

# 多元開放的全球人工智能治理

• 張立健

自基於大型語言模型 GPT-3.5 的 ChatGPT 於 2022 年底上線以來，這款由微軟 (Microsoft) 投資、OpenAI 開發的生成式人工智能帶來的巨大衝擊很快浮現，其一問一答的互動模式在網上迅速流行。此後兩年多，來自矽谷的科技產品的發布一再令大眾驚喜，日常文書處理器以及電子設備都配備了類似 ChatGPT 和 GPT-4 的人工智能 (AI) 助理，只需簡單按鍵，便得到電腦生成的書信草稿、會議記錄、試算表格和演講大綱，處理文書工作更省時省力。主流媒體和社交媒體一直討論其直接或間接的影響，比如此程式可以輔助寫作、研究和整理資料，甚至能像作家般撰寫故事、像導演般製作影片；有人測試過人工智能工具能通過醫生和律師的資格考試，令人擔心連人類的專業工作也可能會被電腦取代；更有人利用相關技術來製作假影音文字以詐騙他人<sup>①</sup>。

儘管科技經常被形容為獨立於社會，無論哪個時代，人們只能不由自主地接受技術帶來的轉變，可是科技與社會研究 (Science and Technology Studies) 學者則持異議，他們認為科技由人所發明，科技亦因其依存於特定的時間和空間，既服務於被設定的目標，亦可能被反向塑造其用途。因此，科技與社會並不是單向的關係，而應該是互相塑造的。已故美國人類學家格雷伯 (David Graeber) 和英國比較考古學家溫格羅 (David Wengrow) 於《萬事揭曉》(The Dawn of Everything: A New History of Humanity) 一書中便指出，雖然科技被認為如此重要，難以否定其在歷史上對社會有決定性的影響力，但不能過份誇大地指稱社會演進全靠科技創新，而忽略了人的參與<sup>②</sup>。人類可以從日常生活或社會儀式中主動將科技分類為合用的與不合用的，就算現時存在潛在威脅到人類生存的智能與機器，也有可能未來被人類主動拋棄。因此，人類不是完全被動地被發明出來的概念或技術系統所控制，不能自拔，而是通常有選擇權，甚至主動設計多餘的科技產物。

本文將按科技與社會互塑的理論，來討論國際形勢與人工智能浪潮的關係，並提出兩個觀點：第一，人工智能的風險不在於商業化或政治化，而在於其封閉性——僅由矽谷科技企業、風險投資公司、英語世界主導；第二，人工智能議題乃全人類的大事，必須全球共議，而非任由美國一霸或中美兩國主導。對此，文末提出東亞社會不應坐以待斃，科技議題尚待更多公民參與，亦有待建立跨國共議、在地調適的多元開放的全球治理機制。

## 一 國際局勢形塑人工智能發展

前美國國務卿基辛格 (Henry A. Kissinger) 生前與美國政治學家艾利森 (Graham Allison) 在《外交事務》(Foreign Affairs) 雜誌聯名發表〈人工智能軍控之路〉(“The Path to AI Arms Control: America and China Must Work Together to Avert Catastrophe”)一文，指出與核武發展不一樣，人工智能並無實體，相對難以觀察，而且其進步之快令學界難以討論管制的空間。人工智能技術研發不只由政府牽頭，企業家、科技專家和企業也在推動，這些個體和私人部門未必會在國家安全、風險和盈利之間取得一個平衡。因此論者認為，人工智

CURRENT ISSUE ARCHIVE BOOKS PODCAST
**FOREIGN AFFAIRS**
NEWSLETTERS LOG IN SUBSCRIBE MENU ☰

# The Path to AI Arms Control

America and China Must Work Together to Avert Catastrophe

HENRY A. KISSINGER AND GRAHAM ALLISON

October 13, 2023



A Tesla robot in Shanghai, July 2023  
ALY SONG / REUTERS

**HENRY A. KISSINGER** is Chairman of Kissinger Associates. He served as U.S. National Security Adviser from 1969 to 1975 and U.S. Secretary of State from 1973 to 1977.

