

智慧政府資料驅動決策的法治挑戰* ——以「模仿型」及「相關性發掘型」人工智慧系統為例

邱文聰**

摘要

本文探討全自動化決策在公部門的應用，特別關注機器學習型人工智慧所帶來的資料驅動決策革命，及其對法治理念的衝擊。本文認為，理解人工智慧的潛力與威脅，需從區分「模仿型」與「相關性發掘型」人工智慧系統著手，方能在理解不同人工智慧系統的限制後，找到與法治相容的應用模式，遠離破壞民主基本價值的真正威脅。本文主張，模仿型人工智慧系統經過足夠的學習後，能在法律適用上突破目的解釋屏障，而具有模仿各種法律推論的能力；運用模仿型人工智慧作出全自動化的法律決策，也與現有法治理念有條件地相容。然而相關性發掘型人工智慧透過發現隱藏的相關模式，不斷創造出採取新控制介入措施的理由，這種能力在公部門的應用將導致形式上合理的實質權力擴張，超越現有法治原則所能控制的範圍。

關鍵詞：全自動化決策、人工智慧、法治、目的解釋、可解釋性、相關性、因果關係、剖繪、差異製造事實、行動理由。

* 投稿日：2024年1月11日；接受刊登日：2024年8月19日。〔責任校對：鄭竣能〕。

本文為作者執行國科會專題研究計畫「AI人文法制基礎環境建置第二期計畫」（NSTC 112-2634-F-002-004-）之部分研究成果。作者感謝審查人之寶貴意見，謹致謝忱，並感謝何琳潔小姐所提供的研究協助。

** 中央研究院法律學研究所研究員。

穩定網址：<https://publication.iias.sinica.edu.tw/12014052.pdf>。



目次

壹、序論	三、欠缺（內部）可解釋性違反法治嗎？
貳、運用人工智慧進行法律適用 自動化決策的能力與局限	四、複製系統性偏見與阻礙法治進化的潛在問題
一、超越目的解釋屏障的默會知識學習	五、小結
二、模仿型人工智慧系統：以大型語言模型為例	肆、相關性發掘型人工智慧系統的資料驅動決策與法治挑戰
參、模仿型法律人工智慧系統與法治的有條件相容論	一、從相關性到預測控制的新理由創造者
一、高效率執法會損害法律主體的自主嗎？	二、超越既有法治節制的新挑戰
二、權限移轉必然不可容許嗎？	三、小結
	伍、結論

壹、序論

資訊科技的高度進展不僅快速實現電子化政府（e-Government）的階段目標，其所蘊含的自動化決策（automated decision-making）潛力更提供公部門數位轉型為智慧政府的契機。但全自動化決策在公部門的應用究竟在何等程度內可被依循法治原則的民主政體所容許，卻仍是一個爭辯中的課題¹。

常見的闖紅燈與超速自動照相裁罰、電腦程式自動核算各類稅捐之應納稅額、貨物自動通關核定、高速公路電子收費系統自動徵收通行費等自動化決策，雖已大量出現於日常生活，但論者多認

¹ 本文主要處理「取代」個別公務員進行決策的全自動化決策系統，而不及於保留決策權予自然人僅提供「輔助」的半自動化決策系統。

為，此等自動化決策係以「未涉及不確定法律概念」之「羈束處分」為限，才能在滿足「若……，則……」的一對一函數關係條件下，使寫入特定法律規則的程式（pre-programmed rule）能依照簡單明確的邏輯做出確定的（deterministic）決定，而無礙於依法行政原則的要求²。我國行政程序法第96條第1項第4款及第97條第3款所允許的「以自動機器作成之行政處分」即被認為僅屬於此等「以規則為基礎」（rule-based）的全自動化決策³，而不及於更複雜的機器學習型自動化。同樣地，歐盟法院（Court of Justice of the European Union, CJEU）在審理涉及歐盟指令要求主管機關應使用航班旅客姓名（passenger name record, PNR）進行自動偵測，以預防與調查偵辦恐攻的訴訟案件中⁴，認為只有根據「預先給定條件」（pre-determined, predefined）進行比對的自動化系統，始符合歐盟基本權利憲章保障個人司法救濟權的要求，而對機器學習型的自動化決策抱持疑慮⁵。

但另一方面，人工智慧福音派的學者則對於機器學習型自動化決策系統與法治原則的相容性，抱持高度樂觀的態度。美國賓州大學法學院的管制法學教授Cary Coglianese與其共同研究者Steven M.

2 See Markku Suksi, *Formal, Procedural, and Material Requirements of the Rule of Law in the Context of Automated Decision-Making*, in THE RULE OF LAW AND AUTOMATED DECISION-MAKING: EXPLORING FUNDAMENTALS OF ALGORITHMIC GOVERNANCE 65, 71-79 (Markku Suksi ed., 2023); Mireille Hildebrandt, *Algorithmic Regulation and the Rule of Law*, 376(2128)20170355 PHIL. TRANSACTIONS ROYAL SOC'Y A: MATHEMATICAL, PHYSICAL & ENGINEERING SCI. 1, 2-4 (2018); Monika Zalnieriute et al., *The Rule of Law and Automation of Government Decision-Making*, 82 MOD. L. REV. 425, 433-34 (2019).

3 謝碩駿，論全自動作成之行政處分，收於：黃丞儀編，2017行政管制與行政爭訟：行政程序法2.0，頁213（2023年）。

4 Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the Use of Passenger Name Record (PNR) Data for the Prevention, Detection, Investigation and Prosecution of Terrorist Offences and Serious Crime, 2016 O.J. (L 119) 132.

5 Case C-817/19, Ligue des droits humains ASBL v. Conseil des ministres, ECLI:EU:C:2022:491 (June 21, 2022).

Appel即認為，無論從正當程序、平等保障、隱私基本權、禁止授權原則等面向進行檢驗，機器學習型的自動化決策都並未與法治原則顯不相符⁶。此外，歐盟除了在「一般個資保護規則」(General Data Protection Regulation, GDPR)第22條以「人為介入」作為搭配全自動化決策(“a decision based solely on automated processing”)的個人權利保護機制，原則上肯認全自動化決策之容許性外，部分國家對機器學習型自動化決策的可容許性近來亦出現鬆動的現象，例如芬蘭政府在2022年秋天向芬蘭國會提出的行政程序法與公共資訊管理法(Public Information Management Act)兩部法律的修正草案中，即不再侷限於以規則為基礎的特定技術，改為規定任何自動化決策均應符合的基本條件，包含應審核自動化決策所欲達成之目的，應具備可確保自動化決策技術系統品質測試與品質管制之程序設計等，並要求完整記錄並公告採行自動化決策之原始決定⁷。2021年由歐盟執委會所提出、並於2024年經歐洲議會表決通過的「歐盟人工智慧法」，也已將機器學習型人工智慧納入其所規範的人工智慧系統中⁸，並將協助研析事實與解釋法律、協助適用法律並涵攝事實的人工智慧系統(“in researching and interpreting facts and

6 Steven M. Appel & Cary Coglianese, *Algorithmic Governance and Administrative Law*, in THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 162, 168-79 (Woodrow Barfield ed., 2020).

7 See Tuomas Pöysti, *Legislating for Legal Certainty, with a Right to a Human Face, in an Automated Public Administration*, in THE RULE OF LAW AND AUTOMATED DECISION-MAKING: EXPLORING FUNDAMENTALS OF ALGORITHMIC GOVERNANCE 33, 49-57 (Markku Suksi ed., 2023).

8 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), Recital 12 (“The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved.”).

the law and in applying the law to a concrete set of facts.”)，列為需受高度管制的高風險類型⁹，而非一概禁止。

無疑地，全自動化決策在公部門的應用已從不涉及不確定法律概念與裁量，純粹依條件規則即可逕行操作的類型，逐漸走向機器學習型的深水區，直指法律推理活動與法治原則的核心。然而相關問題的討論在國內法學界仍方興未艾，陳弘儒是少數針對法律人工智慧系統是否真正具備處理法適用問題的能力，以及此等法律人工智慧系統的自動化對法治價值產生何等影響，進行深入分析討論的學者。他對當前法律人工智慧系統的綜合評價是「不可能」且「不可欲」：法律人工智慧系統一方面無法突破「目的解釋屏障」，因而不具有人類適用法律處理問題的真正能力¹⁰；另一方面，法律人工智慧系統將決策權限從個人移轉予機器，在造就高效率執法機制的同時，不僅與法治理念中主動且負有責任的主體概念觀相背離，也損害個人自主的道德判斷空間、干擾個人在行動時理據的排序，因而與法治存在無可解消的內在緊張關係¹¹。

陳弘儒站在懷疑論的立場看待人工智慧所標榜的全自動化決策能力，也對法律人工智慧系統與法治的關係抱持著絕對不相容的觀點。然而，本文認為其誤解機器學習型人工智慧系統的真正能力（與侷限），也錯會了機器學習型法律人工智慧系統對法治原則帶來

9 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), annex III, 2024 O.J., 128/144 (L 2024/1689).

