人工智能發展帶來的國際法問題及其應對

鹿 璐

(深圳職業技術大學)

摘 要：人工智能技術的快速發展在推動社會變革的同時，也對現有國際法律制度提出了新的挑戰。除國際知識產權法之外，它對國際法中的國家責任原則、國家主權以及國際人道法、國際人權法等都產生了重大影響。因而有必要為人工智能技術的發展進行國際法規制。國家之間應當加強合作，通過簽訂多邊條約與行業規則、國內立法、國際示範法之間的相互影響和相互促進，從而形成相對全面、比較理想的一套國際法律規範，最終實現人工智能在可控制範圍內的安全發展。

關鍵字：人工智能；法律規制；多邊條約；行業規則

The Challenges of Artificial Intelligence Development on International Law and Response

Abstract: The rapid development of Artificial Intelligence (AI) technology has posed new challenges to the existing international legal system while promoting social transformation. Apart from international intellectual property law, it also has significant implications on the principles of state responsibility, national sovereignty, international humanitarian law and international human rights law in international law. Therefore, it is necessary to regulate the development of AI technology through international law. Countries should strengthen cooperation by signing of multilateral treaties and industry rules, enacting domestic legislation, developing international model law, and promoting mutual influence, in order to form a relatively comprehensive and ideal set of international legal norms, and ultimately realize the safe and controlled development of artificial intelligence.

Keywords: Artificial Intelligence; Legal Regulation; Multilateral Treaty; Industry Rules

本文系深圳職業技術大學校級社科類重點項目“人工智能生成物的知識產權保護—國外發展動向與我國應對思考”(項目號:6022310006S)的階段性研究成果。

作者簡介:鹿璐,深圳職業技術大學講師,研究方向為國際法、數據法學。

一、引言

人工智能（AI）技術不同於以往的任何技術，人類不再是生產和創造知識的唯一主體。基於大數據和算法的機器學習和自主選擇系統也改變了行業的運營方式。從未有過任何部門可以像人工智能技術這樣影響著整個世界。

英國大作家狄更斯作品中所言“最好的時代，也是最壞的時代”似乎也可以恰當的用來描述人工智能時代。一方面，它為我們的的生活和工作等帶來了極大的便利；另一方面，它也會給我們帶來許多的麻煩，我們也很可能會面對一個社會問題和安全隱患不斷的糟糕的時代。目前，人工智能已經在全球範圍內開始深度介入人類生活。而人工智能領先技術企業也進一步通過掌控數據資源和領先的智能技術來獲取的更多的經濟和戰略優勢。比如 Google、Meta 等科技權威企業，他們憑借其掌握的智能先進技術的優勢和優良的服務平臺，掌握著全世界用戶的數據信息，從而也進一步鞏固了其領先地位。

隨著人工智能近幾年的迅猛發展，國際社會也面臨著一系列與 AI 相關的法律和道德的挑戰。目前，對人工智能帶來的國際法問題，探討最多的就是知識產權方面的問題，本文不再對此進行闡述，僅擬從人工智能發展帶來的其他國際法問題進行總結，並提出相應的看法。

二、人工智能的責任問題

1. 人工智能系統的自主性和責任問題

近年來，人工智能的發展步入了高速發展階段。從自主駕駛車輛到人工智能機器人，甚至人工智能武器也開始被廣泛應用。這也帶來了一個巨大的社會問題：人工智能的責任問題。這是對人們現有責任機制最大的挑戰。

人工智能的發展與大數據密不可分。信息是決策的基礎，人工智能的決定完全依賴於大數據分析和計算它通過數據訓練來學習複雜的模式和規律，從而做出決策。從理論上，完全可以把人工智能系統認為是獨立的。首先，人工智能具備理性的決策方式，即算法。其次，人工智能的決策方式可以被第三理解，即人工智能的使用者。因此，人工智能是可以自主的，不只是是在行動上可以自主，在算法相關的法律和道德理性層面也可以自主。¹