**GRAHAM ALLISON** is Douglas Dillon Professor of Government at Harvard University. He is the author of *Destined for War: Can America and China Escape Thucydides's Trap?*

[More by Henry A. Kissinger →](#)

[More by Graham Allison →](#)

Share & Download
T

his year marks the 78th anniversary of the end of the deadliest war in history and the beginning of the longest period in modern times without great-power war. Because World War I had been followed just two decades later by World War II, the specter of World War III, fought

Print

前美國國務卿基辛格生前撰文呼籲中美在人工智能軍控問題上合作。(資料圖片)

能的發展需要進行跨越國內外、公私營的全球性討論與分析。當然更重要的是，要在美國國內設立針對人工智能的無黨派偏見的委員會，由企業、國會、軍方和人工智能專家的領袖所組成，來尋求兩黨共識與回應。而在國際上，作為人工智能兩強的美國和中國需要加強合作，召開高峰會議，讓兩國的科學家和政府人員交流意見，然後聯手建立起像國際原子能總署 (IAEA) 般的國際框架，以管制人工智能武器的無序擴散，從而訂立屬於人工智能的國際秩序<sup>③</sup>。正如東京大學工學研究所學者松尾豐於《瞭解人工智慧的第一本書》中一再強調提醒，人工智能技術前景明朗，但必須小心兩項風險：被戰爭利用、被少數人壟斷<sup>④</sup>。實際上，人工智能技術加持下的戰爭可能遠比松尾豐所想的還要可怕，例如情報機構利用社交媒體機器人大軍操縱輿論，掀起足以推翻一個政權的風暴。撇除基辛格的美國本位思維，其關於全球共識、公私合作的原則確實值得推廣，可是仍缺乏國際社會的其他持份者如民間團體的參與。

前《外交事務》主編扎卡利亞 (Fareed Zakaria) 在 2024 年美國總統大選之前出版《革命的年代》 (*Age of Revolutions: Progress and Backlash from 1600 to the Present*) 一書，指出目前我們正處於歷史性的「革命時代」，呈現四方面的重要變化：一、自由市場與貿易受到質疑，因而帶來了不確定性；二、數字革命和人工智能帶來的所謂「第四次工業革命」；三、超越左右派的新運動衝擊舊有國內的政治秩序；四、特朗普 (Donald J. Trump) 試圖回復的過往光榮的美國、崛起中的中國與復仇主義的俄羅斯，三者將挑戰現時國際格局。冷戰結束後，全球政治曾有過相對穩定時期，由於美國成為單極霸權，其所鼓吹的全球化、自由化、民主化成為全球主流思想，而全球化也為人類帶來了不少正面意義，包括壽命延長和全球連結。可是由於科技、經濟與身份認同三股力量合流，帶來了新的政治形態，同時中國、印度、巴西、土耳其亦漸漸崛起，即所謂「其餘國家的崛起」 (the rise of the rest)，動搖了此前的國際地緣政治格局<sup>⑤</sup>。

筆者認為全球的社會政經格局變化，勢將形塑未來的人工智能和機器人創新，尤其會帶來現時難以預測的潛藏風險。以色列歷史學家哈拉瑞 (Yuval N. Harari) 於新作《連結》 (*Nexus: A Brief History of Information Networks from the Stone Age to AI*) 中指出，各國正在發展出自己的人工智能模型，可能產生與冷戰東西方國家之間鐵幕類同的分野，未來可能形成以資訊科技為界線的「矽幕」 (silicon curtain)，對立的兩大數字帝國陣營漸行漸遠；他認為現在是人類從趨同走向趨異的歷史關鍵點。政經局勢多極化，地緣政治對立加劇，進一步刺激各國投入人工智能的軍備競賽，謀求擺脫對競爭對手乃至敵國的依賴；兩大陣營科技方面的硬體與軟體不相容，兩邊的資訊、文化與規範愈走愈遠，勢將帶來文化轉變與暴力衝突。假如理解社會與科技的互動互塑，便可知風險不只來自於科技系統本身，也來自社會結構<sup>⑥</sup>。