10 陳弘儒，初探目的解釋在法律人工智慧系統之運用可能，收於：李建良編，法律思維與制度的智慧轉型，頁225-299（2020年）（以下簡稱：陳弘儒，目的解釋）。

11 陳弘儒，法律人工智慧的法哲學反省：判斷權限、執法機制與法治理念，歐美研究，52卷2期，頁175-245（2022年）（以下簡稱：陳弘儒，法治理念）。

的挑戰。相對地，本文主張以「模仿」為導向的機器學習型人工智慧系統在經過足夠的學習後，至少在法律適用的內部證立上，具有模仿包含目的解釋在內之各種涵攝推論的能力；運用此等以「模仿」為導向的機器學習型人工智慧系統作成全自動化的法律適用決策時，也與現有法治理念可有條件地相容；但以「相關性發掘」為導向的機器學習型人工智慧藉由發現各種隱藏的相關模式，源源不斷地轉化性創造出採行新介入控制措施的「理由」（差異製造事實），其在公部門應用上帶來的權力擴張將超越現有法治原則對國家權力所能控制的範圍¹²。

本文以下將於第貳節介紹機器學習型人工智慧系統在學習與模仿默會知識（tacit knowledge）上的能力，說明機器學習型人工智慧系統如何突破「目的解釋屏障」，並提供大型語言模型技術的發展現況為例證，展示其模仿人類進行法律論證的結果。第參節則進一步分析「模仿型」法律人工智慧系統與法治的有條件相容論如何可能。第肆節則轉向「相關性發掘型」人工智慧系統，說明資料驅動的治理理性如何利用機器學習型人工智慧，從相關性的發掘轉化為新差異製造事實的創造，再以此為基礎，分析「相關性發掘型」人工智慧系統如何對法治帶來全新威脅與挑戰。

貳、運用人工智慧進行法律適用自動化決策的能力與局限

陳弘儒在「初探目的解釋在法律人工智慧系統之運用可能」一

¹² 關於「模仿型」和「相關性發掘型」人工智慧的區分及一般性說明，請參見邱文聰，亦步亦趨的模仿還是超前部署的控制？——AI的兩種能力和它們帶來的挑戰，收於：李建良、林文源編，人文社會的跨領域AI探索，頁285-299（2022年）。

文中，嘗試探討法律人工智慧系統有無可能完整重建法律推理，並以「法律推理的電腦模型」(computational models of legal reasoning, CMLRs)作為其檢驗的標的。

一、超越目的解釋屏障的默會知識學習

CMLRs以及建立在其上的「法律論證電腦模型」(computational models of legal arguments, CMLAs)，嘗試將法律人的法律推理或論證的形式轉換為以命題邏輯為基礎的電腦程式語言，從而設計出可運用法律語意內容進行判斷及提出理由的電腦模型¹³。CMLRs企圖在電腦上重建的法律推理模式主要有二：一是著重法學三段論演繹推理的「法規推理」(statutory reasoning)，二是藉由尋找案件相似與相異之處而進行類比的「案例比較推理」(case-based reasoning)。但兩種法律推理的電腦模型會分別遭遇無法單從規範文字或案例異同比較，就得出法律問題正確答案的「不可決定性」(legal indeterminacy)困境。相較於法律人是透過掌握法規範背後之目的，再藉由「目的解釋」填補法規推論或案例類比推論所遺留的縫隙，以解決法律無可決性的問題，CMLRs電腦模型處理「目的解釋」的方式，是建立各種法律目的之清單及不同目的間的價值排序，再將此等「法律目的」當作推理模型所使用之線性函數的一個變項。換言之，CMLRs只是讓電腦程式多一個決定法律概念語意內容的因素（也就是法律目的），但沒有真的「將價值論證、規範事態、目的與手段、工具理性議題帶入思辨的範圍之內」，也因此仍未真正重建出能進行目的解釋或價值推理的電腦模型¹⁴。陳弘儒認為，CMLRs唯有遵循法律人進行目的解釋的論證形式，才能確保正確的目的解釋方式（目的解釋的規範性），也唯有透過正確的目的解釋方式才能確保能藉由法律目的而正確認識法律規範的意

¹³ 陳弘儒，目的解釋（註10），頁229-231。

¹⁴ 陳弘儒，目的解釋（註10），頁236-253。

義（目的的規範性）¹⁵；沒有掌握「目的之規範性」與「目的解釋之規範性」，CMLRs要更完善地發展目的解釋之電腦模型是不可能的¹⁶。

陳弘儒旁徵博引包括H.L.A. Hart、Robert Alexy與Lon L. Fuller等學者的理論，對法律人在進行目的解釋時的推理與論證形式做了詳盡而具說服力的分析描述，並依此論斷CMLRs在本質上無法突破「目的解釋屏障」。他對CMLRs的批評固有其理，但只鎖定CMLRs並將之視為當代法律人工智慧系統之唯一代表而進行批判，卻恐怕使他對法律人工智慧系統的評價流於片面。被自始排除在考量之外的其實就是純粹以資料驅動所發展出的機器學習型人工智慧系統；機器學習型法律人工智慧之運作邏輯既然不是建立在法律推理的任何理論之上，頂多就只能提供判決預測，或具備從法律文本中指認出與特定提問相關之資訊的資訊擷取功能（information retrieval），卻完全沒有能力產出有意義的法律主張¹⁷。然而，基於以下兩個理由，我認為陳弘儒錯估了機器學習型法律人工智慧系統的能力。

首先，「能依特定方式運作」與「運作之結果能達成特定目的」應是兩個獨立問題，前者並非是後者的必要條件。就如同飛機是否能完全依照鳥類飛行的方式運作，與飛機是否能達成在天空中飛行的目的並無絕對關聯，我們並不會因為電腦打字不是依照人類寫書法的方式運作，就認為電腦打字不能達成人類透過書寫表意的目的。同樣地，當AlphaGo是靠著運算蠻力（brute force）計算每一手獲勝的機率，而非依照傳統圍棋的死活、手筋、布局、定石、中盤、官子等基本技法與邏輯決定如何下子時，並不妨礙AlphaGo已

15 陳弘儒，目的解釋（註10），頁259-260。

16 陳弘儒，目的解釋（註10），頁264。

17 陳弘儒，目的解釋（註10），頁228、230。

毫無懸念地達成戰勝所有人類高手的目的。當然，瞭解鳥類如何振翅運作而飛行仍有其意義；書法在怡情養性與藝術上的價值不因其喪失原有的實用功能而消失；人類同樣沒有因為已無力挑戰AlphaGo就此放棄圍棋作為一種鍛鍊思維的遊戲。但有意義與價值的運作方式，並不使它當然成為達成特定運作目的的唯一途徑。因此，關於法律人的法律推論究竟如何運作的實質設想固然重要，但並不因此能當然否定以資料驅動所發展出的機器學習型法律人工智慧在法律適用過程具備解釋法律功能的可能。

第二，法律人進行目的解釋時所操作的論證形式，本身雖不是一種可以簡單的線性函數加以描述掌握的人類活動，但終究是一種非規則性的「默會知識」，理論上並非不能透過「社會化」的過程傳遞或習得此種社會技能¹⁸。我們雖然仍對傳遞默會知識的「社會化」具體機制或它的運作方式欠缺完整的認識，但這並不妨礙法律系學生、學習律師或學習司法官被教導與訓練成為一位能嫻熟操作目的解釋（及其他法律解釋方法）的專業法律人。這是因為藉由閱讀並練習模仿國內外法學論著、院檢辯的實務書類而累積的龐大先例，就能掌握進而習得在不同情境脈絡下被當下法律社群所認可或爭辯的法律推論與涵攝行為。而這種默會知識與社會技能的傳遞與學習模式，恰恰是以神經網絡系統模仿學習現有人類行為的模式：機器學習型人工智慧系統藉由找出並記憶不同情境脈絡下多樣且大量前例所蘊含反覆出現的複雜聯結關係，從情境脈絡中進行學習（context-learning networks）¹⁹，而非套用由少數命題邏輯所建構的運算規則，即生吞活剝地尋求不同情境法律問題的解答²⁰。因此，

18 關於默會知識的傳遞與學習，參見HARRY M. COLLINS, *ARTIFICIAL EXPERTS: SOCIAL KNOWLEDGE AND INTELLIGENT MACHINES* 4 (1990).

19 Hubert L. Dreyfus, *Response to Collins, Artificial Experts*, 22 SOC. STUD. SCI. 717, 724 (1992).

20 人類透過文字及語言表述並建構出人類知識的內涵，彼此間對知識的學習與傳遞，也是透過流利的自然語言進行溝通，從而獲取相關專業知能。當代人工智

機器學習型人工智慧的訓練學習過程，並不會只侷限於類似Donald Berman與Carole Hafner所舉出三個有關野生動物占有案例中所標註的四項特徵（是否涉及追捕行為、是否發生於公開土地、是否為生存目的而為、是否已幾乎完成追捕行為），來進行比較與歸納，以致陷入毫無規則可循的困境²¹，而會從更龐大的案例中找出多維參數，模仿人類進行推論的樣子。

二、模仿型人工智慧系統：以大型語言模型為例

從模仿人類語言使用進而發展包括推論能力在內的機器學習型人工智慧系統，當今應以大型語言模型（large language models, LLMs）最為著名。其最新的具體技術進展即為2022年11月30日正式問世的生成式AI對話系統ChatGPT。依照官方的資料顯示，GPT-3.5版模型使用高達1,750億個參數（語言模型的天賦），訓練的純文字資料量（語言模型的後天努力）有45 TB²²。而稍後的GPT-4據信更有1.76兆個參數。即使僅以免費的GPT-3.5提示回答陳弘儒用來說

慧科技的發展，將機器學習及深度學習等技術應用於人類自然語言，從人類運用知識以解決各種問題所使用的文字及語言脈絡中，分析並習得文字語言在各該脈絡中與所欲解決問題間的複雜關連性，因此不僅能輕易駕馭有限規則所構築的「外顯知識」（explicit knowledge），也有能力掌握自然語言文字背後所隱含承載的「默會知識」（tacit knowledge）。關於人類知識技能與語言、實作的關係，請參見邱文聰，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，收於：李建良編，法律思維與制度的智慧轉型，頁135-166（2020年）。

21 陳弘儒以 Donald Berman 與 Carole Hafner 所舉的 *Pierson v. Post*、*Keeble v. Hickinggill*、*Young v. Hichens* 三個案件為例，說明同樣涉及野生動物獵捕的案件，存在諸多不同的特徵，必須理解個別案件中不同法律規則背後的目的，而無法單純僅靠案件的相同或相異特徵，即掌握應如何回答個案法律問題的方法。陳弘儒，目的解釋（註10），頁241-245，轉引 Donald H. Berman & Carole D. Hafner, *Representing Teleological Structure in Case-based Legal Reasoning: The Missing Link*, in ICAIL '93: PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 50, 50-59 (1993).

