

人工智慧治理與包容

台灣人工智慧學校基金會秘書長 侯宜秀

當我們談到人工智慧治理時，常著眼於各國關於人工智慧（artificial intelligence, AI）的立法，於是迅速地進入國際間各種規管模式的比較，例如：歐盟採取以「風險分級」架構，在《人工智慧法》（AI Act）中從產品責任出發，全面性地規範各種角色對AI產品和服務需負擔的責任義務，強調消費者人權保障與高風險系統的預防性管理。美國則偏向市場導向與產業自律，主要以自願性標準（如NIST風險管理框架）與行政命令推動治理，著重創新與競爭力，但不同部門或州政府各自發展規範，也導致規範碎片化、難以遵守的困境。中國則由政府主導，結合強制性管理與內容審查，著重國家安全與社會穩定，透過一系列暫行辦法與行政規定快速建立規管框架，要求企業履行審查與風險控制責任。

從這個角度談治理，強調的是政府角色，立法與行政部門制定法律與行政命令，司法部門依據這些立法規範進行判決。這樣的治理框架在處理與科技相關常顯得捉襟見肘、左支右絀，AI立法最顯而易見的挑戰是：AI技術日新月異，立法過程往往趕不上技術演進速度，容易導致法律內容過時。此外，「人工智慧」本身是一個概念模糊的總稱，涵蓋機器學習、深度學習、生成式AI等多種技術，使得精準立法變得困難。技術的複雜性也讓非技術背景的立法者難以掌握其風險與應用情境。

政府與法律規範是治理的重要基礎，但在人工智慧治理中，參與的行動者與規範的形成過程，遠比通過法律更為複雜。

上世紀末，1999年，Lawrence Lessig教授寫了《Code: And Other Laws Of Cyberspace》（注1）這本書。書中描述了虛擬世界規範的構成包括：原始碼、市場、社會規範、以及法律。這四種力量共同塑造了虛擬世界的規範，決定了數位空間中的行動者能做什麼、不能做什麼。

Lessig教授認為，在虛擬世界中**原始碼是所有規範的基石**。程式碼不僅僅是技術層面的指令，它更是架構和限制行為的「建築師」。例如，一個網站的程式碼可以決定使用者是否能匿名發言、訊息是否加密、內容是否能被複製或分享。這些內建的技术限制，直接影響了使用者在虛擬環境中的自由度和互動方式。又以區塊鏈技術為例，其原始碼設計從根本上定義了加密貨幣的交易方式和不可篡改性。當我們登入一個平台，其後台的程式碼早已為我們設定了遊走於其間的邊界。

市場，則透過經濟獎勵和懲罰來規範虛擬世界的行為。例如，一個受歡迎的線上遊戲可能會推出

付費道具或服務，鼓勵玩家投入更多時間或金錢，以此影響他們的遊戲行爲。例如在電商平台上，賣家的評分高低直接影響其銷售額，從而促使賣家提供更好的服務。

此外，在虛擬世界中，我們能夠見證**社會規範在短時間內迅速形成**。這些規範是社群內不成文的行爲準則，透過成員之間的互動和共同認知來形成。例如，在論壇或社群媒體上，使用者會逐漸發展出關於禮儀、討論方式、內容分享界線等方面的共識。違反這些社群規範的行爲，可能會導致被排斥、批評甚至被驅逐出社群。

最後，才是實體世界的法律對虛擬世界的直接干預和規範：我們所熟悉的法律規範手段。

Lessig教授所提出的觀察，在過去近三十年虛擬世界的規範發展中，不斷被印證。時至今日，對於AI的治理，這套框架仍舊適用。自2022年ChatGPT問世以來，實際上有能力與進行規範大型語言模型服務的，是模型開發公司以及服務提供商，以及緩步形成中的某些行業自律規範。由此可知人工智慧的治理仰賴開發者、市場、公民組織、政府的協作，本質上就是一個多方利害因素交互作用的結果，如何引導這些多方利害關係人能朝向共同的方向前進，建構以人爲本的治理機制與規範，不僅是人工智慧治理成功關鍵，更是「包容性原則」的具體落實。

「包容性原則」的建構，是現代人權與發展思想的核心。1948年《世界人權宣言》奠定了人人生而平等、享有尊嚴與基本權利的核心理念，爲後續各項全球治理規範提供價值基礎。進入21世紀，聯合國於2015年提出「不遺漏任何人」（Leave No One Behind）的永續發展目標（SDGs）總綱原則，進一步強調在經濟、社會與科技發展過程中，必須特別關注最邊緣與最弱勢的群體，確保他們能公平地享有資源與機會。

隨著數位科技快速發展，世界資訊社會高峰會（WSIS）於2003年與2005年先後召開，首次將「資訊社會應具備包容性、以人爲中心、強調社會正義與文化多樣性」等原則納入全球共識，爲數位時代的治理價值奠定基礎。資訊科技的快速普及一方面

創造了前所未有的機會，但也暴露出數位落差、資訊不平等與文化邊緣化的風險。

在人工智慧技術迅速發展後，「包容性原則」更在全球多項AI治理框架中被重新強調並制度化，包括OECD《AI原則》、UNESCO《倫理AI建議書》、G7《廣島進程行爲準則》等，皆主張在AI的設計、開發與部署過程中，應納入多元文化、性別、族群、能力與地理背景的觀點，防止演算法偏誤進一步加劇社會不平等。AI技術的黑箱性與規模效應使其對社會結構的影響更加深遠，因此唯有確保包容性，讓不同聲音與需求有機會參與決策與治理，才能讓AI成爲促進社會公義與永續發展的工具，而非不平等的放大器。

