

New Brave World ?

—— AI 與法學、哲學、社會學的 跨界遇合[#]

李建良*

人類運用智慧製造出機器，再讓機器學習運用人類的智慧，展現人類智慧的同時，也開啟了嶄新的未知世界。科技進步與創新，意味機會與風險兩種不同的發展方向，既衝突又相互依存。人工智慧（Artificial Intelligence, AI）無疑是進行中的另一波產業革命，不管稱作 Industrie 4.0¹ 或是 Fourth Industrial Revolution。回顧歷史上交替推進的產業革命，技術革新提高產能、改變人類社會生活，也伴生新的社會問題與管制需求。創新與規範總是攜手並進、胼足同行。規範需要掌握事實、給定理由、設計機制、不斷試誤，返復其道。法學、哲學、社會學等學科的對話與思維相盪，反映出人文社會科學面對科技發展的角色與使命。

有人，就有社會，有社會，就有法。源自於人的科技與社會複合交融，科技的發展與規範不能脫離社會整體的想像與既存的生活脈絡。以自動駕駛為例，從第一輛賓士車問世²、歷福特 T 型車的量產³、再到自動公路系統的研發，以迄 21 世紀自動駕駛汽車的上路，不外是人類以機器「代步」的進化過程。駕

[#] 科技部人文社會科學研究中心於 2018 年 9 月 28 日舉辦「AI 與法律、哲學、社會議題跨領域對談」，分「金融科技」、「自駕車」、「醫療照護」三個場次，邀請法學、哲學、社會學等相關領域的學者專家，針對人工智慧發展的倫理、規範、社會衝擊等相關人社議題，進行跨學門的學術對話與腦力激盪，餘音迴響。本文綜合汲取各場次的主題精華與學者洞見（詳見會後報導），嘗試描繪 AI 的人文社會科學共通議題與未來課題，以策來茲。

* 中央研究院法律學研究所研究員兼副所長、科技部人文司法律學門召集人

¹ H. Kagermann, W.-D. Lukas, W. Wahlster, Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution, VDI-Nachrichten. April 2011.

² 一般認為，現代汽車的誕生是德國工程師 Carl F. Benz（賓士）於 1885 年製造、1886 年取得專利權、裝有馬力汽油機的三輪車。參見 Der Streit um den "Geburtstag" des modernen Automobils, in: <https://web.archive.org/web/20140815102101/http://www.dpma.de/service/klassifikationen/ipc/ipcprojekt/einekurze-geschichtedesautomobils/geburtstagesautos/index.html>.

³ Aldous L. Huxley 於 1932 年出版的寓言小說 *Brave New World*，開頭的對話是：“...years after Our Ford's first T-Model was put on the market.”

駛工具自動化的程度與利弊，一直是人文社會科學思辯科技進化問題中的一環。

AI 的能與不能、該與不該？無可避免地構成了人文社會科學面對科技快速發展的基本課題之一；核心任務有二：一、如何促進 AI 的創新與應用；二、如何規範 AI 帶來的危險與風險。前者牽動人文社會科學的探索主題與範疇；後者反應人文社會科學看待及處理科技問題的能力。與此同時，人文社會科學本身（尤其是研究方法）亦可能受到來自 AI 的左右。總的來說，人文社會科學一方面必須理解、控制或改變 AI 帶來的社會新秩序；另一方面也藉由 AI 反思、發現或發掘原本可能就已經存在於人文社會科學的問題。

一、AI 的發展與規範藍圖

曾幾何時，有智慧的機器（人）正逐漸全面占領人類的生活領域，改變人類的行為模式與思維方法。在這股沛然莫之能禦的 AI 科技浪潮下，交通工具、醫療照護、金融投資的演進與變化，關乎食衣住行、生老病死等切身需求，人民不能不有以感之。除了一般熟知的自動駕駛（自駕車）外，在醫療照護方面，利用電腦實體整合系統（Cyber-Physical System）發展出醫療物聯網技術（Internet of Medical Things, IoMT），讓使用者可以居家透過心跳偵測與睡眠呼吸中止偵測儀器，達到生理量測、健康適能及生活控制，早已不是停留在想像的階段。此外，金融投資對人民的影響程度與層面，尤其不容小覷⁴。不管是金融服務結合科學技術形成的「金融科技」（Financial technology, FinTech），或者是資訊與通訊合體的「資通科技」（Information and Communication Technology, ICT），或是強化金融基礎設施的「普惠金融」（Financial Inclusion），乃至於貨幣的數位化，如比特幣（Bitcoin），帶動起一連串新型態的產業經濟，或建立數位自動交易及支付系統，或透過人工智慧演算法提供交易資訊、投資顧問，或嵌入查帳支援系統，數位化金融環境已然成形，蔚為全面覆蓋之勢。