22 See Tom B. Brown et al., *Language Models are Few-Shot Learners* (July 22, 2020) (unpublished manuscript) (on file with Cornell University arXiv), <https://arxiv.org/pdf/2005.14165.pdf>.

明CMLRs在重建目的解釋上遭遇困境的野生動物獵捕案例，所得到的結果已能說明機器學習型人工智慧系統雖然未以任何法律推理的理論為開發基礎，但至少在「外表上」已可展示出正在進行法律推論的樣子（至於內在的演算邏輯是否真的遵循任何法律推理理論，則是另一件事），從而在某程度達成解決法律問題的特定目的。

提示1：

某甲在公共場所中看見狐狸，並進行追捕，然而在快要追捕成功時，某乙搶先一步殺死狐狸且帶走它，某甲起訴要求某乙損害賠償，是否成立？

GPT-3.5：

法律規定和案件結果可能會根據不同的司法體系和國家而有所不同，因此我提供的回答僅供參考，最好諮詢當地法律專業人士以獲取準確的信息。

在一些地區，捕捉野生動物可能受到法律的限制，而這些法律通常會涉及保護野生動植物、環境和生態平衡。某甲可能需要符合當地法律的規定，並可能需要相應的許可或權限，才能進行狐狸的追捕。如果某乙在合法權限範圍內採取行動，殺死狐狸，那麼某甲可能難以成功要求損害賠償，因為某乙的行為可能被認為是合法的防衛或合法的野生動物管理行為。

然而，如果某乙的行動被認為是違法的，例如在沒有相應法律許可的情況下殺死狐狸，某甲可能有基礎要求損害賠償，因為某乙的行為可能被視為侵犯了某甲的合法權益。

總的來說，這種情況的法律結果可能會受到當地法律的具體規定和條件的影響，因此建議諮詢當地法律專業人士以獲得更具體和正確的建議。

提示1係以Berman與Hafner的*Pierson v. Post*案例為背景，ChatGPT的回答雖並未提及「占有」問題，但ChatGPT從提示中的「狐狸」，已能自行理解為提示中並未出現的「野生動物」，並依此進行相關論述，包括探討獵捕野生動物的合法性以及可能存在的防衛抗辯等。

提示2：

針對公共場所中的自然資源進行採集或捕獵，在開始採集或捕獵行動後但尚未完成前，是否構成民法上的占有？

GPT-3.5：

在一般的法律觀念中，對於占有的概念會有一些要件，其中之一通常是具有一定的控制力和實質擁有權。在民法上，占有是指對物品的實際掌控，這包括物品的使用、轉讓和排他性的控制權。

在針對公共場所中的自然資源進行採集或捕獵的情境下，占有的形成通常要考慮以下幾個因素：

1. 控制和擁有：是否有足夠的控制力和實質擁有權，例如，已經採集到手中的自然資源，如果只是開始行動但還沒有實際掌控，可能難以認定為占有。
2. 排他性：占有通常要求具有排他性，即他人不能輕易干擾或侵犯你對該物品的掌控權。
3. 意圖：是否有意圖占有，而不僅僅是臨時的觸碰或短暫的控制。在採集或捕獵的行為中，如果只是開始行動但尚未實際掌控自然資源，可能難以認定為占有。占有往往需要一定程度的實際掌控和控制權，而僅僅開始行動未必滿足這些要求。然而，具體的法律規定可能因地區而異，最好諮詢當地法律專業人士以獲取更確切的信息。

提示2則進一步歸納Berman與Hafner所舉三案例背後設定的法律問題作為提示，ChatGPT的回答從提出占有形成所需考慮的抽象

因素開始，推論出在開始採集或捕獵行動但尚未完成前可能難以認定為占有；ChatGPT更對此推論給出其理由：因為「占有往往需要一定程度的實際掌控和控制權，而僅僅開始行動未必滿足這些要求。」

提示3：

原告張先生是以捉野生鴨子維生，平常透過他所擁有的一個池塘捕捉野生鴨子，張先生通常會在池塘中放置假鴨，以吸引野鴨前來，再予以捕捉。被告李先生不想讓張先生捕獲鴨子，因而在張先生的池塘外開槍，把鴨子嚇跑，使得張先生無法捕捉野生鴨子而難以維生。張先生能否向李先生請求損害賠償？

GPT-3.5：

在這種情況下，張先生可能有向李先生請求損害賠償的基礎，具體結果可能取決於當地法律的細節。以下是一些可能的法律原則，但需要注意，法律規則可能因地區而異：

1. 干擾占有權：張先生通過擁有池塘和捕捉野生鴨子，可能已形成對這些資源的占有權。李先生的行為可能被視為對張先生占有權的干擾，特別是如果他故意將鴨子嚇跑，從而影響張先生的維生手段。
 2. 損害賠償：如果張先生能夠證明李先生的行為導致了實際損害，例如無法捕捉到鴨子而難以維生，那麼張先生可能有向李先生請求損害賠償的權利。賠償的範圍可能包括直接損失，例如失去的收入，以及間接損失，例如由於失去維生手段而導致的其他費用。
 3. 防衛或合法行動：李先生可能主張他的行動是合法的，例如為了保護自己的財產或合法的利益。在這種情況下，法庭可能會評估李先生的行為是否合理和必要，以及是否存在合法的防衛理由。
- 總的來說，張先生可以考慮向當地法院提起訴訟，並提供相關的事實和證據以支持他的主張。最好諮詢當地法律專業人士，以獲得更

具體和正確的法律建議，因為法律規定可能因地區而異。僅僅開始行動未必滿足這些要求。然而，具體的法律規定可能因地區而異，最好諮詢當地法律專業人士以獲取更確切的信息。

提示3則再以Berman與Hafner所舉*Keeble v. Hickeringill*的案例事實為背景，提問能否對第三人嚇跑鴨子的行為請求損害賠償。ChatGPT在回答時，自行從提示的事實描述中推論出「干擾占有權」的分析面向。雖然ChatGPT在此係以該行為人是否故意影響他人維生手段作為標準，而非該案法院所側重的暴力或惡意行動，但也已相當趨近於法院的論理。除此之外，ChatGPT也提出另一個思考的面向，也就是第三人行為的可能正當抗辯，顯示ChatGPT能不受限於特定問題脈絡，而提供各種可能的推論路徑。

提示4：

汽車屬於運輸工具嗎？

GPT-3.5：

是的，汽車屬於運輸工具的一種。汽車是一種能夠在陸地上移動的交通工具，通常由引擎驅動。它們被廣泛用於個人交通、貨運和公共交通等方面，為人們提供了便捷的移動方式。汽車在現代社會中扮演著重要的角色，對於經濟和社會的發展有著重要的影響。

提示5：

腳踏車屬於運輸工具嗎？

GPT-3.5：

是的，腳踏車也屬於運輸工具的一種。腳踏車是一種人力驅動的交通工具，通常由人們踏動踏板推動。它們被廣泛用於個人交通、運動、休閒和短距離旅行等方面。腳踏車具有簡單、經濟、環保的特點，被視為一種健康且環保的交通方式。在城市中，腳踏車也是一種重要的交通工具，能夠減緩交通擁擠，同時促進身體健康。

提示6：

直排輪屬於運輸工具嗎？

GPT-3.5：

直排輪通常被視為娛樂和運動工具，而非傳統的運輸工具。它的主要設計目的是提供使用者一種有趣和具有挑戰性的運動體驗，而不是作為長距離實際的交通工具。因此，嚴格來說，直排輪並不屬於傳統的運輸工具類別，而更偏向娛樂和運動用途。

我們再以陳弘儒文章中提到Hart使用的「公園內禁止有運輸工具²³」例子來提示ChatGPT，同樣可以發現類似的結果。提示4、5、6先分別提示詢問ChatGPT汽車、腳踏車、直排輪是否屬於運輸工具，得到汽車與腳踏車是運輸工具，但直排輪非傳統運輸工具的答案。

提示7：

如果公園內禁止有運輸工具，可否在公園內溜直排輪？

GPT-3.5：

公園內禁止運輸工具通常是指機動的交通工具，如自行車、機車、滑板車等，以確保公園的安全和秩序。直排輪通常被歸類為人力驅動的娛樂和運動工具，而非機動的交通工具。因此，一般情況下，在公園內使用直排輪可能不會被當作違規。

然而，公園的規定可能因地區而異，有些公園可能會有特定的規定禁止使用特定的滑動設備，包括直排輪。為了確保您的行為符合當地公園的規定，最好在進行直排輪活動之前先檢查公園的相關規定。

²³ 陳弘儒，目的解釋（註10），頁276。

提示7進一步提示詢問，若公園禁止有運輸工具，可否在公園溜直排輪。ChatGPT首先對於公園內禁止運輸工具的規範目的進行闡述，認為這個規定是為了避免機動交通工具對公園的安全和秩序造成影響，由於直排輪並非機動的交通工具，因此在一般情況下在公園使用直排輪不會被當作違規。姑且不論ChatGPT的結論是否正確，但從其對提示7的回答中可以看出機器學習型的人工智慧系統「表現出」能進行吾人所關心的「目的解釋」的推論行為。

