

第二波人工智慧知識學習與生產對法學的挑戰 — 資訊、科技與社會研究及法學的對話

邱文聰*

摘要

當今人工智慧雖還欠缺自我意識與意向性而不能成為法律權利主體，也尚無跨領域通用之能力，但「強/弱 AI」或「通用/限定 AI」的區分，已不足以描述人工智慧從 1950 年代演化至今的轉變。科技與社會研究學者對於人類如何學習、傳遞知識與專業知能的相關研究成果，是理解人工智慧是否能真正取代人類智能的重要切入點。本文考察 Dreyfus 與 Collins 兩位重要學者的相關理論，指出只要機器學習的實作先例涵蓋夠廣，各種身體性默會知識及集體社會性默會知識，就能被第二波人工智慧的學習模式所涵化，與人工智慧是否已具有流利的自然語言能力無關。真正的挑戰在於，依照不同機器學習方法，人工智慧既可能承襲默會知識中的積習與可能的不理性與錯誤，也可能引發人類利用尚無法為現有知識所驗證的預測能力以進行各種控制的慾望。以規範論證為核心活動的法學，則至少在三個面向上可能面臨第二波人工智慧的模仿或參與而遭受衝擊：在外部證立上創造新的差異製造事實、模仿規範論證行為、模仿法律效果裁量行為；法學也將在「效率」、「法與社會的典範移轉」與「控制」三條軸線上，不斷地進行自我辯證。

關鍵字

機器學習、監督式學習、非監督式學習、直覺、默會知識、專業知能、規範論證、涵攝、內部證立、外部證立

* 中央研究院法律學研究所研究員，台灣大學國發所合聘副教授。作者感謝趙若漢先生、何琳潔小姐、顧長芸博士提供的研究協助，並感謝陳瑞麟教授對其中部分論點的啟發。

1. 序論

人類自古雖然就夢想著製造出擁有人類能力的機器，但真正開始以機械模擬人類心靈智能的嘗試，則始於二十世紀的 40 年代。¹ 「人工智慧」從 1950 年代起正式成為計算機科學當中一個被賦予高度期待的研究領域，卻也歷經希望與幻滅的過程。當代的人工智慧則搭上了巨量資料與電腦運算能力大幅提昇的浪潮，在 2010 年起透過機器學習技術而真正蓬勃發展。

然而截至目前為止，宣稱具有人工智慧的智能主體雖已有諸如高等級自駕車（Level 3 以上），與打敗圍棋世界棋王的 Alpha Go 等問世，但也僅能在預先選定的限定領域中運作，而非如人類智能般具有相對廣泛的「通用」性質。除了仍未有真正的通用型人工智慧（Artificial General Intelligence, AGI）²的智能主體被開發出來之外，當代人工智慧縱然進展快速，也仍然如心靈與語言哲學家 John Searle 在 1980 年所做出的診斷與預測一樣，並沒有因為模仿（simulate）人類心智的能力提升，使智慧主體產生自我意識（consciousness）或取得意向性（intentionality）。³ 因此即便在某些外顯的認知功能上，現階段人工智慧所發展出的智能主體已擁有與人類一樣好，甚至更好的能力，模仿（而非真正具有）人類心智能力的人工智能主體，終究只是 Searle 定義中的「弱 AI（weak AI）」⁴。但不僅 Searle 定義中擁有人類完整心智的「強 AI（strong AI）」似乎難以迄及，即使是人工智慧領域奠基者眼中「與人類有相同程度的人工智慧（Human Level Artificial Intelligence, HLAI）」⁴，也仍然還是一個尚未實現而有待追求的梦想。

對比於當前仍以「弱 AI」為限的人工智慧發展現狀，預言法律制度必須修改以承認人工智能主體具有法律上的人格，從而應保障其法律權利並課予其獨立法律責任的觀點，似乎顯得杞人憂天。畢竟，以「意思能力」、「行為能力」、「自主性」為基礎而架構權利行使、義務履行與責任負擔的各種法律制度，不可能將一個欠缺自我意識、欠缺意向性的機器，視為一個具有權利能力與責任能力的適格對象：人工智能主體頂多只是人類的得力助手，掌握最終控制權的仍是人類自己。與人工智能主體有關的權利行使與責任負擔，仍舊應視實際情況分別歸屬於人工智能主體的設計製造者，或者操作使用人工智能主體並仰賴其協助而真正做

¹ 一般認為 Warren McCulloch 及 Walter Pitts 兩人在 1943 年製造出模仿人類神經元的系統是最早的人工智能主體，STUART J. RUSSELL & PETER NORVIG, *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH* 16 (Upper Saddle River: NJ, 3rd ed., 2010)。

² See Cassio Pennachin et al., *Contemporary Approaches to Artificial General Intelligence*, in *ARTIFICIAL GENERAL INTELLIGENCE* 1-30 (Ben Goertzel & Cassio Pennachin eds., 2007).

³ John R. Searle, *Minds, Brains, and Programs*, 3(3) *BEHAVIORAL & BRAIN SCIENCES* 417-457 (1980).

⁴ See John McCarthy, *From Here to Human-level AI*, 171(18) *ARTIFICIAL INTELLIGENCE* 1174-1182 (2007); MARVIN MINSKY, *THE EMOTION MACHINE: COMMONSENSE, THINKING, ARTIFICIAL INTELLIGENCE, AND THE FUTURE OF THE HUMAN MIND* (Simon and Schuster, 2007), Nils J. Nilsson, *Human-level Artificial Intelligence? Be Serious!*, 26(4) *AI MAG.* 68-75 (2005).

出意思表示或事實行為之人。換言之，人工智能主體若非在執行人類意志，就是協助人類形成意志；即使在少數情況下或因第三者的介入而使之處於失控狀態，也同樣能以第三者做為冤頭債主。大體而言，既有法律制度雖還是必須透過保險或補償，填補那些例外無法歸咎於「人禍」的重大損害，但整個法律制度似乎並不需要進行基進的大幅度修正。

然而，除了以能否取得法律人格當作尺度來衡量人工智慧對法律系統的衝擊之外，欠缺自我意識與意向性的「弱 AI」，究竟對法學可能產生什麼影響？要回答這個問題，必須回到人工智慧技術發展本身探詢相關的線索。本文將在簡要介紹人工智慧技術發展的歷程後，進一步從科技與社會研究（STS）對人類知識生產、學習與傳遞的相關理論爭辯，分析第二波人工智慧的技術可能性與其極限。最後再回到法學本身，探討法學在第二波人工智慧嘗試模仿法規範論證行為時，所可能面對的機會與挑戰。

2. 人工智慧的三個發展階段

人工智慧在過去六十餘年的發展過程中，歷經了幾次的典範移轉。美國「國防高等研究計劃署（Defense Advanced Research Project Agency, DARPA）」的報告，依照人工智慧研究者所採用的不同思維或技術方法，將人工智慧的發展區分為三個階段。⁵ DARPA 的報告雖然沒有對各階段的人工智慧發展提供詳盡的文字說明，但 DARPA 所做的分期卻提供一個分析與理解人工智慧技術的有用框架。

DARPA 將第一波人工智慧稱為「手工打造知識（handcrafted knowledge）」的年代。第一波的電腦資訊工程師透過程式語言的符號，將既有的人類知識以「若...則...」的邏輯規則形式在電腦上予以再現（knowledge representation），以便利用電腦的邏輯運算能力，自動處理特定領域當中已經被該人類知識所掌握與回答的問題。第一波的代表人物就是人工智慧領域的開山祖師 John McCarthy, Allen Newell 與 Herbert Simon 等人。對他們而言，能完美表達邏輯規則的符號系統是達成人工智慧的必要與充分手段。⁶ 此一時期人工智慧典範強調符號與規則的特徵，使它常被稱為「符號人工智慧（symbolic AI）」⁷或「規則導向的人工智慧（rule-based AI）」。⁸

⁵ Defense Advanced Research Projects Agency, *Powerful But Limited: A DARPA Perspective on AI* (2017).