如前所述，科技發展伴隨規範需求，AI 亦不例外。規範有習俗、倫理、法律或其他社會規範。於現代法治國家中，法律可說是最重要的社會秩序框架、權利保護機制、資源分配工具，國家運行與社會秩序莫不受到法律的影響與形塑。思考的起點是，如果有所謂的「人工智慧法」，那是創造性、形成性、管制性，還是保護性規範？若是保護性規範，保護的對象是誰？研發者、使用者、第三人，還是 AI 本身？於是乎，AI 是人或物？能否成為另一種享有權利能力

⁴ 君不聞：喪父之痛易解，奪財之恨難消。這句出自馬基維利《君王論》的說法，原文（英譯本）是：“...above all things he must keep his hands off the property of others, because men more quickly forget the death of their father than the loss of their patrimony”.

的「數位人」(Homo Digitalis)？自然是經常被提出的首要法律問題。

在當代民法體系中，「人」其實是一種功能性概念，各國普遍存在的法人制度，或部分國家將動物「去物化」的明文⁵，均為例證。不過，如同法人或動物的「擬人化」有賴法律的規範與承認，AI 是否為「人」並非法理上之必然。從實然到應然，少不了一長串政治、倫理、經濟需求或社會衝擊的論辯。嚴格來說，AI 是否具有主體性的爭議焦點，與其說是否具備「道德適格性」(moral agent)，不如說牽涉 AI 所生利益(利潤)如何歸屬或承載的制度性課題。因此，或許可以(或應該)另闢蹊徑，從法律制度的「目的取向」出發，尋找跳脫法理爭議的解決方案。例如 AI 得否作為著作權「人」，已是時興的法律課題之一，其切入點未必一定要回到民法「人」的概念上，思考方向毋寧可以放在著作權「保護目的」及其制度容量的可能性上。不過，務實以言，無論如何，AI 至少是一種「物」，可以作為「交易客體」。民法體系中有關「物」的規定，理論上均可適用於 AI 身上。至於既有體系構造是否足以因應，例如數位自動交易及支付系統或醫療物聯網之法律關係的得喪變更，就像 AI 發展潛能尚難逆料一樣，有待進一步的觀察與驗證。

權利主體的反向思維是義務或責任的承擔。以自駕車為例，脫離自然人控制的自動駕駛(物)本身是一種「危險源」或「肇因者」，其所生損害或隱藏可能風險的規範，為 AI 發展無可迴避的重要課題之一。傳統責任體系的問題面向大抵有：造成損害的原因是來自「行為」或「物」的瑕疵？對於所生損害的法律效果是刑事處罰(構成犯罪)、損害賠償、損失補償，或排除或預防損害？如果是刑事處罰，制裁的對象是誰？是否以自然人為限？如何認定故意或過失？能否阻卻違法或責任(例如緊急避難)？現行的民法、刑法或行政法的責任體系是否可以若合符節地套用在 AI 身上？或者應該適用既有的危險責任或無過失責任體系？還是乾脆透過損害保險，人身、財產等商業保險或社會保險、責任保險或擔保基金等方式，例如建立自駕車事故受害者補償救濟基金，(事前)分散可預見的危險或不可預見的風險？

從後果考察轉向事前預防的思維取徑，一定程度是從「私法自治」與「事後究責」的範疇遞移或換軌到「公權力介入」的行政管制機制或國家擔保責任。AI 作為一種「物」，例如自駕車、心跳偵測與睡眠呼吸中止偵測儀器等，基本的前提性問題就是：要不要經過(國家)「事前許可」(licensing)才能上路、上市或上線？許可要件為何(例如自駕車失靈提醒系統的必要性)？是否建立「驗證」或