目前輿論普遍忽略科技被壟斷之後衍生的後果。當下，人工智能的對話和網際網絡的搜尋正在改變人類觀看事物、參與和理解世界的方式，此技術由五六間主要來自美國的科技巨頭所主宰，往往以英語定義甚麼是知識、怎樣溝通和如何解決問題；由於演算法和生成式人工智能仰賴數據，而這些經由提取、測量、標記和排序的數據本來帶有某種世界觀、認知偏誤或技術性偏誤，可能放大現時的不公平，進而強化了歧視問題。比如美國華裔科學家李飛飛的ImageNet固然推進了電腦分析圖像和深度學習，可是這套演算法最為人詬病的是，其對種族和國族身份的可量化分類，以生物和文化作為事物的本質和秩序，間接固化了種族之間的差異。故此，機器學習系統按照其設定來建構世界，將偏頗地理解甚麼是常態、異常、差異和他者，同時亦限制了人工智能如何理解世界，進而引起社會不公甚至動盪<sup>⑦</sup>。結合以上討論可知，人工智能的風險不在於商業化，亦不在於政治化，而在於其封閉性，缺乏多方參與、平等共議的平台。

## 二 科技巨頭對人工智能的壟斷

面對人工智能的衝擊，2023年未來生命研究所(Future of Life Institute)發起〈暫停大型人工智能實驗的公開信〉(“Pause Giant AI Experiments: An Open Letter”，以下簡稱〈公開信〉)，呼籲各界深思人工智能滅絕人類的可能性<sup>⑧</sup>。不久之後，美國人工智能安全中心(Center for AI Safety)發布〈人工智能風險聲明〉(“Statement on AI Risk”)，提倡把降低人工智能帶來的滅絕風險列為全球優先事項，因其風險足堪與流行病和核戰爭看齊。該聲明吸引了約六百六十名專家聯署，包括後來獲得諾貝爾物理學獎的辛頓(Geoffrey E. Hinton)、OpenAI的創辦人奧特曼(Samuel H. Altman)等人<sup>⑨</sup>。

從兩次大規模聯署可知，各國專家已相信人工智能的威脅不再只是紙上談兵。即使目下人工智能產品仍存在種種不足，但畢竟也達到了一定程度的自動化學習，按照這種趨勢發展下去，其性能將一日千里，在職場、日常生活乃至各個領域佔據更多、更重要的角色。對於令人們又愛又恨的科技產品，正如祖博夫(Shoshana Zuboff)在《監控資本主義時代》(*The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*)一書中所昭示，這是科技精英悉心打造出來的工具，精準地預測個人消費，潛在地監控人類，像控制木偶般引導人們的行為。因此，論者呼籲人們了解信息不對等、科技巨頭的霸權、虛擬金融所帶來的潛在風險。人工智能來勢洶洶，大眾愈來愈擔心此技術太過強大，將可能取代人類的各種功能與角色，成為一個全知、全能、永在的監控系統<sup>⑩</sup>。