⁶ See NILS J. NILSSON, *THE QUEST FOR ARTIFICIAL INTELLIGENCE: A HISTORY OF IDEAS AND ACHIEVEMENTS* 79-80 (Stanford Univ. Press, 2010).

⁷ See *id.* at 113.

⁸ RUSSELL & NORVIG, *supra* note 1, at 1024.

DARPA 所描述的第一波人工智慧雖然發展出一些以特定知識為基礎的「專家系統」，但在 1980 年代間這個以符號與規則為典範的人工智慧發展模式，即遭遇諸多困境。第一波的研究者逐漸承認人類智能對全局掌握的思考模式與人類行為的複雜程度，遠非幾組邏輯規則與符號有辦法完全掌握。⁹ 以研究現象學與存在主義著稱的哲學家 Hubert Dreyfus，更在第一波發展的初期（1972 年）即毫不留情地潑冷水指出，習得專業知識的過程雖始於對特定知識中基本規則的掌握，但真正成為能流暢地運用專業知識處理問題的專家，則有賴對行動目標與所處文化脈絡等「背景知識（context or background knowledge）」的理解，而後者是一種無意識的直覺（intuition）¹⁰ 或者「邊緣的意識（fringe consciousness）」¹¹，並不能以邏輯符號的方式表達，也無從藉由符號加以儲存或掌握。因此除了少數能夠不受制於具體脈絡，僅透過邏輯法則就能掌握其結構化知識的那些領域之外，單純仰賴符號或規則導向的第一波人工智慧，由於對該知識中包含直覺與邊緣意識在內的更完整部分一無所知，顯然距離 Searle 分類之下能模仿人類思考或行動的「弱 AI」目標，也仍相當遙遠。¹²

Dreyfus 對第一波人工智慧的批評雖然在當下引起不小爭議，但人類知識中具有「不可表達性（inexpressibility）」的成分確實也是第一波人工智慧發展的罩門。DARPA 分類下的第二波人工智慧則相當程度克服了這個困境。DARPA 將第二波人工智慧稱為「統計學習（statistical learning）」的年代，主要仰賴電腦的統計模型從特定領域累積的大量資料中辨識出其中隱含資料與資料間的模式（pattern recognition）。¹³ 第二波人工智慧的「機器學習（machine learning）」方法，不再需要事前掌握一套以符號表示的知識規則，而是透過以下幾個以資料驅動為主的方法，替電腦建立理性的思維與行動模式（thinking and acting rationally）。第一，透過「監督式學習（supervised learning）」，將大量前例（precedents）依照該領域既有的人類知識標註為輸入項資料（問題）與輸出項資料（答案）後，進行正確的輸入與輸出項配對分類，再使機器依據相同的配對分類模式，延伸應用於新的輸入資料，即能得到符合既有知識的正確輸出資料結果，¹⁴ 達到「模仿（emulation）」人類在既有知識下的思維與行動的目的。例如，以大量標註為病灶的醫療影像與健康者之醫療影像作為學習診斷的訓練資料，使人工智能主體能「模仿」擁有醫學知識與豐富診斷經驗的放射線科醫師，對新的醫療影像進行診

⁹ 電腦科學稱之為人工智慧的「限定性問題（qualification problem in AI）」。RUSSELL & NORVIG, *supra* note 1, at 1024.

¹⁰ HUBERT L. DREYFUS & STUART E. DREYFUS, MIND OVER MACHINE: THE POWER OF HUMAN INTUITION AND EXPERTISE IN THE ERA OF THE COMPUTER 28-35 (Blackwell, 1986).

¹¹ HUBERT L. DREYFUS, WHAT COMPUTERS CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON 100-109 (Harper and Row, revised ed., 1979).

¹² RUSSELL & NORVIG, *supra* note 1, at 21.

¹³ Dreyfus 所稱的模式辨認學習，應該是傳統區分「命題性知識」與「技能性知識」之外的第三種知識類型：「模型知識」。關於模型知識請參見陳瑞麟，認知與平價（上），頁 57-95（2012 年）。

¹⁴ See ETHEM ALPAYDIN, MACHINE LEARNING 35-42 (Cambridge, MA: MIT Press, 2016).

斷。第二，透過「強化學習 (reinforcement learning)」，在沒有變項資料配對的前提下，透過不斷試驗後結果成功或失敗的回饋 (rewards and punishment) 中，習得最佳的決策方法。¹⁵ AlphaGo Zero 即是透過此一學習方法，從不斷的對戰經驗而非模仿任何人類或棋譜的過程中，決定每一步的最佳下子策略。¹⁶ 第三，透過「非監督式學習 (unsupervised learning)」，從尚未事先配對分類的輸入資料中，自行發掘各種資料間之相關性 (association)，並建立新的分類特徵 (clustering)，¹⁷ 例如從賣場的消費紀錄中，發現女性突然購買無香料乳液與日後購買孕婦及嬰童用品間具有相關性。¹⁸

從模仿人類思考或行動的「弱 AI」標準而言，DARPA 定義的第二波人工智慧大體已擺脫第一波「老派人工智慧 (Good Old-Fashioned AI, GOF AI)」¹⁹所面臨的一些限制。然而，第二波人工智慧仍高度仰賴人類訓練者選定「訓練」與「應用」的限定背景，因此第二波人工智慧因應不同脈絡而做出回應的能力仍十分有限，一旦資料在不同情境下可能存在多義性或分歧性時，第二波的統計學習就曝露弱點。為此 DARPA 特地描繪了「未來」第三波人工智慧的願景，將之稱為「脈絡適應 (contextual adaptation)」的年代，期許日後人工智慧能自如地在不同脈絡情境中轉換與適用，也能對其做出之決定提供說理與說明 (reasoning and explanation)，從而更接近夢想中「與人類有相同程度的人工智慧 (HLAI)」。²⁰但就本文探索「弱 AI」對法學衝擊影響的目標而言，顯然必須將眼光鎖定在當前第二波的發展。

3. 第二波人工智慧的知識學習與知識生產能力

第二波人工智慧藉「機器學習」直接從大量資料中辨認其中各種資料分布規則或關聯模式的能力，雖然突破第一波人工智慧仰賴符號與外顯知識規則的限制，終究仍因「脈絡適應」能力的不足，在應用範圍上無法成為通用型人工智慧 (AGI)。但除了從「應用範圍」的寬窄 (AGI or narrow AI) 與有無「自我意識/意向性」的 (strong or weak AI) 兩個面向 (見表一)，評斷第二波人工智慧「模仿」人類的能力之外，是否有其他判準更適合於描述第二波人工智慧的能與不能，以區辨其與第一波之差異，並藉以評斷其對人類社會，特別是法律系統中的法學研究帶來的衝擊？科技與社會研究 (STS) 對人類知識如何形成、取得與傳遞的相關研究，提供回答上述問題的可能切入點。

¹⁵ See NILSSON, *supra* note 6, at 515-521.

¹⁶ David Silver et al., *Mastering the Game of Go Without Human Knowledge*, 550 NATURE 354-359 (2017).

¹⁷ See ALPAYDIN, *supra* note 14, at 111-117.

¹⁸ Charles Duhigg, *How Companies Learn Your Secrets*, THE NEW YORK TIMES MAGAZINE (Feb. 16, 2012).