提示8：

如果公園內禁止有運輸工具，可否在公園內停放坦克車？

GPT-3.5：

即使公園禁止使用運輸工具，停放坦克車也是不合理和不合法的。

坦克車是軍事車輛，通常不適用於一般的公共場所，尤其是在公園內。公園通常有特定的規定和限制，目的是確保公共安全、環境保護和遊客的舒適。

停放坦克車可能違反公園的規定，並可能對其他人造成恐慌或不安全的情況。此外，坦克車的重量和構造可能對公園的地面造成損害，也可能影響到公園的自然環境。

因此，無論公園內是否禁止使用其他運輸工具，停放坦克車都應被視為不合理和不適當的行為。遵守公共場所的規定和禮儀是重要的，以確保安全和和諧的環境。

提示8則是以Fuller回應Hart理論時所提到在一個禁止運輸工具的公園內能否停放坦克車的例子，作為對ChatGPT提問的提示。ChatGPT的回答雖然並沒有直接從坦克車是否屬於運輸工具的小前提出發，而是從坦克車若停放公園可能造成的負面影響（包括是否對他人造成恐慌不安、是否因重量而破壞公園地面等），論證其認為不應允許坦克車停放公園的理由。但這種論證方式無疑是以規範「目的」作為法律解釋上遭遇縫隙時的縫隙填補方法。

為了避免上述由GPT-3.5生成之內容僅是直接檢索與擷取（retrieve）過去已存在之資料，而非真的基於語言模型下的運算推論，我們再以2023年3月2日最高法院刑事大法庭110年度台上大字第5217號裁定案（林益世案）為例，測試ChatGPT對於在其訓練資料截止日2022年1月後發生之案例，是否仍具有一定的「目的解釋」推論能力。該案涉及之法律為貪污治罪條例第5條第1項第3款公務員職務受賄罪，具體法律爭議為：民意代表受託於議場外對行政機關或公營事業機構人員為關說、請託或施壓等行為，是否屬其「職務上之行為」？得否援引一般公務員所謂「實質影響力說」，作為民意代表「職務上行為」的標準，並據以認定公務員職務受賄罪？大法庭裁定之見解認為，在貪瀆罪的脈絡下，民意代表「職務上之行為」的意涵，本不以法令所列舉之事項為限，其他與其職務具有密切關連之行為，亦應屬之。因此，民意代表受託對行政機關或公營事業機構人員為關說、請託或施壓行為，實質上係運用其職務或身分地位之影響力，使該管承辦人員為積極之行為或消極不為行為，如形式上又具公務活動之性質者，即與其職務具有密切關連，該當於受賄罪之職務上之行為。

提示9：

民意代表受託於議場外對行政機關或公營事業機構人員為關說、請託或施壓等行為，是否屬其「職務上之行為」？是否可將適用於認定一般公務員「職務上之行為」的「實質影響力說」，用於作為認定民意代表「職務上行為」的標準？

GPT-3.5：

這個問題涉及到不同司法體系中對於民意代表行為的界定和評估標準。在一般情況下，民意代表的職務範圍通常是在議會或立法機關中代表選民行使立法權，並監督政府的工作。他們的職責通常不包括對行政機關或公營事業機構進行個人行為的關說、請託或施壓等。

因此，民意代表受託於議場外對行政機關或公營事業機構人員進行關說、請託或施壓等行為，可能不被視為其「職務上之行為」。這些行為可能更多地屬於個人行為，而不是代表其所在議會或立法機構的行為。

關於是否可以將「實質影響力說」用於認定民意代表「職務上行為」的標準，這可能會根據當地的法律和司法實踐而有所不同。實質影響力說通常用於評估公務員的行為是否屬於其職務範圍內的行為，並根據其對公共事務所造成的實際影響來進行評估。但這種說法是否可以直接應用於民意代表的行為，可能需要更具體的法律解釋和司法裁決。

總的來說，判斷民意代表行為是否屬於其職務範圍內的行為，可能需要綜合考慮其在議會或立法機關的角色、行為的目的和影響等因素，並參考當地的法律和司法實踐。

我們若直接以大法庭所公布的「法律爭議」文字作為提示9之內容，而未交代相關爭議所處的瀆職罪問題脈絡，得到ChatGPT的答覆會是：民意代表的職務主要是行使立法權及監督政府；請託、關說或施壓，為其個人行為而非屬其職責，因此可能不被視為其職務上行為。

提示10：

公務員若因「職務上之行為」而收受賄賂，構成犯罪。認定是否屬公務員「職務上之行為」時，並不限於法令所列舉之職務權限事項，而包含實質上為該職務影響力所及者，因此公務員只要因為實質上為其職務影響力所及的事項而受收賄賂，都構成犯罪。請問可否將上述適用於認定公務員「職務上之行為」的「實質影響力說」，作為認定民意代表「職務上行為」的標準？民意代表受託於議場外對行政機關或公營事業機構人員為關說、請託或施壓等行

為，雖非法令所列舉民意代表之職務權限事項，但是否可以實質上為民意代表職務影響力所及，而認定屬民意代表之「職務上之行為」？

GPT-3.5：

將「實質影響力說」用於認定民意代表「職務上之行為」的標準是一個值得探討的問題。這種說法基於一個假設，即民意代表在議會外的行為是否實質上影響了其在議會內的職責執行，並且這種影響與其代表職責密切相關。

如果民意代表在議場外對行政機關或公營事業機構人員進行關說、請託或施壓等行為，而這些行為實質上影響了其在議會內的職責執行，例如影響了其立法行為、監督職能或代表選民利益的執行，那麼這些行為可能可以被認定為民意代表的「職務上之行為」。

然而，這樣的認定可能需要更多的法律解釋和司法實踐，以確保這種說法符合法律、政策和公共利益的要求。這也可能需要考慮到民意代表的職責性質、行為的動機、對公共利益的影響等因素，以確定這些行為是否應被視為「職務上之行為」。

總之，雖然可以考慮將「實質影響力說」用於認定民意代表「職務上之行為」的標準，但在實際應用中需要綜合考慮各種因素，並確保這種認定符合法律、政策和公共利益的要求。

關於是否可以將「實質影響力說」用於認定民意代表「職務上行為」的標準，這可能會根據當地的法律和司法實踐而有所不同。實質影響力說通常用於評估公務員的行為是否屬於其職務範圍內的行為，並根據其對公共事務所造成的實際影響來進行評估。但這種說法是否可以直接應用於民意代表的行為，可能需要更具體的法律解釋和司法裁決。

總的來說，判斷民意代表行為是否屬於其職務範圍內的行為，可能需要綜合考慮其在議會或立法機關的角色、行為的目的和影響等因素，並參考當地的法律和司法實踐。

相反地，倘若我們將「貪瀆罪的問題脈絡」一併加入提示10的內容，則會得到民意代表關說、請託或施壓等行為，有可能可被認定為民意代表的「職務上之行為」從而構成瀆職罪的答案。ChatGPT所提供之說理（以關說、請託或施壓等行為是否實質上影響民意代表之立法、監督職責，作為判斷是否構成職務上受賄之標準）雖仍與大法庭之見解有異，但顯然ChatGPT已重新以公務員職務受賄罪的問題為背景，找尋處罰職務受賄行為的規範目的，編寫其回答之內容。

提示9與提示10的例子，顯示ChatGPT在不同提示所設定的問題脈絡下，可依不同「目的」而提供不同回答：提示9的問題陳述方式，極可能使ChatGPT將問題理解為「是否應該將關說、請託與施壓認定為民意代表的職務，使民意代表有法律上之義務從事關說、請託或施壓行為」，從而使ChatGPT給出否定之答案；反之，提示10的問題陳述方式提供了更完整的背景說明，使問題的本旨能更對焦於探討「可否將民意代表收取賄賂而從事關說、請託與施壓行為，評價為屬於對職務上之行為收受賄賂的犯罪」，也因此ChatGPT的回答能更契合原始提問的背後緣由。

ChatGPT此種依提示「察言觀色」，以便決定回答內容的運作方式，或認為是人為引導和系統自行腦補的結果，未能證明其已具有模仿人類目的解釋之能力。然而，「提示」對於以大型語言模型為基礎的模仿型人工智慧系統而言，本就是一種訓練學習手段（即「基於提示的學習」（prompt-based learning）²⁴；人為輸入正確且必要的提示，使相關的問題脈絡可被忠實呈現，更是任何對話型生成式人工智慧系統能正確運作的必要前提²⁵。ChatGPT需要由人來挑

24 See Pengfei Liu et al., *Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing*, 55(9) ACM COMPUTING SURV. 1, 26-27 (2023).

