人工智慧監管法律——獨漏自主性武器之規範？

陳建佑*

壹、前言

近來因開放人工智慧（OpenAI）公司之ChatGPT程式受到廣泛使用，「人工智慧」（Artificial Intelligence，以下簡稱「AI」）程式又再次受到世人的高度關注。世上有許多組織例如聯合國、生命未來研究所、歐盟，及AI領域的權威人士例如伊隆・馬斯克（Elon Musk）、山姆・奧圖曼（Sam Altman）、傑佛瑞・辛頓（Geoffrey Hinton）及約書亞・班吉歐（Yoshua Bengio）等，均致力於為AI制定穩固的倫理框架，監督偏見、監控及自動化武器，其中包括現在已發生之嚴重議題（隱私、資訊安全、歧視偏見、智慧財產等），亦聚焦未來之人類生存或戰爭威脅。

特斯拉創辦人馬斯克及許多AI權威人士[1]共同於2023年3月間簽署一封公開信，呼籲全面暫停研發強於GPT-4的AI系統六個月。這封公開信再度引發AI研發的道德與安全問題等關注。曾獲圖靈獎並被譽為AI教父之辛頓博士，於2023年5月2日接受紐約時報訪談時，就警告政府、企業及社會應正視AI將帶來之危險性，並儘速加強AI之安全性及監管[2]。

聯合國教科文組織（United Nations Education Scientific and Cultural Organization, UNESCO）也在2023年3月30日發表聲明[3]，呼籲建立全球道德框架。該組織秘書長Audrey Azoulay建議以2021年11月教科文組織與193個會員國投票通過的《AI倫理全球協議》作為指南[4]，並強調產業自律[5]完全無法避免相關的道德危害，應以《AI倫理全球協議》來

＊本文作者係資鋒法律事務所主持律師、中華亞太智慧物聯發展協會理事。

註1：包括著名書籍《人類大歷史》作家、歷史學者尤瓦爾・哈拉瑞（Yuval Noah Harari）。

註2：Vanessa，〈離開Google是為了說真話，AI教父：普丁拿去用怎麼辦？〉，鏈新聞，2023年5月5日，網址：https://abmedia.io/ai-godfather-leave-google-and-give-warning（最後瀏覽日期：2023年5月8日）。

註3：UNESCO，「Artificial Intelligence: UNESCO calls on all Governments to implement Global Ethical Framework without delay」，聯合國教科文組織新聞稿，2023年3月30日，網址：https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without（最後瀏覽日期：2023年5月11日）。

註4：UNESCO（2022），《人工智能倫理問題建議書》，聯合國教科文組織數位圖書館，網址：https://unesdoc.unesco.org/ark:/48223/pf0000381137_chi（最後瀏覽日期：2023年5月11日）。

註5：微軟公司於2018年所公布之AI六大自律原則：可靠與安全、公平、包容、隱私與保障、透明、負責。

確保未來AI發展能遵循法治體系，並在危害發生時能有問責及補償機制。

美國目前在AI技術應用發展仍領先群雄，但針對AI之監管規範，歐盟之立法例進程較為完善。故本文重點在於從制度面如何規範、引導、鼓勵AI科技應用之良性發展，即有論者所倡導「在科技競爭環境下之人工智慧倫理治理模式建構」意旨[6]。AI技術之應用與發展，終究應回到其倫理治理（Ethical Governance）之根本討論，惟基於其範圍恐過於廣泛、本文篇幅有限，是本文選擇著墨於「AI自主性武器之研發與應用」之議題及歐盟AI法草案之內容，並回頭檢視臺灣之現況與未來。

貳、AI技術之內涵

AI係資訊技術（Information Technology, IT）之一環，係一種系統、程式或演算法（Algorithm），即透過程式設計，使電腦有類似於人類之知識及反應，做到理解人們之表達、學習、推論及解決問題[7]。機器學習（Machine Learnig, ML）是AI之一個分支，涉及很多領域，包括機率論、統計學等多門學科。機器學習是指讓電腦具備自我改進能力及自動學習能力，可根據經驗演化它之行為而自動最佳化下次結果。機器透過以往資料的學習，找到資料之特徵規則後，建立數學統計模型，對之後輸入大量的訓練資料（Training Data）進行分析與判斷之一種技術。機器學習理論主要是設計及分析讓電腦可自動學習之演算法，從資料中自動分析獲得規律或模型，並利用規律或模型對未知資料進行預測之演算法。

深度學習（Deep Learning, DL）是機器學習之一支，它在2006年開始發展，為AI最核心之技術，是一種利用模擬人類神經網路的非線性變換複雜架構的演算法，它是源於類神經網路系統（Artificial Neural Network, ANN）。儘管類神經網路系統的理論從1943年就有數學模型，但後來的發展因為需要大量運算而電腦能力無法跟上，即逐漸被忽視，直到近代因為電腦運算能力增強，彼此合作的多台分散式運算系統架構強大，使如此大量運算而找出模式，變得不再困難。最常用的深度學習演算法就是卷積神經網路（Convolutional Neural Network, CNN）。