¹⁹ JOHN HAUGELAND, *ARTIFICIAL INTELLIGENCE: THE VERY IDEA* 391 (Cambridge, MA: MIT Press, 1985).

| | | 應用範圍 | |
|--------------|------------------|--------------------|----------------------------|
| | | AGI (寬) | Narrow AI (窄) |
| 自我意識/ 意向性 | Strong AI (有) | 仍未實現 | 仍未實現 |
| | Weak AI (無) | 短期內之發展目標 (HLAI) | 目前的人工智能主體 (自駕車、AlphaGo) |

表一 人工智慧的兩種典型分類

3.1 科技與社會研究的理論爭戰：Dreyfus 與 Collins

前述 Dreyfus 對第一波人工智慧的批評，正是根據他對人類如何取得技能 (skill) 與專業知能 (expertise) 的相關研究而來。Dreyfus 以飛行員、汽車駕駛、西洋棋及外語學習等過程為例，將人類取得技能所可能經歷的熟練程度分為五個階段。Dreyfus 指出，從最低階僅能依樣畫葫蘆的菜鳥，到最高階能出神入化的達人專家，兩者間的最大差別是後者可本於不思即能的直覺，應對非結構化的不同情境脈絡，自然做出最適切的決定與處置。²⁰ Dreyfus 認為，專家掌握非結構化脈絡的直覺能力，是在反覆實作的經驗中，將對應於各種情境的最佳決策及行動相連結起來的身體記憶，因此難以表述，也不是第一波仰賴由符號所體現之形式邏輯規則所能窮盡，因此第一波人工智慧從方法上就已注定不可能取得真正的專業知能。相反地，Dreyfus 對於 1980 年代中期之後，由模擬人類神經網絡 (neuron-net, neural network) 的電腦系統，直接建立特定輸入值 (刺激) 與輸出值 (反應) 間配對關聯，藉以模仿人類經驗中各種連結記憶 (associative memory) 的方法，則寄予厚望。²¹ 類神經網路系統的發展雖受限於當時資料蒐集與電腦儲存及運算能力，一開始並未有顯著成果，但 Dreyfus 仍大膽預測，只要能掌握人類經驗的整全 (holistic) 樣態，²² 類神經網絡系統假以時日必能從過去範例抽繹出其中存在的各種關聯模式，成功模擬專家所具有的直覺，畢竟類神經網絡系統本身就是一種學習情境脈絡的系統 (context-learning networks)。²³

相較於 Dreyfus 批判第一波人工智慧，卻樂觀看待日後成為機器學習技術核心的類神經網絡系統，科學知識社會學家 Harry Collins 則不看好第二波人工智慧在取得人類知識與技能的能力上，能大幅超越第一波的侷限。Collins 的專長在研

²⁰ 這五個階段分別是菜鳥 (novice)、進階新手 (advanced beginner)、熟練 (competence)、精通 (proficient)、達人專家 (expert)。DREYFUS & DREYFUS, *supra* note 1012, at 16-36.

²¹ *Id.* at 90-94.

²² DREYFUS & DREYFUS, *supra* note 10, at 90-94.

²³ Hubert L. Dreyfus, *Response to Collins, Artificial Experts*, 22:4 SOCIAL STUDIES OF SCIENCE 717, 724 (1992).

究人類如何取得專業知識，他以超過三十年的時間，長期實際參與觀察重力波物理學的研究，並提出「互動型專業知能 (interactional expertise)」的概念，一方面用以說明他在欠缺「貢獻型專業知能 (contributory expertise)」的情況下，卻能以一名社會學家的身分「參與」重力波物理學研究所需具備的基本專業知能，另一方面也用這組概念區分，規範性地界定公民參與科技政策決策的正當邊界：想要正當地參與科技政策中技術階段的決策，必須要先自我具備「互動型專業知能」，否則就只在政治階段才有參與資格。²⁴ 從 1990 年起至 2018 年，Collins 一共出版了四本探討機器能否取代人類的專書，時間點上剛好橫跨了第一波人工智慧結束與第二波興起之時。

Collins 將人類知識或技能的傳遞與學習模式區分為「算則模式 (algorithmic model)」與「涵化模式 (enculturational model)」兩類。²⁵ 「算則模式」中的人類知識可透過語言被清楚表述，而以類似食譜的形式傳遞；「涵化模式」中的人類知識則不透過語言表達，而是一種非規則性的「默會知識 (tacit knowledge)」，必須透過「社會化」過程才能獲得的社會技能。²⁶ 能以「算則模式」傳遞的知識或技能，充其量僅限於可完全不考慮背景脈絡而具有固定對應關係，且被期待應追求一毫不差的「模版印刷型態之人類行為」(mimeomorphic actions)，例如高爾夫球的揮桿、汽車的換檔、閱兵分列式的動作，均要求做出相同的舉止以完成其行為。反之，舉止與行為間欠缺一對一關係的「社會多形體之人類行為」(polimorphic actions)，其知識與技能之內容無法被形式化，需要因地、因時制宜，也會因不同社會典範、文化與目的而異，因此僅能透過「涵化模式」傳遞或學習，例如，書寫情書、在擁擠的城市街道上安全而符合當地交通文化地駕車、對什麼人應該在何時做出何種適當問候的技能、千變萬化戰場上當機立斷的能力等。²⁷

Collins 認為，以「算則模式」來運作的第一波人工智慧，充其量僅能模仿「模版印刷型態之人類行為」，卻對「社會多形體之人類行為」完全無能以對。是以，第一波人工智慧頂多在人類選擇讓自己的行動像機器一樣運作的領域中，才看起來顯得具有智慧，²⁸ 而多半時候這也只是人類自願地將原本屬於「社會多

²⁴ 以「互動型專業知能」作為界定公共參與科技政策決策的門檻，可能存在一些嚴重的規範性問題，請參見 Wen-Tsong Chiou, *What Roles Can Lay Citizens Play in the Making of Public Knowledge?*, 13(2) EAST ASIAN SCIENCE, TECHNOLOGY AND SOCIETY 257-277 (2019).

²⁵ HARRY M. COLLINS, *CHANGING ORDER: REPLICATION AND INDUCTION IN SCIENTIFIC PRACTICE* 57 (Chicago: Univ. of Chicago Press, 1985).

²⁶ HARRY M. COLLINS, *ARTIFICIAL EXPERTS: SOCIAL KNOWLEDGE AND INTELLIGENT MACHINES* 4 (Cambridge, MA: MIT Press, 1990) [hereinafter COLLINS, AE].

²⁷ Harry M. Collins, Gerard H. de Vries & Wiebe E. Bijker, *Ways of Going On: An Analysis of Skill Applied to Medical Practice*, 22(3) SCIENCE, TECHNOLOGY, AND HUMAN VALUES 267-285 (1997); HARRY M. COLLINS & MARTIN KUSCH, *THE SHAPE OF ACTIONS* 31-54 (1998). Collins 特別說明其使用 poli 做為 polimorphic 字首的原因，蓋其為同時意指 poly-之多數與 polis-之社會性的雙關語。Id. at 33.

²⁸ Collins 在 1990 年時尚未使用 mimeomorphic actions 和 polimorphic actions 的用語，而是

形體之人類行為」以「模版印刷型態之人類行為」取而代之，再由人類自行彌補或修復 (repair) 做為「社會義肢 (social prosthesis)」的機器所無法完成的工作。例如，口袋型電子計算機雖已擁有快速的算數計算能力，但仍仰賴人類依照具體問題自行輸入相關的數字與運算符號才能得出結果，且仍須由人類自行解讀其意義才能應用於實際社會生活；又例如以麵包機預設的少數幾種麵包製作模式，權充一位手藝精湛的麵包師。²⁹ 這些是第一波人工智慧之所能，但也即其極限。

那麼，直接從特定領域累積的大量資料中辨識其中隱含資料與資料間模式的第二波人工智慧，是否就已進化到「涵化學習」模式，而成功地學習「默會知識」，甚至模仿「社會多形體之人類行為」呢？Collins 依舊給出極為保留的評斷。Collins 在 2010 年出版的 *Tacit and Explicit Knowledge* 一書中，將涵化模式賴以運作的「默會知識」與「社會化」作了進一步的分析。Collins 認為源自於 Michael Polanyi 的「默會知識」概念，³⁰ 其實包含了三組不同的意涵。³¹ Polanyi 的「默會知識」雖然以不可表達性 (inexpressibility) 與不可明定 (unspecifiability) 為其核心，例如騎腳踏車的技能，並非熟讀一本詳盡描述腳踏車運作方式與原理的使用手冊即能獲得，而必須自行揣摩與反覆練習。但 Polanyi 似乎也強調拜師（委身權威）學藝與身體揣摩實作的重要性。