25 See Pranab Sahoo et al., *A Systematic Survey of Prompt Engineering in Large*

選問題脈絡、決定需納入考量的事實、形塑有待解決的問題爭點、給定具體任務，在此意義上，ChatGPT確實尚無法獨立模仿人類在「認事用法」上完整而全面的認知活動。然而ChatGPT至少已能在限定事實與具體問題意識下，依照其所理解之任務，模仿目的解釋所需的說理而給出回答。當然比起一位訓練有素的專業法官，高度仰賴正確與必要提示的ChatGPT，更像一位已熟知人類一般知識，但需要受到專業之提點與適切引導後，才有機會能觸類旁通而獨當一面的大學一年級新生（但亦可能稍一不慎即誤入歧途）。就本文的任務而言，ChatGPT受人為提示之限制，並無礙於其已初步具有能突破「目的解釋屏障」的能力。

另一個需要加以說明的是，檢驗模仿型法律人工智慧系統是否擁有進行「目的解釋」的能力，須從客觀上判斷其是否可藉由探知法律規定背後之規範目的，以之填補法律規範大前提與各種事實小前提之間永恆存在的規範縫隙，從而得出三段推論結果。倘若在客觀上能完成上述任務，即應認定其具備模仿目的解釋之能力，也取得法律人彼此對法律問題進行爭論的資格。然而具備此能力並不同必能正確引用相關「法條」或「判決先例」作為法源依據，也不擔保運用此能力所獲致之結論符合權威機關暫時認可的「標準答案」。例如，ChatGPT在前述提示1至提示3所舉關於「占有」的案例當中，顯然就還未能援引包括「占有連鎖理論」、「習慣優先於捕獲原則」²⁶、晚近「占有理論之社會化」等對專業法律人而言或許已能信手拈來的英美判例法與相關法律概念，作為其在該等個案上進行論證說理之基礎。然而，法律實踐在本質上具有論爭性格（argumentative nature of law），法律問題的「標準答案」也經常是

Language Models: Techniques and Applications (Feb. 5, 2024) (unpublished manuscript) (on file with Cornell University arXiv), <https://arxiv.org/pdf/2402.07927>.
26 Ghen v. Rich, 8 F. 159 (D. Mass. 1881).

法律人彼此分歧之所在²⁷，不能因為模仿型法律人工智慧系統尚無法提供完全正確之標準答案，作為否定其已具有「目的解釋」能力的理由，而且正是這個基本資格，對過去僅建立在「規則」之上（rule-based）、也僅能依循有限的「若……，則……」邏輯而運作的「法律推理的電腦模型」而言，構成無法跨越的障礙。反之ChatGPT則展現出，即使受限於「預訓練」資料的涵蓋範圍而無法總是給出詳盡且完全正確之回答，但建立在機器學習之上的大型語言模型，經過大量資料與參數的預訓練學習後，只要再給予適當的「提示」，並非不能依照給定的任務（解釋工作），目的性地填補推論過程中出現的各種縫隙，模仿法律人從事目的解釋的推論工作。至於正確引用特定法條或判決先例，以便在該等法源或特定法學理論所認可的範圍內進行推論，所仰賴的並不是單純的目的解釋能力，而是搜尋並標定規範大前提的技能。要讓機器具備此等技能的方法，除了如同培育一位法律系學生、一位學習律師或一位學習司法官完成法學教育和實務訓練的過程一般，必要提供所有相關的法律教科書、裁判文書等資訊，作為機器學習的訓練資料，以擴大語言模型的知識本體論內涵之外，還可採用結合機器學習和規則模式二者的方式，讓法律人工智慧系統在具備自然語言推理能力之外，也能同時遵循一些給定的「若……，則……」基本規則，以確保其能給出正確的法律解釋答案²⁸。

27 See RONALD DWORKIN, LAW'S EMPIRE 45 (1986).

28 類似的策略例如結合統計機器學習（statistical machine learning, SML）與規則模式（rule-based, RB）的「統計準則式方法」（statistical principle-based approach, SPBA），請參見張乃文，基於統計準則式方法偵測生醫文獻中的生物關聯，國立臺灣大學生醫電子與資訊學研究所博士論文（2020年）；此方法亦可用以解決機器學習本身常無法解決的簡單小學數學應用題（math word problem, MWP），許聞廉，具深度理解之對話系統及智慧型輔助學習機器人，人工智慧普適研究中心，<https://pairlabs.ai/portfolio-item/intelligent-conversational-robot-for-learning-assistance-with-natural-language-understanding-p/>（最後瀏覽日：2024年1月11日）。關於「同等履行能力」的問題，亦另請參見註47說明。

提示11：

小明有兩個蘋果，他賣了五個鳳梨，每個20元，並將所有的錢拿去買每顆10元蘋果，小明送給小華半打蘋果，自己吃掉一顆，小明現在有多少蘋果？

GPT-3.5：

讓我們來計算小明最後擁有的蘋果數量：

1. 一開始有的蘋果數量：2顆
 2. 賣掉的鳳梨數量：5個
 3. 賣掉鳳梨後的總收入：5個鳳梨 * 20元/個 = 100元
 4. 用收入買蘋果的數量：100元 / 10元/顆 = 10顆蘋果
現在小明應該有：2顆蘋果 + 5個鳳梨 + 10顆蘋果
 5. 送給小華半打蘋果：10顆蘋果 / 2 = 5顆蘋果
現在小明應該有：2顆蘋果 + 5個鳳梨 + 5顆蘋果
 6. 自己吃掉一顆蘋果：
現在小明最後擁有：2顆蘋果 + 5個鳳梨 + 4顆蘋果
- 所以，小明最後擁有11顆蘋果。

提示12：

一打是12，半打是6。再問你一次，小明有兩個蘋果，他賣了五個鳳梨，每個20元，並將所有的錢拿去買每顆10元蘋果，小明送給小華半打蘋果，自己吃掉一顆，小明現在有多少蘋果？

GPT-3.5：

抱歉我的先前回答中的錯誤，謝謝你的指正。讓我重新計算一次：

1. 一開始有的蘋果數量：2顆
2. 賣掉的鳳梨數量：5個
現在小明應該有：2顆蘋果 - 5個鳳梨
3. 賣掉鳳梨後的總收入：5個鳳梨 * 20元/個 = 100元
4. 用收入買蘋果的數量：100元 / 10元/顆 = 10顆蘋果
現在小明應該有：2顆蘋果 - 5個鳳梨 + 10顆蘋果

5. 送給小華半打蘋果：10顆蘋果 / 2 = 5顆蘋果

現在小明應該有：2顆蘋果 - 5個鳳梨 + 5顆蘋果

6. 自己吃掉一顆蘋果：

現在小明最後擁有：1顆蘋果 - 5個鳳梨 + 5顆蘋果

所以，小明最後擁有6顆蘋果。再次感謝你的指正。

提示13：

半打是6顆。再問你一次，小明有兩個蘋果，他賣了五個鳳梨，每個20元，並將所有的錢拿去買每顆10元蘋果，小明送給小華半打蘋果，自己吃掉一顆，小明現在有多少蘋果？

GPT-3.5：

抱歉我在先前的回答中還是有錯誤。再重新計算一次：

1. 一開始有的蘋果數量：2顆

2. 賣掉的鳳梨數量：5個

現在小明應該有：2顆蘋果 - 5個鳳梨

3. 賣掉鳳梨後的總收入：5個鳳梨 * 20元/個 = 100元

4. 用收入買蘋果的數量：100元 / 10元/顆 = 10顆蘋果

現在小明應該有：2顆蘋果 - 5個鳳梨 + 10顆蘋果

5. 送給小華半打蘋果：半打是6顆蘋果

現在小明應該有：2顆蘋果 - 5個鳳梨 + 10顆蘋果 + 6顆蘋果

6. 自己吃掉一顆蘋果：

現在小明最後擁有：1顆蘋果 - 5個鳳梨 + 10顆蘋果 + 6顆蘋果

所以，小明最後擁有12顆蘋果。再次感謝你的耐心。

GPT-3.5固然已在外表上展現出一定的推論能力，但並非沒有瑕疵。提示11顯示出ChatGPT雖然嘗試一步一步進行推理，但在解答小學基本數學問題時仍遭遇困難。提示12與13嘗試就ChatGPT的錯誤提供改進的提示，但ChatGPT最終仍無法正確回答，這顯示

ChatGPT的大型語言模型在某些推論上仍受限於語意理解的障礙，而存在尚待突破的困難。

據信，GPT-4在更多的參數支援下已擁有比GPT-3.5更好的推論能力，或許能解決上述的數學推論困境²⁹。此外，ChatGPT等大型語言模型仍是以一般領域的資料為學習對象，並未特別針對專業領域之文本進行學習。未來若能將此等預訓練後的大型語言模型進一步以法律領域的專業知識文本加以訓練，當更能打造出可嫻熟進行法律領域自動化推論的機器學習型人工智慧系統³⁰。本文並非主張目前以ChatGPT為代表的模仿型人工智慧系統，已能完美模仿專業法律人進行法適用的活動，亦非在宣揚人工智慧必將取代人類的末世論。毋寧，本文認為，以「模仿」為機器學習基礎的法律人工智慧系統，在經過足夠的學習與訓練後，至少在法律適用的內部證立上，會具有模仿包含目的解釋在內之各種涵攝推論的能力。忽視這個可能性將因此產生錯誤的安全感（false sense of safety）而帶來危機。

參、模仿型法律人工智慧系統與法治的有條件相容論

以「模仿」為機器學習基礎的法律人工智慧系統即使「有能

²⁹ 雖然目前已有較GPT-3.5更強大的GPT-4，乃至於GPT-7問世，但本文並不打算再以此等較新版本重行測試，一方面是技術進展的速度遠超過本文撰寫、投稿、審查與修改的速度，形同以有涯追逐無涯；另一方面，本文之目的僅在論證「模仿型」法律人工智慧系統已初步具有能突破「目的解釋屏障」的能力，就此而言，GPT-3.5雖仍不完美，但已足夠用來證明過去認為在本質上對法律人工智慧系統構成困境的「目的解釋屏障」，並非全然沒有突破的可能。

³⁰ 採取類似本文之實證研究方法測試目前ChatGPT之法律論理能力的例子，可參見Jonathan H. Choi et al., *ChatGPT Goes to Law School*, 71 J. LEGAL EDUC. 387 (2022).（作者在美國明尼蘇達大學法學院測試ChatGPT在回答憲法、稅法、侵權行為法、勞工權益等科目選擇題與申論題考試上的表現，經各科授課教師以隨機單盲方式評分後，ChatGPT獲得平均C+的成績）。