正因為卷積神經網路的出現，一舉將辨識分類的準確率提升至90%以上，所以AI在2015年之後快速進入商業化，即AI應用落地。近年又研發出生成式AI（Generative AI）即人工智慧生成內容，又稱AIGC（AI Generated Content）。生成式AI是人工智慧中的一個分支，主要用於創造性之工作，例如文章、影像、音樂甚至程式生成等。生成式AI主要依賴於深度學習技術，其中最常見的是生成式預先訓練模型（Generative Pre-trained Transformer, GPT），用於自然語言處理、圖像處理、音頻處理等各種生成式任務。它最主要的特點就是技術上可幫助模型

註6：李崇僖（2020），《人工智慧競爭與法制》，第8頁，翰蘆。
註7：林東清（2018），《資訊管理：e化企業的核心競爭能力》，第93頁以下，臺北：智勝。

更加直覺地輸入指令（Prompt）與數據（Data），而更完善使用者之體驗。

從2021年底至今，因Midjourney、DALL-E、Stable Diffusion、ChatGPT及微軟Office 365 Copilot等生產應用工具之大爆發，AI再次成為熱門焦點。當然，AI在自主性武器之研發腳步亦從未停歇。因此，AI相關之監理規範再次受到重視與關切。

參、無所不包之AI法律監理問題及其反思

在瞭解AI技術與應用發展後，吾人可爬梳目前面臨到之新型態法律問題，大致可分成九大類型[8]，包括：

一、人工智慧倫理

人工智慧之規範框架及制定標準[9]，包括在司法實務上或各國自主性武器之研發應用，亦為本文之重點討論所在。

二、個資保護及資訊安全

AI數據蒐集須符合歐盟GDPR等國際標準。G7成員國義大利曾於2023年3月間命ChatGPT下線，並調查ChatGPT可能違反個資法之情形，儘管隨後已解除禁令，然此項舉止仍引發歐洲其他國家隱私監管單位啟動相關調查。

三、數據財產權[10]

考慮在民法中新增數據財產權，理由係數據為人工智慧時代的「石油」，在商業上已成為交易的客體，應確立數據法律上之財產權地位。

四、法律體系變革

人工智慧發展導致傳統法律有調適之必要，不僅是民法上，包括刑法、保險法、道路交通等法規範。

五、智慧財產權

藉由人工智慧所創作之新型態著作權[11]及專利權問題。

六、金融科技

人工智慧在金融科技方面的使用，或人工

註8：朱宸佐（2023.5.6），〈人工智慧時代的新型態法律問題〉，發表於：《新型態人工智慧法律問題》，台北律師公會律師在職進修系列課程，第13頁，臺北。

註9：陳建佑（2021），〈人工智慧法律的現在與未來〉，《月旦會計實務研究月刊》，第41期，第30-36頁。

註10：朱宸佐（2020），〈人工智慧時代數據財產權的保護路徑〉，《人工智慧與法律衝擊》，第165-187頁，元照。

註11：陳建佑（2020），《人工智慧著作權法及管理規範之研究》，國立臺灣科技大學管理研究所EMBA碩士在職專班碩士學位論文。
陳建佑，〈AI生成的圖片、文字著作權到底該歸誰？這可能取決你的指令強弱〉，INSIDE硬塞的網路趨勢觀察，2023年2月20日，網址：
https://www.inside.com.tw/article/30775-ai-law（最後瀏覽日期：2023年5月8日）。

智慧在區塊鏈及智慧合約的應用，均涉及金融監理是否規範及其介入之強度[12]。

七、出口管制

在中、美科技戰爭下，人工智慧成為重點管制之技術[13]。

八、競爭法

單一行業數據蒐集範圍過大，恐涉及反壟斷或不當競爭，例如媒體議價權之立法例及相關討論。

九、勞資爭議

人工智慧取代大量勞動力，或用於評估決定員工去留，均可能導致相關之勞資糾紛。

著名書籍《人類大歷史》作家、歷史學者尤瓦爾·哈拉瑞（Yuval Noah Harari）日前在《經濟學人》雜誌上發表一則標題為〈人工智慧已入侵人類文明之操作系統〉（Yuval Noah Harari argues that AI has hacked the operating system of human civilisation）[14]之文章，指出AI之大規模毀滅性武器，可以摧毀人類之精神與社會世界，然我們現在仍可以且必須迅速規範出新的AI工具，首要重要步驟就是要求在將強大的AI工具投放到公共領域之前，國家、政府與企業應進行嚴格的安全檢查[15]。該文重點摘要如下：「自1945年以來，我們知道核技術可以為人類提供廉價能源，但也可能在物理上摧毀人類文明。因此，我們重塑了整個國際秩序，以保護人類，並確保核技術主要用於良善的事情。我們現在必須面對的是，一種新的大規模毀滅性武器，AI可以摧毀我們的精神與社會世界。我們現在仍然可以規範新的AI工具，但我們必須迅速行動。與核武器無法創造更強大的武器不同，AI可以自我創造出，呈指數級增長的、更強大的AI。第一個至關重要的步驟是，要求在將強大的AI工具投放到公共領域之前進行嚴格的安全檢查。正如製藥公司在研究新藥物的短期與長期副作用之前不能推出新藥物一樣，科技公司在讓新的AI工具變得安全之前不應推出它們。我們需要一個類似食品藥品管理局的新技術機構，而且期望它早就已經在運作。在公共領域放緩AI的導入速度，會讓民主國家落後於更無情的專制政權嗎？恰恰相反。不受規範的AI部署