Collins 認為「默會知識」的概念不夠精確，其中包含數種不同的內涵，乃將之一分為三，分別為：「關係性的默會知識 (relational tacit knowledge, RTK)」是特定社會關係下偶然形成對外不被明白說出而以秘密形式存在的特定知識或技能，例如祖傳秘方；「身體性的默會知識 (somatic tacit knowledge, STK)」則是由身體藉由實作而獲得的知識或技能，騎腳踏車或飛行員的技能即屬之；「集體社會性的默會知識 (collective tacit knowledge, CTK)」則是一種鑲嵌於社會生活形式之中，與「社會多形體之人類行為」有關的知識或技能。這三種默會知識雖然都不是以語言表述之形式存在，難以被直接形式化與規則化，但卻非完全不可能轉化為可被言說的「外顯知識 (explicit knowledge)」。存在於社會關係之中屬於「關係性默會知識」的秘密，不太可能從社會中完全消除，但本質上並非不可表述或明定，因此一旦秘密被揭露，即可轉變為外顯知識。³² 儲存於身體記憶中的「身體性默會知識」則因其本身就是一組一組因果序列的知識，具有被科學性地理解並表述為外顯知識的可能，除了受實際工程技術或材料性質所限之外，身體藉由

machine-like actions 或 behavior-specific action 指稱後來的 mimomorphic actions，以 regular 或 ordinary actions 指稱後來的 polymorphic actions。COLLINS, AE, *supra* note 26, at 32-42.

²⁹ HARRY M. COLLINS, *TACIT AND EXPLICIT KNOWLEDGE* 71-72, 143, 170 (Chicago: Univ. of Chicago Press, 2010) [hereinafter COLLINS, TEK].

³⁰ MICHAEL POLANYI, *PERSONAL KNOWLEDGE* (London: Routledge & Kegan Paul, 1958). Polanyi 原是一位化學家，後轉行當起了哲學家。

³¹ 其他學者也多認為「默會知識」概念具有多義性，see, e.g., Klaus Nielsen, *The Concept of Tacit Knowledge-A Critique*, 5(2) *OUTLINES* 3-25 (2002).

³² COLLINS, TEK, *supra* note 29, at 97-98.

實作所取得的技能並非不可被機器所「複製」，同時 Collins 也承認，透過類神經網絡系統所使用的正負向回饋的制約訓練（此即為本文前述第二波機器學習中的「強化學習」方法），人工智慧技術確實已有能力習得「身體性默會知識」。³³ 但 Collins 堅持認為，到目前為止我們僅知道人類在「實作」之外，主要是靠著學習特定社會之「語言」並使之達到流利（fluency）的程度，來完成獲取「集體社會性默會知識」所需的「社會化」，但對於「社會化」的具體機制為何，仍欠缺完整的認識，因此「集體社會性默會知識」是三種默會知識當中最難以轉換為外顯知識者。³⁴

基於上述對「默會知識」的拆解分析，Collins 認為人工智慧是否能獲得與「社會多形體之人類行為」有關的「集體社會性默會知識」，將取決於人工智慧能掌握多少「語言」能力。Collins 在 2018 年出版的 *Artificial Intelligence* 一書中，提出一個三段論的前提假說，並依其經驗研究成果做出以下主張：(1) 除非電腦能完全嵌入一般的人類社會當中，否則將不能學會流利的自然語言以通過嚴格版的圖靈測試，從而取得類人類的完整智能；(2) 受限於當下技術的緩慢進展，目前電腦仍未能完全嵌入一般的人類社會當中。³⁵ Collins 認為，目前機器學習所進行的「模式辨認（pattern recognition）」仍或多或少須仰賴「現有人類知識」進行訓練資料的篩選、詮釋與分類。³⁶ 從這個前提，Collins 完成了以下的推論：欲學習現有的人類知識，必須先具備能「以流利的自然語言進行溝通」的「互動型專業知能」；要能以流利的自然語言進行溝通就必須先「浸淫在專家團體的對話當中」；必須「進入專家團體」才有辦法真正浸淫在專家團體的對話；「取得專家團體的信任」是進入專家團體的必要條件；取得專家團體的信任仰賴包括身體互動、小團體、專業、原創性等目前的機器仍欠缺的條件；僅透過網路上的各種訊息進行機器學習的（第二波）人工智慧，仍無法真正取得進入專家團體所需的信任，也因此無法浸淫在專家間的對話情境，從而無法真正取得學習人類知識所需的「互動型專業知能」。³⁷

3.2 語言、實作及人工智慧的能與不能

Collins 將「實作」與「語言」二分，主要受到維根斯坦後期哲學中「生活形式（forms of life）」概念的影響。³⁸ 但 Collins 以「實作」與「語言」分別對應「身

³³ *Id.* at 117, 160, 162.

³⁴ COLLINS, TEK, *supra* note 29, at 138, 161. 陳瑞麟則反對 Collins 以默會知識解釋實驗知識的傳遞。請參見陳瑞麟，*認知與平價*（下），頁 413-415（2012 年）。

³⁵ HARRY M. COLLINS, *ARTIFICIAL INTELLIGENCE: AGAINST HUMANITY'S SURRENDER TO COMPUTERS 2* (Cambridge, UK: Polity Press, 2018) [hereinafter COLLINS, AI].

³⁶ Collins 甚至不認為機器學習有完全不仰賴任何先知識而進行真正「由下而上（bottom-up）」學習的可能。See COLLINS, AI, *supra* note 35, at 110-130.

³⁷ COLLINS, AI, *supra* note 35, at 149, 166, 172.

³⁸ Harry Collins, *Language and Practice*, 41(2) *SOCIAL STUDIES OF SCIENCE* 271-300 (2011); COLLINS, AI,

體性默會知識」與「集體社會性默會知識」，並將人工智慧能否成功模仿人類的關鍵置於其「語言」能力，則與他所提出的「貢獻型專業知能」與「互動型專業知能」區別有關。簡言之，Collins 認為即使欠缺參與實踐科學實作的「貢獻型專業知能」，也能單獨藉由流利地以特定專業領域語言進行溝通的「互動性專業知能」，完整理解此一專業領域的生活形式。³⁹

乍看之下，Collins 對「語言」以及「集體社會性默會知識」的重視，是他與 Dreyfus 意見分歧的根源。Collins 曾明言批評 Dreyfus，認為他過分重視「身體」與「實作」，忽略了「語言」與「互動型專業知能」在取得人類知能上的重要性，也因此高估第二波人工智慧透過類神經網絡系統（以及後來的深度學習）模仿人類的能力。⁴⁰ 對 Collins 而言，第二波人工智慧充其量或許能習得「身體性默會知識」，但難以再進一步地全面社會化而獲取人類知能中最為關鍵的「集體社會性默會知識」。然而，Dreyfus 雖未區分身體「實作」與進行溝通所需的「語言」能力，但 Dreyfus 也暗示，「從先例中學習」的方法能否成功地建立不同情境脈絡下刺激與反應的連結記憶，關鍵在於提供訓練的先例能否充分掌握人類經驗的整全 (holistic) 樣態。⁴¹ 換言之，Dreyfus 強調「整全性」做為其推論條件，對 Collins 批評的可能回應是：倘若提供作為訓練的先例能整全地將 Collins 所稱的「集體社會性默會知識」也包含在內的話，並不存在根本性的理由懷疑「從先例中學習」的方法不能習得「集體社會性默會知識」；Collins 認為人工智慧必須「先」擁有流利的自然語言能力，才能藉由「互動型專業知能」與人類的專家進行建設性的交流溝通，進而掌握模式辨認學習「前」就需要的完整人類知識；但機器學習本身就是學習包括各種默會知識在內之人類知識的方法，不需要「先」取得其他的能力。Collins 與 Dreyfus 的差異與其說是前者重視「語言」和「集體社會性默會知識」，後者卻只在意「實作」與「身體性默會知識」，不如說前者認為「語言」是取得「集體社會性默會知識」的必要手段，而後者認為「實作」中即包含「集體社會性默會知識」，可從完整展現「實作」之先例中直接學習而得。

事實上，第二波人工智慧的機器學習雖常仰賴自然語言作為學習之「媒介」，但卻並非總是如此。以「監督式學習」為例，除非所發展的人工智慧技術本身就結合著語言系統，例如智慧語音助理，否則一般都直接以人類知識所指涉的現象為訓練材料，不再透過自然語言的媒介。例如，流通於專業放射線科醫師社群之內的醫療影像診斷知識，難以被形式化與規則化，性質上為 Collins 所稱之「身體性默會知識」與「集體社會性默會知識」，長久以來無法有效地由第一波的「算

supra note 35, at 54, 63.