自亞裏士多德提出道德責任理論以來，人們可通過控制條件與認知條件來歸因某項行

1 鄔桑：《人工智能的自主性與責任》，《哲學分析》2018年8月，第125-134頁。

動的责任主體。即如果你在“知道自己在做什麼”這種認知條件下，對操作有足夠的控制權，並實施了操作，即使你是操作的代理人，也必須承擔全部責任。在傳統判定責任歸屬的方式中，人造之物常被認為無法理解道德與法律，因而也不可能承擔各種責任。而人工智能能夠具備法律與道德的自主性，然而，就算可以證明人工智能具有自主性，對其進行歸責也存在著極大的困難。

2. 機器決策和法律責任的界定難題

決策一直是人類獨有的行為。在傳統的法律框架下，法律責任通常也是針對人類行為的。然而隨著智能社會的到來，人工智能算法開始進入人類社會生活並對傳統的決策方式、呈現形態和運行邏輯重新定義。如今，人工智能在決策和執行任務方面的應用日益廣泛。人工智能不僅用來輔助人們進行判斷，具有自主和決策的人工智能也開始大量應用於社會各個領域。

人工智能在全面改變我們生活的同時，也會給國家決策和執行系統帶來重大變化，最終對國際法產生重大影響。比如，國家機構和政府機構出於自身目的和需要而使用人工智能時，也會產生其是否應該為人工智能的決策和行為承擔法律責任的問題。這個問題不僅在國內法上有重要意義，在國際法上同樣具有重要意義。因為對於國家機構或政府機構所使用的人工智能做出的決策和行為，可能會有人從中獲利或受到損害，從而引發國家是否應該為此承擔國際法責任的問題。

國家應當履行其國家義務，並承擔因國家行為違背國際義務而產生的國家責任。² 然而，國家責任歸因問題是國際法中尚未解決的最重要的問題之一，³ 同時也成為眾多糾紛中的主要爭議焦點。特別是對如何區分國家機構與非國家機構，以及非國家機構的行為何時可歸因於國家的問題，一直以來爭議不斷。隨著人工智能的出現，現在需要關注的不是“非國家”主體，而是“非人類”主體的行為在國際法上的規制問題。

目前，已有學者開始研究人工智能的判斷、決策和行為可能會引發的國家責任問題。例如，國家機關利用人工智能主動實施違反國際法的行為（如利用人工智能設備進行拷問），或者國家機關沒有對其所使用的人工智能的決策、行為進行適當規制，從而導致發生了違反國際法的行為（如在意識到可能會對附近平民造成危險的情況下，仍舊部署無人攻擊裝

2 張乃根：《試析〈國家責任條款〉的“國家不法行為”》，武漢科技大學學報（社會科學版）2004年第2期，第72-75頁。

3 參見 Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), American Journal of International Law, 96 (4), p.91-108

置)，這些都可能引發國家責任。⁴

在國際法上，對責任主體進行追責是一個複雜的問題，而人工智能的應用，進一步增加了解決這個問題的難度。對於機器決策的責任界定來講，透明度和可解釋性是關鍵因素。可解釋性是指能夠理解和解釋機器決策的過程和依據，而透明度則是確保決策過程對外部觀察者是可見和可理解的。為了應對機器決策的法律責任問題，可能需要更新和完善現有的法律框架以適應技術的發展，明確機器決策的責任範圍和界限。

而探究與人工智能相關的國家責任問題時，承擔法律責任的法律人格主體問題，也成為新的爭議。在國際社會探討關於人工智能的國際法規時，也有必要考慮人工智能與國家責任法方面的關聯性，即“非人格”主體的行為是否會引起國家責任的問題。總的來說，在研究這些問題時，需要綜合考慮技術、倫理、法律和社會因素，以確保人工智能系統的合理使用，並明確機器決策背後的法律責任。

3. 人工智能責任追責的國際合作和司法途徑

人工智能追責是一個重要且複雜的問題，它需要國際合作和司法途徑等多渠道來解決。

在國際合作機制方面。國際社會普遍認識到人工智能技術的全球性和跨境性特徵，因而已經開始推動建立一些國際合作機制來解決人工智能的追責問題。比如，聯合國人權理事會和其他國際組織已經開始討論人工智能倫理和法律框架的發展，並要求促進國際合作以指定共同的准則和標準。