註12：Perry，〈金管會8月出台金融科技發展方案2.0，將AI技術納管，訂定指導原則〉，鏈新聞，2023年4月20日，網址：https://abmedia.io/financial-supervisory-commission-is-expected-to-regulate-ai-technology（最後瀏覽日期：2023年5月8日）。

註13：李崇僖，前揭註6，第224頁。

註14：哈拉瑞，〈Yuval Noah Harari argues that AI has hacked the operating system of human civilisation〉，經濟學人，2023年4月28日，網址：https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation?fbclid=IwAR3SuJoK-kIX6j9gKUpjQV4frg0cSegixJ2u30BjgGltjCB-PqJuh8tejI0（最後瀏覽日期：2023年5月11日）。

註15：無獨有偶，研發出ChatGPT程式之開放人工智慧（OpenAI）公司共同創辦人山姆・奧圖曼（Sam Altman）亦有類似之警語。

將造成社會混亂，這將有利於獨裁者，並破壞民主」。

從歐盟、聯合國或台灣目前關於AI相關規範草案或計畫報告[16]裡，似乎對於自主性武器並無明確提及或管制，大多係致力於為AI制定穩固的倫理框架，例如倫理、數據、環境生態、性別、文化、教育、經濟、勞動、健康或社會福祉，惟似乎獨漏自主性或自動化武器之相關管制規定，故自然引起AI倫理道理界針對未來之人類生存或戰爭威脅之關注。

肆、AI戰爭時代已經來臨：國際法有何作用？

因AI科技屬於破壞式技術之性質，不僅會顛覆全部產業或國防之發展，且因攸關產業競爭力、決定軍事戰略優勢，以及挑戰民主自由價值，所以各國都大舉投資在AI科技之研發，冀望能拔得頭籌，取得優勢或競爭力。

基此，在AI時代裡不論是產業、利益團體或政府部門，或是一般個人，都難以逃脫AI所帶來種種的社會、倫理或法律面向衝擊，如何在研發、設計與運用AI的時候加以規範，在民主制度的世界中就涉及「公平」、「責任」、「透明度」等考量因素與爭議。因為AI最大特性就是出現「黑箱效應」，即指研發者或使用者都已經無法清楚理解或解釋其創造或使用AI是如何運作或得出結果，而AI此種不確定或無法完全控制的技術與狀態，就需要加以被規範。

所謂AI之「自主性武器系統」（Autonomous Weapon Systems），係指能夠在無人干預情況下獨立搜索、識別並攻擊目標之新式武器，包括目前某些防禦武器所具有能夠攔截來襲之飛彈、火箭彈與炮彈，或飛機的自主模式，均屬於自主性武器之雛形，而其在隨後發展中是否能夠區分平民與軍事目標亦受到國際人道組織之關注。然而，自主性武器系統及其管制規範體系，卻向來是《國際人道法》領域頗為棘手之議題（即難以再遵循或符合國際人道法之關鍵規範──「區分平民與戰鬥員之原則」），像此類仰賴大數據與演算法所驅動之自主性武器已明顯挑戰國際法之秩序。即使係《日內瓦公約》之「比例原則」──禁止使平民生命受損失、受傷害、物件受損害，或三種情形均有且與預期的具體直接軍事利益相比損害過分之攻擊，然而此亦僅為一種主觀決定，且客觀上必須根據個案具體情況判定。

自從2014年以來，簽署1983年《聯合國特定傳統武器公約》（Convention on Certain Conventional Weapons）的國家，就一直在討論對致命性自主武器系統的可能限制。聯合國並在2017年首度針對全自動武器召開會議並提出警告：有鑑於機器人武器系統的快速發展，自主性愈來愈高，迫切需要國際社會就設限達成共識，以解決基本的法律及道德疑慮。

註16：行政院112年4月7日院授科會科辦字第1120012011號函核定，《臺灣AI行動計畫2.0》，網址：https://digi.nstc.gov.tw/File/7C71629D702E2D89（最後瀏覽日期：2023年5月14日）

曾有論者專文[17]指出，軍用無人系統因其零傷亡及對複雜任務的良好適應性，已被視為是近年的革命性武器，一旦廣泛使用將明顯改變現有的武裝衝突型態，甚至改變攻伐的本質。該文亦提出四項觀察及評價：（一）演算法驅動之武器運作，恐會導致「自主性」意涵之認定產生浮動性；（二）武器系統可能誤判軍事目標，乃至科學證據的可信賴性與可解釋性存疑；（三）AI參與之自主性武器系統，如何符合日內瓦公約及其附加議定書規定意旨；（四）如何釐清國家責任成立與否的解釋論疑義，也是附帶衍生難題。

如果越來越多有心人士使用AI自主性武器攻擊或侵略他國，則遭攻擊或受侵略之國家是否在此等領域應取得領先優勢？生命未來研究所曾在2017年間發表一封公開信，呼籲聯合國禁止所謂的「殺手機器人」（Killer Robot，此係對自主性武器的別稱），認為致命之自主性武器極可能引發戰爭型態的第三次世界大戰，當時此封公開信則獲得人工智慧圈內逾百人之連署，包括馬斯克及辛頓等人。