雖然科技商業化無可厚非，但新興的風險投資行業助長了少數美資科技巨頭獨大，並持續在全球擴張。互聯網的雲端數據庫、應用程式、通訊制式

和硬件基礎建設等領域的不公平競爭，成為科企的主要權力來源，它們更在2008年金融海嘯之後量化寬鬆和低息投資的環境下，成為受風險投資公司和創業板追捧的企業。科技巨頭如谷歌 (Google) 或 Facebook 母公司 Meta 不會輕易放棄對人工智能領域的經年投資，讓盈利的良機付之東流。一方面，政府、議會和政黨受制於科企的商業表現，科企的影響力甚至足以癱瘓國家的部分政治功能。比如2018年揭發的Facebook與劍橋分析 (Cambridge Analytica) 的醜聞，就是一個重大的資訊安全教訓——Facebook在蒐集用家資訊後，擅自用作其他用途，影響了2016年特朗普首屆當選的美國總統大選和英國退出歐盟公投的結果。另一方面，在經濟和技術進步的包裝下，科技的壞處往往被專家淡化以至合理化，為了促進經濟發展，要求社會大眾接受未知的風險，以技術事先決定社會走向。然而，技術帶來的風險如影隨形，似乎時刻在警惕大眾要有風險意識，於是不安感驅使大眾尋找替罪羔羊。就像2022年谷歌對話程式語言模型 (LaMDA) 所引發的事件：工程師表示該系統擁有像八歲小孩般的感知 (sentient)，該發言流出後引起外界震驚，最後谷歌讓工程師免職來「揸黑鍋」，但相關爭議不會就此停歇<sup>⑩</sup>。這類事件只是轉移了公眾的注意力，減輕對人工智能風險的戒心，卻錯失跟科企對話的契機，並降低社會集體決策的可能。

另一個關注點是機器自動化取代人力勞動，造成大規模「科技性失業」的結構性問題，這也與資本操作、企業營運方式等相關。先是高盛 (Goldman Sachs) 2023年報告指出由於全球勞動效率提升、勞動成本減少，估計未來十年人工智能將使全球國內生產總值 (GDP) 按年增長7%，為全球帶來七萬億美元經濟收益；若生成式人工智能被全面推廣到職場，將會取代三億份全職工作。後是世界經濟論壇 (World Economic Forum) 同年發表報告，預計全球在未來五年內將會有1,400萬個工作職位消失，這將徹底擾亂現時的勞動市場<sup>⑪</sup>。有論者質疑自動化帶來的一切以成本計算、替人人力的傾向，數以千萬計的勞工被機器取代，這可能是無法解決的挑戰。當支持者認為機器人能提供更便宜的服務，而反對者認為機器人正搶走所有工作，社會可能陷於兩難<sup>⑫</sup>，我們將失去與人工智能合作走向理想未來的機會。

當然，由恐懼科技到信任科技之路仍很漫長，單靠簡單的政治承諾和商人的花言巧語是遠遠不夠的，依靠集體反思與決策，更有可能讓人安心走向人工智能、大數據、機器人和元宇宙的科技社會。要讓大眾安心交出人臉特徵、指紋、個人資料、手機通訊和行蹤等資訊，其實最理想的辦法是提供調研和數據，讓大眾正確認識風險，以民主方式協助社會作出選擇。比如德國民眾曾經反對谷歌地圖 (Google Map) 拍攝到自己的面容和房子外觀，認為侵犯私穩。經過訴訟，谷歌地圖的德國街景圖片在2022年再次更新。在爭議過程中，谷歌完善了其私隱政策，德國民眾亦慢慢接受了新科技的發展。由此可見，如要避免衝突，無論是政府、私營企業還是科學界，都必須在公共領域中公開決策過程和接受討論，這項原則在面對人工智能帶來的風險時同樣適用。

### 三 跨國共議人工智能治理問題

社會上有一種聲音認為，只要確保科技得到最大化利用，盡量減少其潛在的負面影響，人類便應該毫無保留地接受所有新技術，並相信科技是解決許多全球問題（從社會不公、食品安全到氣候變化）的關鍵方法。可是如前文綜述，毫無保留地擁抱科技對人類共存會有巨大風險。因此，在人工智能正式全面進佔社會生活之前，我們應該未雨綢繆，減少風險，避免因循守舊，以免日後釀成像福島核災般無可挽救的人禍。與其坐以待斃，不如因應時代的挑戰及時行動，在進入所謂「後人類」(posthuman) 的時刻之前，反覆思量何去何從，以知識和經驗來集體設計全球治理的規範。除了個人私隱需要得到系統性的設計與保護（像歐盟於2016年實施的資安法例《一般資料保護規則》[*General Data Protection Regulation*]<sup>④</sup>）之外，各方也需要找到合適的分享、使用方法，讓公私機構、個人都可以享受到科技的益處。與其被動接受科技，不如設立包括數據使用倫理在內的道德倫理規範，既讓科技服務人類，也教育人類如何使用科技。