在國際標準和准則方面。如制定和推廣國際標準和准則，以引導人工智能開發和應用，並明確責任追究的原則和機制。國際組織如聯合國、歐盟和國際標準化組織（ISO）應在此方面發揮出重要作用。

在國際法律框架方面。國際法律框架的構建將會對解決人工智能追責問題方面起到重要作用。國家人權法、國際知識產權法和國際人道法等領域的法律原則與法律規範都可以為人工智能追責提供法律幫助。例如，國際人權法可以用來保護人類在與人工智能系統交互時的權利和自由，國際知識產權法可以保護人工智能創新和知識產權權益保護，國際人道法可以用於確保在使用人工智能系統時要遵守國際人道原則。通過制定國際公約、協議和條約，從而確保人工智能的合法、道德和負責任的發展與使用。

最後是跨國司法合作方面。由於人工智能技術的跨境性特點，跨國司法合作對於解決

4 “The law of state responsibility is concerned with the determination of whether there is a wrongful act for which the state is to be held responsible, what the legal consequences are, and what reparations should be taken.” Denise Garcia, “Future arms, technologies, and international law: Preventive security governance”, *European Journal of International Security*, 1(1), 2016, p. 104

人工智能追責問題至關重要。國際刑事法院、國際仲裁機構和跨國司法合作協定等機制都可以促進不同國家之間的信息共享、證據收集和司法合作，以便有效地追究人工智能的責任。同時，各國也需要指定國內立法，並加強監管，從多維度確保人工智能追責的執行。

人工智能法律責任是一個複雜的問題，需要多方合作和綜合性的解決。國際合作和司法途徑等的發展可以在其可持續發展和責任感知方面發揮重要作用。

三、人工智能對國際人權的影響

1. 人工智能對人權的影響

人權來源於人的固有尊嚴，世界人權宣言也將人的尊嚴和固有權利看作世界和平發展的基礎。人工智能技術的發展，對人權也產生了廣泛的影響。既有其有利的一面，也帶來了很大的挑戰。人工智能技術的發展對人權在如下方面已經出現一定影響：⁵

首先是隱私權方面。人工智能技術需要海量數據來進行深度學習，數據采集雖然是碎片化的數據提取，然而大數據技術通過對這些數據進行整合，就可以獲取到精準的個人隱私信息，從而導致個人隱私數據的泄露。因此需要相應的法律和政策來平衡隱私權和人工智能的發展問題。

其次是平等權方面。人工智能系統因可能受到數據偏見和算法偏見的影響，導致對某些群體的不平等待遇。例如，在招聘、信用評估和司法系統中使用的人工智能系統可能存在性別、種族或其他方面的偏見，從而對特定群體造成不公平的影響。

第三是自由表達權方面。人工智能技術可以用於內容過濾、言論審查和網絡監控，這可能會對自由表達權產生限制。

第四是勞動權方面。人工智能的廣泛應用可能導致自動化和機器取代人類勞動力，對就業和勞動權產生影響。需要制定政策和一定機制來應對人工智能對就業市場的沖擊，並為受影響的人群提供培訓和轉型機會。

最後在公平正義方面。人工智能系統的決策過程需要公平和透明，以確保公平正義的實現。例如在司法系統中使用的人工智能算法，需要可解釋性和公正性，以避免對某些群體的不公平對待。

人工智能對人權的影響需要全球範圍內的合作和共同努力來解決問題。在人工智能的發展和應用過程中，需要制定相應的法律、政策和倫理准則，來平衡技術進步和人權保護

5 胡獻紅：《人工智能的國際倫理探討：聚焦隱私與數據保護問題》，《新媒體與社會》2021年第1期，第3-20頁。

的關係，確保人工智能的發展能夠符合普遍人權價值和原則。

2. 人工智能在社會偏見和歧視等方面的潛在風險

人工智能技術可以成為社會發展向善的推力，但若在使用時未能充分考慮它們對人權的影響，那也有可能產生壞的甚至可怕的影響。比如在公平正義方面，人工智能系統的決策過程需要公平和透明，以確保公平正義的實現避免對某些群體的不公平對待。本部分著重闡述人工智能在社會偏見和歧視等方面存在的潛在風險。