⁵ 例如德國民法第 90a 規定：「動物非物(Tiere sind keine Sachen)。…」

「認證」制度？相對應的問題清單還有：許可的對象（義務人）是 AI 的原始製造者，還是使用者（例如第三方安裝自駕車系統）？AI 作為一種「行為」（程式設計、系統建置），如何建立有效的監督或管控機制？像是金融創新的監理政策與法制，或自駕車的駕照許可制？等等諸類，不一而足。此外，自駕車操作者的地位？個人資料搜集的告知同意程序及其效力？自駕車事故的責任歸屬原則？自動駕駛演算法的（強制）監督必要性？如何監督？研發者的義務？製造商的義務？操作者的義務？駕駛的義務？何種情況下必須由人類駕駛接管人工智慧？對於受自駕車影響者的警告義務與警告範圍？都是 AI 法制不可想像其不存在需要相應整備課題。德國聯邦交通部成立的自動駕駛倫理委員會（Ethics Commission: Automated and Connected Driving），於 2017 年 6 月提出「自動駕駛倫理綱領」（Ethical rules for automated and connected vehicular traffic），提列出 20 點準則，對於上述問題作出初步的回應與規範框架，或許可以作為法學、哲學與社會學對話的示範領域與思考起點。

二、AI 的兩難課題與抉擇

有人類，就有兩難問題，不是假設性的腦力激盪困局⁶，就是實境式的存亡兩難抉擇⁷，自古有之，電車難題（Trolley Problem）產生於人類發明電車之後⁸，今則是討論自動駕駛的常見問題模式。假設某種「不可逆」的異常情境，古典的提問是：當駕駛發覺異常狀況時，只能選擇撞上前方五人，或轉向撞上左方三人，應如何選擇？或者，當駕駛發覺異常狀況時，只能選擇駛向懸崖，或轉向撞上左方三人，應如何選擇？

放在刑法的思維脈絡，何種選擇（如果可以選擇）不罰？刑法學者通常認為，「人為駕駛」在面臨上述生命與生命衝突的抉擇難題時，不得以犧牲他人生命的方式保全他人或自己的生命。因為人的生命質量無法權衡，要求無辜第三

⁶ 例如康德 1797 年《道德形上學》（*Die Metaphysik der Sitten*）一書論及「緊急權」（Nottrecht）暨「必要之權」（*Ius necessitatis*），提到：在發生船難時，二人同陷於生命危險之際，一人為解救自己的生命，將另一人推落唯一的救生浮板，不能以刑法相責。參見 Immanuel Kant, *Die Metaphysik der Sitten*, 1797, Nachdruck 1990, S. XLI.

⁷ 例如 1884 年的 R v. Dudley and Stephens (14 QBD 273 DC) 一案，四名船員遇海難逃上救生艇，其中兩名船員為了存活，殺死另一名船員，並食用之，引發緊急避難界線的倫理與法律難題。

⁸ 現代「電車難題」常以 Philippa Ruth Foot 於 1967 年提出的命題為版本。在此之前，在刑法學界早有各種類似的討論，例如 Karl Engisch, *Untersuchungen über Vorsatz und Fahrlässigkeit im Strafrecht*, 1930, S. 288.

人為他人犧牲自己的生命，超出社會連帶義務的界限。

較諸人為駕駛的危疑實況，無人操控汽車（第四級與第五級的自動駕駛）遇上兩難困境時的「選擇」，不是來自「當下」的駕駛「人」，而是由程式設計者「事先」決定自駕「車」在面臨「未來」兩難困境時的制約反應。倘若程式設計者指示自動巡航系統遵守「最小損害原則」與「優先保護乘客原則」，自駕車因此轉彎撞死他人，程式設計者或使用者是否負殺人罪或過失致死罪的刑責？自動駕駛與人為駕駛在電車難題的法律評價基準，是否不同？目前仍然是開放性的倫理與法律難題。

AI 的需求與規範的需求，互為因果，如何規範建立在規範的理由基礎上。倫理評價在人工智慧規範中可以扮演「定錨」的功能，但倫理評價如何形成？遵循誰的倫理評價？Stuart J. Russel 從電腦科學家的角度出發，指出 AI Community 的共同難題是：人類如何製造合乎人類價值觀的人工智慧？這道大哉問潛藏了「人類價值觀」是什麼的前提性難題，由來已久，效益主義與自由主義之爭，只是代表性主題之一。如果要在自動駕駛車上內建一套倫理系統，那應該是什麼樣的倫理系統？

