 1-15.

器系統」指涉武器系統一旦啓動，僅得針對人類操作員所選擇之特定軍事標的執行任務；其次，所稱「人類監控自主性武器系統」係指系統仍受人類監控，人類操作員得介入或干預武器系統之運作，並於系統錯誤時終止操作；復所謂「完全自主性武器系統」則界定爲人類完全脫離智能決策系統，意即可在無任何人類操作員干預情況下選擇並攻擊目標之應用型態。人權觀察組織及美國國防部類型化自主性武器之對應關係，請見表 1 所示。

問題在於，美國國防部《3000.09 指令》僅就自主性武器系統賦予概念上之區分，並未詳盡列舉相應武器型態與種類。揆諸實際，武器系統其功能及是否存在人爲介入，兩者間往往互有疊合，甚或難以截然劃分。具體而言，以無人機掛載反裝甲飛彈用於實施作戰爲例，反裝甲飛彈固然爲合法武器，且無人機用於偵蒐亦屬合法，惟二者一旦結合，所形成之新型複合式武器系統究應如何評價仍不無疑義，亦凸顯美國國防部《3000.09 指令》之規範盲點。

儘管如此，以「自主性」之強弱消長作爲界定標準仍廣泛爲各國與國際組織所採用。以紅十字國際委員會爲例，其即以人道考量爲中心，廣納法律及科學專家之意見。2016 年由紅十字國際委員會所出版，且廣爲其後學界所引用之《自主性武器系統：增強武器關鍵功能之自主性》專家報告指出，人類自主性與機器人自主性互屬迥不相侔的兩個概念，兩者之核心差異在於前述「道德能動性」(moral agency)之哲學命題。[30] 人類因擁有自由意志，故而能於道德與法律層

30. International Committee of the Red Cross, *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons* (Geneva: International Committee of the Red Cross, 2016), pp. 57-59. (statements of representatives of the Ministry of Defence of the UK arguing that LAWS means different things to different people, and that different militaries employ unique approaches to assess and regulate new weapons systems).

面構成一個主體；然機器人因無自由意志，其運作及行動所依據者僅爲數學運算結果，故無法構成一個獨立的道德主體。儘管各類演算法一日千里，致令機器學習結果在表面上趨近於人類學習結果，惟兩者於本質上仍是大相逕庭。[31]

紅十字國際委員會於報告中闡釋，若自主性武器系統將摧毀或傷及人類生命實乃不可迴避之事實，則應確保人類自主性武器具備適當且實質、有意義地控制其最終決策，因而呼籲各國於研發、製造或使用智能軍事武器之際，應確保對自主性武器系統保留適當之人爲控制。基於國際人道法及國際人權法之核心法益保護目的，必須確定相關自主性武器系統遵循各該規範，武器系統尤須就國家或個人之法律問責機制，令違反國際法而研發、設計、部署、使用或操作任何自主性武器系統者承擔相關法律責任。[32]

表 1　以自主性程度區分的自主性武器系統架構

自主性程度	人權觀察組織	美國國防部
弱 ↕ 強	有大量人爲因素介入 (human-in-the-loop)	半自主 (semi-autonomous)
	人類可監控並撤銷決定 (human-on-the-loop)	人類監控自主 (human-supervised autonomous)
	未有人爲因素介入 (human-out-of-the-loop)	完全自主 (fully autonomous)

資料來源：Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," p. 480。

31. Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," *Houston Journal of International Law*, Vol. 40, No. 2, February 2018, pp. 461-530.

32. International Committee of the Red Cross, *Autonomous Weapon Systems: Implications Of Increasing Autonomy in the Critical Functions of Weapons*, p. 7.

2. 學理上類型化之嘗試

爲因應自主性武器系統其定義內涵過於抽象化、特徵化、概括化致失焦之虞，另有學者依自主性程度之高低層級，將各該武器系統予以階層化，請見表 2。

表 2 依據自主性程度自低至高之武器系統階層化

1	手槍	單手即可發射之短管槍支
2	對人用地雷	埋設於地面上或地面下之爆炸性地雷，由重量或壓力觸發
3	機關槍	僅須按動扳機即能快速連續發射子彈之自動槍械
4	步哨槍 （例：方陣近迫武器系統 (Phalanx CIWS)）	可自動瞄準並向被感應器所探測目標發射之武器
5	武裝無人機 （例：MQ-1 捕食者、MQ-9 收割者）	執行情報、監事和偵察任務之遠程駕駛航空器，裝備有地獄火飛彈
6	反導彈防禦系統 （例：C-RAM）	具高度與弧形彈道之導彈，最初以動力和導向，在重力作用下落擲於目標
7	滯空型彈械 （例：以色列的哈比 (Harpy)）	一種武器系統類別，彈械於目標區周圍徘徊一段時間並搜索目標，一旦發現目標即發動攻擊
8	終端高空防禦飛彈 （例：THAAD）	攔截大氣層內外短至中程彈道導彈防禦系統

資料來源：Raine Sagramsingh, “Lethal Autonomous Weapons Systems: Artificial Intelligence and Autonomy,” April 2019, *WISE*, <https://wise-intern.org/wp-content/uploads/2019/04/Raine_S_-FinalPaper.pdf>。

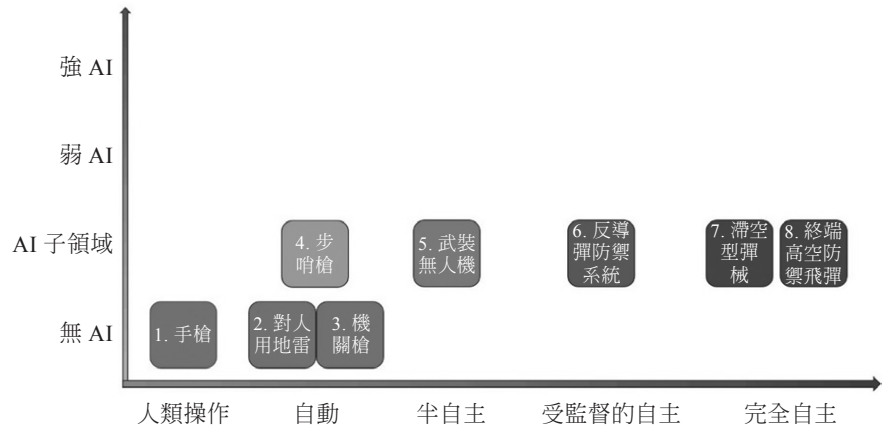

圖 1　自主性武器系統與 AI 類型對應座標軸

資料來源：Raine Sagramsingh, "Lethal Autonomous Weapons Systems: Artificial Intelligence and Autonomy"。

從圖 1 可知，對比於當前「弱 AI」(Weak Artificial Intelligence; Artificial Narrow Intelligence) 或「AI 子領域」占據大宗之自主性武器系統現狀，抑或所謂「泛」AI 武器系統於武裝衝突實務應用上仍盤據主流地位，擁有完整人類心智的「強 AI」[33] 似乎尚難以企及，亦呼應學說觀察，完全自主性武器系統迄今尚未能運用於戰場上。[34]

33 主要以具備自主意識和意向性與否作爲區別實益所開展的體系。提出 AI 存在強弱之別者乃美國哲學家瑟爾 (John R. Searle)，其學說指出「強 AI」(Strong Artificial Intelligence; Artificial General Intelligence) 係指電腦除具備自主判斷能力外，情感、個性及社交等自我人格意識的附加功能乃最大特色；相對而言，「弱 AI」固然具備模擬人類行爲逕爲判斷和決策之能力，然未具備人類內生性之自我認知和思考能力。請見 John R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, Vol. 3, No. 3, September 1980, pp. 417-424。

34 Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham: Ashgate, 2009), pp. 33-35.

值得注意者，穆爾 (Nicholas W. Mull) 結合觀察自主性武器與機器學習，精闢地指出各方於界定自主性武器系統時所犯之邏輯謬誤，他主張：武器系統之「自主性」尚不得與「自動性」等量齊觀。舉例而言，傳統地雷與觸發陷阱所用引爆裝置 (trip wire)，一旦偵測到特定變化即會自動啓動，惟此些自動武器背後之交戰決定權仍屬於使用及設置其之「人類」。[35] 然而穆爾提出警示，即便人類能保持對此類武器一定程度之審查權，惟人類介入行爲的能動性實已或深或淺因人類過度倚賴機器運算結果，而受到機器及其演算法影響，恐難以發揮實質道德能動性，或可謂此等武器運作模式中之人類監督恐已流於表象。[36] 穆爾精準指出，在人類過度仰賴機器爲特定功能履行之進程中，已導致利用自身自由直覺與經驗以全權判斷並執行攻擊指令之弱化與退化。[37] 以 1988 年溫賽尼斯號 (U.S.S. Vincennes) 導彈驅逐艦誤擊伊朗民航機事件爲例，雖然當時驅逐艦之電腦已正確判斷該架伊朗民航機正在上升中，然後續人爲疏失卻導致艦上人員誤判該機意欲對艦

35. Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," p. 477-478.

36. 應附帶指出者，穆爾將機器學習之特質歸納爲以下三點：監督式學習 (supervised learning)、強化學習 (reinforcement learning) 及非監督式學習 (unsupervised learning)。首先，所謂「監督式學習」係指機器被給予若干樣本以完成某些預期結果，例如令電腦於多張照片中辨識出含有某人之照片爲何；其次，所謂「強化學習」係指將多種演算法應用於機器上，令其得以學習如何求得最佳解方，並據此決定欲採取若干行動；最後，所謂「非監督式學習」指涉即便未有外部給予之預設目標時，演算法仍能令機器學習並發展出各種處理方法。請見 Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," pp. 478-479。

37. Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," pp. 484-485.

艇發動攻擊，故而決定先行擊落之。[38]

（二）以智能決策迴路程序控制界定之觀點

1. 人權觀察組織

有別於上述論點集中於「自主性」意涵之闡述，人權觀察組織所出版之人權分析報告 (*Losing Humanity - The Case Against of Killer Robots*) 則另闢蹊徑，轉而依據人類智能決策迴路程序控制之消長分類自主性武器如下：

(1) **有大量人為因素介入武器**：人權觀察組織對於「有大量人為因素介入」之界定，略以：「能夠追蹤或辨識潛在標的、能對人類操作員發布提示、能優先處理被選定之標的、能選擇攻擊時機，或能對被人類操作員所選定之標的及個人發出最終指令之系統。」[39] 如熱感應飛彈、水雷和魚雷等，此些武器全然根據人類事先給定模式以偵測並攻擊軍事標的。據此以觀，其主觀道德能動性最終仍歸屬於人類，而非機器。[40]

---

38. William M. Fogarty, *Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988* (Washington, D.C.: U.S. Department of Defense, 1988), pp. 1-53.

39. 當前具備部分自主性功能之武器可類屬為「有大量人為因素介入」者，其態樣涵蓋多種感測器融合 (sensor-fused) 武器、飛彈、徘徊型武器與魚雷彈等。其中，較具爭議者係南韓之三星哨兵武器系統，該系統部署於南北韓間之非軍事區一帶，並以人員為攻擊標的，其具備「人類可監控並撤銷決定」之任務執行功能，但仍歸由人類操作員以遠端遙控以對標的展開交火。相關文獻請見 Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," pp. 485-486。

40. Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (New York: Human Rights Watch, 2012), pp. 2-3, *Human Rights Watch*, <https://www.hrw.org/report/2012/11/19/losinghumanity/case-against-killer-robots>.