根據聯合國之報告指出2020年3月利比亞內戰中，利比亞政府軍對ISIS使用無人機，並且該無人機可以在「無需操作者與彈藥之間聯通數據的情況下攻擊目標，可以在沒有人為下令的情境下自動判斷發動攻擊」，該報告宣稱，這完全有可能是「史上第一個由演算法操縱來殺害人類的無人機」。

以色列國防軍2020年5月空襲加薩走廊時，運用無人機蜂群來「定位、辨識並攻擊」哈瑪斯武裝分子，這是全球首例無人機蜂群參與實戰，也是首場違反國際法無差別攻擊的「AI戰爭」。

根據科技媒體INSIDE報導，南丹麥大學（University of Southern Denmark）的博德（Ingvild Bode）表示，在烏俄戰爭中已呈現無人機軍事化發展的白熱化，烏克蘭使用致命的AI無人機防抗俄軍，為該等技術提供前所未有的試驗場；而俄軍亦以「伊朗見證者-136型」（Shahed-136）無人機群，猛攻烏克蘭的關鍵基礎設施。

美國白宮科學與技術政策辦公室曾在2016年5月間舉行四場公開研討會，進行與AI相關的政策討論，認為AI科技的發展如同過去任何具有變革意義的科技發展（例如工業革命、資訊革命）相同，勢必影響人類的工作、經濟、安全等面向，所以認為應該要有具體的管制藍圖。嗣後在2020年1月發布《AI應用管制指引》，為美國政府機關起草AI規範並進行管制時提供指引，更清楚羅列10項AI原則要求各單位遵守。

根據媒體報導[18]，美國白宮在2023年5月4日透過新聞稿發布「美國政府關鍵與新興技

註17：林昕璇（2021），〈AI自主性武器系統在國際法上適用之研析〉，《軍法專刊》，第67卷第4期，第20-44頁。

註18：James，〈拜登怕AI危及國安！緊急召集微軟、Google、OpenAI進白宮商談人工智慧風險〉，動區動趨BLOCKTEMPO，2023年5月5日，網址：https://www.blocktempo.com/us-announces-standards-strategy-for-critical-technology/（最後瀏覽日期：2023年5月8日）。

術國家標準戰略」，並表示該戰略之目標，係加強美國保護美國消費者技術之基礎，以及美國在國際標準發展中之領導地位和競爭力，該戰略將促進對已確認關鍵領域之「標準化前研究」投資，鼓勵私營部門、學術界參與研究，針對培訓進行投資，並確保完整性與包容性。其中列為優先事項之技術，包括人工智慧與機器學習。

事實上，美國國防部許多專家計畫（例如：對大數據與機器學習的使用、演算法作戰跨職能團隊）的推動，均有賴像Google這類的企業支持，因為它們近年來已累積建造深度學習系統所需的專業與基礎架構，這也是五角大廈建立新科技的典型方法──與民間企業合作，最終基於人工智慧道德倫理的問題考量，而部分計畫受到阻礙[19]。有論者稱此乃「科技不作為」與軍用AI商業化的內在阻力[20]。然美國政府為維持其競爭優勢，而對AI自主性武器之相關管制，係毫無任何誘因促使其著墨。不過縱使如此，美國國防部仍有一個「DOD 3000.09指令」[21]，作為其AI武器管制與研製之自律規範，尤其在俄烏戰爭後且我國國防部及中科院近來也有研發自主性武器之今日，可供台灣作為參考；甚至美國國防部在2023年1月間有修訂該自律規範之指令內容[22]，盡量符合學者強調之AI倫理道德原

註19：凱德．梅茲（2022），《AI製造商沒說的秘密：企業巨頭的搶才大戰如何改寫我們的世界》，第十六章第3頁，時報文化。

註20：法蘭克．巴斯夸利（2023），《二十一世紀機器人新律：如何打造有AI參與的理想社會》，第249頁，左岸文化。

註21：美國國防部（2023），〈DoD Directive 3000.09: Autonomy in Weapon Systems〉，網址：https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.。（最後瀏覽日期：2023年5月15日）

註22：美國國防部「DOD3000.09指令」內容重點整理：

A.自主武器系統定義，係指能夠在沒有人類直接控制的情況下執行目標選擇和攻擊行動的武器系統。

B.遵循國際人道法，即使用自主武器系統應符合國際人道法之原則規定，並尊重人權、遵循國際人道法公約與協議。

C.監督和控制，即應建立有效之監督和控制機制，確保自主武器系統之操作與使用符合國家政策及法律要求。人類操作者應保持適當之監督權限，能夠介入並決定自主武器系統之使用。