³⁹ Harry Collins & Robert Evans, *Expertise Revisited, Part I—Interactional Expertise*, 54 *STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE PART A* at 113, 115-116 (2015); Harry Collins, Robert Evans, & M. Weinel, *Expertise Revisited, Part II – Contributory Expertise*. 56 *STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE PART A*, 103-110 (2016).

⁴⁰ COLLINS, TEK, *supra* note 29, at 147.

⁴¹ DREYFUS & DREYFUS, *supra* note 10, at 28, 90-92.

則模式」將相關知識傳授予機器，但透過專業醫師的影像標註，第二波的監督式學習得以在機器上訓練出相當於專業放射線科醫師（甚至更少失誤與更有效率）的醫療影像診斷能力，卻不必「先」幫機器培養出與專業放射線科醫師進行溝通所需要的「互動型專業知能」。又例如，在繁忙的街道上行車一直是 Collins 心目中「社會多形體人類行為」的代表，而第二波人工智慧的自駕車對各種街道狀況的辨識訓練，雖仍需要人類的標註與調教，但同樣不以自然語言為媒介，而是從人類駕駛的「實作」中直接學習，去發展高等級的自駕車。至於第二波的「強化學習」與「非監督式學習」同樣也不需仰賴自然語言，即可訓練出具有強大能力的 AlphaGo Zero 與預測消費行為的系統。

第二波人工智慧的發展顯然已超出 Collins 的理論認識，不僅大多時候並不仰賴自然語言能力，諸多例證更顯示，只要機器學習之先例包含充分的實作足以反映「社會多形體人類行為」的多樣性，第二波人工智慧也能學會包括「集體社會性默會知識」與「身體性默會知識」在內的各種人類知能。關鍵僅在於：做為訓練的「實作」先例是否「足夠整全」地包含多數可能的「社會多形體人類行為」，以使人工智慧機器在不完全相同但相類似的情境中，仍具有進行推理的能力。雖然一味宣揚人工智慧福音不值一哂，但僅因第二波還不完全能以流利語言與專家溝通，就認為沒有資格稱為人工智慧，並沒有回答第二波究竟可能帶來哪些衝擊的問題。

真正的挑戰來自於以下兩個面向。首先，在第二波中明確以現有知識做為學習準據的「監督式學習」，既然以現有分類知識為典範，所訓練出來的人工智慧即使能模仿人類操作與應用現有知識，卻不可能超越現有知識的框架與極限，成為一位能自主引入原創性元素、對現有知識進行修正、改造或演繹的人類專家。此外，當「監督式學習」以「過去之人類決策行為」為學習對象（例如過去被認為成功而足堪典範的人事聘用、入學許可、司法判決等決策），承襲各種「身體性默會知識」與「集體社會性默會知識」的人類決策被涵化於機器的同時，也可能將深植於「身體」與「社會集體」中的歧視或偏見一併複製。承襲傳統及各種默會知識的人工智慧，也將承襲其積習與可能的不理性與錯誤。原本人類透過「打破成規（rule-breaking）」締造文明進化，藉由新的問題框架（framing）與重新協商「後設科學價值決定（metascientific value judgement）」等途徑改變知識內容甚至移轉科學典範的可能性，在第二波將現有知識與成規立為典範的學習模式下，並不存在。

其次，藉由第二波中的「強化學習」與「非監督式學習」方法，已有能力發展出「超越」現有人類知識侷限的人工智慧，例如，AlphaGo 與 Alphago Zero 的創新棋步，迫使圍棋界必須改變思維，甚至轉而向電腦拜師學藝。雖然 Collins 認為，以機器學習方法發展出的各種人工智慧下棋系統，其實只是靠著電腦的運算

蠻力 (brute force)，並非真的模仿人類下棋時帶有創意的思考模式後贏得勝利，可說勝之不武，也不能算是真的在「下棋」。⁴² 但是當「耕耘機比人更會耕田」成為現實時，質疑耕耘機並非真的在「耕田」並不會改變工業革命所提供的「效率」對第一級產業造成的社會衝擊。當人工智慧透過非監督式學習，宣稱比老闆更能預測一個應徵者是否適任、貸款申請者是否能準時還款時，質疑其欠缺可資驗證的「基準事實 (ground truth)」，⁴³ 或者呼籲「別向機器投降」，⁴⁴ 都很難抵抗人類自身想利用人工智慧所提供「超越現有知識」的預測能力進行各種「控制」的慾望。

總括來說，Collins 對人工智慧是否已能取代人類的判斷或許不盡正確，但其提出的諸多概念工具卻有助於我們更細緻地分析人工智慧發展的現狀。從 Collins 與 Dreyfus 的交鋒以及吾人對第二波不同技術方法的理解，我們可以得出以下的描述性結論：只要機器學習的實作先例涵蓋夠廣（而非先取得流利的自然語言能力），過去在第一波人工智慧下難以被「算則模式」所掌握，而與「模版印刷型態之人類行為」或「社會多形體人類行為」有關的各種「身體性默會知識」及「集體社會性默會知識」，就能被第二波人工智慧所涵化；但承襲人類傳統及各種默會知識的人工智慧，不僅同時承襲其積習與可能的錯誤，也可能因無法自我超越而使仰賴此等人工智慧的人類，陷入文明進化遲滯的危險。另一方面，非建立在現有分類知識之上的「強化學習」與「非監督式學習」，雖具有發展超越現有知識框架之人工智慧的潛能，但亦可能引發人類利用尚無法為現有知識所驗證的預測能力進行各種控制的慾望。那麼，上述的第二波人工智慧對法律系統中的法學研究可能帶來什麼樣的衝擊？

4. 第二波人工智慧對法學的衝擊：以法律規範論證為中心

對法規提出具體主張與建議，一直是法學的主要工作，而這樣的工作則必須建立在法律規範論證之上，⁴⁵ 也就是針對法規應如何制訂、修改或解釋適用的主張，以提出理由的方式來加以證立。分析第二波人工智慧對法學的可能衝擊，不可迴避地必須從法律活動參與者所從事的法律規範論證行為著手。借用 Collins 的說法，「規範論證行為」就是構成法學之所以是一種特殊生活形式 (form of life) 而與其他社會生活形式有所區別的構成性行為 (formative actions)。⁴⁶ 法

⁴² COLLINS, AI, *supra* note 35, at 83-85; COLLINS, TEK, *supra* note 29, at 106-110.

⁴³ 因欠缺基準事實驗證導致的真實爭議可參見 CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 3-11 (New York: Crown, 2016).

⁴⁴ 這是 Collins 在 *Artificial Intelligence* 一書中的規範性訴求，如該書之副標題：Against Humanity's Surrender to Computers。

⁴⁵ 請參見顏厥安，*規範建構與論證—對法學科學性之檢討*，收於：*規範、論證與行動：法認識論論文集*，頁 18 (2004 年)。

⁴⁶ COLLINS & KUSCH, *supra* note 27, at 8-12.

