

《歐美研究》第五十卷第二期 (民國一〇九年六月), 137-145
DOI: 10.7015/JEAS.202006_50(2)0001
<https://euramerica.org>

專題序 —歐美 AI 的發展與挑戰*

洪德欽

中央研究院歐美研究所
E-mail: dchorng@sinica.edu.tw

蔡政宏

中央研究院歐美研究所
E-mail: chtsai917@gate.sinica.edu.tw

壹、AI 的崛起與衝擊

人工智慧 (Artificial Intelligence; AI) 的發展與計算機密切相關。英國科學家圖靈 (Alan Turing) 於 1936 年依據抽象計算模式，提出一種數學邏輯機，事後發展成為電腦，而被尊稱為電腦之父。圖靈於 1950 年發表一篇〈計算機與智慧〉論文，以解決其所提出一個基本假設問題：「機器能思考嗎？」(Can machines think?)，並據此對機器模仿人類智慧進行深度思考及系統論述。圖靈在文中提出的模仿遊戲 (The Imitation Game) 設想及測試方式，形成著名的

© 中央研究院歐美研究所
責任校對：蔡旻芳、王崇懿

* 本文獲得《歐美研究》季刊主編鄧育仁研究員提供寶貴建議，並由蔡旻芳小姐協助文書工作，謹此一併敬謝。

「圖靈測試」(Turing, 1950: 433)。本篇前瞻性論文以及機械思想實驗奠定了 AI 的理論基礎，圖靈因此也被視為 AI 之父。

AI 自 2010 年以來在深度學習等方面有重大突破，結合數據科技等發展，已逐漸應用於物聯網、區塊鏈、電子商務、工業自動化、智慧製造、機器人、自動駕駛、金融科技、智慧電網、遠距或精準醫療、精緻農業、文藝創作、智慧交通、智慧城市、生物科技、環境、電子、資訊、軍事及其他產業，或災害防治、瘟疫防控、環境保護、生態多樣性或永續發展等不同領域，形成影響深遠之關鍵科技。AI 科技之發展與應用，不僅會帶來產業創新及新的經濟功能與社會服務，同時也會影響人類之現行生產方式、經濟結構與生活方式。AI 在歐美皆被列為一項關鍵性策略產業，深受歐盟及美國之重視，形成一項國際主流議題，具有產官學之重要性 (European Union, 2020: 1-2; National Science and Technology Council, 2016: 3-5)。

AI 時代，人類生活與工作方式，將有哪些社會變遷與歷史意涵，各國經濟結構、法律體系與政策制度另有哪些創新與改變。AI 的衝擊涉及法律、哲學、社會、科技、產業及政策等問題，例如：歐美 AI 倫理原則、法律改革、立法哲學、組織建制、政策論證、風險規範、安全控制、法理原則、大數據資料庫、演算法、智慧財產權、補貼、AI 歧視或不平等、生物辨識系統與 AI 電子監控、侵權責任、保險、個人隱私、網路安全等問題，另對工作機會、所得分配、公平競爭、治理效率、AI 使用及政策的合法性、正當性、透明性及如何課責、相關國際組織或國際條約、爭端解決等議題，對各國管制、民主、法治及國際規範皆構成嚴峻挑戰。

我國目前積極推動「5+2 產業創新推動方案」、「臺灣 AI 行動計畫」、生醫產業及精緻農業等政策，以期技術創新、結構調整與產業升級。歐盟與美國的 AI 發展較為先進，針對 AI 的發展機會與

潛在問題，所從事的哲學省思、政策論證、法律建構等皆有諸多值得借鑑之處，以因應 AI 帶來的潛在法律問題、經濟結構調整與社會變遷等挑戰，這是中央研究院歐美研究所於 2019 年 5 月 9 日至 10 日舉辦「歐美 AI 的發展與挑戰」跨學門學術研討會之緣起與目的。

本次研討會共有 14 篇論文宣讀，另由日本東京大學大學院情報學環教授，同時也是日本高級智能計畫中心 (RIKEN Center for Advanced Intelligence Project; RIKEN-AIP) 計畫主持人佐倉統教授發表一場主題演講：“AI in Cultural Contexts: Toward Comparative Studies between Euro-American and East-Asian Countries, with a Short Introduction of Current Conditions in Japan”。會議論文宣讀人針對 AI 特定議題，進行深度反思及論證，並提出一些前瞻性洞見及建議，以解決 AI 衍生的潛在風險及相關問題。會議論文包括：AI 倫理準則設計、醫療照護機器人、人工道德、圖靈測試、人工意識、風險預測、AI 法律推理及判斷、研究倫理、自駕車法律及管制模式、隱私及個資保護、數位版權、人工智慧創作之著作權可行性等題目。這些議題皆是大哉問之前沿性問題，值得從事跨學門研究及深度討論，以收集思廣益之綜效。