D.風險評估，即在部署自主武器系統之前，必須進行全面風險評估，包括技術可行性、法律倫理風險、戰略風險等。評估結果應該用於指導決策及政策制定。

E.安全措施，即應採取必要安全措施，保護自主武器系統免受未經授權之存取和操控，包括加密通訊、身份驗證與授權機制等。

F.須符合「知情同意」原則，即在執行自主武器系統行動時，必須獲得相關人員之知情與同意，特別是在涉及平民人口或友軍部隊之情況下。

G.智慧倫理與訓練，即應加強智慧倫理研究與訓練，以確保自主武器系統之設計與使用符合倫理與道德準則。操作者應接受相應之培訓，以適應使用自主武器系統之挑戰。

H.透明度與回報，即需要建立透明之機制，向相關利害關係方提供有關自主武器系統之資訊，並回報系統之效能、使用情況與潛在風險。

I.國際合作，即鼓勵與盟國與國際組織合作。

則即安全、透明、登錄機制及國際合作[23]。

在自主性武器被視為美國及中國等軍事大國之間軍備競賽時，紐西蘭政府早已決定對此採取完全禁止自主性武器之立場。紐西蘭政府曾聲明：「當一台機器能夠在完全沒有任何人類介入的情況下，做出識別和攻擊人類目標的決定，那將非常令人擔憂」，並質疑「自主性武器違背國際人權法律的可能性，尤其是在保護平民、軍事行動的合理性等基本原則上」[24]。除紐西蘭外，奧地利、愛爾蘭、墨西哥與智利等國家亦支持自主性武器之禁令。

吾人明白AI科技係屬中立，然此科技中立係選擇性的？抑或假性的？在沒有一個強而有力之國際法拘束力或制裁規範下，AI科技尤其在自主性武器之研發應用，或是戰爭攻防之使用上，今後均面臨此種雙面刃之難題。聯合國可譴責某個國家，但其權威卻時常被藐視。因為欠缺全球治理權威，自主性武器議題出現各種主義或論調。廢止主義者試著透過國際條約禁止殺手機器人；現實主義者則主張國家必須儲備先進的軍事技術，以免被邪惡的競爭對手佔得先機或遭到威脅；AI提倡者聲稱自主性武器將減少戰爭的可怕，因為可預設該精準度可減少誤殺或誤傷；懷疑者認為未來還有很長一段路要走，嚇阻理論家則是擔心，即使自動化戰爭變得更人道，也不能讓武裝衝突太容易發生，以免強國利用自己的技術優勢來支配其他國家。因此，AI法及資訊法專家法蘭克．巴斯夸利教授就主張「必須要求指名機器人背後的控制者及擁有者」，而有責任歸屬的形式，才能有效阻止AI的軍備競賽[25]。

隨著AI軍事應用逐漸成為全球新挑戰，包括美國與中國在內之60多個國家（但不包括俄羅斯與以色列）於2023年2月中旬在荷蘭海牙（Hague）高峰會上，均簽署一項聲明文件，支持使用負責任的軍事用途AI武器。儘管這項聲明文件不具任何國際法律之拘束力，惟仍被外界視為有可能朝向制定AI國際武器條約的第一步，此亦為國際社會首次舉辦這類的會議。然而另人憂心的是，目前仍很少有國際法規來定義需要人類參與軍事用途AI武器的程度，若科技達到完全不需人類參與的程度，則我們將可能看到AI如何開始決定人類的生死。

伍、從歐盟AI法來看自主性武器規範之適用可能性

歐盟號稱AI及其監管為其首要任務。歐盟針對AI技術帶來的劇變與影響，近年陸續提出相關準則或策略報告，以達成未來法規調適的目標。歐盟在2019年4月委由專家委員會

註23：舒孝煌，〈美國國防部更新自主化武器指導方針〉，國防安全雙週報，2023年2月22日，網址：https://indsr.org.tw/respublicationcon?uid=12&resid=1942&pid=3802&typeid=3（最後瀏覽日期：2023年5月15日）。

註24：關鍵評論網，〈軍事大國都在搶佔優勢、紐西蘭卻推動自主性武器禁令，紐官員憂量產恐降低戰爭門檻〉，2021年11月30日，網址：https://www.thenewslens.com/article/159667（最後瀏覽日期：2023年5月11日）。

註25：法蘭克．巴斯夸利，前揭註20，第224頁。

提出《可信賴的人工智慧倫理準則》（Ethics Guidelines for Trustworthy AI）[26]，說明4項倫理原則及7項要求（風險控制、透明、個資及隱私保護、社會責任、增進人類福祉、問責制、無偏見歧視）[27]，並表示AI發展到自主決策時，傳統民事歸責理論可能有所不足，於是針對「嚴格責任」之適用進行評估，可能修訂歐盟法律，且考慮是否需要訂立新的法規。

是以，繼歐盟於2019年發布《可信賴人工智慧倫理準則》後，歐盟委員會啟動一種三管齊下的立法框架來監管AI，包括與AI相關的民事責任規則及通用產品安全條例，以支持建構可信賴的AI倫理規範。

歐盟執委會於2020年2月發布兩份資料，分別是《人工智慧白皮書》（White Paper on Artificial Intelligence）[28]，就政府機關而言，應該如何規劃未來的AI發展法律管理框架？而歐盟《人工智慧白皮書》提出以「風險」為基礎的方法（risk-based approach）作為日後法規管理框架，人工智慧在高風險領域的應用原則受法律較高度的監管，以確保政府監理干預合乎比例。對企業來說最關心的就是如何判斷哪些AI的應用會被認定是高風險？歐盟執委會提出主要應從安全性保護及消費者權利與基本權的觀點來看待，並可以下述二個標準依序檢視：