2021年9月15日，聯合國人權高級專員辦事處發布了一份報告，分析了人工智能如何影響人們的隱私權和健康權、受教育權、行動自由權等權利。⁶ 報告中提到許多人因為人工智能而受到不公正的對待。不可否認，人工智能為人類提供了極大的便利服務，但人工智能在很難被注意的情況下發生大量侵犯人權的可能性也是不可否認的。這樣一來，就可能使某些群體邊緣化，導致程序上和結果上的不公，從而加劇現有社會的失衡。具體來說潛在的風險如下：

首先是數據偏見方面的潛在風險。人工智能系統的訓練需要大量數據，如果是訓練數據本身存在偏見或不完全代表多樣性群體，那麼人工智能系統可能會產生和放大這些偏見。例如，在使用人工智能系統進行招聘和貸款決策時，如果訓練數據傾向於某些特定群體，那麼可能會對其他群體產生不公平的對待。

其次是算法偏見的潛在風險。人工智能系統所採用的算法和模型可能存在潛在的偏見。這種算法的設計和參數設置可能會對特定群體產生不公平的結果。例如，面部識別算法可能會對某些種族或性別的人群識別準確率較低，從而造成不公平的結果。

第三是自我強化偏見的潛在風險。人工智能系統通過學習和反饋機制不斷優化自身，但如果系統在早期階段就收到了偏見的影響，那麼就可能在後續的學習和決策中保持和放大這種偏見。這種自我強化的過程也可能導致不公平和歧視的結果。

第四是透明度和解釋性的潛在風險。許多人工智能系統的決策過程對用戶和受影響的個體來說是不透明的，這使得很難理解系統是如何作出決策的，也無法解釋為何產生了某個結果。這種缺乏透明度和解釋性可能增加了系統的不可信度和難以追溯性，使得偏見和歧視難以檢測和糾正。

因此，要應對這些潛在的風險，需要採取一系列措施。如：要確保訓練數據的質量和

6 <https://news.un.org/zh/story/2021/09/1091022#:~:text=%E8%81%94%E5%90%88%E5%9B%BD%E4%BA%BA%E6%9D%83%E9%AB%98%E7%BA%A7%E4%B8%93,%E4%BB%A5%E5%8F%8A%E8%A8%80%E8%AE%BA%E8%87%AA%E7%94%B1%E6%9D%83%E3%80%82>

多樣性，避免數據集中的偏見和不平衡。在設計和優化算法時，要考慮公正性和可解釋性需求。同時制定相關的法律和監管框架，明確人工系統應遵循的法律制度和標準。

3. 人工智能背景下維護人權的國際法規範和機制

在人工智能背景下，維護國際人權的國際法規範和機制涉及到一下幾個方面：

首先，在人工智能背景下，利用現有的國際法規範和機制對維護人權至關重要。《世界人權宣言》是一項具有普遍意義的國際人權法律文件，它確立了人權的普遍原則，其中包括言論自由、隱私權和平等權等，這些原則對於確保人工智能系統在設計和使用時符合基本的人權原則至關重要。

此外，聯合國也制定有多項國際人權法公約，包括《公民權利和政治權利國際公約》和《經濟、社會和文化權利國際公約》等。這些公約明確規定了個人的基本權利，如隱私保護、言論自由、教育權利等，這些規定也都可以適用於人工智能系統的設計和使用過程中，使之被規範。

第三，國際組織和機構制定的一系列倫理准則和指南。這些倫理准則和指南可以引導人工智能技術的發展和應用符合倫理原則和人權機制。例如，聯合國教科文組織的《人工智能的倫理和社會影響框架》，歐洲委員會的《人工智能倫理指南》等。