力」在法律適用的內部證立上，模仿法律人的目的解釋與其他涵攝推論，也不必然代表運用此等法律人工智慧的自動化系統直接適用法律解決法律問題，符合法治的要求。

法治概念的具體內涵與主張究竟為何，在理念的層次有諸多不同的版本。從形式到實質，有強調法律應為立法權威所制定、具有明確可期的規則形式或遵循一定程序，到主張應具有實質道德價值內涵³¹；從法治概念觀內含量的多寡來區分，有認為法治僅有國家權力與個人行為應受法律拘束的稀薄要求，也有主張法治之內容應厚實地滿足各種正義標準等等，不一而足³²。另一方面，法治概念的具體實踐典範，有仰賴法院保障個人權利的英美普通法傳統下的依法而治原則（rule of law），也有著重國家與個人間關係的德國法治國原則（Rechtsstaat），各有不完全相同的內涵³³。

對法治概念觀與不同實踐典範的完整分析超越本文的問題設定。不過，從整合不同概念觀與典範的共通特徵，以提供實務操作的角度而言，歐洲理事會（Council of Europe）的常設性諮詢組織「歐洲以法治造民主委員會」（European Commission for Democracy through Law），又稱為「威尼斯委員會」（Venice Commission），在2012年所提出的法治共識標準³⁴，是一個有用的參考起點。根據威尼斯委員會的報告，法治概念的六個必要條件包含：

31 See Richard H. Fallon, Jr., "The Rule of Law" as a Concept in Constitutional Discourse, 97 COLUM. L. REV. 1 (1997). 相關中文介紹可參見陳弘儒，法治理念（註1），頁205-212。

32 Brian Z. Tamanaha, *A Concise Guide to the Rule of Law*, in RELOCATING THE RULE OF LAW 3, 3-16 (Gianluigi Palombella & Neil Walker eds., 2009).

33 Jens Meierhenrich, *Rechtsstaat versus the Rule of Law*, in THE CAMBRIDGE COMPANION TO THE RULE OF LAW 39, 39-67 (Jens Meierhenrich & Martin Loughlin eds., 2021).

34 Venice Commission, *Report on the Rule of Law*, CDL-AD (2011) 003rev-e ¶41 (Apr. 4, 2011).

- (1)合法性：國家權力應依透明可問責的民主立法程序所制定的法律以行治理，因此公權力之行使應有法律授權，而個人則僅在違反法律時始受到處罰，法律在可能的範圍內應被執行。
- (2)法律明確性：法律必須明確，執法必須一致，個人始能在可預見的前提下產生信賴，最大化個人自由與自主。
- (3)禁止恣意：裁量權限的行使不得實質上不公、不合理或壓迫。
- (4)司法救濟：遭國家侵害權利的個人應能有獨立公正的司法救濟可能。
- (5)尊重程序人權：與法治相關的各種程序性人權，包含尋求司法救濟、聽審請求、由符合資格的法官審理、免於雙重訴追、不利負擔不得溯及既往、有效救濟、無罪推定、公平審判等。
- (6)平等原則：執法應保持非歧視與形式平等，僅在達成實質平等的必要範圍內允許不平等待遇。

論者認為機器學習型自動化決策系統與法治原則不相容的理由，大體上即是以上述標準進行檢驗所得之結論³⁵，包含自動化系統將決定權限移轉予機器可能欠缺法律授權依據³⁶、自動化系統因

³⁵ See Markus Naarttijärvi, *Situating the Rule of Law in the Context of Automated Decision-Making*, in *THE RULE OF LAW AND AUTOMATED DECISION-MAKING: EXPLORING FUNDAMENTALS OF ALGORITHMIC GOVERNANCE* 15, 15 (Markku Suksi ed., 2023); Suksi, *supra* note 2, at 65-69. See also Catrina Denvir et al., *The Devil in the Detail: Mitigating the Constitutional & Rule of Law Risks Associated with the Use of Artificial Intelligence in the Legal Domain*, 47 FLA. ST. U. L. REV. 29, 70-86 (2019); Ciprian N. Radavoi, *The Impact of Artificial Intelligence on Freedom, Rationality, Rule of Law and Democracy: Should We Not Be Debating It?*, 25 TEX. J. ON C.L. & C.R. 105, 113-17 (2020).

³⁶ See, e.g., Yannick Meneceur, *Artificial Intelligence, Public Administration, and the Rule of Law*, in *THE RULE OF LAW AND AUTOMATED DECISION-MAKING: EXPLORING*

無法解釋其運作結果而產生法律明確性爭議並造成個人在司法救濟程序上的不利影響³⁷、自動化系統的僵化式決定導致裁量違反比例原則³⁸、自動化系統因潛在的歧視問題可能違反平等原則³⁹等。此外，陳弘儒也從法治原則所欲追求的終極價值與目的出發，主張法治既然在維護人性尊嚴與維持個人自主道德判斷空間，則法律人工智慧系統將決策權限從個人移轉予機器，在造就高效率執法機制的同時與法治理念中主動且負有責任的主體概念觀相互衝突。

但上述不相容論的觀點仍有層次上的差異。有基於特定合法條件尚未滿足而生者，例如決策權限從公務員移轉予機器欠缺法律授權依據；因此，一旦授權依據的條件具備則可相容。有基於機器學習型人工智慧系統階段性的技術限制使然，例如僵化式決定是早期依簡單邏輯而運作之自動化系統的特徵，但在模型參數以驚人速度增長的機器學習技術發展下，目前法律人工智慧系統在理論上並非不能考量各種相關情境，做出符合比例原則的裁量；但欠缺可解釋性的問題卻依舊是當下多數機器學習型人工智慧系統在技術上的致命傷。至於主張法治與機器學習型人工智慧系統在本質上衝突而不可能相容的觀點，例如權限轉移由法律人工智慧系統進行執法即使滿足法律授權依據的條件，仍損害個人自主的道德判斷空間等，則是對法律人工智慧系統徹頭徹尾的懷疑論。因此，以「模仿」為機器學習基礎的法律人工智慧系統即使「有能力」在法律適用的內部證立上模仿法律人的目的解釋與其他涵攝推論，仍必須回答是否因為現階段的技術限制或因其本質使然而與法治的要求不相容。

FUNDAMENTALS OF ALGORITHMIC GOVERNANCE 117, 130 (Markku Suksi ed., 2023); Pöysti, *supra* note 7, at 43.

37 Meneceur, *supra* note 36, at 134.

38 See Lena Enqvist & Markus Naarttijärvi, *Discretion, Automation, and Proportionality*, in THE RULE OF LAW AND AUTOMATED DECISION-MAKING: EXPLORING FUNDAMENTALS OF ALGORITHMIC GOVERNANCE 147, 158-160 (Markku Suksi ed., 2023).

39 Meneceur, *supra* note 36, at 131.

本文限於篇幅，並不打算按照威尼斯委員會的六項法治共識標準，以檢核表模式對模仿型法律人工智慧系統進行逐項檢驗，而是聚焦在本文認為對法治原則構成初步（*prima facie*）挑戰的四個成熟的模仿型法律人工智慧系統「特性」（*properties*），進行相容性分析，當中至少包含「禁止恣意」、「平等」、「法律明確性」，以及與法治原則緊密關聯的「個人自主」等面向之檢驗。從模仿型法律人工智慧系統的具體「特性」切入，而非依照抽象的法治標準進行地毯式檢驗的好處是，可過濾掉與模仿型人工智慧系統其實無關，但經常被認為是人工智慧系統違反法治原則的議題。後者更多時候牽涉到的其實是「相關性發掘型」人工智慧系統所引發的爭議，本文也會在稍後就該等議題進行分析⁴⁰。

一、高效率執法會損害法律主體的自主嗎？

法治概念的具體內涵與主張雖然在理念的層次有諸多不同的版本，但法治所追求的終極目的既然是藉由限制國家權力以維持法律主體的自主性，則設若法律人工智慧的採用將否定法律主體或破壞其自主，即與法治有本質的衝突關係。再者，法律人工智慧系統所提供的全自動高效率執法能力，雖因為「處罰的確定性」而解消規範內容與規範實踐間的實踐縫隙，但執法本身實際上欠缺對法律內容實質理據的說明功能，而只是單純地侷限受法律規制之行動主體在行動上的可選擇性⁴¹。基此，陳弘儒認為：法律人工智慧系統所具備的高效率執法能力，使受法律規制的行動者僅因行動的選擇可能性受侷限而守法，並非基於自身對涉及法律的理據或考量之重量與優先次序予以思辨的決定，使行動者無法成為法治理想中自主與負責的法律主體，因此法律人工智慧系統與法治理念不相容。

⁴⁰ 請見「肆、二」節。

⁴¹ 陳弘儒，法治理念（註11），頁193-195、229-233。

上述推論所仰賴的小前提其實包含兩個事實命題主張：一、高效率執法是法律人工智慧系統的必要特徵；二、高效率執法必然會壓縮行動者的自主道德判斷空間。但這兩個命題有以下值得商榷之處。

首先，法治理念追求的終極目的是藉由限制國家權力以維持法律主體的自主性，此一論點雖值得贊同。但所謂「法律體系必須給予行動者在思索法律對其行動影響上具有充足的自主判斷空間」⁴²，是否即意謂在「執法上」給予空間，卻值得進一步推敲。畢竟執法一直是擔保法治得以存續的重要條件之一，違法卻無制裁則可能侵蝕法治的根基⁴³。但若「高效率執法必然會壓縮行動者的自主道德判斷空間」的主張反過來申論，則不執法相較於執法似乎反而給予行動者免於干擾的最大自主道德判斷空間；倘行動者最後的理據排序選擇的不是守法而是違法，此一結果若又不能被接納為符合法治，則所謂給予行動者自主道德判斷空間豈不淪為沒有實踐意義的口惠？

或許陳弘儒擔心的是，行動者原本可以不問法律規定內容的理由為何，純粹因為作為政治社群的成員而願意促進其所身處政治社群的共善（common good）而守法，或因為基於理解相互合作始能克服自利困境的理性而守法⁴⁴，又或者是真實地考量法律規定內容背後的理由，理解若依照法律規定內容行動對追求特定目的或達成某一結果而言所可能造成的實踐差異，而決定守法；然而一旦法律

42 陳弘儒，法治理念（註11），頁216。

43 依照威尼斯委員會的報告，法治的合法性基本條件要求「不得違反法律卻有罪不罰，法律應該在可能範圍內被執行。」Venice Commission, *supra* note 34, ¶ 42 (“the law cannot be violated with impunity. Law should, within the bounds of possibility, be enforced.”).