一、首先是判斷AI是否被應用於風險較可能發生的行業，如醫療、交通運輸、能源產業及部分的公部門。上述行業清單應隨實務上相關發展而定期被重新檢討及修正

二、其次則是當在特定行業中以特定方式應用AI時，將有極高的機率發生風險

就此，七大工業國（G7），包括正式成員國（英國、加拿大、法國、德國、義大利、日本及美國），外加非正式成員歐盟之數位部長，於2023年4月30日在日本群馬縣舉行會議，針對AI產業商討風險性（risk-based）監管政策，並達成初步共識為「負責任之AI應用」，鼓勵AI研發同時加強各項風險控管[29]。由上可知，世界各國也在參考與追隨歐盟之立法例。

歐盟執委會於2021年4月21日提出《人工智慧法》（Artificial Intelligence Act，以下簡稱「AIA」）草案[30]，其篇章架構如下表所示。這是全球第一個概括性、對於非特定商

註26：2019《Ethics Guidelines for Trustworthy AI》，網址連結：https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=8248（最後瀏覽日期：2021年4月6日）。

註27：歐盟《可信賴的人工智慧倫理準則》之七項要求：人類自主性與監控、技術穩健性與安全性、隱私與資料治理、透明度、保持多樣性、不歧視與公平、社會與環境福祉及問責制。

註28：EU Commission, WHITE PAPER on AI-A European approach to excellence and trust, Brussels, 19.2.2020 COM (2020) 65 final.

註29：陳穎芃，〈對AI採取風險性監管G7成員國達共識〉，工商時報，2023年4月30日，網址：https://ctee.com.tw/news/global/854062.html（最後瀏覽日期：2023年5月11日）。

註30：歐盟議會已於2023年5月9日通過AIA所有修正案全文，並安排於2023年6月13日完成最後三讀投票，為全球第一部人工智慧監理法規。

品或特定領域，針對AI之現象及風險所做之立法嘗試，因AI科技之應用，將對於公共安全、公共衛生、基本權，造成危害與風險，故AIA將AI科技當成一項商品或服務進行管制，並仿效歐盟商品安全法之管制模式，針對不同級別之風險採取不同強度之管制[31]。

表：2021年版歐盟AIA，共13篇、86筆條文[32]

第一篇	總則（GENERAL PROVISIONS） 立法意旨、適用範圍、定義
第二篇	被禁止實施之人工智慧實施（PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES）
第三篇	高風險之人工智慧系統（HIGH-RISK AI SYSTEMS） 分類、要件、風險管理系統、數據管理、技術文件、記錄保存、向使用者提供資料、資料透明性、人為監督、準確性、穩健性、網路安全、提供者、使用者、其他當事人義務、標準、合格評定
第四篇	某些人工智慧系統之透明性義務（TRANSPARENCY OBLIGATIONS FOR CERTAIN AI SYSTEMS）
第五篇	支持創新的措施（MEASURES IN SUPPORT OF INNOVATION）
第六篇	治理（GOVERNANCE）
第七篇	歐盟獨立的高風險人工智慧系統之數據庫（EU DATABASE FOR STAND-ALONE HIGH-RISK AI SYSTEMS）
第八篇	上市後監控、資訊共享、市場監督（POST-MARKET MONITORING, INFORMATION SHARING, MARKET SURVEILLANCE）
第九篇	行為的規範（CODES OF CONDUCT） 鼓勵與促進有意使AI系統自願適用的行為規範之制定
第十篇	保密與處罰（CONFIDENTIALITY AND PENALTIES）
第十一篇	權力的委任與委員會程序（DELEGATION OF POWER AND COMMITTEE PROCEDURE）
第十二篇	最後條款（FINAL PROVISIONS）

資料來源：本文整理

AIA草案中首先就AI系統（AI system）定義相當寬廣，凡使用一種或多種其附件[33]所列之機器學習、邏輯與知識方法或統計學方法，並可以對一組人為給定的目標，產出任何結果例如內容、預測、建議或決定之軟體或程式，均屬之。

註31：歐盟AI法制草案，除AIA係整體框架性之管制規範（前端管制AI與預防損害）外，另有後端配套之《AI責任指令》（AI Liability Directive）與《產品責任指令修正》，前者係針對侵權行為損害賠償責任，後者係針對產品責任與消費者保護規範。

註32：參「Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS」，網址：https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206（最後瀏覽日期：2023年5月15日）。

註33：AIA附件所列AI開發技術和方法包括但不限於：機器學習之監督式、非監督式、強化式及深度學習等各式方法。

關於AIA這部法律之立法目的，係為可保障人們、企業安全與基本權利，而對於AI之實行、創新等做出規範，在人們可以信任之狀態下在各個領域運用AI。故其適用相當廣泛。AIA已於2022年12月初時通過歐盟理事會（The Council of the European Union）之立法程序，本文撰稿時已進入歐洲議會（European Parliament）之最後審查程序[34]。

歐盟AIA草案依循《人工智慧白皮書》以「風險」為基礎的方法（riskbased approach），將AI系統分為四個風險層級：不可接受的風險（Unacceptable risk）即禁止使用的AI、高風險（High-risk）、有限風險（Limited risk）及最低風險（Minimal risk），不同風險層級的AI系統需遵循不同的規範，而AIA草案主要係針對「不可接受的風險」及「高風險」AI進行規範。