在透明度和問責機制方面，在前一部分已有闡述，此處不再展開。此外，還有數據保護和隱私權等，將會在下一部分進行詳細闡述。

四、人工智能對數據安全的影響

1. 人工智能對數據安全的威脅

信息時代，海量數據跨境流動成為常態。而隨著大數據和人工智能技術的廣泛應用，機器學習和深度學習所需的數據更加驚人，因此數據安全變得更加迫切。

對個人數據而言，隨著信息技術的發展，公民的隱私數據被越來越多的記錄，掌握大數據的企業越來越容易獲得更多的個人信息數據。企業通過處理海量數據，通過算法對數據進行相關分析和挖掘，甚至於一些看似無關緊要的隱私數據經過算法的計算，就能精確定位到特定的個人隱私信息。由於個人隱私保護意識的缺乏，掌握數據企業保護力不強，相應法律的不健全以及監管技術的落後，帶來隱私保護和數據安全的隱患。特別是當這些數據使用目的被改變，甚至作為資源被交易，就很容易被不法使用，從而侵犯個人隱私。比如用戶信息被販賣、隱私被敲詐勒索、銀行卡被盜刷、跨境電信詐騙等常見問題，以及近期剛剛出現的 AI 換臉詐騙等，都是由於信息的泄露而造成的。同樣，人工智能通過大

數據技術對數據的整合與分析，也可以得出一些企業商業秘密。隨著大數據時代的到來，人工智能與算法技術加持下，甚至存在著危害國家安全的可能。如美國軍方認為，一款名為 Strava 的跑步類 APP 公布的用戶使用軌跡，有可能對美軍在在阿富汗和敘利亞的數個軍事基地造成泄密。⁷ 這表明信息保護不僅是個人、企業、國家多層面的問題，也是一個需要國際社會共同協作來解決的問題。

2· 跨境數據流動的法律問題

數據本身的特點，使得跨境數據傳輸極其便利，因而要通過地域來進行制約很難實施。實踐中，被稱為“史上最嚴數據法案”的歐盟《通用數據保護條例》（“GDPR”），就曾經對 Google 公司就違反歐盟 GDPR 相關規定而予以巨額罰款。因而，數據跨境流動不光涉及一國的國內法規制，也會涉及其他國家和地區的法律法規。尤其是西方發達國家，更早意識到數據安全的重要性，因而也對數據的保護標準會更為嚴格。如歐盟、美國等可以借助“長臂管轄”擴大其執法權利，甚至搶奪他國的“治外法權”。上文提到的 GDPR 就大幅擴展了管轄範圍，從過去的“屬地”向“屬人”轉變。其管轄範圍並不僅局限在成員國內部，而是可以擴大到所有在歐盟設立有辦公地點，或企業在提供產品或服務的過程中涉及到處理歐盟境內個體的個人數據，都可以受到該條例的規制，並且還規定了極高的罰款數額。因此，人工智能涉及的數據保護問題不容忽視。跨境數據流動存在有以下幾個方面的已有或潛在的法律問題：

第一是數據保護和隱私。不同國家對數據保護和隱私的法律要求和標準存在著差異。在跨境數據流動中，如何保護這些數據的隱私以及確保符合相關法律要求，這是一個重要問題。

第二是數據主權。不同國家對數據主權的理解和立法也會存在差異。比如一些國家主張對本國境內生成的數據擁有絕對的主權，這可能也會對數據的跨境流動和使用帶來限制和障礙。

第三是跨境數據的轉移限制。如果一些國家制定法律和規定，限制或控制個人數據或敏感數據的跨境轉移，這可能要求進行跨境數據流動時遵守特定的審查程序或獲得相關的許可或批准。

第四是法律的管轄和適用法律。跨境數據流動涉及到不同國家的法律體系，因而確定適用的法律管轄和適用法律可能是一個複雜的問題。不同國家的法律體系和法律規則可能會對數據的適用、處理和保護提出不同的要求。

為了解決這些問題，國際社會正在努力制定跨境數據流動的法律框架和協議。例如，一些國際組織和國家之間已經達成了數據保護和隱私保護的協議，以促進跨境數據流動並

7 https://www.sohu.com/a/220076719_99939413

確保數據的合法和安全傳輸。

3. 國際數據保護法律框架和協議

近年來，全球掀起完善數據安全保護體系浪潮。數據安全保護體系包括數據采集、利用規範和跨境流動安全保護規範。各國基於不同的利益考量，採取不同的立法政策，保護本國利益，並試圖引領跨境數據流動國際規則。

比如歐盟的《通用數據保護條例》（“GDPR”），成為公認的全球個人信息保護最有影響力的代表；隨後是美國加州的《加州消費者隱私法案》（“CCPA”）；巴西的《一般數據保護法》以及泰國的《個人數據保護法》等，都深度參考了GDPR的條款。總體而言，GDPR對各國的隱私保護立法都產生了重要影響。