(2) **人類可監控並撤銷決定武器**：人權觀察組織對於「人類可監控並撤銷決定」之操作型定義，略以：「人類操作員得以介入武器系統所做出之決定並監督之，人類決策在最終意義上可凌駕於系統決定。」[41] 學者有認，儘管此類武器予人一種「機器受到人類有意義的控制」之錯覺，惟該觀點與實際情形有所出入，蓋人類實已將（至少是部分的）生殺大權「委託、交付」予機器判斷，並由其執行之；而此甚易導致人類於缺乏實證——抑或有反證——之情況下，過度依賴甚或盲從機器判斷，造成所謂之自動化偏差 (automation bias)。[42]

(3) **未有人爲因素介入武器**：人權觀察組織對於「未有人爲因素介入」之概念意涵，略以：「作爲機器使用之完全自主系統及參與相關職能之自主權，包括動能打擊之發動。」[43] 完全自主性武器若成立，意謂人類武力使用之決策權限全部授權或委諸自主性武器系統執行，遂使其與區分原則、比例原則、馬爾頓條款及攻擊預防原則等國際人道法所強調的諸項原則產生規範扞格，成爲人權觀察組織持續關注的焦點。[44]

---

41. Human Rights Watch, *Losing Humanity: The Case Against Killer Robots*, pp. 2-3.

42.「人類可監控並撤銷決定」類型之自主性武器包含反飛彈防禦系統、反火箭防禦系統，例如美國海軍之神盾系統及陸基愛國者飛彈防禦系統。此等武器於軍事上一般統稱爲「反軍事設備武器」，其他系統諸如方陣近迫武器系統 (Phalanx Close-In Weapon System)，雖具備偵測、追蹤、識別、選擇或攻擊標的之部分自主性，惟交火指令仍須由人類操作員許可方能執行。相關文獻請見 Nicholas W. Mull, "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," pp. 484-485。

43. Human Rights Watch, *Losing Humanity: The Case Against Killer Robots*, pp. 2-3.

44. Human Rights Watch, *Losing Humanity: The Case Against Killer Robots*, pp. 23-26.

2. 自主性的連續性與人類判斷之空間

從上述可知，近代學者常囿於相對簡單之「有大量人為因素介入」、「人類可監控並撤銷決定」或「未有人為因素介入」等三類自主性武器系統背景下，討論無人武器系統之人類判斷等級，然其實未揭示人類或機器於武裝衝突期間對於理解環境時所面臨之挑戰。[45] 爰此，為解決高度流動且環境複雜之「人類可監控並撤銷決定」系統，即需有不斷變化之人類判斷。由此可見，自主性可沿一連續線發展，而於不同點上執行之武器系統則可能適用不同的人類判斷。復因人類有權決定何時、何地部署或加入參數，故嚴格言之，並無眞正「未有人為因素介入」之武器系統存在。[46]

上述人權觀察組織推導出無眞正「未有人為因素介入」武器系統之觀察，在規範論上與德國及法國的官場立場頗為相似。德國於 2021 年 GGE「新興科技領域之致命自主性武器系統」會議中提出之官方意見表示，德國與法國均不應發展、生產製造、取得或部署完全未有人為因素之自主性武器系統，在德國的定義中，此指涉完全脫離人類控制和指揮之操作系統。[47] 析言之，德國與法國仍沿用了「自主

45. Dan Saxon, "A Human Touch: Autonomous Weapons, Directive 3000.09, and the 'Appropriate Levels of Human Judgment over the Use of Force'," *Georgetown Journal of International Affairs*, Vol. 15, No. 2, Summer-Fall 2014, pp. 103-104.

46. Dan Saxon, "A Human Touch: Autonomous Weapons, Directive 3000.09, and the 'Appropriate Levels of Human Judgment over the Use of Force'," pp. 103-104.

47. Peter Beerwerth, "National Statement by Germany Group of Governmental Experts on 'Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)'," August 12, 2021, *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2021/08/Germany.pdf>; German Federal Foreign Office, "German Commentary on Operationalizing All Eleven Guiding Principles

性」(autonomy) 此一類型化區辨要素，將致命武器分爲「完全自主」和「部分自主」兩大體系。前者得以在沒有人類干預的情況下選擇目標並發動攻擊，以及修改其戰略任務。而被德國和法國認爲應該完全禁絕之；後者之部分自主系統則在由人類操作員定義的框架內選擇和攻擊目標，但無法自行做出更深遠的決策，此一類型應有適度規範之必要以確保合乎國際人道法。[48]

（三）特徵式列舉之觀點

與美國相去不遠，中國對自主性武器系統一貫採取矛盾立場，一方面表態禁絕自主性武器的使用 (use)，但卻支持其研發 (development)，似可見軍事準則與現代技術進步之政治修辭密不可分。實則，此模稜兩可態度有助於可操作之戰略平衡在形塑國際監管過程中維持開放性。[49] 2017 年中國國務院發布之《新一代人工智能發展規劃》明確指出：「建立倫理道德多層次判斷結構及人機協作的倫理框架。……積極參與人工智慧全球治理，加強機器人異化和安全監管等人工智慧重大國際共性問題研究，深化在人工智慧法規、國際規

---

at a National Level as Requested by the Chair of the 2020 Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW)," June 24, 2020, *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2020/07/20200626-Germany.pdf>.

48. Elisabeth Hoffberger-Pippan, Vanessa Vohs, & Paula Köhler, "Autonomous Weapons Systems: UN Expert Talks Facing Failure Time to Consider Alternative Formats," July 2022, *SWP*, <https://www.swp-berlin.org/publications/products/comments/2022C43_AutonomousWeaponsSystems.pdf>.

49. Congressinal Research Service, "International Discussions Concerning Lethal Autonomous Weapon Systems," *Congressinal Research Service*, No. IF11294, February 2023, pp. 1-3.

則等方面的國際合作，共同應對全球性挑戰。」[50] 此亦沿襲中國素來在 AI 政策採取之多邊主義。[51]

2018 年，中國進一步對於何謂預防性控制之立場提出嚴謹闡述：「新興技術之風險影響應客觀、公正且充分予以討論評估。在進行相關討論前，不應存在任何預設前提或預先設定結果之干擾，否則恐阻礙人工智慧技術之發展。」[52] 此等軍事友好政策，似乎說明中國對於防止 AI 軍備競賽預防原則 (Precautionary Principle) 採取消極保留態度。

根據中國之觀點，自主性武器應包括但不限於下述五項基本特徵：1. 致命性：意指足夠之有效彈藥和手段係屬致命；2. 自主性：執行任務過程中未存在任何人為干預和控制；3. 不可逆性：意指一旦啓動，即無法終止該設備；4. 無差別性：意指該設備不論條件、情境和目標為何，皆將會執行殺戮和瞄準之任務；5. 進化性：意指透過與環境之互動，該設備可自主學習，以超出人類預期之方式擴展其功能和能力。[53]

---

50. 中華人民共和國國務院，〈國務院關於印發新一代人工智能發展規劃的通知〉，《國務院文件》，國發〔2017〕35 號，2017 年 7 月 20 日。

51. Thomas Christian Bächle & Jascha Bareis, "Correction: Autonomous Weapons as a Geopolitical Signifier in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies," *European Journal of Futures Research*, Vol. 10, No. 1, December 2022, p. 14.

52. United Nations, "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects."

53. United Nations, "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects."

二、分析與檢討

綜上論結，各國與非政府組織認定自主性武器系統之概念範疇，構成發展相關國際規範之核心，因事涉「自主性」、「人類介入程度」及何謂「有意義之人爲控制」等意涵界定，且攸關因果關係及戰爭責任歸屬之判斷疑義，看法分歧實不足爲奇。誠如學者穆爾之剴切觀察，自主性武器系統一貫以來的爭議既然不脫權責歸屬之劃定，對於人類介入程度及武器啓動後之人爲控制程度，當屬嗣後可否納入國際人道法之癥結。再者，對於建構「有意義之人爲控制」的監督課責，原則上亦應透過各國合意並凝聚形塑國際規範上之共識；職此，應考量兼顧該 AI 自主性武器系統於道德與法律上之潛在利弊，以及人類於若干範圍內得將武力使用之決策權限全權委諸自主性武器系統執行，並綜合參酌國家安全與生命價值維護之衡平，始得爲妥適判斷。

參、《特定常規武器公約》之發展演替與重要里程

聯合國人權理事會 (United Nations Human Rights Council, HRC) 於 2013 年召開之審查會議中，由聯合國特別報告員海因斯 (Christof Heyns) 所撰寫之評估無人機影響報告中，特別強調了致命的自主性武器系統，認爲部署自主武器不僅須升級所用武器種類，亦須改變使用此些武器之人員身分，且隨著對致命自主機器人技術之疑慮日增，武器與戰士間之區別或將趨於模糊。該份報告更強調，機器人於處理定量問題方面固然卓有成效，然於處理人類生活時所須進行之定性評估能力則甚爲有限。[54]

54. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," in Elisabeth Johansson-Nogués, Martijn C. Vlaskamp, & Esther Barbé, eds., *European Union Contested Foreign Policy in a New Global Context* (Berlin: Springer, 2019), p. 135.

「制止殺手機器人運動」(UK Campaign to Stop Killer Robots, CSKR) 即以前述報告爲契機，倡議禁止使用致命自主性武器系統之新規範。儘管自由國際主義認爲軍備控制規範其傳播實屬線性，惟不乏實證案例顯示軍備控制規範之傳播遠非如此：致命自主性武器系統其概念化毋寧係一個不斷內在生成變遷之共識凝聚過程。故而 2013 年聯合國人權理事會議事期間，法國在歐盟駐日內瓦代表團支持下，建議將辯論場域轉移至《特定常規武器公約》之審查會議，藉以確保系爭公約能隨著新戰爭型態之演變爲適度活用。[55]

職此，聯合國於 1980 年通過《特定常規武器公約》及三個議定書，該公約於 1983 年 12 月 2 日生效，以禁止或限制被認爲會造成不必要傷害或無差別傷害之武器，保護平民免於武裝衝突，確保戰鬥人員免於遭受超出合理範圍之傷害爲宗旨。[56]《特定常規武器公約》後於 2014 年啓動爲期四天之非正式專家會議，藉由一般性規定的彈性框架，以一系列議定書將多數具體規範準則涵化爲原始條約之一部；[57] 其中存在如程序性機制、無等級制度等制度性工具，足以鼓勵並促成有效討論。[58] 在此機制中，國家及非國家行爲者藉由非正式會議平臺

55. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," p. 136.

56. 目前《特定常規武器公約》締約方總數爲 127 個，包括美國、英國、德國、法國、日本、俄羅斯等主要軍事大國，以及中國與印度。在《特定常規武器公約》架構下，締約方及非政府組織自 2014 年起召開相關會議，圍繞致命自主性武器系統之管制展開討論。

57. H. Müller & C. Wunderlich, *Norms Dynamics in Multilateral Arms Control* (Athens: The University of Georgia Press, 2013), pp. 337-366; T. Gehring, *Dynamic International Regimes: Institutions for International Environmental Governance* (New York: Peter Lang, 1994), pp. 1-515.

58. T. Risse, "'Let's Argue!': Communicative Action in World Politics," *International Organization*, Vol. 54, No. 1, Winter 2000, pp. 1-39.

互動磋商，產出對於規範之理解與解釋，進一步形塑規範內涵，成功將該議程設定由過往強權壟斷之單邊主義導向多邊主義。

《特定常規武器公約》及後續審查會議作爲一個彈性之國際合作機制實有其優勢，其特徵係藉由一系列協議將特定規範添加至原始條約內。因每回合國際談判皆爲國家間之權力博弈與權力互動，在創造一系列國際規範派生協議之過程中，將權力較量之博弈場域由自主性武器系統轉移至談判回合，更可靈活應對武器技術之未來發展及其可能引發之弊端。[59] 以《特定常規武器公約》之《第一附加議定書》規定爲例，該議定書甫提出後即獲得大量支持，原因在於該限制並未對各締約方在武器使用上或其所涉軍事利益上構成負面影響或實質減損。[60] 然而，並非所有規定都能如此順利地被各國所接受。舉例而言，《第二附加議定書》便因涉及軍事機密而在談判上遇到不小的阻力。[61] 迄至 1990 年代中期因地雷的氾濫，使得《特定常規武器公約》之修訂議程再次獲得一定的重視，各國希冀藉此將地雷對無辜平民造成的殺傷及影響降至最低。[62] 各國也較爲願意將武器限制納入考慮，蓋此種遍及範圍廣泛且深遠的潛在危機，可能導致其必須負擔更多的損害賠償，此意謂經濟戰略的價值權衡亦開始成爲各締約方的考量因素，進一步奠定《特定常規武器公約》於武器控管議題上作爲重要國

59. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," pp. 136-137.
60. J. Mcclelland, "Conventional Weapons: A Cluster of Developments," *International & Comparative Law Quarterly*, Vol. 54, No. 3, July 2005, pp. 755-767.
61. J. Mcclelland, "Conventional Weapons: A Cluster of Developments," pp. 755-767.
62. J. Mcclelland, "Conventional Weapons: A Cluster of Developments," pp. 755-767.

際法淵源之地位。[63]

綜言之，得益於上述動態循環之對話機制，復以其成員組成結合如軍事人員、國際法專家、學術界人士及機器人研究人員等多方利害關係人群體，針對致命自主性武器系統的風險與管制，得以激發更爲活絡之規範形成場域。在此歷程中，專家會議之議論焦點始終不脫兩項爭點：其一，自主性武器系統於若干程度上能夠符合國際法？其二，何謂必要且有意義之人爲控制？茲分析如下。

一、2015 年至 2016 年：「有意義之人爲控制」與「人機關係」的正面交鋒

由德國於 2015 年及 2016 年主持之《特定常規武器公約》非正式會議，強調平民豁免權等基本規範之重要性。代表團一致宣布致命自主性武器系統應遵守國際人道法，此等規定涵蓋諸如《日內瓦公約第一附加議定書》（*Additional Protocol I to the Geneva Conventions*，以下簡稱第一附加議定書）第 48 條、第 51 條第 3 項所揭櫫之區分平民與戰鬥人員之原則、第 51 條第 5 項及第 57 條之比例原則，乃至於第 58 條 (c) 款揭櫫之攻擊預防原則。[64]

事實上，不獨國際法學界抱持人道主義觀點，部分 AI 與機器人研究者亦做如是觀點：渠等由科學證據角度，試圖將科學社群對 AI 研究之專業意見納入議程設定及規範形成過程，所提出公開信指摘當前技術已達可部署完全自主性武器系統之階段，故呼籲各國應採行預防性禁止措施，以避免出現「人類無法控制」之武器。[65] 申言之，橫

63. J. Mcclelland, "Conventional Weapons: A Cluster of Developments," pp. 755-767.

64. United Nations, *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)* (Geneva: United Nations, 2016), pp. 1-16.

65. "Autonomous Weapons: An Open Letter from AI & Robotics Researchers,"

跨 2015 年至 2016 年訂約過程，其所著眼之「有意義之人爲控制」基本立場已蔚爲主流。[66]

儘管如此，如法國與美國等軍事大國相繼對此一議題設定表示疑義。法國認爲「有意義」一詞缺乏精確性及技術準確性，無法保證人爲控制於武器系統均能適用；美國則傾向將人爲控制概念定義爲「人類對武力使用之適當判斷水平」(appropriate levels of human judgment over the use of force)。[67] 爲弭平歧見並凝聚共識，德國則主張以「適當之人類判斷及參與」(appropriate human judgment and involvement) 界定「有意義之人爲控制」，俾作爲衡平性之判斷指標。[68]

對於何謂「有意義之人爲控制」之歧見，導致與會代表轉而將爭論焦點朝向「人機關係」(human-machine relationship) 傾斜。[69] 此一文本設計之討論過程反映各國代表團之共識與需求，且彰顯「人爲控制」此一規範性概念須被編纂爲一個可操作性之組織原則，藉以保持規範之靈活性，並允許其意義隨武器系統之不同態樣而演變。[70]

February 9, 2016, *Future of Life*, <https://futureoflife.org/open-letter-autonomous-weapons>.

66. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," p. 137.

67. U.S. Department of Defense, *Department of Defense Directive (2017)*, Number 3000.09: Autonomy in Weapon Systems, November 10, 2017, pp. 1-15.

68. United Nations, *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, pp. 1-16.

69. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," p. 137.

70. E. Rosert, "How to Regulate Autonomous Weapons: Steps to Codify Meaningful Human Control as a Principle of International Humanitarian Law," January 1, 2017, *JSTOR*, <https://www.jstor.org/stable/resrep14276?

學理再次直指無論是「自主性」之爭，抑或對「人爲控制」意涵界定之僵持不下，皆彰顯出此兩者間並非一個膠著恆定之固定點，而毋寧係隨著各該利害關係團體於特定議題互動而動態起伏之浮動概念。換言之，自主性武器非爲僵化之「自主性」機械化釋義，而係來回擺盪於遠程操作之「系統—受監督的系統—無監督的系統」之譜系間，且持續巡弋定位點之浮動性規範標的。[71] 基此，於此軍武科技驅動下不斷演變及進化之場域中，如何將基本規範轉化爲標準化程序之規範性原則一事仍具談判空間，[72] 攸關後續規範形成和文本制定。

依循前開分析，各國代表團開始將「人機關係」視爲辯論中的一個參考點，而「人爲控制」正透過審議討論漸次發展爲一項新的組織原則；此項組織原則除了體現不僅止一種解釋存在之操作空間，更會構成認知落差，從而導致分歧。[73] 在談判回合中，如何將基本規範轉化爲標準程序成爲各國可談判之事項，對於分析人類控制如何塑造或定義存在之規範性問題而言，益發重要。[74]

《特定常規武器公約》締約方於 2016 年 11 月進行第五次審查會議時，一致通過成立 GGE，希冀加強審議致命自主性武器系統領域之新興技術，並就造成人身威脅之自主性武器之發展予以監管，不限成員名額之 GGE 便於此理念下誕生。嗣後在 GGE 討論議程中，各國持續關注「人爲控制」之重要性及國際規範機制之侷限性等議題；

---

seq=1>.

71. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," p. 137.

72. A. Wiener, *A Theory of Contestation* (Basingstoke: Springer, 2014), pp. 37-39.

73. B. Jose, *Norm Contestation: Insights into Non-conformity with Armed Conflict Norms* (Basingstoke: Springer, 2018), p. 28.

74. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," pp. 137-138.

至此爲止，儘管各國對於自主性武器之定義仍難以劃一，管制框架亦難以達成共識，然對於武器系統之潛在開發及其使用應服膺既有國際法規範（尤其國際人道法）一事，可謂具備高度共識。[75]

二、2018 年至 2019 年：三種倡議之醞釀成熟

2018 年及 2019 年之 GGE 中，「人爲控制」開始由組織原則上升爲具廣泛道德影響之規範，並聚焦於「人爲控制」其規範性意涵闡釋之互動結構；儘管研究者試圖提出不同理論予以詮釋，然仍未發展出廣爲各國所認同採納之周延學說。GGE 於 2018 年發布報告揭櫫三種規範模式供各界參考，分別爲：（一）訴諸現有國際人道法；（二）以硬法 (hard law) 形式強行禁止或限制自主性武器之發展；（三）以軟法 (soft law) 結構訴諸各國共識，使其逐漸展現一致化方向等三種標準化程序。[76] 就此，除締約方向來普遍認同適用國際人道法以外，第三種以軟法訴諸各國共識的例子亦可見於《第二附加議定書》之技術性附件中，其指出：「根據本議定書產生對地雷區域地雷和陷阱位置進行記錄的義務時，下列指導方針應列入考慮：……關於預先計畫的地雷區域與大規模預先計畫使用陷阱的情況。」(Whenever an obligation for the recording of the location of minefields, mines and booby traps arises under the Protocol, the following guidelines shall be taken into account: 1. With regard to pre-planned minefields and large-

75. United Nations, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems* (Geneva: United Nations, 2018), pp. 1-21；林昕璇，〈AI 自主性武器系統在國際法上適用之研析〉，頁 33-34。

76. United Nations, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, pp. 14-15.

scale and pre-planned use of booby traps:.....) [77] 採納所謂「自願最佳實踐」的規範模式，然而 1996 年 5 月通過之《第二附加議定書》嗣後則歷經從軟法推展至以硬法形式強行限制軍武發展的轉變歷程。如在《修訂第二議定書》的技術附件中，這些指導方針從「應列入考慮」(shall be taken into account) 修正爲「應被落實」(shall be carried out in accordance with the following provisions)，致使前開準則增添一層非自願性之義務色彩，[78] 反映各國對此前之自願最佳實踐的國家實踐效果有所疑慮。

儘管前述三項規範路徑頗具見地，惟各國對此仍有高度爭議，且呈現各陣營之對峙性。首先，澳洲、以色列、美國、英國、韓國及俄羅斯所組成之聯盟支持上述「訴諸現有國際人道法」之倡議，認爲基於《第一附加議定書》第 36 條所定之新武器審查原則，已確保某種程度之「人爲控制」，亦即既存國際人道法即足以處理自主性武器系統所衍生之問題。[79] 但一個重要的情勢轉變在於，俄羅斯提出異議，

77. United Nations, *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects* (Geneva: United Nations, 1980), p. 2; International Humanitarian Law Databases, "Technical Annex," October 10, 1980, *International Committee of the Red Cross*, <https://ihl-databases.icrc.org/en/ihl-treaties/ccw-protocol-ii-1980/technical-annex?activeTab=undefined>.

78. United Nations, *Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on 3 May 1996* (Geneva: United Nations, 1996), p. 10; United Nations, "Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as Amended on 3 May 1996 (Protocol II to the 1980 Convention as amended on 3 May 1996)," May 3, 1996, *United Nations*, <https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.40_CCW%20P-II%20as%20amended.pdf>.

79. Esther Barbé & Diego Badell, "The European Union and Lethal Autonomous

主張若未來各國欲採取以硬法強制性或軟法途徑規範瞄準階段之「人爲控制」，其將退出該輪談判。[80]

相形之下，以奧地利、巴西及智利爲首之聯盟鑑於自主性武器嚴重牴觸現有國際人道法和公約，再次強調「有意義之人爲控制」原則。爰此，制定一項具有法律約束力之國際協議並就此展開談判，貫徹瞄準過程中確保人爲控制之規定實已刻不容緩，藉以禁絕瞄準過程中缺乏人爲控制之武器系統。[81]

由於雙方矛盾分歧與衝突對立，導致談判過程中之僵局和窒礙難行。爲緩和上述游離於兩極間之矛盾被進一步激化，在法國及德國主導下，第三組國家建議採行軟法結構之倡議，其所提出之政治宣言冀求藉由較爲緩和的規範以凝聚共識並推動機制發展。[82] 此陣營所持依據在於，根據《第一附加議定書》中所定義之完全自主致命武器系

Weapons Systems: United in Diversity," pp. 137-138.

80. The Russian Federation, "Considerations for the Report of the Group of Governmental Experts of the High Contracting Parties to the Convention on Certain Conventional Weapons on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems on the Outcomes of the Work Undertaken in 2017-2021," June 2021, *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2021/06/Russian-Federation_ENG1.pdf>.

81. United Nations Office for Disarmament Affairs, "Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS)," September 2020, *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2020/09/GGE20200901-Austria-Belgium-Brazil-Chile-Ireland-Germany-Luxembourg-Mexico-and-New-Zealand.pdf>.

82. Kenneth W. Abbott & D. Snidal, "Hard and Soft Law in International Governance," *International Organization*, Vol. 54, No. 3, Summer 2000, pp. 421-456.

統，其中人爲控制本係定義爲一個模稜兩可之組織原則，故透過政治宣言將此不具約束力的政治措施和軟法逐步推進法制化過程乃當然之理。

此宣言並得到歐盟代表團之附議，於嗣後《特定常規武器公約》的歷次會議中，被視爲希望制定禁止條約國家與支持現狀國家光譜兩端間之第三路線。在規範效力上，由於該政治宣言僅爲重申《第一附加議定書》之規範原旨，故屬不具規範拘束力之文本。[83] 在該文本中，「人爲控制」作爲一具可解釋操作空間之調節性組織原則，並藉由《第一附加議定書》第 36 條規定落實程序審查，確保武器審查過程及決定均保有透明度，同時建立政治與法律及個人與國家之問責機制。該做法亦體現此第三路線係採取較複雜化之軟法結構，例如以行爲守則形式實行具政治約束力之措施、於公約機制下設置專家委員會等柔和勸進模式，審議與致命自主性武器系統相關之技術發展。各陣營所代表之三種倡議路線請見表 3，又各締約方對於是否應對 AI 新武器實施預防性禁止的立場請見表 4。

表 3　各陣營所支持之三種倡議路線

	維持現狀派	強行規定派	法國及德國陣線
規範性基礎原則	平民豁免原則	平民豁免原則	平民豁免原則
組織標準	區分原則、比例原則、攻擊預防原則、人爲控制原則（英美兩國觀點：瞄準階段之人爲控制）	區分原則、比例原則、攻擊預防原則、人爲控制原則（擊殺階段之人爲控制）	區分原則、比例原則、攻擊預防原則、人爲控制原則（法國：人類指揮官負責；德國：有效之人爲控制）

83. Ministerie van Buitenlandse Zaken, *Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems in the Context of the Objectives and Purposes of the Convention* (Hague: Ministerie van Buitenlandse Zaken, 2017), pp. 1-3.

	維持現狀派	強行規定派	法國及德國陣線
實體法依據	訴諸現有國際人道法	以硬法形式化強行禁止規定以限制自主性武器之發展	透過軟法結構訴諸各國共識的凝聚
建制依據	暫停或禁令爲時過早且毫無根據，蓋人爲控制於《第一附加議定書》第 36 條中已獲致保障，該條規範要求各國核實武器是否爲該議定書或任何其他國際法規則所禁止	自主性武器嚴重牴觸現有國際人道法和公約，故應制定一項具法律約束力之國際協議並就此展開談判，貫徹瞄準過程中缺乏人爲控制的武器系統以保證人爲控制之禁止規定有其必要性	鑑於《第一附加議定書》所定義之完全自主致命武器系統，其中人爲控制被定義爲一個模稜兩可之組織原則，故應透過政治宣言將此不具約束力之政治措施和軟法逐步推進法制化
代表國家	澳洲、以色列、美國、英國、韓國和俄羅斯	奧地利、巴西和智利	德國、法國

資料來源：作者整理自 Esther Barbé & Diego Badell, “The European Union and Lethal Autonomous Weapons Systems: United in Diversity,” p. 139。

表 4 預防性禁止自主性武器之各國立場

立場	國家
支持	阿爾及利亞、阿根廷、奧地利、玻利維亞、巴西、智利、哥倫比亞、哥斯大黎加、古巴、吉布地、厄瓜多、埃及、薩爾瓦多、迦納、瓜地馬拉、梵蒂岡、伊拉克、約旦、墨西哥、摩洛哥、納米比亞、紐西蘭、尼加拉瓜、巴基斯坦、巴拿馬、秘魯、烏干達、委內瑞拉、辛巴威
反對	澳洲、法國、德國、印度、以色列、俄羅斯、南韓、西班牙、土耳其、英國、美國
其他	中國

資料來源：Congressinal Research Service, “International Discussions Concerning Lethal Autonomous Weapon Systems,” p. 1。

三、2019 年迄今：爲新興科技軍事武器提供妥適之框架準據

由於 UNCCW 與致命武器系統的新興科技發展呈現高度的相關性，GGE 於 2019 年就「致命武器系統領域的新技術問題」歸納 11 項指導原則，茲臚列如下：[84]