本期專號收錄經由匿名審查通過而獲刊登的六篇論文，涵蓋四篇哲學及二篇法律學門文章，呈現我國專家學者在 AI 集體創作的重要研究成果。以下簡要說明這幾篇論文的研究重點及主要發現。

貳、專號文章亮點

機器能有思想嗎？這是人工智能哲學的核心問題。第一篇趙之振論文〈戴維森論圖靈測試〉，重返經典的圖靈測試，透過闡釋與辯護戴維森在圖靈測試上的觀點，指出在何種條件之下可把思想歸屬給機器。作者提出兩點戴維森帶給人工智能研究上的啟示：第一、

「我們無法僅僅從外顯行為來決定機器能否思想，思想之條件比圖靈所預設來得複雜得多，因此圖靈的模仿遊戲之進路是不可取的」（頁 172）；第二、「機器能否被視為思想者，主要不是取決於機器與人類在外顯行為有多相似，而是取決於人類有多大的意願先行把機器看作與自身（作為思想者）相似」（頁 172）。本文並延伸探討將機器「視為」思想者（對比於機器「是」思想者）所帶來的問題。

機器能有道德嗎？若能，人工道德與人類道德有何異同？第二篇何宗興論文〈反思機器人的道德擬人主義〉，探討人工道德主體的本質。假設機器人可以通過道德版圖靈測試（Moral Turing Test），那麼機器人就可被視為道德主體。此種行為主義的道德主體觀，將連結至作者文中所提到的道德擬人主義：「既然機器人是跟人類一樣好地遵守道德規則，那麼適用於人類的道德規則就適用於機器人，這就代表凡是對人類是道德上正確的，對機器人也是」（頁 201）。作者以機器人應不應該阻止人類自殺為例，透過史特勞森的反應態度概念，指出道德擬人主義無法成立，並建議機器人的道德行為應該比人類受到更嚴格的限制。

第三篇洪子偉論文〈淺論 AI 風險預測的規範性爭議〉，探討「分析過去巨量資料以預知未來風險」所引發的規範爭議，包括「與演算法相關」以及「與資料相關」兩方面的爭議。作者透過層層辯證，最後做出四點宣稱及建議：第一、AI 預測的不可解釋性是來自於人類的認知限制；第二、AI 的黑箱問題在人類大腦上也有，兩者是程度而非種類上的差異；第三、目前仍未有一致性的理由來反對以 AI 預測來預防人類造成的風險；第四、使用 AI 預測結果時依自主性原則為佳。

第四篇甘偵蓉、許漢論文〈AI 倫理的兩面性初探：人類研發 AI 的倫理道德與 AI 的倫理規範〉，分別探討 AI 研發及使用等利益相

關人員所應依循的倫理規範，以及 AI 系統本身所應依循的倫理規範。作者分析了三個國際組織所公布的 AI 倫理準則，指出其不足，並提出底線式道德觀作為 AI 的兩倫理規範面向的共同起點，其中包含二項消極義務（「無條件的禁止所研發的 AI 內建程式為：(1) 將人類只當工具來使用，包括以傷害生命健康來威脅人類；或 (2) 對人類生態系統可能帶來毀滅性的傷害」[頁 268]）與一項積極義務（「AI 應以滿足人類基本且一般的生活需求作為優先研發順序」[頁 268]）。

第五篇陳弘儒論文〈初探目的解釋在法律人工智慧系統之運用可能〉，研究目的在解釋人工智慧系統適用之可能性。法律推理向來是法律人應具備的重要能力，法律推理能力是否可透過演算法得到解決，攸關人工智慧是否可以具備法律推理能力。本論文依據法律推理電腦模型，從目的解釋的論證形式、工具理性運用及目的規範來源等三大面向，論證並說明目的解釋作為電腦模型遭遇的一些困境，導致目前仍無法有效將目的解釋適用在法律人工智慧系統。

本文論證「目的如何產生規範性」，為法律推理的目的解釋提供反思，並說明法律推理的不確定性。另外，本文另透過「目的」概念分析為法律推理電腦模型提供哲學反省，釐清其適用界線及相互關係。本論文乃人工智慧法理學的基礎研究，從事前沿性及創新性研析，具有很高學術價值。

第六篇洪德欽論文〈歐盟自動駕駛車之發展策略與法律規範〉，研析歐盟自駕車的發展策略及法律規範，以了解歐盟自駕車發展過程遭遇哪些問題，採取哪些因應政策？又如何透過法律加以規範，尤其如何由法律從事自駕車安全及倫理的管制。本文並以歐盟為個案研究，深度研析歐盟自駕車安全要求，以及智慧交通系統之規範及建構。歐盟自駕車建制有哪些特徵、意涵、影響，同時又能提供