社會總體福祉極大化的說法，人們並不陌生，但什麼是社會總體福祉？如何極大化？如果在一般民生議題上沒有答案，何以在安裝自動駕駛「倫理系統」的時候就可以找到問題的解答？如果人類自己都沒有固定的價值觀，AI 又如何會有自己的價值觀？除非 AI 比人類有智慧！也因此，呼之欲出的弔詭議題將會是：究竟是人類給定 AI 以人類的價值觀，或者可以透過 AI 解決人類無法解決的價值觀難題？

換個角度提問，AI 科技發展帶給人類的正面利益與潛在威脅為何？能否及如何辨識、防範、抉擇？價值或產值、工具技術或體系結構如何權衡，孰輕孰重？例如不同於汽車發明的目的主要在於以動力取代人力，自駕車的研發是為了提高交通安全、減少肇事率、解決交通阻塞、提升能源利用效率？還是只為了提高舒適性、減少人工成本、增加經濟效能？

又如 FinTech 或 ICT 的發展，無疑可以提高交易效率、避免投資風險、降低交易損失、穩定經濟市場。但是否同時造成軍備競賽、浪費資源、甚至操縱市場，或信用危機或「裸貸」亂象等，甚至於社會層面因數位落差而形成另一波數位正義課題。

再如智慧醫院或醫療物聯網的建構，整合了醫學與電資系統，使服務模式進入居家環境，提供貼心、迅速的醫療與健康照護，解決高齡社會帶來的醫療困境，同時打破 cure 與 care 的傳統僵化界線與模式。醫療資訊的雲端化，不受

限於時間或空間，可能藉此找到最佳的診療與照護方法。反面以觀，是否會成為某種形式的「隔空」取藥療治，顛覆「望聞問切」的醫療理念？奠基在信賴慣性的醫病關係（某種社會關係）是否也會為之丕變？「教考訓用」的醫學傳承網絡是否斷裂或無法銜接？更不用說，當 AI 長驅直入私領域的時候，個人的敏感資訊是否被看不見的眼一覽無遺？

如果說 AI 的影響無遠弗屆，個人資料的問題就是無所不在！例如偏鄉或弱勢族群在享受普惠金融服務的同時，可能已經大量傳遞出個人的數位資料。AI 暨 machine learning 需要大量資訊作基礎，因此在各領域發展勢將面臨一項共通問題，並且形成某種程度的悖論：人工智慧資料越少，越無法深度學習；人工智慧越是可以深入學習，對個人資料的侵害卻是越大。當前舉世注目的歐盟個人資料保護規則（General Data Protection Regulation, GDPR）等相關規定，能否及如何在 AI 創新與個資保護之間劃線、界分、定出平衡點，舉世關注。在大數據與數位時代下，我們面臨的困難是：我們並不清楚當下及未來，個人資料如何被侵害？被侵害到何種程度？當我們享受資訊快速便利的同時，正在一點一滴地失去我們的個人資料，兩難之間，如何選擇？還是，其實我們已經沒有選擇的自由！？

三、人的智慧、理性與尊嚴

道德經云：「智慧出，有大偽」。西文 Intelligence 蛻變自拉丁文 intellegere，由 inter 與 legere 二字組成，原意是「問與問之抉擇」。人工合成的智慧是否大偽似真，全憑人如何設定目的、價值抉擇與應變改正。值此全球鼎革之際，AI 可能帶來思維典範的轉型、生活方式的遷易、市場機制的變化、專業圖像的重構、人際角色的消長、乃至於社會結構的位移。人類其實比機器更需要 deep learning，還有 deep thinking。

莫忘人生而自由平等的初衷！AI 的進境應該是脈絡有序的社會調理，而不是全面性的制度顛覆。犯錯乃人的本性之一，面對錯誤的態度與方法則是人類文明的一環。AI 時代的到來，意味的是提高人類不犯錯的機會，還是讓人類找到不犯錯的藉口？

人有智慧，此所以人為萬物之靈；人是理性？則有歧見，尚無定說；人有尊嚴，規範應然⁹，不可侵犯。科技來自人性，也會改變人性。如何確保人的理性與尊嚴，必須由 AI 世代人人共同面對，這當然也是人類代代的恆久課題。

⁹ 司法院釋字第 603 號解釋。