學生活形式的「構成性行為」是否以及在何種程度內可被人工智慧模仿或取代，將關係著法學在第二波人工智慧的衝擊下可能產生的變化。

依照 Robert Alexy 的分析，法律規範論證可分為外部證立（external justification）與內部證立（internal justification）二者。⁴⁷ 法律規範論證的邏輯結構是由法規構成要件、特定案例事實與法律效果三者所構成。一般將此一由法規構成要件之大前提，到特定案例事實之小前提該當法規構成要件，再到法律效果結論的論證過程稱為「法學三段論」。其中，外部證立所要確認的是做為大前提之法規範本身的正確性或真實性，而內部證立所關心的則是將特定案例事實歸屬於法規構成要件而得出法律效果的邏輯推論有效性。⁴⁸ 法學三段論的最簡單形式，就是在外證立上發現/認識有效的法律規範，⁴⁹ 並在內部證立上確認特定案件事實吻合法律規範構成要件要素，以得出法律規範所連結的特定法律效果。⁵⁰ 乍看之下，法學規範論證是一個可以被高度形式化與規則化的行為，相關的法學知識似乎可以透過「算則模式」傳遞與學習。

然而，法律人在進入實際案例的操作後，多半都能很快理解這個法學三段論的簡單形式幾乎很少在真實世界中發生。王鵬翔就指出，法規構成要件與案件事實的描述間總是存在著各種「裂縫」，需要透過「涵攝」來加以填補，才能完成法律適用的工作。⁵¹ 至於「裂縫」的產生，一方面是因為有效的法律規範究竟為何，並不是透過簡單經驗觀察就能一望即知的事實（plain fact），而是需要透過詮釋加以論證的規範主張，這是法律的論爭性格（argumentative）使然。⁵² 另一方面，真實世界的複雜事實如何被裁剪成與法律規範相關的案件事實，本身就需經過挑選、證明與詮釋，經裁剪後的案件事實描述（通常以某甲、某乙如何如何的方式表達）也往往不完全等同於構成要件。「涵攝」就是一種透過規範論證填補「裂縫」的行為，找出為了填補構成要件與案件事實間之裂縫，需要加入哪些額外（大）前提才能確保內部證立符合推論邏輯，並提出「理由」論證這些額外前提的正確性（外部證立）。至於當法律效果存在裁量空間，亦即單一法律規範可對應複數之法律效果時，理論上亦是考量案件事實的各種特徵之後，依同樣的方式，藉由涵攝論證構成要件該當之特定案件事實，應該對應何種法律效果。

⁴⁷ ROBERT ALEXY, A THEORY OF LEGAL ARGUMENTATION: THE THEORY OF RATIONAL DISCOURSE AS THEORY OF LEGAL JUSTIFICATION 221 (Ruth Adler & Neil MacCormick trans.)(Oxford: Clarendon Press, 1989)(1978). 中文另請參照顏厥安，〈法、理性與論證—Robert Alexy 的法律論證理論〉，收錄於氏著《法與實踐理性》，頁 146（1998 年）。

⁴⁸ 王鵬翔，論涵攝的邏輯結構—兼評 Larenz 的類型理論，成大法學，9 期，頁 20（2005 年）。

⁴⁹ 認識及描述有效的法律規範是法釋義學的主要工作之一。王鵬翔、張永健，經驗面向的規範意義—論實證研究在法學中的角色，中研院法學期刊，17 期，頁 221-225（2015 年）。

⁵⁰ 王鵬翔稱之為「法律適用的簡單模式」。王鵬翔（註 48），頁 8-12。

⁵¹ 王鵬翔（註 48），頁 12-13。

⁵² See RONALD DWORKIN, LAW'S EMPIRE 3-11, 31-37 (Cambridge, MA: Belknap Press, 1986).

可用來證成某個規範主張的規範性理由，傳統上認為包括對規範構成要件所使用之文字做出文義解釋的語意規則、對立法者意旨提出說明的立法史料、可作為理解法規範內涵參考座標的外國法規範或實施狀況的描述、乃至於對法規範所追求目的之價值理解。然而近年來，實證科學所提供的經驗論據具有補充修正既有法規範，甚至發現新的規範前提的功能，逐漸受到重視。⁵³ 王鵬翔與張永健更認為，「差異製造事實（difference-making facts）」能透過因果關係的刻畫，指出一個行為會造成何等差異，因此「差異製造事實」可藉由對特定的規範命題提供結果論或目的論的說明而成為規範理由；至於差異製造事實的「發現」，則可透過法實證研究所作出的因果推論而來。⁵⁴

然而，從事觀察性、實驗性研究或統計的後設分析以「發現」新的差異製造事實，向來不是傳統法學訓練的重點。法學訓練及多數的「內在觀點」法學研究基本上是透過類似算則模式的方式，取得關於法律規範構成要件要素的去脈絡化抽象知識後，學習並探討如何填補存在於法規範構成要件與案件事實描述、法律效果間的各種「裂縫」，進行以涵攝為核心的規範論證。形成法學特殊生活形式（form of life）的此種規範論證活動，主要透過呈現在法學論著與法院判決中的語言與文字進行。法律人藉由閱讀國內與國外法學論著及法院判決所累積的浩瀚先例，學習被當下法學社群所認可的涵攝行為，或批判其中存在的各種瑕疵，並嘗試將此等能力，創造性地展延應用於新的社會事實。這使得法學的「實作」與法律「語言」，存在高度重疊的關係。參與法學的「實作」必需先掌握與法律參與者以流利「語言」進行溝通的能力；不存在沒有法律「語言」就能參與法學「實作」的可能性。⁵⁵ 至於如何將真實世界的複雜事實裁剪成與法律規範相關的案件事實描述，也就是「裁決事實（adjudicative facts）」⁵⁶的認定，牽涉到證據取得、證據評價、甚至不同領域之專業知識（domain knowledge），其重要性雖然在法律實務上不亞於規範論證（甚至更為重要），但一般並未成為台灣法學訓練之內容，亦鮮少成為法學研究之對象。

法學研究的上述結構與特性，使得第二波人工智慧有可能在以下三個面向上參與或模仿法學的相關活動，從而對法學帶來衝擊：創造新的差異製造事實、模仿規範論證行為、模仿法律效果裁量行為。

(1) 創造新的差異製造事實

⁵³ 例如邱文聰，被忽略的（立法）事實—探詢實證科學在規範論證中的可能角色兼評釋字第584號解釋，臺大法學論叢，37卷2期，頁233-284（2008年）

⁵⁴ 王鵬翔、張永健（註49），頁233-239，246-264。

⁵⁵ 簡資修教授批評，部分法律經濟分析研究者，因未能將現實法律納入考量，僅以「黑板經濟學」分析法律，將無法對法學作出真正的貢獻。簡資修，科斯經濟學的法學意義，載於：經濟推理與法律，頁3,4（台北：元照，三版，2014年）。其所批判者即是因為欠缺參與法律「實作」所需的「語言能力」所致。

⁵⁶ 關於裁決事實的簡單說明請參見邱文聰（註53），頁244-247。

如前所述，第二波人工智慧能以「非監督式學習」方法，發現超越現有人類因果知識所能充分說明的資料「相關性」，例如發現購買無香料乳液與懷孕具有相關性，發現購買家具保護貼與貸款償還履約遵從度具有相關性等。⁵⁷ 此等「相關性」因為欠缺因果關係的知識，並不足以直接作為發現「差異製造者(difference maker)」的證明：絕對沒有辦法說購買無香料乳液導致了懷孕，也沒有辦法說購買家具保護貼是還款與否的差異製造者；但卻可能在「結果論證」下，⁵⁸ 成為氾濫而不假思索地創造新「控制手段」以證成各種「控制目的」的誘因：如果向購買無香料乳液的女性消費者促銷婦嬰用品，將能增加銷售該等用品成功的機會；如果不貸款給沒有購買家具保護貼的貸款申請人，將可以降低銀行呆帳的風險。當「非監督式學習」方法被廣泛地運用在發現各種超越人類現有知識能說明的「相關性」時，就難以避免被用來擴張人類社會中的各種控制慾望。換言之，利用第二波人工智慧提供新的差異製造事實的目的，主要並非基於以機器取代人力的「效率」考量，而是藉由預測以進行「控制」。這對於將法學自外於社會價值與目的選擇的法學工作者而言，並不會造成太大的影響；但對於將節制各種權力當作是當代法學價值基礎的法學工作者而言，卻帶來莫大的挑戰。