在各國社會擁抱科技前行的同時，當務之急是建立多方參與的機制來回應本地需要，以協調異議與反抗，比如各國的工會和工人會否團結起來反對資本和機器，以期阻擋進一步的自動化搶走更多工人的飯碗，出現像英國工業革命後雪萊 (Mary Shelley) 科幻小說《科學怪人》(*Frankenstein*) 對科學進步所衍生之傲慢的控訴，以及放火燒毀機械的盧德主義者 (Luddites) 社會運動<sup>⑤</sup>。近年美國的編劇和演員、配音員等已經就人工智能廣泛應用於電影工作而走上街頭抗議<sup>⑥</sup>。在機器人支援下，各國社會的長者照顧場所、幼兒託管中心、教育培訓機構是受惠還是受害呢？他們會因為科技浪潮而提升生活水平和福祉嗎？此外，我們會否因而活得更孤獨、更乏味？因此，每一個社會中的政府政黨、私企、公民社會和宗教文化機構等持份者應該集中討論，革新國家、科技、經濟連結的系統和角色，協商屬於本地社會的願景和公私平衡的安排。

筆者認為，建立安全有效、能夠保障權利、值得信賴的跨國科技系統至關重要，該系統不能為少數人、私企或政府政黨所壟斷。2023年5月七大工業國 (G7) 首腦峰會所制定的「廣島人工智能進程」(*Hiroshima AI process*)，亦與〈公開信〉持相同意見，指出政府需在人工智能的開發中扮演中介角色，訂立開發者可做和不該做的規則，且對其他政府如何回應這一波人工智能的衝擊提供建議<sup>⑦</sup>。2024年3月聯合國亦就建立人工智能規範達成決議，而歐盟於同月通過的《人工智能法案》(*Artificial Intelligence Act*) 是全球首部人工智能法案，被視為人工智能全球治理框架的里程碑<sup>⑧</sup>，兩份跨國文件可以作為供世界參考的範本，踏出技術規管的第一步。

此外，人工智能既要討論倫理，亦要討論價值，力求科技與社會共同達致可持續發展。正如意大利經濟學者馬祖卡托 (Mariana Mazzucato) 指出，

1947年被聯合國採用的國民所得 (national income) 計法成為戰後世界主流，可是這種會計制度只顧及單一的指數式增長，而忽略了經濟的其他可能<sup>⑨</sup>。與此同理，現時人工智能的算法會否給予某一價值過高權重，忽視各個社會的差異，成為另一種扭曲社會價值系統的工具？筆者認為，人工智能和機器人也涉及很多基本價值，諸如民主、自由、平等、人權。機器及其智慧能為人類帶來福祉是最重要的，能造就更有趣的生活、更多元的參與，不助紂為虐，更不應為世界帶來戰爭、暴力、分裂、不公。就價值的規範而言，反面例子是，美國科企IBM設計的穿孔卡機協助德國納粹黨登記和識別猶太人，該企業賺錢之餘卻間接助長極權監控<sup>⑩</sup>。科技缺乏對價值的重視只有一個結果，便是增加不公平、不公正的社會與擴大不健康的文化，放大權力濫用、貪污腐敗、環境破壞、貧富懸殊、以強凌弱的現狀。人工智能當然會面臨同樣的情況。正如沃爾許 (Toby Walsh) 於《2062》(2062: The World That AI Made) 中提出，應當確保以後的人工智能系統反映人的價值，並且確保「數位人」(Homo digitalis) 比「智人」(Homo sapiens) 更符合倫理<sup>⑪</sup>。凡此種種，我們是否準備好迎接智能與機器人科技對人類的基本價值帶來的挑戰？