在國際數據保護法律框架中，歐盟《通用數據保護條例》是當前最具代表性的硬法規範之一，而經濟合作與發展組織（以下簡稱經合組織，OECD）制定的《關於隱私保護和個人數據跨境轉移指南》（簡稱《OECD指南》）和亞太經濟合作組織（簡稱亞太經合組織，APEC）制定的《跨境隱私保護規則》（簡稱《隱私框架》）以及二十國集團的相關成果文件等，則構成了當前國際數據保護軟法實踐的基礎框架。這些國際數據保護軟法框架，與近年不斷興起的區域貿易協定，共同形成了當今國際社會多層次的實施機制。⁸

當然，在數據保護這一問題上，國際社會一直面臨著數據的本地存儲和數據自由流動之間的價值選擇。數據保護問題的背後是搶占新時期全球經濟和技術革命先機和話語權的重大問題。同時，基於各國國內數據治理的法律規則之間存在相互沖突的現實，在現階段國際軟法可能是在國際層面對數據保護的一個相對較好的選擇。

五、自主武器系統對國際人道法的影響

人工智能技術的發展使得從國家決策者、國家軍隊、國內民意這三個層面減少了對發起戰爭的制約，一定程度上也降低了進行戰爭的門檻；大量無人武器的投入使用可以使戰爭中的傷亡預期大大降低；這些都會能降低決策者發動戰爭顧慮。目前，人工智能自主武器系統的出現已經引發了國際人道法中相關原則和規定的解釋和適用的爭議。

1. 自主武器系統的威脅

當前，針對國際法下的人工智能治理的討論主要集中於對人工智能武器的規制。例如，在俄烏沖突中雙方對於面部識別技術和無人機的應用。以“無人機”等“致命自主武器”為

8 敖海靜：《數據保護的國際軟法之道》，《法商研究》2022年第2期，第159-160頁。

代表的人工智能應用在國際法中的合法性與責任歸屬等方面均有相應的討論，其中國際人道法面臨的挑戰最為突出。人工智能武器可能動搖人類主體可歸責性這一現代人道法的核心。⁹

根據國際人道法原則，任何戰鬥行動都必須遵守區別待遇原則，即區分和區別對待平民和非參戰人員。然而，自主武器系統的決策過程可能缺乏足夠的判斷力和辨別能力，導致平民和非參戰人員的風險增加。此外，國際人道法還要求使用武力時的必要性和比例原則的要求，即只有在必要的情況下才能適度使用武力。自主武器系統的自主決策能力可能導致過度使用無力或無法準確評估和應對戰鬥的複雜環境情況，從而違反必要和比例原則。

此外，自主武器系統的決策過程使得確認責任和追究責任變得困難。如果自主武器系統導致違反了國際人道法的規定，如何追究責任和確保合適的司法程序仍然是個難題。對於人工智能自主武器系統造成的損害，責任主體應該是武器的制造者還是使用者？

隨著人工智能技術的廣泛應用，武器的作戰效能得到顯著提升，同時也大大減少了自己一方的人員傷亡，因而各國更增加了將其應用於武器的意願。然而，交戰雙方都有權選擇自己的作戰方式和手段，這也是國際人道法的基本原則。¹⁰ 根據國際條約，所有準備用於武裝沖突的新武器都必須符合國際人道法的規定，¹¹ 這其中自然也包括使用人工智能技術的先進武器或者武器系統。

傳統的國家安全准則已經不能滿足日益增長的安全需求，為此，全球應重新審視人類安全，並將其置於最高的地位，重視其作為決策的重要參考。¹² 同時，為了保護全球安全，必須建立更嚴格的全球安全標準，包括推動起草相關國際條約等以設立相應的國際安全條款，如嚴格劃分致命性自主武器和非致命性自主武器，並加以禁止和限制。¹³ 人工智能武器的出現使得戰爭從傳統的單兵作戰轉變為多兵作戰，其具備的自主性、機動性、可控制範圍、可被用於多個場景以及可以被多個技術組合所控制，其在為一方提供了多樣化保障的同時，也給國際人道主義法帶來挑戰。由於自主武器系統其具體表現形式和技術的難以確定，使得我們仍需要綜合考慮多個技術組合來分析各種自主武器系統對違反國際人道法