(2) 模仿規範論證行為

以涵攝為中心的規範論證顯然不是算則模式可以輕易掌握的「模版印刷型態之人類行為」，而是一種充滿身體性與集體社會性默會知識的「社會多形體之人類行為」。然而法學規範論證之「實作」與「語言」高度重疊的特性，使得第二波人工智慧有機會從大量法學論著與法院判決的資料中，將各種不同案件事實描述與特定法規構成要件進行正確與否的配對，藉由「監督式學習」取得法律人從事涵攝所需要的身體性與集體社會性默會知識。愛沙尼亞法務部在 2019 年開始推動由「機器人法官」審理小額訴訟（7000 歐元以下）的試辦計畫，⁵⁹ 已經開始實際應用第二波人工智慧的此一模仿能力。其主要目的乃在於利用人工智慧所提供的「效率」，消化大量訴訟案件的壓力。

然而以人工智慧模仿規範論證行為，進而取代真人法官從事判決工作的作法，受限於法規範論證之實作與語言必然與特定法域之法規範及其所使用之語言有關，不同國家所發展模仿規範論證行為的人工智慧系統，較難產生跨國應用的可能。除此之外，以人工智慧取代真人法官的應用，也遭受諸多更根本的質疑。

⁵⁷ See Charles Duhigg, *What Does Your Credit-Card Company Know About You?*, THE NEW YORK TIMES MAGAZINE (May 12, 2009).

⁵⁸ 結果論證就是以採取某個手段可能產生的結果，來說明為什麼應該（或不應該）採取這個手段的論證方式。關於「結果論證」的說明請參見王鵬翔、張永健（註 49），頁 242-243。

⁵⁹ See Eric Niler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, WIRED (Mar. 25, 2019, 7:00 AM), <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.

⁶⁰ 就第二波人工智慧技術本身所存在的限制而言，即使第二波人工智慧確實可能習得目前法學社群從事規範論證所需的各種「身體性默會知識」及「集體社會性默會知識」，但就如同前已提及，機器學習在承襲一個社群之默會知識的同時，也就承襲其積習與可能的錯誤。第二波人工智慧即使已能熟練於當下的各種司法見解，也因無法自我超越而不可能創造推翻先例以促成人類文明進展的可能。倘若 *Brown v. Board of Education*⁶¹一案是由第二波人工智慧的機器人法官來審理，其結果將不會是終結「隔離而平等」的舊傳統，而是繼續肯認 *Plessy v. Ferguson*⁶²的合法性。⁶³ 換言之，利用人工智慧模仿規範論證行為以促成司法「效率」的同時，可能必須付出長久無法移轉過時典範的代價。

(3) 模仿法律效果裁量行為

法律效果的裁量是廣義涵攝活動的一部分，也因此法律效果的裁量行為也具有被第二波人工智慧的「監督式學習」方法所涵化的可能。目前司法院所推出的「量刑資訊服務平台」雖然仍稱不上是真正的人工智慧系統，但藉由統計分析過去之判決如何就不同犯罪類型，⁶⁴ 依據量刑因子決定刑罰種類與刑度，此一量刑系統乃相當於學習過去法官的量刑行為（假設法官確實只考慮犯罪行為與量刑因子），藉以預測（或指引）未來法官行為的一套準人工智慧系統。其背後想像無非是希望在「限定的量刑因子」範圍內，盡量將法律效果的裁量，規則化為具有一對一對應關係的「模版印刷型態之人類行為」，以求裁量結果能在形式上維持一致。

值得一提的是，人工智慧量刑系統在美國也發展出另一種令人擔憂的樣貌。例如，美國威斯康辛州所使用的 COMPAS 系統，藉由 137 項人格特質及社經資料，訓練出一套被告再犯風險的人工智慧預測軟體，作為法官量刑之依據。⁶⁵ 此一系統相較於司法院目前之量刑系統看似更為先進，卻隱含著混淆刑罰的行為主義與行為人主義的危險，更已從追求法官裁量行為符合形式平等，轉變為再犯控制之目的。當第二波人工智慧的學習與分析對象，可以很容易地從「法官的行為」移轉到「被告」身上時，將人工智慧應用於法律效果裁量就不得不更謹慎為之。

⁶⁰ Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242 (2019).

⁶¹ 347 US 483 (1954).

⁶² 163 US 537 (1896).

⁶³ 關於人工智慧應用於律師所提供之其他法律服務所可能產生之問題，請參見 Frank A.

Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEORGE WASHINGTON LAW REVIEW 1 (2019).

⁶⁴ 目前提供包括妨害性自主罪、不能安全駕駛罪、槍砲案件、幫助詐欺罪、毒品案件、竊盜案件、搶奪罪暨強盜罪、殺人案件、普通傷害罪、肇事逃逸罪等十種罪名的量刑資訊服務。

⁶⁵ *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

5. 結論

人工智慧從 1950 年代發展迄今，已歷經重大的典範移轉。從仰賴程式語言符號，將既有人類知識之邏輯規則透過手工打造的方式在電腦上予以再現的第一波，到直接從大量資料的統計學習中辨識出隱含規則的第二波，人工智慧雖仍未具有人類的意向性，也還沒有發展出能在不同脈絡間轉換通用的能力，但單純以其有無人類自我意識與應用範圍寬窄為基準，已不足以適當描述第二波人工智慧所帶來的新挑戰。科技與社會研究對於人類知識如何形成、取得與傳遞的相關理論，則提供一個可能的分析框架。

依照 Dreyfus 的理論，第一波人工智慧的困境在於其所使用之規則導向學習方法，無法真正取得專家藉由連結記憶掌握非結構化脈絡的直覺能力。Collins 則以第一波仰賴「算則模式」充其量僅能模仿「模版印刷型態之人類行為」，說明第一波的困境在於其完全無法習得與「社會多形體之人類行為」有關的各種身體性與集體社會性默會知識。Dreyfus 與 Collins 除了以不同的說理對過往第一波的失敗做出解釋，也對第二波初步之發展與未來的展望，分別提出了評斷與預測。Dreyfus 認為採用類神經網絡方法的第二波人工智慧，將能從不同情境脈絡的「實作」範例中，抽繹出各種關聯模式，成功模擬專家透過連結記憶所建立起的直覺能力。反之 Collins 則悲觀地認為，由於目前仍無法使機器完全融入正常人類社會，機器仍無法發展出流利的自然語言能力與專家進行溝通，在欠缺此種「互動型專業知能」的情況下，類神經網絡與機器學習充其量將只能習得身體性的默會知識，卻仍將對於模仿「社會多形體之人類行為」所必要的集體社會性默會知識一籌莫展。

然而，Collins 似乎並未理解，無論是第二波的「監督式學習」、「非監督式學習」或者「強化學習」，都不以「先」取得自然語言能力為必要。事實上正如 Dreyfus 理論所暗示的，只要機器學習的實作先例涵蓋夠廣，各種「身體性默會知識」及「集體社會性默會知識」，就都能被第二波人工智慧的學習模式所涵化。第二波人工智慧的問題並不在於其還未能以流利語言與專家溝通，因而沒有資格稱為人工智慧；第二波的真正問題在於，一方面以現有知識做為學習準據的「監督式學習」，雖已能習得各種默會知識，卻無法超越現有知識的框架與極限，也將承襲既有的積習與可能的不理性與錯誤；另一方面，「非監督式學習」與「強化學習」雖已有能力發展出超越現有人類知識的預測能力，卻也可能因而引發人類利用尚無法為現有知識所驗證的預測能力以進行各種控制的慾望。