#### 四 想像屬於東亞的人工智能發展

以上引述的觀點和提及的私企大都來自美國，不少探討源於在矽谷的實務研發、進修經驗和訪談觀察。我們不應該將人工智能這種非人類系統看成與人類相同的智慧，甚至看成獨立、自然存在的事物，而應注意建基於其背後的社會、文化、歷史與政治。正如前文提及的《連結》一書指出，人工智能與像「外星來客」般的「非人類智慧」(Alien Intelligence) 不同<sup>⑫</sup>。筆者認為，現時人工智能幾乎等於「美國智慧」(American Intelligence)，但人工智能不應該只有單一的版本。

也許來自東亞的研究者可以追問，到底東亞的經驗能否改善歐美等國 (尤其是英語世界) 初創的技術系統呢？對於東亞，人工智能似乎是「從天而降」的外來力量，尤其依照英語習慣設計的 ChatGPT 等人工智能程式與東亞的社會結構與文化有不少差異；東亞不是單一語言和族裔的地區，而且政治、經濟與社會傳統和西方國家不一樣。西方主流社會相對較為擁抱基督教的價值，但是在東亞地區傳統上更崇尚儒家、佛法、道家，以及其他來自亞洲的價值，當中便有英語人工智能暫時難以轉譯的思想和習慣。如現時百度推出的生成式人工智能「文心一言」、Sony 的「相棒」(Aibo) 機器狗，便反映出科技亦有必要根植於本土。前者以簡單中文文句為主的素材來訓練人工智能，以便華人家用使用其服務；後者以日語下達溝通指令，機器狗的動作有如鞠躬，乃依據日本文化習慣來設計的。反過來，兩個例子也說明，技術系統與文化息息相關，本地社會中的生活、工作和習慣，能為技術系統提供學習、調適和優化的機會。

當然，現時東亞各國意圖在政經動盪中力爭下一個時代的優勢，實在無可厚非。日本的「社會5.0」<sup>⑳</sup>和德國的「工業4.0」<sup>㉑</sup>，以及中國的「『十四五』智能製造發展規劃」等計劃，不約而同地提出大力發展人工智能、機器人、物聯網等尖端科技，掀起了智能社會的發展熱潮。凡此種種，東亞是否準備好迎接類似的新技術促成的變革？當重視科研成為東亞社會的國策，社會各界亦會引進並普及深度學習、大數據等最新科技，短期而言可能會帶來一種明顯的功能價值，引導經濟、社會走向電子化、虛擬化和自動化的環境。不過，東亞各國政府近年提倡的新科技政策，如推動以氫氣等替代能源取代汽油、引進機器人照護、推進邁向人工智能社會等，如推行失當，可能導致政府在日後政治形勢或選舉中失勢，甚至令現時的政商精英集團失信倒台，換來反全球化、反機械化、反數碼化的保守政治勢力登上廟堂。

追趕歐美的科技變遷和全球化，並不意味着進行見風使舵式的轉向，也要審視科技和人工智能學者一切未有堅實根據的預言、不顧現實的未來秩序設計和並非建基於已有基礎上的構想。人工智能無法一勞永逸地解決貧富懸殊、人口老化、經濟失衡、社會流動停滯和環境污染等問題，但至少能給現在極為牢固的經濟秩序和邏輯帶來一點創新。我們可探索人工智能帶來的經濟轉型機遇，以解構市場秩序，修正經濟失衡，提高社會流動性，拉近貧富、城鄉差距。我們有理由期望東亞憑藉台積電和三星 (Samsung) 等企業的晶片製造能力，鴻海、日立、三菱等工業製造經驗，以及中國大陸龐大的供應鏈和勞動力，加上從香港、北京、東京、首爾到台北的頂尖科研大學機構提供的人才，並運用東亞乃至亞洲龐大的人口基數優勢，帶領世界走上物聯網、大數據和智能應用的智慧城市社會。這樣一來有助扭轉東亞人才流失、資金外流、產業轉移的風險，二來能夠為就業轉型以及下一代企業家和從業員拓展新的區域市場機會。