（一）國際人道法全面適用於所有的武器系統，涵蓋可能開發和使用的致命自主性武器系統。

（二）人類必須對做出使用武器的決定負責，不能將其責任推託給機器。而此必須將機器的整體使用壽命進行考量。

（三）人機互動得以多種形式存在，並在武器生命週期的數個階段實施，應確保其基於自主性武器系統的新興科技之潛在使用符合國際法，尤其是國際人道法。在確定人機互動的品質與程度時，應將一系列因素納入考量，包括作戰環境及全體武器系統的特色與能力。

（四）須依據可適用的國際法以確保在特定常規武器公約框架內開發、部署和使用任何新興武器系統之責任，包括於責任鏈內透過負責人的指揮與控制下操作系統來確保責任歸屬。

（五）根據各國在國際法下的義務，在研究、開發、獲取或採用新武器、作戰手段及方法時，必須確定其使用是否在某些或所有情況下會被國際法禁止。

（六）開發或獲取基於致命自主性武器系統領域之新興科技的新武器系統時，應考慮實體安全、適當之非實體保障措施（包括防止駭客或數據欺騙的網路安全）、被恐怖組織獲取的風險及擴散的風險。

---

84 United Nations, *Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems* (Geneva: United Nations, 2019), pp. 1-14.

（七）風險評估及緩解措施應成爲任何武器系統中新興科技的設計、開發、測試和部署週期的一部分。

（八）在堅持遵守國際主義法和其他國際法義務的同時，應對自主性武器系統使用於新興科技時加以考慮。

（九）在制定潛在的政治措施時，不應將自主性武器系統的新興科技擬人化。

（十）在《特定常規武器公約》範圍內進行的討論和採取任何潛在的政治措施，不應對和平利用 AI 自主技術方面的進展或獲得此類技術造成妨礙。

（十一）《特定常規武器公約》爲在《日內瓦公約》的目標及宗旨範圍內對自主性武器系統之新興科技問題提供了適當的框架，而該公約強調軍事必要性及人道主義考量之間應取得平衡。

由上述 11 項指導原則可見，國際人道法的遵循似成爲 GGE 頗爲青睞的指標，在多項指導原則中都可見其與既有《特定常規武器公約》的對話痕跡，惟亦有挑戰者指出，國際人道法揭櫫的原則固然被廣泛接受，但亦應適度檢視必要性原則和最小武力原則適用在 AI 自主性武器系統的相容性。特別是布蘭查德 (Alexander Blanchard) 和塔奧德 (Mariarosaria Taddeo) 觀察到，戰場上操作員原本即需具備：（一）認識軍事目標的能力；（二）認識納入特定軍事攻擊之成功機率的計算之能力。然而當 AI 自主性武器系統某種程度上對設計者或操作員而言係不可預測的時候，何謂軍事必要性和最小武力在判斷上即會遭致困難。[85] 又值得深究者，如果上述國際法淵源均試圖各自從不同的指標去預測 AI 自主性武器的潛在風險，又該如何在去蕪存菁後將其逐步建構爲一個整合性的遵循指標，殊值探究。

---

85 Alexander Blanchard & Mariarosaria Taddeo, “Jus in Bello Necessity, The Requirement of Minimal Force, and Autonomous Weapons Systems,” pp. 295-298.

四、分析與探討

綜言之，「有意義之人爲控制」此一概念之出現，對自主性武器系統之權責歸屬問題製造曖昧不明且難以嚴謹區隔之規範形成空間，成爲各國話語權競逐的角力場域。亦即，人爲控制業已成爲各國代表及各陣營慣常引用之依據，並基於自身內國利益提出不同解讀。然如同前述分析，自 2014 年的第一次非正式專家會議開始，與會代表即已對於自主性武器系統適用既有國際人道法規範形成高度共識，而在 GGE 2019 年爲 AI 自主性武器系統揭示之三大原則——區分原則、比例原則及攻擊預防原則——適用過程中，有無可能產生與現行規範不相吻合的情形，以及如何在去蕪存菁後，奠基於前述國際法之理論與實踐基礎，就 AI 新興武器逐步建構一個共同價值維繫之規範性遵循指標，遂成爲值得關注之重要課題。

肆、國際人道法與特定常規武器公約適用於 AI 自主性武器系統之規範性省思

一、規範性省思之一 UNCCW 與國際人道法之兼容並行[86]

（一）區分原則

區分原則 (the principle of distinction) 乃係軍事行動中攻擊目標

86. 國際法涉及戰爭行爲規範，學理上素來主張應切割爲下列三個層面分別討論：合法訴諸武裝衝突 (Jus ad bellum; the right to resort war)、戰時法 (Jus in bello; laws of war or laws of armed conflict) 及戰後法 (Jus post bellum)。首先，「合法訴諸武裝衝突」是指在國家或國際組織行使武力時，必須遵循國際法律規範與要件，以及明確界定何種情況下禁止、允許和限制使用武力。此一面向涵蓋《聯合國憲章》第 2 條第 4 項所規定的禁止使用武力原則，也包括合法使用武力的例外情況，如自衛權及經聯合國安全理事會授權的武力行使；復「戰時法」的核心目標在於規範

之行爲規範，旨在禁止使用無差別之殺傷性武器或濫殺無辜。此原則奠基於 1977 年《第一附加議定書》第 48 條規定：「爲確保對平民及民用物之尊重及保護，交戰各方無論何時均應在平民與戰鬥員、民用物與軍事目標之間予以區別，衝突一方之軍事行動僅應以軍事目標爲對象。」[87] 同法第 50 條至第 52 條則具體區辨軍事目標之對象及物體，並對武裝衝突中之交戰方式與攻擊進行限制及規範。[88]

復根據《第一附加議定書》第 51 條第 3 項規定：「平民應加以保護，惟對於平民直接從事敵對行動或直接參加敵對行動時，喪失其受保護之資格，則可例外予以攻擊。」此處所定之「直接從事敵對行動」究何所解？以色列最高法院 *The Public Committee Against Torture in Israel v. The Government of Israel* 乙案中曾就無人機狙

武裝衝突中交戰方的整體行爲，以明確界定其權利和義務。它透過法律規定武裝衝突狀態中的作戰方法、途徑和武器使用，旨在降低戰爭造成的非人道損害。在戰爭或武裝衝突爆發時，所有交戰各方無論其戰爭目的爲何，都必須遵循國際規範或習慣國際法，以保護平民及未參加戰鬥或已退出戰鬥的戰鬥員。目前的戰時法律譜系建立在海牙公約體系和日內瓦公約體系的基礎上。其具體界定了可以使用的武器種類、作戰方式和手段等細節，同時確保對平民和非戰鬥人員的保護於法律上所享之人道權利保護。請見 Anthony Clark Arend & Robert J. Beck, *International Law and the Use of Force* (London: Routledge, 1993), pp. 1-3。國內文獻請見林昕璇，〈AI 自主性武器系統於國際法適用上之研析〉，頁 28；李鈺翎，《從國際法論人工智能軍事武器之發展與挑戰》（臺北：東吳大學法學院碩士論文，2018 年），頁 75-79。

87. 此段原文爲：“In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.” 關於區分原則之國內文獻，請見趙國材，〈論國際人道法適用於內戰之發展〉，頁 122-153。

88. 林昕璇，〈AI 自主性武器系統於國際法適用上之研析〉，頁 28-29。

殺行動適法與否予以評價；[89] 該案採取「功能性」之解釋取向，以「市民是否正行使戰鬥員之功能角色」(performing the functions of combatants) 作爲權衡基準，並將「功能角色」之範疇予以進一步擴張性解釋：指涉對象非僅限於「攻擊前」、「攻擊時」及「攻擊後」裝備武器，更擴張及於提供服務予非法武裝集團及自發性參與武裝衝突等市民，[90] 均含括爲得予以合法攻擊之標的群。以色列最高法院進一步將參與武裝衝突之「時點」予以擴張解釋，導致一般市民本來依據《第一附加議定書》第 51 條第 3 項除直接參加敵對行動或直接參加敵對「行動時」之外，均一概得享有之實質豁免保障，將會因該個人於武裝衝突持續狀態進行中之某一時點「曾經」參與某一敵對行動而告消滅。[91]

承上述，以《日內瓦公約》爲首之國際人道規範，於多數情形下充其量僅能爲目的性框架規定，無法就各國於從事敵對行爲中應採取之手段或措施、平民是否直接從事敵對行動等諸多細節鉅細靡遺窮盡規範，此等事項與權衡民用物體之軍事用途無異，均需高度複雜之分析。換言之，由《第一附加議定書》第 41 條第 2 項第 2 款可知，在「區分原則」要求下，針對自主性武器系統須能判定特定之攻擊目標爲平民或戰鬥員一節，目標客體之形狀或大小固然可以數據量化，然區別「平民」與「直接參與敵對行動之平民」、區別「戰鬥員」或「投降之戰鬥員」，抑或區別「一般民用物體」與「具有軍事用途之民用物體」，量化數據所能發揮之功能甚爲有限，必須仰賴人類於採取敵

89. Kristen E. Eichensehr, "On Target? The Israeli Supreme Court and the Expansion of Targeted Killings," *The Yale Law Journal*, Vol. 116, No. 8, June 2007, pp. 1873-1881.

90. Kristen E. Eichensehr, "On Target? The Israeli Supreme Court and the Expansion of Targeted Killings," pp. 1873-1881.

91. Kristen E. Eichensehr, "On Target? The Israeli Supreme Court and the Expansion of Targeted Killings," pp. 1873-1881.

意行動瞬間，解讀敵方外在舉動所隱含意圖，並綜合衡酌現況始能爲之。此一判斷，尤重經驗嫻熟之人類感官及處理能力，職是之故，巨量數據所驅動與預判之自主性武器，能否掌握非量化分析，非無再商榷之餘地。

更爲迫切的問題，毋寧係自主性武器既依賴人類提供數據並以程式預設執行任務，除非自主性武器於本質上無法接受或依照人類提供正確數據以執行任務，否則其於軍事目標上之判斷失誤，是否及如何對於幕後人爲設計者之錯誤，予以課責或歸咎，容有疑義。[92] 進一步言，規範上倘若意圖以《第一附加議定書》第 52 條及第 54 條作爲規範依據，在交戰情境下權衡判定民用物體是否作爲軍事用途，俾決定是否得以攻擊，所需分析及注意之複雜程度，遠高於單純指定軍事目標或戰鬥員爲攻擊對象。

（二）比例原則

揆諸國際人道法所設立之機制標準，比例原則 (the principle of proportionality) 強調國家武力行使時須衡酌軍事目的，具體盱衡系爭攻擊行爲所致平民死傷或民用物之受損，確保目的與手段間具備相稱性與衡平性，抑或應確保兩者間不得有明顯失衡。[93] 此原則於國際人

92. Mary Manjikian, "Conflicts, Cohesion, and Comrades in Arms: Social Implications of Robotics in the Military," in Ryan Kiggins, ed., *The Political Economy of Robots: Prospects for Prosperity and Peace in the Automated 21st Century* (Berlin: Springer, 2017), pp. 249-269.

93. Ian Henderson & Kate Reece, "Proportionality Under International Humanitarian Law: The Reasonable Military Commander Standard and Reverberating Effects," *Vanderbilt Journal of Transnational Law*, Vol. 51, No. 3, May 2018, pp. 835-836; Judith Gail Gardam, "Proportionality and Force in International Law," *American Journal of International Law*, Vol. 87, No. 3, July 1993, pp. 391-413。國內文獻請見趙國材，〈論國際人道法適用於內戰之發展〉，頁 122-153。

道法作爲基礎規範指針，具舉足輕重之地位。

該原則體現於國際人道法禁止在武裝衝突期間爲「不加區別（不分皂白）之攻擊」(indiscriminate attack)，而所謂「不加區別之攻擊」係規定於《第一附加議定書》第 51 條第 5 項 b 款：「（所謂不加區別攻擊係指）可能附帶使平民生命損失、使平民受傷害、使民用物體受損害，或三種情形均存在之情狀，且與預期的具體和直接軍事利益相權衡，會造成不合乎法益相稱性的損害之攻擊。」而對於此種攻擊，該議定書中之第 57 條第 2 項 a 款 iii 即加以禁止：「不得實施可能附帶使平民生命受損失、使平民受傷害、使民用物體受損害，或三種情形均存在且與預期的具體和直接軍事利益相互權衡，會造成過分損害之攻擊。」上述二條文，即是將習慣國際法中之比例原則予以條約化。值得注意者，即便於國際人道法層次，比例原則仍與於一般法律領域中適用類同基準，重點繫諸於「可能造成附帶損害」與「欲達成之軍事利益」之權衡結果是否符合法益相稱性、戰鬥員是否選擇適當且必要手段，並已達正當武力行使之發動門檻，乃至是否能比較攻擊合法目標所致軍事利益與平民附帶損害間具備「法益相稱性」等規範性之法原則。

國際軍事法權威施密特 (Michael N. Schmitt) 指出，附帶損害之評估辦法既能藉由公式計算，且對於尚未發生之攻擊損害及期待利益，以客觀條件及各項科學標準予以權衡比較，自然亦能藉由預測型演算法預先設定於自主系統，並能與人類計算獲致相同分析結果；自主性武器系統能勝任運算任務，自屬無疑，惟套用公式運算附帶傷亡之數據及可能性，並不等同於作戰方法符合比例原則。[94] 此處須注意者，施密特認爲由演算法所主導之「附帶損害預測模型」(Collateral

94. Michael N. Schmitt & Jeffrey S. Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict," *Harvard National Security Journal*, Vol. 4, No. 2, February 2013, pp. 253-257.