我國行政院國科會在2024年7月15日科會前字第1130048999號所公布的「人工智慧基本法」草案第三條立法理由中載明：「人工智慧研發與應用需公平、完善且演算法應避免產生偏差或歧視之結果，爰參考美國二〇二二年AI權利法案藍圖（Blueprint for an AI Bill of Rights），於第六款定明公平與不歧視原則（Fairness and Non-discrimination），強調應重視社會多元包容，避免產生偏差與歧視等風險。」（注2）

但究竟「包容性原則」這個抽象概念，應該如何具體落實？勞動部「1955 勞工諮詢申訴專線」智慧化的過程，是一個可參考的案例。

「1955勞工諮詢申訴專線」是由台灣勞動部於2001年所設立，原始構想是爲了解決當時國內勞工在勞資爭議、勞動條件及職場安全等問題上，無法即時獲得協助與正確資訊的困境。該專線一開始僅提供國語諮詢服務，後來因應外籍移工人數大幅增加與多語需求上升，陸續擴充包含印尼語、泰語、越南語、英語等多語客服服務，並由伊甸基金會承接部分運作與客服人力管理工作。伊甸也特別聘用具備語言能力與專業知識的身心障礙者參與專線服務，讓「1955專線」同時成爲政府推動**身心障礙者就業支持政策的實踐場域**。

在COVID-19疫情期間，這個專線突然面臨極大的挑戰，單日來電一度飆破三萬通，致專線嚴重塞

爆，許多求助者無法接通，客服人員工作量過大，尤其身心障礙同仁的疲勞與壓力亦顯著增加，而有了是否可以用人工智慧技術解決問題的討論。

客服服務因為其**重複性、標準化、高頻率**的任務特性，一直被認為是最會廣泛運用人工智慧技術的應用場景，早在生成式AI技術出現之前，就已經有許多實際運用案例。因此當勞動部在面對「如何用人工智慧技術解決1955服務量能不足？」這個問題時，其實有兩種選擇路徑，一是以AI取代既有的客服人員，另一則是以輔助既有客服人員為主的輔助型AI服務。前者以技術為導向在經濟效率上有可能是最佳解，後者則是在提升效率的同時也兼顧客服同仁的權益，是落實包容原則的路徑，但同時也將面臨許多人機協作、人人（系統開發者、系統使用者、系統擁有者、終端客戶）溝通、磨合的挑戰。

最終勞動部2021年的1955專案以「人機協作」為核心架構，而非以AI全自動化取代人工，採取人機互補的架構。設計了以下的服務架構：（注3）

- 24小時智慧文字客服：自動回應常見問題，即時提供基礎諮詢與指引，減少人工負擔，確保全天候服務可用性。
- 真人文字客服（正常上班時間）：由客服專員（含身心障礙者）透過線上文字交談處理較複雜的查詢與個案需求，提升使用者滿意度。
- 真人電話諮詢（正常上班時間）：針對需要即時互動或複雜溝通的案件，提供專人電話支援，以確保精準解決問題。

該專案透過AI處理常見問題與標準回應，讓真人客服專注於複雜案例與高互動需求的服務工作，發揮人機各自優勢。這種設計不僅保留了人的價值，也提升整體工作品質與服務彈性，是「以人為本」AI應用的重要體現。

更重要的是，專案不僅讓身心障礙者成為AI的使用者，更讓他們實質參與AI的設計與訓練。客服人員不只是操作機器的人，也投入AI訓練資料的蒐集、標註與驗證，協助構建知識庫系統。一位參與者甚至成為台灣首位全盲AI數據標註師。透過這樣

的參與式資料治理設計，身障者從被動使用者轉變為主動貢獻者，也藉此提升數位素養與職場自信，實現真正的「賦能」（empowerment）。

技術可近性（accessibility）亦是包容性AI不可或缺的條件。1955智能客服系統依循無障礙網站設計標準（AA級）建構，支援螢幕閱讀器、語音輸出與鍵盤操作等功能，確保視障者可順利操作並參與系統維護。這不僅符合《身心障礙者權利公約》對數位平權的期待，也實踐了「設計即包容」的理念。

透過AI的引入，身障客服人員不僅保住工作，甚至透過再培訓與參與技術開發，逐步轉型為新型數位人力。根據相關報導，自專案導入後，部分客服人員的薪資在三年間成長約25%，服務量則提升超過兩倍，展現出AI不僅不排除弱勢，更能擴大其參與與影響力。

包容性原則在人工智慧發展中，是確保科技公平、預防差異擴大的關鍵，不僅在法制上必須納入以做價值宣示，更要邁向制度實踐。從人民個人能力建構出發，透過多元參與機制、包容性設計、與權益救濟等具體措施，讓不同背景與弱勢群體能在AI發展中被聽見、被納入、被保障，確保AI科技進步真正服務所有人，而非複製甚至擴大既有的不平等，透過建構包容的AI治理，建構AI時代多元包容的社會。■

參考文獻

注1：Lessig, L. (1999). *Code: And other laws of cyberspace*. Basic Books.

注2：國家科學及技術委員會（2024）。預告制定「人工智慧基本法」草案（中華民國113年7月15日，科會前字第1130048999號）。公共政策網路參與平臺。<https://join.gov.tw/policies/detail/4c714d85-ab9f-4b17-8335-f13b31148dc4>

注3：丁玉珍（2025）。AI應用於勞動業務創新轉型。《國家人力資源論壇》，（35），2025年4月25日。考試院。https://www.exam.gov.tw/NHRF/News_EpaperContent.aspx?n=3778&cs=49777&ctype=D78770E01F7001AC