同時，東亞社會不必只仰賴美國、歐盟或聯合國制定的規範，因為完全按照其標準與價值，不易坐享其成，大抵只能削足適履。在全球共享經驗的基礎上，或許能從中國大陸、台灣、日本、韓國的科技競爭和合作中，找到一套合乎東亞歷史、宗教、價值、語言、社會的人工智能技術及其治理辦法。比如在2024年由新加坡研發的SEA-LION出台，此大型語言模型乃集合東南亞十一種語言、多元語境文化的一體化網絡。韓國科技巨頭Kakao最近亦宣布與OpenAI達成合作協議，共同開發針對韓國市場的人工智能產品Kanana。假如在不久的將來，東亞國家能夠後來居上，提供不一樣的「亞洲智慧」(Asian Intelligence)，誠為美事。

當然一如其他社會改革計劃，智能社會也許馬到功成，也許引狼入室，也許「種瓜卻得豆」，其所期許的革命性影響暫時仍在計劃中。長遠來看，人工智能將為我們帶來不可知的潛在風險，可能產生社會反作用力，或許要數十年或幾代人以後方可得出定論。除了要慎思如何應對技術層面的風險之外，也要小心背後私企壟斷和地緣政治的結構力量，包括東亞社會該注意自身的深厚淵

源與利益關係。東亞社會與其被動地接受來自美國私企或地緣政治形塑人工智能，不如在國內持份者之間達成共識，再尋求區內組織協定和跨國機構協調。這些來自東亞社會的科技發展經歷，甚至有可能給全人類提供寶貴經驗。

## 五 結語

在送龍迎蛇之際，科技界好不熱鬧。美國新政府上台後，先是公布美企甲骨文(Oracle)、OpenAI與日企軟銀集團(SoftBank)將組成「星際之門」(Stargate)的合資企業，計劃投資五千億美元於人工智能基礎設施；繼而中國初創公司「深度求索」(DeepSeek)開發的人工智能大模型橫空出世，在全球引起極大衝擊。中美兩國似有加劇人工智能競賽之勢，甚至引發在人工智能領域全面「脫鉤」的擔憂，未來事態發展仍有待進一步觀察。

全球人工智能治理需要達致各個國家、地區和社會內部的多方共識，找到平衡科技與社會的出路，以免置科技於人類的對立面。由於社會與科技互相塑造，當人工智能帶來不可持續的惡果時，不應強調技術問題靠技術解決，理應通過社會選擇予以排除。要確保社會和科技之間的良性互動，我們要有既符合科學而又在地化的理解，才能建立更可靠、更合用的科技系統。當科技議題有更多公民參與，對於科技的監管有望建立跨國平等共議和多元開放的協調機制，未來貴乎能從中找到東亞的新位置和人類的角色，一起尋求平等自由而有尊嚴的生活。

### 註釋

① “Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI” (26 March 2023), [https://youtu.be/L\\_Guz73e6fw](https://youtu.be/L_Guz73e6fw); Sam Altman, “Planning for AGI and Beyond” (24 February 2023), <https://openai.com/blog/planning-for-agi-and-beyond>.

② 格雷伯(David Graeber)、溫格羅(David Wengrow)著，林紋沛譯：《萬事揭曉：打破文明演進的神話，開啟自由曙光的全新人類史》(台北：麥田出版，2024)，頁708。

③ Henry A. Kissinger and Graham Allison, “The Path to AI Arms Control: America and China Must Work Together to Avert Catastrophe”, *Foreign Affairs*, 13 October 2023, [www.foreignaffairs.com/united-states/henry-kissinger-path-artificial-intelligence-arms-control](http://www.foreignaffairs.com/united-states/henry-kissinger-path-artificial-intelligence-arms-control).