9 Chantal Grut., *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, *Journal of Conflict and Security Law*, (18) 2013, p. 73

10 1907年海牙《陸戰法規和慣例規章》第22條最早規定了這項原則，1977年《1949年8月12日日內瓦第四公約關於保護國際性武裝沖突受難者的附加議定書》（《第一附加議定書》）第35條第1款對此亦有規定。

11 參見1977年《1949年8月12日日內瓦第四公約關於保護國際性武裝沖突受難者的附加議定書》（《第一附加議定書》）第36條。

12 Kaja Kowalczywska, *The Role of the Ethical Underpinnings of International Humanitarian Law in the Age of Lethal Autonomous Weapons Systems*, *Polish Political Science Yearbook*, 48(3), 2019, p.464-475

13 張衛華：《人工智能武器對國際人道法的新挑戰》《政法論叢》2019年第4期，第144-155頁。

的風險，並從國際層面進行規範和限制，以便未來對安全的、可控制的使用人工智能武器提供參考依據。並且，隨著人工智能的高速發展，未來人工智能武器在“智能化”和“自主性”方面的不斷改進，未來自主武器系統可能會對戰爭法和國際人道法產生前所未有的影響，它們違反國際人道法的風險將會大大提高。

傳統國際人道法旨在禁止使用過度殘忍、給戰鬥員造成不必要痛苦的作戰手段和武器。雖然人工智能武器具體“智能化”和“自主性”的特徵，但是也應當受到相應的約束交戰雙方的國際人道法規範的約束，以確保他們的安全性。具體來說，隨著人工智能武器的出現，它們對國際人道法中相關的“法律主體”的規定也構成了一種新的挑戰。即武器這個原本的國際人道法上的“法律客體”，似乎因為在某種程度上具備“法律主體”的資格與能力而需要受到國際人道法的約束。這對於國際人道法來說，顯然是一個新的問題。

2. 國際法對自主武器系統的限制和規範

目前，國際法尚未制定專門針對自主武器系統的具體法律框架和規範。然而，一些國際法規定可以適用於自主武器系統，並對其開發、部署和使用提出了一些限制和規範。

國際人道法是設計武裝衝突和戰爭時行為規範的法律體系。包括《日內瓦公約》和《日內瓦公約附加議定書》等。人道自主武器系統的設計和使用必須遵守這些原則和規定。對於一些國際法律文件禁止或限制使用的特定武器，這些禁止和使用也可以適用於某些形式的自主武器系統，特別是那些具有殺傷力和不可控制行為的系統。

國際人權法確保了人權的保護和尊重。雖然國際人權法主要適用於平時的和平環境，但也可以適用於武裝衝突和戰爭情況下的行為。自主武器系統的開發和使用必須遵守國際人權法，以保障人權的尊重和保護。

3. 加強國際合作的必要性

在非政府組織的敦促下，《特定常規武器公約》論壇自 2013 年以來就一直在日內瓦討論禁止自主武器系統的問題。參與國之間似乎一致認為，在武裝衝突中使用物理，包括使用致命武力，必須始終受到有意義的人類的控制，而完全自主的武器系統應當被新的國際法律文件所禁止。¹⁴ 通過對人工智能武器系統進行“有意義的”控制，不僅可以使其受到國際人道法的規制，而且還可以在一定程度上緩解人們對人工智能武器的擔心：即自主人工智能武器的出現是否會構成對違反國際人道法的行為進行追責的阻礙？畢竟，現階段，無論自主性人工智能武器其的智能化程度有多高，仍然還只是一臺機器，把侵犯人道主義

14 Thomas Burri, *Internal Law and Artificial Intelligence*, German yearbook of International Law 2017, vol.60, p.

的罪責歸於一臺機器並沒有現實意義；而其自主性也可能成為操縱者逃避責任的托詞。

在聯合國裁軍委的一次會議中，政府專家組強調了致命性自主武器系統需要被關注，並指出它們可能會給國際安全帶來潛在威脅。對此，他們给出了一些建議。例如，由於致命性自主武器系統技術方面的軍備競賽，要防止擴大不同國家之間的技術差距，以及提高使用武力的標準等；同時也建議國家應當承擔起相應的法律責任，並討論參考《特定常規武器公約》議定書的形式，制定具有約束力的法律文件，限制致命性自主武器的使用，以維護人類安全。