已能涵化身體性與及集體社會性默會知識的第二波人工智慧，至少在三個面向上，可能以其能增進「效率」或提供「控制」為由，參與或模仿法學的規範論證行為，因而對法學帶來衝擊：第一，「非監督式學習」方法所發現之資料間

「相關性」雖不足以直接證明「差異製造者」之存在，卻可能透過「結果論證」，促成在外部證立上創造各種以「控制」為目的之法規範的誘因；第二，法規範論證之「實作」與「語言」高度重疊的特性，使「監督式學習」具有取得法律人從事涵攝所需要的身體性與集體社會性默會知識的可能，但受限於既有知識之框架，利用人工智慧模仿規範論證行為雖或能帶來司法「效率」，卻也可能必須付出長久無法移轉過時典範的代價；第三，協助法律效果裁量的人工智慧，一方面可能以法官行為為學習對象，並以追求裁量結果能達成形式平等為目的，另一方面卻也可能以法律效果施加之對象（例如被告）為分析之客體，而以確保法律效果能達成控制作用為其目標。總的來說，第二波人工智慧對法學之可能影響，未來將繼續在「效率」、「法與社會的典範移轉」與「控制」這三條軸線上，不斷地進行辯證。結果如何，雖仍未可知，卻深深影響著法學的未來。

參考資料

1. 中文部分

- 王鵬翔（2005年），論涵攝的邏輯結構—兼評 Larenz 的類型理論，成大法學，9期，頁 1-45。
- 王鵬翔、張永健（2015年），經驗面向的規範意義—論實證研究在法學中的角色，中研院法學期刊，17期，頁 205-294。
- 邱文聰（2008年），被忽略的（立法）事實—探詢實證科學在規範論證中的可能角色兼評釋字第 584 號解釋，臺大法學論叢，37卷2期，頁 233-284。
- 陳瑞麟（2012年），認知與平價（上）（台北：臺大出版中心）。
- _____（2012年），認知與平價（下）（台北：臺大出版中心）。
- 簡資修（2014年），科斯經濟學的法學意義，載於：經濟推理與法律，頁 3-25（台北：元照，三版）。
- 顏厥安（1998年），〈法、理性與論證—Robert Alexy 的法律論證理論〉，收錄於氏著《法與實踐理性》，頁 95-212。
- _____（2004年），規範建構與論證—對法學科學性之檢討，收於：規範、論證與行動：法認識論論文集，頁 3-35。

2. 外文部分

- ROBERT ALEXY. 1989. A THEORY OF LEGAL ARGUMENTATION: THE THEORY OF RATIONAL DISCOURSE AS THEORY OF LEGAL JUSTIFICATION 221 (Ruth Adler & Neil MacCormick trans.)(Oxford: Clarendon Press).
- ETHEM ALPAYDIN. 2016. MACHINE LEARNING (Cambridge, MA: MIT Press).
- Wen-Tsong Chiou, 2019. *What Roles Can Lay Citizens Play in the Making of Public Knowledge?*, 13(2) EAST ASIAN SCIENCE, TECHNOLOGY AND SOCIETY 257.
- HARRY M. COLLINS, CHANGING ORDER: REPLICATION AND INDUCTION IN SCIENTIFIC PRACTICE (Chicago: Univ. of Chicago Press, 1985).
- _____. 1990. ARTIFICIAL EXPERTS: SOCIAL KNOWLEDGE AND INTELLIGENT MACHINES (Cambridge, MA: MIT Press).
- _____. 2010. TACIT AND EXPLICIT KNOWLEDGE (Chicago: Univ. of Chicago Press).
- _____. 2011. *Language and Practice*, 41(2) SOCIAL STUDIES OF SCIENCE 271.
- _____. 2018. ARTIFICIAL INTELLIGENCE: AGAINST HUMANITY'S SURRENDER TO COMPUTERS (Cambridge, UK: Polity Press).
- Harry Collins & Robert Evans. 2015. *Expertise Revisited, Part I—Interactional Expertise*, 54 STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE PART A at 113.
- Harry Collins, Robert Evans, & M. Weinel, *Expertise Revisited*. 2016. *Part II – Contributory Expertise*. 56 STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE PART A at 103.
- HARRY M. COLLINS & MARTIN KUSCH. 1998. THE SHAPE OF ACTIONS.
- Harry M. Collins, Gerard H. de Vries & Wiebe E. Bijker. 1997. *Ways of Going On: An Analysis of Skill Applied to Medical Practice*, 22(3) SCIENCE, TECHNOLOGY, AND HUMAN VALUES 267.
- Defense Advanced Research Projects Agency. 2017. Powerful But Limited: A DARPA

Perspective on AI.

- HUBERT L. DREYFUS. 1979. WHAT COMPUTERS CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON (Harper and Row, revised ed.).
- _____. 1992. *Response to Collins, Artificial Experts*, 22:4 SOCIAL STUDIES OF SCIENCE 717.
- HUBERT L. DREYFUS & STUART E. DREYFUS. 1986. MIND OVER MACHINE: THE POWER OF HUMAN INTUITION AND EXPERTISE IN THE ERA OF THE COMPUTER (Blackwell).
- Charles Duhigg. 2009. *What Does Your Credit-Card Company Know About You?*, THE NEW YORK TIMES MAGAZINE (May 12, 2009).
- _____. 2012. *How Companies Learn Your Secrets*, THE NEW YORK TIMES MAGAZINE (Feb. 16, 2012).
- RONALD DWORKIN. 1986. LAW'S EMPIRE (Cambridge, MA: Belknap Press).
- JOHN HAUGELAND. 1985. ARTIFICIAL INTELLIGENCE: THE VERY IDEA (Cambridge, MA: MIT Press).
- John McCarthy. 2007. *From Here to Human-level AI*, 171(18) ARTIFICIAL INTELLIGENCE 1174.
- MARVIN MINSKY. 2007. THE EMOTION MACHINE: COMMONSENSE, THINKING, ARTIFICIAL INTELLIGENCE, AND THE FUTURE OF THE HUMAN MIND (Simon and Schuster).
- Klaus Nielsen. 2002. *The Concept of Tacit Knowledge-A Critique*, 5(2) OUTLINES 3.
- Eric Niler. 2019. *Can AI Be a Fair Judge in Court? Estonia Thinks So*, WIRED (Mar. 25, 2019), <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.
- Nils J. Nilsson. 2005. *Human-level Artificial Intelligence? Be Serious!*, 26(4) AI MAG. 68.
- _____. 2010. THE QUEST FOR ARTIFICIAL INTELLIGENCE: A HISTORY OF IDEAS AND ACHIEVEMENTS (Stanford Univ. Press).
- CATHY O'NEIL. 2016. WEAPONS OF MATH DESTRUCTION (New York: Crown).
- Frank A. Pasquale. 2019. *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEORGE WASHINGTON LAW REVIEW 1.
- Cassio Pennachin et al.. 2007. *Contemporary Approaches to Artificial General Intelligence*, in ARTIFICIAL GENERAL INTELLIGENCE 1 (Ben Goertzel & Cassio Pennachin eds.).
- MICHAEL POLANYI. 1958. PERSONAL KNOWLEDGE (London: Routledge & Kegan Paul).
- Richard M. Re & Alicia Solow-Niederman. 2019. *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242.
- STUART J. RUSSELL & PETER NORVIG. 2010. ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (Upper Saddle River: NJ, 3rd ed.).
- John R. Searle. 1980. *Minds, Brains, and Programs*, 3(3) BEHAVIORAL & BRAIN SCIENCES 417.
- David Silver et al.. 2017. *Mastering the Game of Go Without Human Knowledge*, 550 NATURE 354.

3. 判決

- Plessy v. Ferguson, 163 US 537 (1896).
- Brown v. Board of Education, 347 US 483 (1954).
- State v. Loomis, 881 N.W.2d 749 (Wis. 2016).