Damage Estimation Methodology) 全然取代比例原則的適用，將會導致此一原則之質變。即使其並不因此一概否認演算法在武裝衝突脈絡下可勝任某些運算任務，惟必須強調的是，比例原則所訴求法益相稱性之背後蘊含道德價值秩序遠超過量化數據之權衡；同理可證，一臺機器亦絕無可能移植複製人類決定發動攻擊與否的行爲決策是否合乎比例原則之主觀判斷與心理歷程。[95]

（三）攻擊預防原則

攻擊預防原則 (precautions in attack) 亦屬評估軍事行動適法性要件之一。揆諸《第一附加議定書》第 57 條第 1 項即指出：「軍事行動之執行，自始至終皆應注意避免波及平民或民用物體。」而同條第 2 項更明定：「指揮官計畫或決定攻擊時，應盡一切可能，確定目標屬軍事性質而非屬平民；應採取一切可能預防措施及選擇適當之作戰方法手段，以避免或減少在任何期間之附帶傷亡利益；禁止發動可能會造成平民附帶傷亡超過可明確預期之直接軍事利益的攻擊。而攻擊若一旦發動，如發現非屬軍事目標，或預期可能造成過分之平民附帶傷亡損害，即應取消或終止攻擊。」武器系統於執行任務之際，能否意識到可能波及之平民或民用物體並加以迴避？其執行任務所採取之手段及其選擇手段背後之運算與決策過程，能否通盤納入所有可能之預防措施，以避免或降低可能造成之傷亡或損害？一旦發起攻擊，如攻擊標的物爲非軍事目標，是否能即時調整其辨識結果，進而中止攻擊行動？對於自主性武器系統面對軍事利益之主觀判斷已然難以期待，而於權衡後選擇作戰方法，以避免或最小化附帶傷亡，更有賴戰

95. Michael N. Schmitt & Jeffrey S. Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict," pp. 254-256; Marco Sassóli, "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified," *International Law Studies*, Vol. 90, April 2014, pp. 331-332.

場實況之現場判斷。面對瞬息萬變之戰場實況，自主性武器系統僅以事先決定的運算數值並依據公式計算，顯然在本質上無法勝任對於採取預防措施之法規範要求，亦即無法避免或最小化附帶傷亡。

換言之，基於演算法之自主性武器系統，其研發須予以審慎設計並深入測試、應對，俾使風險最小化。而對於此種規範上之法益相稱性判斷，必須持續來回擺盪於「人工智慧自主性武器系統對於社會帶來的益處」和「可能造成之風險甚至危害」權衡之中，且須回返「義戰」(just war) 理論，[96] 訴諸武力理由與軍事干預的手段正當性，審慎「評估武力行使所招致之惡，是否能爲武力行使所能達成目的之利，而加以平衡」；[97] 兩者間之相互競合，基於禁止殺人誡命所衍生而來之不可量化原則及人命之不可衡量性，不宜僅以量化方式做出高低不同之評價。而此點於武裝衝突現場下達判斷之際尤爲重要，指揮者應遵循以國際人道法爲主導性之價值，就軍事必要性、軍事優勢、平民

96.「義戰」源自基督教傳統之自然法觀念，彰顯武力行使及戰爭行爲應服膺上帝之神聖旨意。公元 4 世紀時，奧古斯丁 (St. Augustine) 曾述及：“Just wars are usually defined as those which avenge injuries, when the nation or city against which warlike action is to be directed has neglected either to punish wrongs committed by its own citizens or to restore what has been unjustly taken by it.....” 主張義戰之發動須是爲處罰從事不法行爲，且須具備正當及正義理由始能進行戰爭行爲。有「國際法之父」之稱的格勞秀斯則將戰爭行爲予以理論化，認爲義戰須爲出於自衛、保護財產及懲罰不法行爲等，始具正當性。國內相關文獻請見楊永明，〈武力威脅與國際法：中共武力威脅臺灣之國際法分析〉，《臺大法學論叢》，第 30 卷第 5 期，2001 年 9 月，頁 33-55；田力品，〈由比例原則檢驗武裝衝突理由及手段之正當性〉，《軍法專刊》，第 59 卷第 5 期，2013 年 10 月，頁 181-182。

97. 田力品，〈由比例原則檢驗武裝衝突理由及手段之正當性〉，頁 180-192；田力品，〈國際人道法對武器使用之限制〉，《軍法專刊》，第 59 卷第 3 期，2019 年 6 月，頁 121-144。

附帶性損害等法律規範意涵，進行個案性之裁量評估。[98]

上述見解亦獲致其他國際法學者之共鳴，蓋武力行使之要件涵攝，更加倚賴戰場實況和相關個案脈絡之掌握，以便即時進行適度調整，故須因地制宜、依賴人類之經驗法則。AI 演算法固可提升計算附帶傷亡之精確性，然軍事利益之判斷與決定，顯然無法由實驗室工程師預先賦予一定數值，而須仰賴戰鬥員盱衡戰場當下種種情形方能做出合乎實況之判斷。[99]

二、規範性省思之二：人機協作之軍事決策

（一）人機協作之基本概念

誠如前述，軍事決策涉及高度不確定性及龐雜資訊，復基於近年 AI 技術之迅即發展，預測型演算法被用以協助軍事決策的場合與日俱增。如前所述，迄今尚未發展出嚴格意義下可在武裝衝突情境中獨占軍事決策自主能力之武器系統，目前占據大宗之弱 AI 及 AI 子模式自主性武器系統仍須協同人類相互合作，方能達成任務。爲發展 AI 與人類間有效分工模式，人類須對 AI 具相當程度之「信任校正」(trust calibration)，並時時評估 AI 達成目標之性能及強、弱項等數值，俾適時針對雙方協作模式進行必要之調整。基此，以范登博斯 (Karel van den Bosch) 等爲代表之學者咸認，人類須與 AI 密切聯繫互動，始得互相合作依賴。而透過連續且循環不間斷之人類與 AI「協作行動」(joint activity) 及其回饋，人類方能對 AI 系統之能力與限制

98. Ashley Deeks, Noam Lubell, & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States," *Journal of National Security Law & Policy*, Vol. 10, No. 1, February 2019, pp. 14-15.

99. Marco Sassóli, "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified," pp. 323-325.

具備更完善之掌握與瞭解。[100]

范登博斯等學者將目光朝向人機協作之制度設計，另關於人機協作模式中之信任校正、其程度（或量值）設定是否恰當、是否及如何調整等問題意識，亦日漸吸引其他學者之關注。

（二）人機協作之流程步驟

如圖 2 所示，范登博斯等學者根據 AI 可解釋性程度之高低，將人機協作劃分爲單向操作階段 (uni-directional)、雙向操作階段 (bi-directional) 及人機操作階段 (human-AI collaboration) 等三個階段。[101] 首先，第一階段屬「單向操作階段」，其目的在於令人類使用者更加瞭解 AI 系統之運行與決策過程。對人類而言，此一階段之 AI 仍處於「黑箱」之中，其透明度尚待提升，俾益更有效之利用；第二階段則進入「雙向操作階段」，於此階段，人類不僅可主動向系統要求資訊解釋，系統本身亦可提出解釋或建議（如偵測到錯誤時）；換言之，AI 系統已具備分析情況之能力，並能爲使用者提供可理解之解釋；最後，由第二階段發展至第三階段則正式邁向「人機操作階段」，顯示人類與 AI 終能眞正達到協作，雙方對彼此狀態皆具適當瞭解，並能隨時調整適應；遇到緊急狀況時，亦能重新分配工作，並透過一次次互動產生之回饋而日漸改善，[102] 反映出學理就實踐上述《特定常規武器公約》中「人機關係」程序性原則所爲之規範性與技術性闡釋。

---

100. Karel van den Bosch & Adelbert Bronkhorst, "Human-AI Cooperation to Benefit Military Decision Making," May 25, 2018, *NATO Science & Technology Organization*, <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-S3-1.pdf>.

101. Karel van den Bosch & Adelbert Bronkhorst, "Human-AI Cooperation to Benefit Military Decision Making."

102. Karel van den Bosch & Adelbert Bronkhorst, "Human-AI Cooperation to Benefit Military Decision Making."

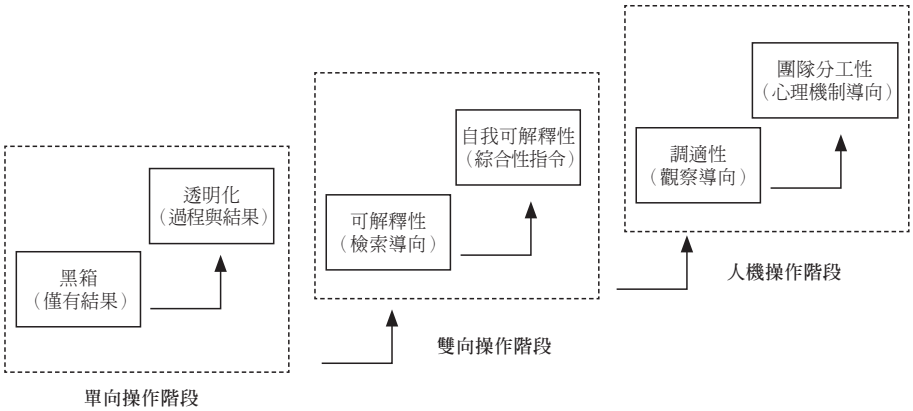

圖 2 人機協作之流程步驟

資料來源：Karel van den Bosch & Adelbert Bronkhorst, “Human-AI Cooperation to Benefit Military Decision Making”。

三、規範性省思之三：使用 AI 武器方法之限制

（一）馬爾頓條款

承前所述，《特定常規武器公約》中之強行規定派，主張以硬法形式強行介入自主性武器之發展，其見解著眼於如何避免「完全自主性武器」之形成及其可能產生之負面影響，更是攸關國際人道法之旨趣甚鉅。

此等規範進路強調自主性武器研發和方法之從嚴限制，「人道原則」(principle of humanity) 及「公眾良心之指責」(dictates of public conscience)——亦即所謂「馬爾頓條款」(Martens Clause)——如何適用於 AI 軍武科技，不僅牽涉制度選擇，更是牽動未來國際規範秩序對新武器系統擴散之關鍵問題。在軍事電腦決策之審查確有其必要性之思潮影響下，紅十字國際委員會於 2006 年出版《爲符合第一附加議定書第 36 條之武器、戰場實踐、方法、措施之法律審查準則》（*A Guide to the Legal Review of Weapons, Means and Methods of Warfare:*

Measures to Implement Article 36 of Additional Protocol I of 1977，以下簡稱 ICRC Guide），納入《第一附加議定書》第 36 條規定之新興武器審查義務。ICRC Guide 指出，若武器之使用與人道原則及公眾良知原則相衝突，即便該武器本身不受任何特定國際法規則禁止或限制，仍應遵循人道原則及公眾良知原則。[103] 新興軍武科技在研發上之人道與公眾良知，緣於《1899 年海牙第二公約》(*Hague Convention (II) of 1899*) 前言所載：「於有更完備之戰爭法規範形成前，各締約國均承認以下權宜之聲明，在公約規則未規範之情況下，平民及交戰方仍受國際法原則之保護及規範，此乃基於該原則係源自文明社會之習慣、人道法及公共良知之要求。」固然「馬爾頓條款」本身並未明文指涉特定武器之使用方法，然迄今仍爲國際人道法適用科技中立原則最古典之示範，該等規定藉由公共良知爲規範要素，要求軍備競賽中無論武器之技術特徵如何演進，皆應受限於人道原則及公眾良心原則。[104]

另武裝衝突中交戰方之作戰方法及手段之規範遵循，揆諸《1868 年聖彼得堡宣言》(*St. Petersburg Declaration of 1868*) [105] 序言所載：

---

103. Kathleen Lawand, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977* (Geneva: International Committee of the Red Cross, 2006), pp. 15-16.