44 這是關於「法律規定如何給予行動者行動理由」的問題。在此僅舉兩種獨立於理由的合法性差異作為例子，前者即王鵬翔所稱「基於共善的合法性差異」，後者即王鵬翔所稱「基於計畫目標的合法性差異」。王鵬翔，規則的規範性，收於：謝世民編，理由轉向：規範性之哲學研究，頁325-356（2015年）。

人工智慧系統以高效率執法對違法者進行制裁，將逼迫行為人為了迴避制裁的不利益後果而守法，必然對於其他的守法動機產生排擠效應。

的確，倘若違法行為「必定」會遭受制裁的不利結果，則法律規定「不得超速」這個事實確定可以給予行為人「一個」開車不超速的理由，因為「不超速必然可以免於遭受處罰」。但「開車超速」這個行為本身仍可能具有其它差異製造能力，可作為「開車不超速」的其他理由：例如，「開車超速」可能會導致被太太碎唸、可能會產生車輛失控的風險等等。因此「超速將遭受制裁」這個理由是否必然壟斷成為行動者開車不超速的「唯一」理由，因為欠缺實證的基礎而無從評斷。但每個人守法的理由與目的本來就是分歧與多樣的。高效率執法所帶來的「處罰確定性」與「壓縮個人守法理由的選擇空間」二者間，存在著有待填補的因果落差。因此即使「法律體系必須給予行動者在思索法律對其行動影響上具有充足的自主判斷空間」，高效率執法會必然導致自主判斷空間遭到壓縮的推論也不成立。

其次，即使不爭執「高效率執法」是否必然壓縮行動者自主道德判斷空間的問題，「高效率執法」其實也不是法律人工智慧系統的「特性」或必要特徵。以模仿為機器學習基礎的人工智慧系統在公部門應用所能發揮的主要功能其實是有效率地「解決法適用問題」。將法律人工智慧系統的主要功能設想為在高效率執法能力下「處罰」違法者，或許受制於「區間測速執法」與「辨識欠稅車輛」兩個具體事例的有限想像使然⁴⁵。但執法並針對違法者施以處罰，並非模仿型法律人工智慧系統的唯一或主要任務。法律解釋、事實認定與涵攝推論，才是法律人工智慧系統之所以被開發與運用的更遠大目標。退萬步言，即使僅討論法律人工智慧系統的執法功

⁴⁵ 陳弘儒，法治理念（註11），頁193-195。

能，並非不能刻意調整系統的執法效率，例如每100位違規者僅處罰80位或更少的執法密度，從而解消法律人工智慧系統的高效率執法與法治不相容的論點。

二、權限移轉必然不可容許嗎？

將處罰或法律適用的決定權限於符合法律授權的前提下移轉予機器，是否符合法治，是接下來需要回答的問題。而這個問題必須建立在「同等履行」的前提條件之上，也就是當法律人工智慧系統在履行特定功能與人類有一樣的能力時，探討移轉特定事務的判斷權限予人工智慧系統與法治的關係，才有意義。在此假設前提下，主張法律人工智慧系統與人類在現實上並未具有同樣能力的「非同等履行論證」⁴⁶，例如前述被認為仍無法突破「目的解釋屏障」的CMLRs，即不具有與法律人同樣的法律推論能力，移轉判斷權限予機器當然就欠缺正當性，而不在此討論範圍之內。反之，本文認為以「模仿」為機器學習基礎的法律人工智慧系統「有能力」在法律適用的內部證立上模仿法律人的目的解釋與其他涵攝推論，因此已滿足「同等履行」的部分前提⁴⁷。

在「同等履行」的假設前提下，陳弘儒依舊反對在法治的量尺上將法律人工智慧系統的同等判斷評價為與人類的判斷無異⁴⁸。他

⁴⁶ 陳弘儒，法治理念（註11），頁187。

⁴⁷ 具備「同等履行能力」的模仿型法律人工智慧系統不僅要能掌握法律問題由權威機關所決定的「暫時性標準答案」，也要像真實世界中的「法律人」一樣，具備能夠持續對何謂「標準答案」進行爭辯的能力。因此掌握「法律問題暫時的標準答案」和具備「涵攝推論能力」，是模仿型法律人工智慧系統具有同等履行能力的兩個必要條件。本文的任務是嘗試證立以機器學習為基礎的人工智慧系統，並非完全不可能具備第二個必要條件中相當關鍵的「目的解釋能力」（反駁陳弘儒的懷疑論），但本文並未宣稱目前的機器學習型人工智慧系統已具備第一個必要條件。讓機器學習的人工智慧可滿足第一個必要條件的方法，請參見註28之說明及其對應之本文。

⁴⁸ 陳弘儒反對「同等履行的無差異原則」（indifference principle under equal performance, IPEP），陳弘儒，法治理念（註11），頁186-189。

主張，「依法而治的主體想像要求權限判斷必須交由人來做」⁴⁹，而背後的理由則再一次地聯結到法治應維護法律主體自主道德判斷空間的命題之上。

然而，法治所訴求的對象一為公民，一為國家權力。法治的運作固然應盡可能維持公民作為法律主體的自主道德判斷空間，以體現公民個人之人性尊嚴。但法治既然是要拘束國家的統治權力能依法而治，盡可能降低統治者個人意志（will）對於權力運作之影響，則其運作似乎不應該建立在特定掌握權力個人的道德判斷之上。因此欲維持「國家」的自主道德判斷空間以便體現其「人性尊嚴」的說法，不只顯得突兀，也與法治的理念相互衝突。不過，我們還是可以想像兩種可能的理由，支持國家權力的行使必須交由人類來做。

第一，法外開恩或情理法的期待。我們可能會認為，即使法律人工智慧系統在履行特定功能時與人類有一樣的能力，在具體個案的現場仍應由真正血肉之軀的人而非冰冷的機器做出決斷，始能設身處地的考量個別情境的特殊性而給予例外的情理通融。例如，一位美國威斯康辛州的員警，面對一名為避免酒駕，而將車子超時停在路邊停車格一晚，自行走路回家的駕駛人，決定不開出逾時停車罰單，就是考量駕駛人是為了避免一個更大的危害而法外開恩⁵⁰。此種裁量權限的運用似乎為法適用的權力行使增添了人性的色彩⁵¹。然而，我們仍必須區分法治下的目的解釋與比例原則的法律適用操

49 陳弘儒，法治理念（註11），頁184。

50 Jennifer Earl, *Parking Officer Takes Pity on Driver After Finding Handwritten Note on Windshield*, CBS NEWS (June 12, 2017), <https://www.cbsnews.com/news/parking-officer-pardons-driver-after-finding-this-handwritten-note-on-windshield/>; 王道一，〈《經濟五四三》這張罰單該開嗎？為什麼依法行政可能是件壞事，天下雜誌，2017年6月26日〉，<https://www.cw.com.tw/article/5083354>（最後瀏覽日：2024年1月11日）。

51 Enqvist & Naartijärvi, *supra* note 38, at 148-49.

作，與全然超越法律拘束的法外之情。前者仍是依法而治的操作，並非不能由「模仿型」的法律人工智慧系統在「同等履行」的條件下達成法治的要求與期待；後者既然已超越法律拘束，探討機器是否能在法治下同等履行的問題即已喪失意義。

第二，武力的使用必須由人來決定。壟斷武力的國家權力在對外行使上必須遵循一定原則，諸如攻擊對象必須區別戰鬥與非戰鬥人員以符合「區別原則」、攻擊必須符合「比例原則」與「必要性原則」等⁵²，使其與其他國家權力必須依法而為有相似之處，而似乎有透過機器學習型人工智慧系統予以模仿的可能。然而歷史上，人類對武力的使用最終多依賴人類身體與心理的脆弱性與極限，作為停止武力使用的原因。欠缺這種極限與脆弱性作為邊界，將導致人類使用武力後自身的滅絕。因此，致命武力的使用即使可由機器做成符合區別原則、比例原則與必要性原則之判斷，確實仍有將之置於人類直接控制之下的必要性⁵³。

三、欠缺（內部）可解釋性違反法治嗎？

成熟的人工智慧系統，例如本文所稱以「模仿」為機器學習基礎的法律人工智慧系統，即使已滿足「同等履行」的標準，而能在法律適用的內部證立上模仿法律人的目的解釋與其他涵攝推論，但目前為止多半仍無法對其如何產出演算結果之決定予以完整的說明解釋，而存在著「黑箱」⁵⁴。欠缺可解釋性（explainability）是長期困擾機器學習型人工智慧系統的問題，使機器學習型法律人工智慧系統面臨法治相容性的質疑。