④ 松尾豐著，江裕真譯：《瞭解人工智慧的第一本書：機器人和人工智慧能否取代人類？》(台北：經濟新潮社，2016)，終章。

⑤⑥ 札卡瑞亞(Fareed Zakaria)著，江威儀譯：《革命的年代：從十七世紀至今的全球化、科技化、地緣政治的衝擊》(台北：天下文化，2024)，頁26；18。

⑦⑧ 哈拉瑞(Yuval N. Harari)著，林俊宏譯：《連結：從石器時代到AI紀元》(台北：天下文化，2024)，頁440；263。

- ⑦ Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021), 136-49.
- ⑧ “Pause Giant AI Experiments: An Open Letter” (22 March 2023), <https://futureoflife.org/open-letter/pause-giant-ai-experiments>.
- ⑨ “Statement on AI Risk”, [www.safe.ai/statement-on-ai-risk](http://www.safe.ai/statement-on-ai-risk).
- ⑩ Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (London: Profile Books, 2019), 8-12.
- ⑪ Khari Johnson, “LaMDA and the Sentient AI Trap” (14 June 2022), [www.wired.com/story/lamda-sentient-ai-bias-google-blake-lemoine](http://www.wired.com/story/lamda-sentient-ai-bias-google-blake-lemoine).
- ⑫ “Generative AI Could Raise Global GDP by 7%” (5 April 2023), [www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent](http://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent); “Future of Jobs Report 2023: Insight Report” (May 2023), [https://www3.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_2023.pdf](https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf), 6.
- ⑬ 巴斯夸利 (Frank Pasquale) 著，李姿儀譯：《二十一世紀機器人新律：如何打造有AI參與的理想社會？》(台北：左岸文化，2023)，頁259。
- ⑭ *General Data Protection Regulation*, <https://gdpr-info.eu>.
- ⑮ 〈演員挺編劇！21世紀好萊塢最大罷工：串流與AI如何餓死編故事的人？〉(2023年7月19日)，報導者網，[www.twreporter.org/a/hello-world-2023-07-19](http://www.twreporter.org/a/hello-world-2023-07-19)。
- ⑯ “G7 Hiroshima Leaders’ Communiqué, ‘Hiroshima AI Process’” (20 May 2023), [www.whitehouse.gov/briefing-room/statements-releases/2023/05/20/g7-hiroshima-leaders-communication](http://www.whitehouse.gov/briefing-room/statements-releases/2023/05/20/g7-hiroshima-leaders-communication).
- ⑰ 參見“Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development: Draft Resolution” (11 March 2024), <https://digitallibrary.un.org/record/4040897?ln=en&v=pdf>; “Artificial Intelligence Act: MEPs Adopt Landmark Law” (13 March 2024), [www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law](http://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law)。
- ⑱ Mariana Mazzucato, *The Value of Everything: Making and Taking in the Global Economy* (London: Penguin Books, 2018), 229-69.
- ⑲ David M. Luebke and Sybil Milton, “Locating the Victim: An Overview of Census-taking, Tabulation Technology and Persecution in Nazi Germany”, *IEEE Annals of the History of Computing* 16, no. 3 (1994): 25-39.
- ⑳ Toby Walsh, *2062: The World That AI Made* (Victoria, Australia: La Trobe University Press, 2018), 130-36.
- ㉑ 2016年，日本政府提出「社會5.0」計劃，在時任日本首相安倍晉三的領導下，與東京大學和日立等企業協調，打算利用物聯網、人工智能、機器人技術和大數據，來建立一個平衡經濟發展、社會環保問題與以人為本的超級智能社會。參見“What Is Society 5.0?” (22 January 2016), [https://www8.cao.go.jp/cstp/english/society5\\_0/index.html](https://www8.cao.go.jp/cstp/english/society5_0/index.html)。
- ㉒ Klaus Schwab, *The Fourth Industrial Revolution* (Geneva: World Economic Forum, 2016), 12. 世界經濟論壇創辦人史瓦布 (Klaus Schwab) 提出世界正面臨「第四次工業革命」，其核心科技為：量子電腦技術、物聯網、區塊鏈、人工智能和機械人、虛擬和擴增實境、3D打印技術、新型材料和納米材料、生物技術、神經技術、能源採集、儲存和輸送、地理工程和太空技術。