104. 黃居正，〈與人工智慧相關的國際法議題—從國際人道法到生命體法〉，劉靜怡主編，《人工智慧相關法律議題芻議》（臺北：元照出版社，2018 年），頁 218-238。

105. 俄國沙皇亞歷山大二世 (Alexander II of Russia) 於 1868 年在聖彼得堡召集人類史上首次爲討論戰爭法而舉行之國際會議，會後通過協議即爲《1868 年聖彼得堡宣言》；該宣言訴求應禁止不必要加重傷者痛苦之武器使用與開發，使用該等武器亦屬違反人道原則。參見姜皇池，《國際公法導論》（臺北：新學林，2013 年），頁 759。

「人類文明之進步應使戰爭所致災難盡可能減輕；削弱敵方軍事力量爲交戰國於戰爭時致力完成之唯一合法目的；爲達到此一目的，盡可能使最多數人失去戰鬥能力，即爲已足；但凡對於失去戰鬥能力之人使用武力，造成不必要之痛苦、必然致死或突然加劇痛苦之武器，均屬違反國際人道法則。」此即現今國際人道法中所謂「禁止使用不必要痛苦或過分傷害」核心價值之體現。此一規範之遵循亦可望針對國際上尚未成熟開發的軍事科技，做出超前之預防性規範作用，同時亦宜作爲因應迅即發展之軍備競逐及高度不可預見性之演算法開發是否合法之檢驗標準，並限縮軍武科技之研發與使用均不得違反人命不可量化及生物體脆弱性之倫理要求。[106]

（二）軍事演算法之審查義務

1.《第一附加議定書》之新武器審查義務

法律審查制度之目的在於避免部署爲特定國際法律規則所禁止或限制之武器，或者無法遵守規範敵對行爲主要規則之武器。根據《第一附加議定書》第 36 條規定：「在研究、開發、獲取或採用新式武器、作戰手段或方法時，締約方有義務確定其使用在某些或所有情況下是否會受到本議定書或適用於該締約方的其他國際法規則之禁止。」作爲審查軍事演算法之國際法淵源，此一規定係由《第一附加議定書》第 82 條所定之「各締約國及衝突各方之法律顧問參與義務」延伸而來。復觀諸《1868 年聖彼得堡宣言》序言所載：「人類文明之進步應使戰爭所致災難盡可能減輕；削弱敵方軍事力量爲交戰國於戰爭時致力完成之唯一合法目的；爲達到此一目的，盡可能使最多數人失去戰鬥能力，即爲已足；但凡對於失去戰鬥能力之人使用武力，造成不必要之加劇其之痛苦、必然致死或突然加劇痛苦之武器，均屬違反國際人道法則。」此一原則亦經嗣後《1899 年海牙第二公約》

106. 黃居正，〈與人工智慧相關的國際法議題－從國際人道法到生命體法〉，頁 231、236-237。

與《1907 年海牙第四公約》(*Hague Convention (IV) of 1907*)「陸戰法規及慣例公約附件」(Convention Respecting the Laws and Customs of War on Land) 第 23 條肯定。[107] 殆至 1949 年《日內瓦公約》第 4 條揭櫫：「權利交戰方選擇及採用戰爭手段及方法之數量『並非無上限』。」亦彰顯締約國應負評估正處於開發階段及既已使用武器合法性之責任。爾後 2003 年於紅十字會與紅新月會國際聯合會第二十八屆國際會議中，再次肯認：「新的武器、手段和方法戰爭『應該接受嚴格和多學科之審查』，且此類審查『應涉及包括軍事、法律、環境和健康相關之綜合評估。』」[108]

前揭美國國防部《3000.09 指令》第 4 條則就自主性武器功能及其相關測試認證程序明文規範，藉以控制可能衍生之系統性風險。根據該條第 a(2)(a) 項規定，設計人員於開發武器系統功能之際，必須確保對下列兩面向予以穩定控管：第一，確保軍事指揮官及相關操作人員對武器使用仍具一定程度之決策權；其次，須整備防止竄改 (anti-tamper mechanisms) 之安全機制及便利人爲控制之介面。[109]《3000.09 指令》第 4a(1) 項復明定，武器系統即便完成設計，仍須循序定時接受兩項考評認證——設計程序及有效性認證 (verification and validation)、測試與評估 (test and evaluation) 程序認證——藉以控管可能產生之系統性風險，且於系統設計經修改之際，負擔再次接受認證之義務。[110]

---

107. 田力品，〈由比例原則檢驗武裝衝突理由及手段之正當性〉，頁 180-192；田力品，〈國際人道法對武器使用之限制〉，頁 121-144。

108. International Committee of the Red Cross, *Declaration Agenda for Humanitarian Action Resolutions* (Geneva: International Committee of the Red Cross, 2003), pp. 2-6.

109. U.S. Department of Defense Directive, *Department of Defense Directive 3000.09: Autonomy in Weapon Systems (2012)*, pp. 1-15.

110. U.S. Department of Defense Directive, *Department of Defense Directive*

儘管《第一附加議定書》第 36 條是否已然成爲國際習慣法容有爭議，惟美國和以色列等非該附加議定書之簽署國仍已將軍事武器法律審查制度化及內國法化。美國國防部發布之《測試與評估管理指引（第六版）》(*Test and Evaluation Management Guide*) 指出：「測試和評估義務，在嚴謹性上，應達致足適提供決策者基本數據、評估技術性能參數的實現效益，並確定系統是否在預期用途中運行有效、合適、可存續性和安全的實施強度。測試和評估的執行，與建模和仿眞相結合的測試和評估的實施應促進學習、評估技術成熟度和可交互運作性、促進與實戰部隊的整合，並根據威脅評估判讀敵對方的軍事實力。」[111]

2. 紅十字國際委員會發布之新武器審查準則

依循 ICRC Guide 之說明：「審查人員首先應評估是否存在禁止審查中之武器或其某些用途之具體條約或習慣法規定。優先性上應先審酌系爭武器是否與具體禁令或限制明顯相牴觸之情事。倘若未與各該特定國際條約相牴觸，續而應依循國際人道法基本原則對武器進行評估。」[112] 該報告指出下列三大準則係爲構成新武器法律審查之基

3000.09: Autonomy in Weapon Systems (2012), pp. 1-15.

111. U.S. Department of Defense, *Test and Evaluation Management Guide (Sixth Edition)* (Washington, D.C.: U.S. Department of Defense, 2012), pp. 1-250.

112. ICRC Guide 於此例示臚列之國際條約略以：1907 年《關於敷設自動觸發水雷公約》（《海牙第八公約》）(*The Laying of Automatic Submarine Contact Mines*)(*Hague Convention VIII*)、1925 年《日內瓦議定書：禁止於戰爭中使用窒息性、毒性或其他氣體和細菌作戰方法議定書》(*Geneva Protocol: Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or other Gases, and of Bacteriological Methods of Warfare*)、1975 年《禁止細菌（生物）及毒素武器之發展、生產及儲存以及銷毀此類武器公約》(*Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction*)、1976 年《禁止爲軍事或任何其他敵

礎：(1) 禁止使用《第一附加議定書》第 35(2) 條及《海牙公約》第 23(e) 條所規定之「具有造成過分傷害或不必要痛苦性質之彈藥、材料及戰爭方法」；(2) 禁止使用性質上不能區分軍事目標及民用物體之武器，或未能按照《第一附加議定書》第 48 條及第 51 條所定得以貫徹區分原則之武器；(3) 根據《第一附加議定書》第 35 條第 3 款及第 55 條規定，禁止使用對自然環境造成廣泛、長期和嚴重破壞之武器。[113]

3. 技術審查與法律審查之匯流？

儘管挑戰嚴峻，學者維斯特納 (Tobias Vestner) 與羅西 (Altea Rossi) 於評估可解釋性 AI (explainable AI) [114] 適用於軍事武器演算法之可行性時，指出一個甚爲關鍵的觀察：可解釋性 AI 之監管舉措若欲在 AI 新武器上發揮加強演算法透明性之課責效果，則必須正視嗣後技術應用和法律評價可能產生難以截然二分之融合併行現象。[115]

對目的使用改變環境技術公約》(*Convention on the Prohibition of Military or any other Hostile Use of Environmental Modification Techniques*) 及 1980 年《特定常規武器公約》等。請見 Kathleen Lawand, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, pp. 10-11。

113. Kathleen Lawand, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, pp. 15-17.

114. 國外法學文獻中，對可解釋性 AI 之定義性闡述略以：「被視爲理解演算法內部運作提供洞見，並得以提供對於人類而言具可理解性之近似估計值 (approximations)。」請見 Sandra Wachter, Brent Mittelstadt, & Chris Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology*, Vol. 31, No. 2, Spring 2018, p. 850。

115. Tobias Vestner & Altea Rossi, "Legal Reviews of War Algorithms," *International Law Studies*, Vol. 97, February 2021, pp. 541-548.

申言之，設計程序及有效性認證程序與法律審查傳統上係分開進行，此揆諸美國國防部指令「國防採購系統」(The Defense Acquisition System) 將技術和法律評估作爲武器系統採購之兩個不同步驟進行審查自明。[116] 武器測試及技術評估係提供評測武器性能之實證數據，而法律專家則另將其作爲法律審查基礎。審查之數據範圍涵蓋武器準確性、可靠性、性能及故障率等。

然而，對於以機器學習爲基礎之武器系統，維斯特納與羅西主張未來 AI 新武器之技術驗證程序應同時涵蓋「技術層面評估」及「法律評價」之雙重意涵。[117] 其原因在於，正如前文一再檢證，運用巨量數據或演算法之自主性武器與傳統武器之最大區別，在於前者係透過預先輸入程式與指令，決定人類擬授權武器執行攻擊之標的群，以及此項預先決定程式是否與日內瓦公約爲首之戰爭法行爲規範吻合。故可預期者，將來自主性武器於其設計階段，區分原則、比例原則及攻擊預防原則皆將被轉化爲技術參數並嵌入至系統中。據此，維斯特納與羅西論證道：一旦系統習得法規範準據，遵循此些標準即「等同」一項技術任務。關於驗證過程——亦即確保系統符合開發者之概念描述和規範——是否符合現行交戰規則之規範性要求，將有極大可能納入技術評估之射程範圍，從而導致技術評估與法律審查兩者匯流混同。[118]

正因設計程序與有效性認證及測試與評估程序或將產生與法律審查合二爲一之趨同化，致使未來將法律知識與前揭設計程序與有效性認證及測試與評估程序過程等監管機制技術專長相結合的必要性隨

116. U.S. Department of Defense, *Department of Defense Directive 5000.01: The Defense Acquisition System (2020)*, September 9, 2020, pp. 1-17.

117. Tobias Vestner & Altea Rossi, "Legal Reviews of War Algorithms," pp. 541-548.

118. Tobias Vestner & Altea Rossi, "Legal Reviews of War Algorithms," pp. 547-548.

之升高。未來監管 AI 於大規模殺傷性武器之適法性監督，如欲發揮成效，亟須仰賴法律專家參與涉及數據篩選之程序與有效性認證程序（《第一附加議定書》第 36 條、第 82 條參照）。在此一程序中，技術專家及法律專家分而治之，前者掌理監督程序主要流程之踐行；後者則負責評估於此些階段所觀察到行爲及結果是否符合國際人道法之實定法規範。[119]

伍、結論

隨著預測型演算法興起，AI 自主性武器系統開發與應用相關爭議逐漸浮上檯面。本文旨在以國際人道法及其關聯法律淵源作爲論述開展之規範憑藉，探究 AI 自主性武器對國際法秩序肇致之挑戰與調適模式。本文發現，承襲義戰傳統變遷而來之區分原則、比例原則及攻擊預防原則，於檢驗 AI 自主性武器系統與現行國際戰爭法之規範扞格趨於明顯；而權衡 AI 新武器之衡量基點，已由過往學說與實務強調之自主性程度高低，逐漸向 AI 複合型武器之人機關係及人機協作等爭議傾斜。人類對於自主性武器系統之監控應持續存在，人類監管者亦須確信全階段武器系統均能遵守國際法及適用之交戰規則。而

119 Tobias Vestner & Altea Rossi, “Legal Reviews of War Algorithms,” pp. 547-548。值得注意者，從人爲監管的面向上，亦有另一派見解提出一種基於工程學、社會技術和治理視角的全面人類監控框架 (comprehensive human oversight)，用於控制自主性武器系統。渠等主張在技術、社會技術和治理層面對自主性武器系統展開全面的人類監督，從而確保其可控性和責任性。亦即以演算法開發進程的時間爲橫軸、系統開發的內部至外部發展出分別包括事前監督、回饋機制、持續控管及事後監督與審查等監督流程。囿於篇幅本文未及深究，具體監管框架請見 Ilse Verdiesen, Filippo Santoni de Sio, & Virginia Dignum, “Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight,” *Minds and Machines*, Vol. 31, No. 1, March 2021, pp. 150-153。

於複雜任務及有限自主性技術中，爲恪遵國際法義務，須於下述軍事行動三個階段中有「適當程度之人爲判斷」介入或干涉：一、計畫階段：當人類選擇使用何種自主性武器系統時；二、選擇後／攻擊前：須對個案任務中所分配之武器系統做出人爲注意層級之決定時；以及三、在攻擊前、中及後：須即時具體輸入人爲判斷時。