其中，「不可接受的風險」是將其對於人們安全、生計與基本權利等具有顯而易見的威脅的AI系統，是不被允許且被禁止運用的。例如，像是可操控人類行為且規避使用者自由意志的AI（歐盟所舉的例子是，使用語音輔助的玩具及鼓勵未成年人的危險行為），目前這樣的例子並不多，但只要是非以人本概念或以人類為工具的AI就不允許其運用。

被認定為「高風險」的AI（僅得作為輔助使用）運用，在所列的第八項用途：利用AI於司法與民主程序。在司法系統中，關於審判的主體是以法官為核心，強調以人類依據立法者所制定的法律，以正常邏輯推理及價值判斷，理性地做出合法且合理的裁判。此為人類所專屬，法律的體系不同於數學計算，並非套用一定公式就可以得出必然且絕對的答案，對於事實的合理認定、法律靈活的運用及法律價值的實現，是AI所無法取代的價值。AI的運用僅得在於查找及分析資料所用，或者部分機械式的簡單程序，擔任輔助的角色。

AIA適用主體係於歐盟市場上AI系統之所有提供商與使用者，即旨在為歐盟各成員國創建一個全面、統一的AI監管框架，對歐盟成員國具有直接的法律效力。毫無疑問，AIA之影響係極為巨大的。從AIA可以看出，歐盟在監管模式上更加側重「綜合全面」的頂層監管框架。從業務角度，以AIA為基礎對AI提供商、服務主體與使用者做出明確的基礎規定，如同當年的歐盟一般資料保護規則（GDPR）影響力，AIA將會在這基礎上，再逐步拓展適用主體於歐盟以外之區域、國家與企業，以循序漸進方式一步一步加強全部AI之監管。

然而AIA在適用範圍方面似乎仍有限，即在「專為研究國防與國家安全目的而開發之AI系統」遭排除在外[35]，換言之，例如AI自主性武器之研發，只要係基於國防或國安目的之

註34：參〈Council of the EU Proposes Amendments to Draft AI Act〉，網址：https://www.wsgr.com/en/insights/council-of-the-eu-proposes-amendments-to-draft-ai-act.html#2（最後瀏覽日期：2023年5月11日）。

註35：在AIA草案第2條第3項關於適用範圍有明文排除之：「This Regulation shall not apply to AI systems developed or used exclusively for military purposes.」。

研發，均不受管制或監理。

惟本文認為，上開條文規範應仍容有解釋空間。若反面解釋該等規定，關於「不可接受的風險」（對於人們安全、生計與基本權利等具有顯而易見的威脅），或「高風險」（僅得作為輔助使用）等AI系統，此類技術應用若具有侵入性，包括已違反歐盟價值觀及侵犯基本權利，抑或造成人身安全或基本權利負面影響者，例如國家利用自主性武器主動攻打他國，則該等AI自主性武器之研發或應用，即非「基於國防或國安目的之研發」，故應仍受AIA甚至國際法規範之管制或監理，違者應給予相關嚴厲制裁。

就此，本文認為為確保任何戰爭機器人或自主性武器系統之問責性，上開武器之系統在技術上應具有「可追溯性」，即試著透過AI系統行動的特徵與已知之控制者連結，以解決「責任歸屬問題」，自主性武器均應表明其創造者、控制者或擁有者之身份，此亦為戰爭之基本原則，若有違反將受到嚴厲制裁[36]。

申言之，基於歐盟AIA所課予之資訊透明責任，全面將進口商、通路商，甚至包括使用者，均清楚定義為佈建者（Deployer）之概念，是以，舉凡任何與自主性武器產品或服務相關者，包括研發上需提供大數據或AI演算模型，則AIA適用主體涵蓋AI武器系統資訊之原始提供商與使用者（即使用AI系統之廠商為終端客戶），此佈建者就可能係原廠建構AI系統之企業例如Google等[37]，理應受到規範。

陸、代結論──台灣AI基本法草案？

全球面對AI時代發展，對於法制層面及環境上的需求益增，尤其AI在智慧財產法制上的爭議討論度愈來愈熱烈。然而，台灣當前似乎仍欠缺前瞻性的遠見來面對一波波AI的浪潮，我們冀望能透過新的法規範思維來建構適切的實務運作環境，特別是AI科技法律面分析（政府作為或規範、業界應注意事項或管理方向），需要法律界與各界搭起溝通與討論的橋樑。

AI可透過演算法去自主產出領域知識，且只要輸入資料就或下達指令可達成結果。我們的世界已面臨必須就AI做出的某項決策，該如何描述該AI的意思表示、應由誰去承擔AI決策的風險，或應如何分擔此種風險？此外，AI演算法的監管也迫在眉梢，因為AI生成結果均有可能侵犯智慧財產權（包括著作權及專利權等）、資訊安全。而隨著AI時代造成資料經濟市場的轉變，也逐漸地改變法律體系及人民的法意識。

台灣產業AI技術主要落在模式辨識、語意分析、知識系統，故資料蒐集及數據運用相關法規應優先考量，作為我國資料法制的大方向；且我國產業主要以中小企業為主，主

註36：法蘭克．巴斯夸利，前揭註20，第241頁。

註37：Google與其他科技業者掌握美國人工智慧大多數人才，上開科技業者雖非傳統之軍事承包商或武器製造商，惟因多年來發展與累積深度學習系統所需之專業基礎架構，實質上早已涉足軍事相關領域之消費性科技服務或產品。