臺灣哪些參考及啟示。

歐盟自駕車乃以技術創新及安全要求為發展策略，以倫理人權及歐盟理念作為法理基礎及公共利益之論述，以 AI、智慧交通系統、數位市場、個資保護及網路安全等作為必要性配套措施。歐盟自駕車相關法律，主要包括：2018/858 規則、2010/40 指令 (ITS 指令)、2016/679 規則 (GDPR)、2019/103 指令 (汽車保險) 及 85/374 指令 (產品責任) 等，兼顧科技創新、官方管制、安全要求、產業競爭力、公共信任及公共利益等面向。自駕車的健全發展並建立大眾信賴，預期對傳統車輛及交通系統產生一項典範移轉，同時對人類生活方式帶來重大社會變遷。

安全是自駕車發展的先決條件。歐盟對自駕車採取高標準安全要求，其原因是確保用戶生命及財產安全，建立用戶及歐洲消費大眾之「信心」，這是歐洲自駕車是否可以大量上路使用之核心問題。其次，歐洲汽車產業長期以來已建立高品質之信譽。高標準安全才足以延續其汽車產業之品質保證及信用聲譽，並持續歐洲汽車產業在國際市場之競爭力及領先地位。歐盟 2007/46 指令及 2018/858 規則針對自駕車許可前及上市後之測試、檢查及評估，皆有詳細規定，其主要目的在於確保自駕車高標準的功能、品質及安全等要求。

對臺灣的啟示方面，我國目前已無具規模性汽車產業，但是，自駕車之發展必須結合 AI、智慧交通系統、數位科技、5G、電子通訊等基礎建設。臺灣在電腦、晶圓、IC、IT、電子通訊、精密機械等領域頗為發達，仍有機會針對自駕車之光學雷達、通聯器、黑盒子、感測器等，以及 AI、5G 等零件，發展自駕車感測、定位、監控、決策及控制技術相關核心零件產業，具有潛力成為自駕車全球生產、供應及價值鏈的重要供應國家。所以，我國仍應重視歐盟自駕車的法律與政策，尤其歐盟自駕車安全要求及技術標準，結合臺

灣電子、通訊及晶圓優勢科技及精密工業，發展具有較高附加價值的自駕車核心零件及支援系統設備等，諒對我國產業技術創新及經濟成長等有所貢獻。

參、AI 的美麗新世界

AI 的發展，將促進產品的創新，提高企業經營效率，增加人類生活的便利性。然而，AI 的發展，如同其他具有重大突破之新興科技，往往伴隨一定之科學不確定性以及潛在風險。AI 是一種具有雙重面向及多重用途之科技，如果進一步與生物科技或基改技術相結合，產生的生物 AI、AI 生物或生物 AI 人等，是否形成如同希臘神話「人獸合體」生物，而令人不安。事實上，圖靈在 1950 年〈計算機與智慧〉論文之結論，已預設機器與最靈敏知覺器官 (the best sense organs) 結合的可能性 (Turing, 1950: 460)。據此，AI 所形塑的未來世界，是否變成如同狄更斯 (Charles Dickens) 在《雙城記》(A Tale of Two Cities) 所言，「那是最好的時代，也是最壞的時代；那是光明的社會，也是黑暗的社會。」

依據科學中立原則，AI 如果僅是科技創新，將如同機器人，其好壞主要取決於人類之使用。然而，鑑於 AI 技術的專業性及複雜性，以及資訊不對稱性，AI 的力量變大時，仍有需要受到必要規範，以避免 AI 歧視、壟斷、不當使用或濫用。另外，AI 也須事先從事哲學反思及相關論證，以界定 AI 的規範原則及管制模式。哲學反思與法律規範在 AI 發展過程，扮演一個重要角色。哲學反思可以論證 AI 的倫理原則，作為法律、政策、制度的法理基礎及管制指引。法律原則可以規範 AI 延伸相關問題，調合科學創新與風險管制，形塑一個可以信賴的 AI 制度。哲學與法律因此是各國 AI 建制的思想指

導與基礎建設，在各國甚至形成一種理論競爭、法律競爭及制度競爭。

AI 的發展與應用，除了以科技創新為主軸作為驅動力量之外，必須同時兼顧以倫理道德為內涵，以及法律管制為規範，共同架構為 AI 建制的三大支柱。質言之，AI 發展及建制必須以人為本，誠如《尚書·大禹謨》所言，天工開物，必須連結「正德、惟和」的理念，才足以達成「利用厚生」的結果。《尚書》之古典智慧似已為 AI 指出一條正確的發展方向。凡此，在在顯示哲學與法律跨界整合研究的重要性，以提高「AI 理性」，期使「AI 之美」(The beauty of AI) 對人類生活福祉與文明進步有所貢獻。

參考文獻

- European Union. (2020, February 19). *White paper on artificial intelligence—A European approach to excellence and trust*, COM (2020) 65 final. Retrieved from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- National Science and Technology Council. (2016, October). *Preparing for the artificial intelligence*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 236: 433-460. <https://doi.org/10.1093/mind/LIX.236.433>