追究戰爭罪犯的國際責任在國際法層面本身就是困難的，而人工技術發展帶來的自主武器系統使得追責變得更為困難。自主武器系統的出現，涉及到了更多的參與方。一旦其發生嚴重違反國家人道法的事實，如殺害投降的戰俘，無端攻擊平民等，這將會使得追責不再局限於政府機構、軍隊、指揮官、戰鬥員等，還有部署自主武器系統的國家，甚至人工智能自主武器系統的設計和生產者也可能會牽涉其中。

儘管人工智能自主武器具備一定的獨立性，但對其進行懲罰並沒有真正的現實意義，關鍵還是要考慮如何填補自主性人工智能武器的使用所帶來的國際規範之歸責的空白問題。在當前的技術水平和法律狀況下，繼續由人類對人工智能武器進行“有意義的”控制，是解決有關問題的一種解決方法。但是，在當前人工智能技術飛速發展，人工智能武器的自主性必將繼續提高的大背景下，如何實現“有意義的”人類控制，是各國在研發人工智能武器時應當優先考慮的問題，各國都需要認真對待。這個問題很難有一個統一的標準答案，只能寄託於各國在研發和部署人工智能自主武器時，能夠嚴格依照國際人道法進行審查。

六、人工智能發展之國際法應對

毫無疑問，國際社會有必要對人工智能進行國際法規制，而主權國家治理的有效性也使得人工智能的國際法治理具有可能性。

在具體如何對其進行國際法規制方面，有以下幾點建議：

首先，由於人工智能技術已經在全球範圍內的許多領域被使用，因而唯有通過制定國際標準對其提供統一的指導。而在變革過程中的國際法領域，由於新的國際規則的形成存在著困難，為了有效治理人工智能這一新事物，可以先用國際層面的行業規則等“軟法”形式進行確認，發揮其先導或補充作用。同時，在人工智能領域制定國際軟法雖然能夠解決技術問題和一定的法律和倫理問題，但由於其不具有約束力，因而也需要相關的國內法予以規制和保護。

其次，要想解決國家之間對人工智能使用產生的爭議、存在的潛在影響以及需要面對的沖突，最好的解決辦法，締結多邊條約是較好的解決辦法。並且，最初的“軟法”經過長期的實踐，可以將其中的部分被廣泛接受的軟法納入立法從而轉變為“硬法”。通過多邊條約與行業規則、國內立法、國際示範法之間的相互影響和相互促進，從而形成相對全面、比較理想的一套國際法律規範。¹⁵ 最終實現人工智能在可控制範圍內的安全發展。

人工智能的發展已經成為不可阻擋的趨勢，其發展更是凸顯了科技與法律的矛盾。人工智能技術的發展宗旨應是維護全人類的共同福祉與長遠利益，因此其設計及應用要符合人類基本倫理道德，維護人的尊嚴和權利等，堅持以人為本；同時，發展人工智能技術的目的是要服務於人類，確保人工智能系統的安全、有效和可控性，保護公民個人隱私，保護數據安全；也要確保人工智能算法的可追溯性與透明性，最大限度減少算法歧視；構建合理使用機制，使全世界更多的人可以從人工智能技術的發展中受益，避免數字鴻溝的出現，從而加大世界的不平等。

3、人工智能技術要用於和平發展目的。倡導和平使用人工智能技術，防止開展致命性自主武器軍備競賽，危害世界和平。

4、人工智能技術的運用，應符合《聯合國憲章》的宗旨以及各國主權平等、和平解決爭端、禁止使用武力、不干涉內政等現代國際法基本原則。

地球是全人類的。世界各國在發展人工智能技術的同時，也要建立一個公平公正的合作機制，並探索利用現有的國際法基本原則和准則建立一套適用於人工智能領域的全球性法律規範。

15 沈偉：《數字經濟背景下的人工智能國際法規制》，《上海財經大學學報》2022年第5期，第123-137頁