要研發重點通常為產品研發、效能提升等，不像國際企業如Microsoft、Google、Meta等皆已注意到人工智慧倫理問題，因此可能忽略相關倫理議題的重要性，故須由台灣政府協助推動，以利未來產品銷往國際時符合國際法規及標準，並提升其競爭力。我國科技部在2019年間催生一部《人工智慧科研發展指引》，其內容與歐盟2019年《可信賴人工智慧倫理準則》大同小異。由於目前我國尚無AI專法，因此尚無法透過法律規範產業適用，然仍可協助產業運用倫理準則，可從規模較大的產業進行推動，或透過政府資助的科技計畫先行推廣。

台灣前幾年曾有立委提出《AI發展基本法》草案[38]，就是為落實「透明、可控制性、隱私保護」等原則，嗣後不了了之。目前諸多廠商均陸續推出各式AI應用，各國政府更已開始針對AI新興科技擬定法規，此時台灣政府更應儘速回應AI新科技的高速發展，擬定合適的法規框架，以避免面臨開放範圍太小、不符合經濟效益，以致於過度限縮產業發展的窘境。

隨著這波ChatGPT等盛行風潮，在台深耕AI法制多年之人工智慧法律國際研究基金會，亦提出一版AI基本法草案，其主要重點如下表。

表：2023年版人工智慧法律國際研究基金會之AI基本法草案重點

1.AI倫理原則法制化 明定AI之研發及利用應以人為本，普惠人民及永續發展為目標。
2.以國家高度主導之產業發展策略 AI應由國家主導政策並擬定發展計畫，落實人才培育、建立基礎設施及推動產學合作。
3.強調社會公平與弱勢保障 AI可能造成社會資源分配不均，政府應落實對於需要協助族群、勞工權益及公平交易秩序之保護。
4.重視隱私及個人資料保護 為維護人民隱私權益，政府應對AI開發與應用所需之資料蒐集、處理及利用，建立必要的保護及監督機制。
5.建立完善的人工智慧監理沙盒 政府應設置創新實驗環境，提供相關研發與利用之安全場域及實驗空間，以妥善評估創新技術之潛在效益與風險。
6.依AI產品或服務之風險高低區分管制嚴寬 政府應建立AI研發及利用之風險評估及監管機制，並依風險高低與程度進行適當管制，以平衡新科技之風險。

資料來源：本文整理[39]

雖然上開草案用意良善，惟本文建議如能參照前揭歐盟AIA關於自主性武器規範有疏漏之處，並予以增加相關章節內容，應更能符合台灣現況及未來之需求。

註38：前立法委員許毓仁等21人、委員鄭麗文等20人提出草案版本名稱均為「人工智慧發展基本法草案」，台灣民眾黨黨團提出草案版本名稱為「人工智慧發展法草案」，略有不同。

註39：立法院第10屆第7期財政委員會第8次全體委員會議，《為因應金融業導入人工智慧（Artificial Intelligence）科技（例：ChatGPT）如何跨部會強化數位基礎建設、擴展金融科技監理路徑與完善資安法規，以加速我國金融科技發展》，法務部專題報告，第3-4頁。網址：https://ppg.ly.gov.tw/ppg/SittingAttachment/download/2023041383/307003120421082100 2.pdf（最後瀏覽日期：2023年5月15日）

尤有甚者，自主性武器系統之研發與擴散，為各國發展科技軍事武器及國際人道法帶來巨幅轉變，其內涵似非既有國際法規範及特定常規武器公約所能完整詮釋。基此，有論者主張以限制無過失責任作為國家責任的主觀歸責門檻，透過此一高度歸責門檻之設定，能敦促國家對AI新興科技軍事武器的使用更加審慎以對[40]。本文十分贊同，因為或可供國內法在制定AI基本法規範之際，增添一些更符合本國甚至國際需求之相關規定。申言之，領導人應根據國際人道法之基本原則──「軍事必要性」，試圖全面禁止某些殺戮功能或方法，亦要求指揮官在追求成實務與人性的責任之間取得平衡。而人性等概念的模糊性，正是AI演算法邏輯的障礙，需要藉由人類的價值觀去明確判斷、抉擇及觸發相關行為。

誠然，早有論者[41]在多年前撰文點出，自主性軍事武器系統之科技發展事實，與國際規範適用或法律評價上，形成相當大之落差與爭議；其研究結論係以國際社會現實較可能達共識之規範制度為基底，再以國際規範與原則為基礎法律規範框架，進而定義自主性武器系統、自主程度分級、武器適法性標準與探討，再歸納既存法對於自主性武器系統合法使用、《國際人道法》之遵守、軍備競賽以及法律究責等之規範。本文認為，在台海情勢詭譎多變、AI發展日新月異、自主性武器系統發展成熟之際，台灣應就此多加研究相關規範與準備因應之道。或許能藉由管制AI武器之治理方式與規則，有助於將潛在戰爭轉變為維護和平之契機。

註40：林昕璇，前揭註17，第20-44頁。

註41：李鈺翔（2018），《從國際法論人工智能軍事武器之發展與挑戰》，東吳大學法律研究所碩士學位論文。