鑑於各國對「有意義之人爲控制」之解釋論齟齬叢生，就自主性武器系統之權責歸屬問題製造曖昧不明且難以嚴謹區隔之規範形成空間，各國以《特定常規武器公約》爲主戰場，秉持自身觀點之差異化解讀彰顯國家間之權力落差與對立齟齬，此說明兩項趨勢。第一，從國際安全以觀，制定 AI 國際武器條約抑或對軍事用途 AI 武器的立場設定，將逐漸被解讀爲何人掌握定義遊戲規則的能力，其就具有建立 (standard setting) 及議題設定 (agenda setting) 的政治權力博弈場域。換言之，在 AI 新武器冉冉崛起擴散過程中，如何透過參與《特定常規武器公約》這個全球最關鍵的國際規範治理平臺，創造並主導一系列的派生規則和相關派生附加議定書，不僅囿於在不同國家間進行協調或統一管制規則，毋寧更應被視爲一種科技競逐及國家戰略意涵的政治抉擇。值得強調者，如前所述，有別於其他國際規範往往彰顯參與國之間的權位不對等和連帶產出的不對稱談判權力，[120] 得益於《特定常規武器公約》的議事傳統上所蘊含的動態循環之對話機制，復以其成員組成結合如軍事人員、國際法專家、學術界人士及機器人研究者等多方利害關係人群體，就致命自主性武器系統的風險與管制，可望激發更爲活絡且平等之規範形成場域。然而這個動態而不同生成內涵的場域是否能重複不斷地取得國際肯認，藉以轉化爲強制性之國際遵循義務，甚至進一步將該公約強固爲未來全球制定 AI 新武器監管規則的威權性與正當性，仍有待未來研究進一步深入檢證。

120 曾雅眞，〈不對稱地位的法制化強固：NPT 建制核武國核保防特權的深化與影響〉，《政治學報》，第 72 期，2021 年 12 月，頁 29-30。

第二，本文亦再次證實 AI 軍事應用與既有國際法不相吻合且亟需調適此一事實。國際法有關武力行使之規範發展至今，似已形成法規範體系及軍事實際應用之矛盾衝突。此等應然和實然之扞格，未來武器一旦具備完全自主性，便意謂人類將無須、甚或無從再對軍事任務中之資料蒐集、分析回饋，乃至最終決策等程序，進行實質監控或干涉。自主性武器系統於整體運作上極易脫離人類控制，並專斷完成高階軍事任務之決策與執行；而其系統背後所涉重要情報（如機器學習過程、與各類軍事參數之監控蒐集程序等），恐怕連設計者亦難以完全掌握。本文亦發現，基於目前尚未發展出完全自主的武器系統，大部分弱 AI 及 AI 子模式自主性武器系統仍需與人類相互合作才能完成任務。人類需要對 AI 進行信任校正，不斷評估其性能和優缺點，並根據需要進行調整。范登博斯等學者認爲，人類需要與 AI 密切聯繫和互動，以便實現更好的協作。而此項人機協作的制度設計，將著重於信任校正和其程度的設定及如何進行調整。

除此之外，既有的國際法亦言及新興軍事科技在研發上之人道與公眾良知的重視。其中，《特定常規武器公約》等國際法律文件定有限制措施，以避免完全自主性武器的形成及其可能的負面影響。以《特定常規武器公約》之《第一附加議定書》規定爲例，其在被提出後即獲得大量支持，究其根本原因，乃是該限制並沒有對各締約方在武器使用上構成影響，其軍事利益並無實質上減損。然並非所有規定都能如此順利地被各國所接受，復又因公約條款欠缺強制性而有難以法制化遵循義務之憾。此外，《1868 年聖彼得堡宣言》強調在武裝衝突中，交戰方之作戰方法及手段之規範遵循，禁止使用不必要之痛苦、必然致死或突然加劇痛苦之武器。上開法律文件的制定和實施，體現了國際社會對人道和公眾良知的關切和重視。又，《1868 年聖彼得堡宣言》、《海牙公約》和《日內瓦公約》等國際法律文件也明確肯定了人道主義原則，即在戰爭中要盡可能減少對人類造成的傷害和痛苦，同時削弱敵方軍事力量是交戰國唯一合法的戰爭目的。

最後，《第一附加議定書》第 36 條所定之新武器審查義務及透明化要求，其法律審查制度的目的在於防止使用被禁止或限制的武器，或者使用違反敵對行爲規範的武器。這個規定乃係由《第一附加議定書》第 82 條所訂定之「各締約國及衝突各方之法律顧問參與義務」所延伸而來。此一審查機制之規範性要求，根據本文剖析，爲美國國防部和紅十字國際委員會所承襲沿用。其所涵蓋範疇包括設計程序與有效性認證及測試與評估，此一國際法淵源雖尚未達致習慣國際法或強行法之位階，但仍可望爲各國爭相競逐之新武器研發與擴散提供近似正當法律程序之框限性準據。

本文藉由提煉《特定常規武器公約》之歷屆談判歷程中各國針對自主性武器系統未來發展與潛在風險所提出之改革建議，將其透射至現有國際人道法及相關法律淵源，探討兩者如何調適兼容，進而獲致新生變遷動力，最後以主張「強化 AI 複合型武器之人機協作」，併同「貫徹《第一附加議定書》第 36 條所定之新武器審查義務」，作爲未來軍事演算法監管機制之規範性建議。惟須強調者，新興科技軍武的全球治理核心關懷，毋寧是如何讓道德與科技在天平上重新被適切評價，以確保 AI 軍武科技的發展符合人類價值觀並遵從國際法原則。鑑諸 AI 及其衍生運用，當前一般認爲仍係作爲一個在定義內涵與界線範疇上仍持續不斷變動而未臻明確的技術集合體，解釋論或立法論上均仍存在高度不確定，連帶致令國際上未能形成共識機制，乃一項極具新穎性和與未知程度甚高的規範秩序與規則體系。本文的付梓，一方面意味著 AI 全球監理相伴而生的必然宿命——規範性的監管步伐往往落後於技術發展之所謂「柯林格里奇困境」(Collingridge Dilemma)，[121] 但也正是因爲這種難以協調的困境，彰顯建構 AI 自主

121.「柯林格里奇困境」，係由柯林格里奇 (David Collingridge) 於 1980 年所提出，係從科技的社會控制的觀點強調前瞻技術之雙重約束困境。蓋技術革新前期的問題與風險往往難以預測，在無法獲得衍生風險及影

性武器系統風險之國際治理，需仰賴當前各國審酌考量一致性共識要素，國際社會亦需加強合作，制定適應性強且明確的規範標準誠屬刻不容緩的課題，本文即在於分析當前學術文獻，針對國際戰爭法需要適應新的戰爭威脅與攻伐型態一節，提出作者之觀點與發展建議。惟國家責任之歸責基準的釐清、AI 在軍事情報和戰略決策中引發之隱私爭議，本文囿於篇幅及確保本論文的論述緻密性，亦僅能有待日後研究另行處理析論。

收件：2022 年 5 月 23 日
修正：2023 年 10 月 16 日
採用：2023 年 12 月 3 日

響的必要資訊下，人類將動輒陷入「我們可以控制卻不知該控制什麼；當創新技術已在市場上占有穩固地位，在其所生影響隨著技術的發展而逐漸明朗時，我們知道該控制什麼卻已陷入難以控制之困境」。郭戎晉，〈論人工智慧技術應用、法律問題定位及監管立法趨勢—以美國實務發展爲核心〉，《成大法學》，第 39 期，2020 年 6 月，頁 173-235；David Collingridge, *The Social Control of Technology* (London: Cambridge University Press, 1980), pp. 1-200。

參考文獻

中文部分

專書

姜皇池，2013。《國際公法導論》。臺北：新學林。

專書論文

黃居正，2018。〈與人工智慧相關的國際法議題—從國際人道法到生命體法〉，劉靜怡主編，《人工智慧相關法律議題芻議》。臺北：元照出版社，頁 218-238。

期刊論文

田力品，2013/10。〈由比例原則檢驗武裝衝突理由及手段之正當性〉，《軍法專刊》，第 59 卷第 5 期，頁 180-192。

田力品，2019/6。〈國際人道法對武器使用之限制〉，《軍法專刊》，第 59 卷第 3 期，頁 121-144。

林韋仲、廖宗聖，2019/6。〈致命自主武器發展之國際法管制〉，《台灣國際法學刊》，第 15 卷第 2 期，頁 9-32。

林昕璇，2021/8。〈AI 自主性武器系統於國際法適用上之研析〉，《軍法專刊》，第 67 卷第 4 期，頁 20-44。

郭戎晉，2020/6。〈論人工智慧技術應用、法律問題定位及監管立法趨勢—以美國實務發展爲核心〉，《成大法學》，第 39 期，頁 173-235。

郭雪眞，2011/12。〈人道軍事干預與國際人道法：美國反恐戰正中關達那摩灣 (Guantanamo Bay) 被拘禁者釋憲案例分析〉，《復興崗學報》，第 101 期，頁 15-40。

陳建佑，2023/6。〈人工智慧監管法律—獨漏自主性武器之規

範？〉，《全國律師》，第 27 卷第 6 期，頁 37-50。
曾雅眞，2021/12。〈不對稱地位的法制化強固：NPT 建制核武國核保防特權的深化與影響〉，《政治學報》，第 72 期，頁 1-42。
楊永明，2001/9。〈武力威脅與國際法：中共武力威脅臺灣之國際法分析〉，《臺大法學論叢》，第 30 卷第 5 期，頁 33-55。
趙國材，2010/8。〈論國際人道法適用於內戰之發展〉，《軍法專刊》，第 56 卷第 4 期，頁 122-153。

學位論文

李鈺翎，2018。《從國際法論人工智能軍事武器之發展與挑戰》。臺北：東吳大學法學院碩士論文。

官方文件

中華人民共和國國務院，2017/7/20。〈國務院關於印發新一代人工智能發展規劃的通知〉，《國務院文件》，國發〔2017〕35 號。

英文部分

專書

Arend, Anthony Clark & Robert J. Beck, 1993. *International Law and the Use of Force*. London: Routledge.
Boulanin, V. & M. Verbruggen, 2017. *Mapping the Development of Autonomy in Weapon Systems*. Solna: SIPRI.
Collingridge, David, 1980. *The Social Control of Technology*. London: Cambridge University Press.
Gehring, T., 1994. *Dynamic International Regimes: Institutions for International Environmental Governance*. New York: Peter Lang.
Jose, B., 2018. *Norm Contestation: Insights into Non-conformity with Armed Conflict Norms*. Basingstoke: Springer.

Krishnan, Armin, 2009. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Farnham: Ashgate.

Lawand, Kathleen, 2006. *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*. Geneva: International Committee of the Red Cross.

Müller, H. & C. Wunderlich, 2013. *Norms Dynamics in Multilateral Arms Control*. Athens: The University of Georgia Press.

Wiener, A., 2014. *A Theory of Contestation*. Basingstoke: Springer.

專書論文

Barbé, Esther & Diego Badell, 2019. "The European Union and Lethal Autonomous Weapons Systems: United in Diversity," in Elisabeth Johansson-Nogués, Martijn C. Vlaskamp, & Esther Barbé, eds., *European Union Contested Foreign Policy in a New Global Context*. Berlin: Springer. pp. 133-152.

Chengeta, Thompson, 2022. "Is the Convention on Conventional Weapons the Appropriate Framework to Produce a New Law on Autonomous Weapon Systems?" in Frans Viljoen, Charles Fombad, Dire Tladi, Ann Skelton, & Magnus Killander, eds., *A Life Interrupted: Essays in Honour of the Lives and Legacies of Christof Heyns*. South Africa: Pretoria University Law Press. pp. 379-397.

Manjikian, Mary, 2017. "Conflicts, Cohesion, and Comrades in Arms: Social Implications of Robotics in the Military," in Ryan Kiggins, ed., *The Political Economy of Robots: Prospects for Prosperity and Peace in the Automated 21st Century*. Berlin: Springer. pp. 249-269.

期刊論文

Abbott, Kenneth W. & D. Snidal, 2000/Summer. "Hard and Soft Law in International Governance," *International Organization*, Vol. 54, No. 3, pp. 421-456.

Altmann, Jürgen & Frank Sauer, 2017/9. "Autonomous Weapon Systems and Strategic Stability," *Survival*, Vol. 59, No. 5, pp. 117-142.

Bächle, Thomas Christian & Jascha Bareis, 2022/12. "Correction: Autonomous Weapons as a Geopolitical Signifier in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies," *European Journal of Futures Research*, Vol. 10, No. 1, pp. 1-18.

Berman, Emily, 2018/1. "A Government of Laws and Not of Machine," *Boston University Law Review*, Vol. 98, pp. 1277-1355.

Blanchard, Alexander & Mariarosaria Taddeo, 2022/5. "Jus in Bello Necessity, The Requirement of Minimal Force, and Autonomous Weapons Systems," *Journal of Military Ethics*, Vol. 21, No. 3-4, pp. 295-298.

Crootof, Rebecca, 2015/8. "The Killer Robots Are Here: Legal and Policy Implications," *Cardozo Law Review*, Vol. 36, pp. 1837-1915.

Crootof, Rebecca, 2016/5. "War Torts: Accountability for Autonomous Weapons," *University of Pennsylvania Law Review*, Vol. 164, No. 6, pp. 1347-1402.

Deeks, Ashley, 2018/3. "Predicting Enemies," *Virginia Law Review*, Vol. 104, No. 8, pp. 1529-1592.

Deeks, Ashley, Noam Lubell, & Daragh Murray, 2019/2. "Machine Learning, Artificial Intelligence, and the Use of Force by States,"

Journal of National Security Law & Policy, Vol. 10, No. 1, pp. 1-25.

Eichensehr, Kristen E., 2007/6. "On Target? The Israeli Supreme Court and the Expansion of Targeted Killings," *The Yale Law Journal*, Vol. 116, No. 8, pp. 1873-1881.

Ekelhof, Merel, 2019/3. "Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation," *Global Policy*, Vol. 10, No. 3, pp. 343-348.

Ferguson, Andrew Guthrie, 2015/1. "Big Data and Predictive Reasonable Suspicion," *University of Pennsylvania Law Review*, Vol. 163, No. 2, pp. 327-410.

Gardam, Judith Gail, 1993/7. "Proportionality and Force in International Law," *American Journal of International Law*, Vol. 87, No. 3, pp. 391-413.

Geist, Edward Moore, 2016/8. "It's Already Too Late to Stop the AI Arms Race—We Must Manage It Instead," *Bulletin of The Atomic Scientists*, Vol. 72, No. 5, pp. 318-321.

Haas, Michael Carl & Sophie-Charlotte Fischer, 2017/8. "The Evolution of Targeted Killing Practices: Autonomous Weapons, Future Conflict, and the International Order," *Contemporary Security Policy*, Vol. 38, No. 2, pp. 281-306.

Haner, Justin & Denise Garcia, 2019/9. "The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development," *Global Policy*, Vol. 10, No. 3, pp. 331-337.

Henderson, Ian & Kate Reece, 2018/5. "Proportionality Under International Humanitarian Law: The Reasonable Military Commander Standard and Reverberating Effects," *Vanderbilt Journal of Transnational Law*, Vol. 51, No. 3, pp. 835-855.

Johnson, Aaron M. & Sidney Axinn, 2013/8. "The Morality of Autonomous Robots," *Journal of Military Ethics*, Vol. 12, No. 2, pp. 129-141.

Korać, Srđan T., 2018/1. "Depersonalisation of Killing: Towards a 21st Century Use of Force 'Beyond Good and Evil'?" *Philosophy and Society*, Vol. 29, No. 1, pp. 49-64.

Lehr, David & Paul Ohm, 2017/12. "Playing with the Data: What Legal Scholars Should Learn About Machine Learning," *UC Davis Law Review*, Vol. 51, pp. 653-717.

Matthias, Andreas, 2004/9. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology*, Vol. 6, No. 3, pp. 175-183.

Mayer, Michael, 2015/7. "The New Killer Drones: Understanding the Strategic Implications of Next-generation Unmanned Combat Aerial Vehicles," *International Affairs*, Vol. 91, No. 4, pp. 765-780.

Mcclelland, J., 2005/7. "Conventional Weapons: A Cluster of Developments," *International & Comparative Law Quarterly*, Vol. 54, No. 3, pp. 755-767.

Mull, Nicholas W., 2018/2. "The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It," *Houston Journal of International Law*, Vol. 40, No. 2, pp. 461-530.

Risse, T., 2000/Winter. "'Let's Argue!': Communicative Action in World Politics," *International Organization*, Vol. 54, No. 1, pp. 1-39.

Sassóli, Marco, 2014/4. "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and

Legal Issues to be Clarified," *International Law Studies*, Vol. 90, pp. 307-340.

Saxon, Dan, 2014/Summer-Fall. "A Human Touch: Autonomous Weapons, Directive 3000.09, and the 'Appropriate Levels of Human Judgment over the Use of Force'," *Georgetown Journal of International Affairs*, Vol. 15, No. 2, pp. 100-109.

Schmitt, Michael N. & Jeffrey S. Thurnher, 2013/2. "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict," *Harvard National Security Journal*, Vol. 4, No. 2, pp. 231-281.

Schulzke, Marcus, 2013/6. "Autonomous Weapons and Distributed Responsibility," *Philosophy & Technology*, Vol. 26, No. 2, pp. 203-219.

Searle, John R., 1980/9, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, Vol. 3, No. 3, pp. 417-424.

Sparrow, Robert, 2016/3. "Robots and Respect: Assessing the Case Against Autonomous Weapon Systems," *Ethics & International Affairs*, Vol. 30, No. 1, pp. 93-116.

Verdiesen, Ilse, Filippo Santoni de Sio, & Virginia Dignum, 2021/3, "Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight," *Minds and Machines*, Vol. 31, No. 1, pp. 137-163.

Vestner, Tobias & Altea Rossi, 2021/2. "Legal Reviews of War Algorithms," *International Law Studies*, Vol. 97, pp. 508-555.

Wachter, Sandra, Brent Mittelstadt, & Chris Russell, 2018/Spring. "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology*, Vol. 31, No. 2, pp. 842-887.

研討會論文

Burton, J. & S. R. Soare, 2019/5/28-31. "Understanding the Strategic Implications of the Weaponization of Artificial Intelligence," paper presented at 11th International Conference on Cyber Conflict (CyCon). Tallinn: NATO Cooperative Cyber Defence Centre of Excellence. pp. 1-17.

官方文件

Congressional Research Service, 2023/2. "International Discussions Concerning Lethal Autonomous Weapon Systems," *Congressional Research Service*, No. IF11294, pp. 1-3.

Fogarty, William M., 1988. *Investigation Report－Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988*. Washington, D.C.: U.S. Department of Defense.

International Committee of the Red Cross, 2003. *Declaration Agenda for Humanitarian Action Resolutions*. Geneva: International Committee of the Red Cross.

International Committee of the Red Cross, 2016. *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*. Geneva: International Committee of the Red Cross.

Ministerie van Buitenlandse Zaken, 2017. *Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems in the Context of the Objectives and Purposes of the Convention*. Hague: Ministerie van Buitenlandse Zaken.

Sayler, Kelley M., 2020. *Artificial Intelligence and National Security*.

Washington, D.C.: Congressional Research Service.

United Nations, 1980. *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which may be deemed to be Excessively Injurious or to have Indiscriminate Effects*. Geneva: United Nations.

United Nations, 1996. *Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as Amended on 3 May 1996*. Geneva: United Nations.

United Nations, 2016. *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*. Geneva: United Nations.

United Nations, 2018. *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. Geneva: United Nations.

United Nations, 2019. *Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. Geneva: United Nations.

United Nations Group of Governmental Experts, 2018. *Emerging Commonalities, Conclusions and Recommendations*. Geneva: United Nations.

U.S. Department of Defense, 2012/11/21. *Department of Defense Directive 3000.09: Autonomy in Weapon Systems (2012)*, pp. 1-15.

U.S. Department of Defense, 2012/12. *Test and Evaluation Management Guide (Sixth Edition)*. Washington, D.C.: U.S. Department of Defense.

U.S. Department of Defense, 2017/11/10. *Department of Defense Directive 3000.09: Autonomy in Weapon Systems (2017)*, pp. 1-15.

U.S. Department of Defense, 2020/9/9. *Department of Defense*

Directive 5000.01: The Defense Acquisition System (2020), pp. 1-17.

雜誌

Acheson, Ray, 2018/8/30. "New Law Needed Now," *CCW Report*, Vol. 6, No. 9, August 30, 2018, pp. 1-7.

網際網路

2016/2/9. "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," *Future of Life*, <https://futureoflife.org/open-letter-autonomous-weapons>.

2020/6/2. "New SIPRI and ICRC Report Identifies Necessary Controls on Autonomous Weapons," *SIPRI*, <https://www.sipri.org/media/press-release/2020/new-sipri-and-icrc-report-identifies-necessary-controls-autonomous-weapons>.

Beerwerth, Peter, 2021/8/12. "National Statement by Germany Group of Governmental Experts on 'Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)'," *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2021/08/Germany.pdf>.

van den Bosch, Karel & Adelbert Bronkhorst, 2018/5/25. "Human-AI Cooperation to Benefit Military Decision Making," *NATO Science & Technology Organization*, <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-S3-1.pdf>.

Chavannes, Esther, Klaudia Klonowska, & Tim Sweijs, 2020/2/3. "Governing Autonomous Weapon Systems," *The Hague Centre for Strategic Studie*s, <https://hcss.nl/wp-content/uploads/2021/01/

HCSS-Governing-AWS-final.pdf>.

German Federal Foreign Office, 2020/6/24. "German Commentary on Operationalizing All Eleven Guiding Principles at a National Level as Requested by the Chair of the 2020 Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW)," *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2020/07/20200626-Germany.pdf>.

Groll, Elias, 2013/8/8. "The Sudden and Unexpected Return of the Drone War," *Foreign Policy*, <https://foreignpolicy.com/2013/08/08/the-sudden-and-unexpected-return-of-the-drone-war/>.

Hoffberger-Pippan, Elisabeth, Vanessa Vohs, & Paula Köhler, 2022/7. "Autonomous Weapons Systems: UN Expert Talks Facing Failure Time to Consider Alternative Formats," *SWP*, <https://www.swp-berlin.org/publications/products/comments/2022C43_AutonomousWeaponsSystems.pdf>.

Horowitz, Michael C. & Paul Scharre, 2015/3/13. "Meaningful Human Control in Weapon Systems: A Primer," *Center for a New America Security*, <https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf>.

Human Rights Watch, 2012. *Losing Humanity: The Case Against Killer Robots*, New York: Human Rights Watch, *Human Rights Watch*, <https://www.hrw.org/report/2012/11/19/losinghumanity/case-against-killer-robots>.

International Humanitarian Law Databases, 1980/10/10. "Technical Annex," *International Committee of the Red Cross*, <https://ihl-databases.icrc.org/en/ihl-treaties/ccw-protocol-ii-1980/technical-

annex?activeTab=undefined>.

Purkiss, Jessica & Jack Serle, 2017/1/17. "Obama's Covert Drone War in Numbers: Ten Times More Strikes Than Bush," *The Bureau of Investigative Journalis*, <https://www.thebureauinvestigates.com/stories/2017-01-17/obamas-covert-drone-war-in-numbers-ten-times-more-strikes-than-bush>.

Roff, Heather M. & Richard Moyes, 2016/4/11-15. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons," *Article 36*, <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>.

Rosert, E., 2017/1/1. "How to Regulate Autonomous Weapons: Steps to Codify Meaningful Human Control as a Principle of International Humanitarian Law," *JSTOR*, <https://www.jstor.org/stable/resrep14276?seq=1>.

The Russian Federation, 2021/6. "Considerations for the Report of the Group of Governmental Experts of the High Contracting Parties to the Convention on Certain Conventional Weapons on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems on the Outcomes of the Work Undertaken in 2017-2021," *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2021/06/Russian-Federation_ENG1.pdf>.

Sagramsingh, Raine, 2019/4. "Lethal Autonomous Weapons Systems: Artificial Intelligence and Autonomy," *WISE*, <https://wise-intern.org/wpcontent/uploads/2019/04/Raine_S_-FinalPaper.pdf>.

United Nations, 1996/5/3. "Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as Amended on 3 May 1996 (Protocol II to the 1980 Convention as amended on 3 May 1996)," *United Nations*, <https://www.un.org/en/

genocideprevention/documents/atrocity-crimes/Doc.40_CCW%20P-II%20as%20amended.pdf>.

United Nations, 2018/4/11. "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects," *United Nations*, <https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/CCW_GGE.1_2018_WP.7.pdf>.

United Nations Office for Disarmament Affairs, 2020/9. "Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS)," *United Nations Office for Disarmament Affairs*, <https://documents.unoda.org/wp-content/uploads/2020/09/GGE20200901-Austria-Belgium-Brazil-Chile-Ireland-Germany-Luxembourg-Mexico-and-New-Zealand.pdf>.

International Norms and Regulations for AI Autonomous Weapons Systems: Reflections and Prospects

Hsin-hsuan Lin

(Assistant Professor, Department of Political Science,
National Cheng Kung University)

Abstract

With the rise of big data-based predictive algorithms, controversies have arisen over the development of AI autonomous weapons systems, which have emerged as a new focal point in the study of international law. This paper examines how these new systems are and ought to be applicable to the international legal order on the basis of international humanitarian law as well as other relevant laws. Focusing on the texts of international law, this paper compares how autonomous weapons system is defined by the U.S. Department of Defense, China, Germany, International Committee of the Red Cross and Human Rights Watch according to the degree of autonomy and intellectual decision-making loop. This paper then unpacks the process of negotiations and milestones of the Convention on Certain Conventional Weapons. It demonstrates that several proposals put forth by the contracting parties to the Convention have reached an impasse in negotiations, making it difficult to adequately address the potential risks posed by autonomous weapons systems. Consequently, this paper proposes several regulatory insights regarding the control over military algorithms by implementing

meaningful human control and accountability mechanisms.

Keywords: Autonomous Weapons Systems, AI, Convention on Certain Conventional Weapons, Additional Protocol I to the Geneva Conventions, Inspection of Weapon-related Algorithms