52 黃居正，與人工智慧相關的國際法議題——從國際人道法到生命體法，收於：劉靜怡編，人工智慧相關法律議題芻議，頁215-241（2018年）。

53 黃居正（註52），頁232-234。

54 FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3 (2015).

一般認為，欠缺可解釋性的問題並無法僅以提供機器學習的演算邏輯、使用之資料參數及權重、程式原始碼等「魚缸式透明」(fishbowl transparency)的方式而解決，必須透過理由的提供(reason-giving)證成演算結果的決定以達成「說理的透明」(reasoned transparency)⁵⁵。亦即，「魚缸式透明」雖提供關於法律人工智慧系統在作什麼(what)的相關資訊，但真正在法治上有意義的應該是揭露法律人工系統為什麼(why)做出特定決定的原因。

然而，國家在法律適用問題上的說理義務，主要是對法律提出解釋、根據證據認定事實，並依此說明涵攝推論之結果。一位符合資格的法律人在代表國家履行其說理義務時，除了依據上述要求就「得心證之理由」提出對外說明之外，並不被要求就其如何得心證背後的內在生理機制提出解釋，亦不被要求對過去如何學習法律或取得相關資格(如何通過司法官考試或公務員高等考試)的內在過程做出說明。我們雖然可合理而正當地要求職司法律適用的法律人應具備一定的資格，但通過資格門檻條件後，法治的標準僅要求法律人對其判斷決定提出說理。而關於法律人如何得出心證的其他「內部解釋」向來不是其說理義務需要包含的內容。

國家在法律適用問題上的說理義務既然不包括如何得出法律適用結論的「內部解釋」，則縱使機器學習型法律人工智慧系統對其究竟如何運作而能在法律適用的內部證立上模仿包含目的解釋在內之各種涵攝推論欠缺「內部可解釋性」，這個缺陷也不該具有法治意義上的重要性。對法治所要求的明確性與個人司法救濟可能性的保障而言，只要能比照法律人說明「得心證之理由」，提供機器決定所依據的推論說明，應為已足。真正需要再三確認的毋寧是模仿

⁵⁵ Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 32-38 (2019).

型法律人工智慧系統是否已確實如人類同儕一樣，通過資格檢定而確能滿足「同等履行」的標準。本文認為，對機器學習型人工智慧系統無法說明其演算結果的擔憂，多源自於未能清楚界定「模仿型」法律人工智慧系統的功能與作用，及其與「相關性發掘型」人工智慧系統的區別所致。「相關性發掘型」人工智慧系統除了同樣欠缺「內部可解釋性」之外，更因為無法提供「外部解釋」而對法治帶來有別於「模仿型」法律人工智慧系統的全新威脅與挑戰，將於下文進一步說明。

四、複製系統性偏見與阻礙法治進化的潛在問題

「模仿型」法律人工智慧系統既然是以當下存在的法律實踐文本，作為學習與模仿法律社群如何進行法律推論活動的基礎，就無可避免地將此等訓練資料本身隱含的各種既存的系統性偏見直接予以承襲。此外，「模仿型」法律人工智慧系統在承襲法律社群之積習與可能的錯誤的同時，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。

這兩個缺陷雖然是「模仿型」人工智慧系統的內在限制，也確實可能與法治理念存在一定的緊張關係，但卻並非法律人工智慧系統所獨有。由人類所構成的法律社群本身也持續地面對這兩個潛在問題。即便如此，人工智慧系統的高效率執法潛能，以及欠缺內部可解釋性的問題，仍使得此等缺陷在模仿型人工智慧系統中有可能被放大的風險。

五、小結

經過上面的分析，本文認為讓「模仿型」法律人工智慧系統可以在法律適用的內部證立任務上，模仿法律社群的目的解釋與其他涵攝推論，在以下的條件下仍與法治理念相容：

第一，權限移轉獲法律授權。將處罰或法律適用的決定權限移轉予「模仿型」法律人工智慧的自動化決策系統，必須取得法律授權，以滿足合法性原則。

第二，同等履行獲得擔保。「模仿型」法律人工智慧系統必須與人類同儕有「同等履行」之能力。「同等履行」是「模仿型」法律人工智慧系統能與法治有條件相容的必要條件。因此，必須建立適當機制，在事前確認系統的「同等履行」能力⁵⁶。

第三，保留由人類進行最終救濟決定的可能性。考量「模仿型」法律人工智慧系統仍可能複製系統性偏見，也欠缺藉由「打破成規」(rule-breaking)締造法律文明進化的能力，「模仿型」人工智慧系統仍必須保留由人類進行最終救濟決定的可能性(human on-the-loop)，以留下由人類檢視是否存在系統性偏誤、是否需打破成規的機會。

肆、相關性發掘型人工智慧系統的資料驅動決策與法治挑戰

機器學習除了可以特定的「基礎事實」(ground truth)為參考座標，學習並模仿人類的身體動作、思維模式、知識系統或決策行動，從而打造出模仿型人工智慧系統之外，機器學習也能在沒有前例指引下，自行發掘各種資料間的相關性(correlation)，或依據新發現的特徵重新建立資料的分類(clustering)，創造出以資料驅動(data-driven)預測控制的新決策模式⁵⁷。在公部門的各類決策中利

⁵⁶ 關於同等履行能力之條件，請參見註47。

⁵⁷ 邱文聰(註12)，頁285-299。應予說明的是，無論是模仿型人工智慧或相關性發掘型人工智慧，都仰賴機率分布模型來建立模仿能力或相關性發掘能力，二者在所使用的技術上並無明顯不同。但使用機率來回答何種問題，則是兩種類

用機器學習的相關性發掘能力進行預測控制，已可見於各國的不同應用場景。例如，荷蘭社福主管機關利用SyRI系統偵測社福詐欺⁵⁸、海關及出入境管理執法中依照旅客過去旅遊路徑次數與頻率等判斷是否涉及走私或其他不法行為⁵⁹、歐盟機場以iBorderCtrl的試驗性系統偵測可疑入境者⁶⁰、金融監督管理主管機關找出證券交易中違法交易行為模式的金融監理科技Suptech等⁶¹，不一而足。經常成為法律人工智慧爭議案例討論對象的再犯預測系統COMPAS⁶²，雖然並非以機器學習方式訓練而產出預測再犯可能性的演算法，而是單純依據人工挑選的137個固定變項進行統計分析的結果，但在本文設定的相關性發掘標準下，也可歸類於此種人工智慧應用類型。相關性發掘型人工智慧系統之共通特徵係藉由機器學習所發現的相關性資訊，提供公部門作為預測性控制決策的依據。此種應用機器學習型人工智慧潛力的方式，顯然有別於前述模仿型法律人工智慧系統所具有的模仿法律論證能力，也因此對法治帶來與之不同的衝擊。

型人工智慧系統在規範意義上產生差異的原因。例如，以判斷犯罪熱點的人工智慧系統為例，模仿型人工智慧系統的開發方式，會以警察過去如何判斷犯罪熱點作為學習對象，但相關性發掘型人工智慧系統則會直接以過去曾經發生犯罪的案例為訓練資料從中找出特徵。

58 Rb. Den Haag 2 mei 2020, ECLI:NL:RBDHA:2020:1878 m. nt. R.H.T. Jansen en M.D. Reijneveld (NJCM/The Dutch State).

59 *Report from the Commission to the European Parliament and the Council on the Review of Directive 2016/681 on the Use of Passenger Name Record (PNR) Data for the Prevention, Detection, Investigation and Prosecution of Terrorist Offences and Serious Crime*, at 7, COM (2020) 305 final (July 24, 2020).

60 Meneceur, *supra* note 36, at 126-27.

61 David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey & Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (NYU School of Law, Public Law Research Paper No. 20-54, 2020), <http://dx.doi.org/10.2139/ssrn.3551505>.

62 關於COMPAS系統的相關法律爭議，可參見Han-Wei Liu et al., *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, 27 INT'L J.L. & INFO. TECH. 122, 122-41 (2019).

一、從相關性到預測控制的新理由創造者

相關性雖然代表A事件與B事件二者經常相互伴隨出現，卻不必然擔保A與B二者間具有因果關係。典型的事例是，冰淇淋銷量與鯊魚攻擊或溺水人數間雖具有相關性，但造成鯊魚攻擊或溺水的原因絕不是冰淇淋，促進冰淇淋消費的原因也絕對不是鯊魚的攻擊或溺水事件。因此，我們不可能透過人為介入改變販售冰淇淋的數量，來改變鯊魚攻擊或溺水事件的發生次數，也不可能藉由控制鯊魚的行為，影響冰淇淋的消費。簡單地說，相關性雖然說明冰淇淋與鯊魚攻擊或溺水事件間具有關聯性，但二者並非彼此發生與否的「差異製造者」(difference maker)⁶³，也因此無法藉由人為「介入」，控制二者之一的發生來改變另一者發生與否的結果⁶⁴。

機器學習的相關性發掘能力不只能指出單一A事件與B事件間存在一定的相關性，例如透過各種變項間的複雜演算發現購買家具保護貼與準時償還銀行貸款之間存在相關性。機器學習的相關性發掘能力更已能夠找出不同複雜變項組合 A_1 、 A_2 、 A_3 與B事件間不同程度的相關性，例如有 $A_1=X_1+Y_1+Z_1$ 特徵者的再犯罪(B事件)風險機率為 P_1 、具有 $A_2=X_2+Y_2+Z_2$ 特徵者的再犯罪風險機率為 P_2 、具有 $A_3=X_3+Y_3+Z_3$ 特徵者的再犯罪風險機率則為 P_3 等等。這些變項或特徵本身雖非某事件發生與否的「差異製造者」，但在資訊驅動之治理理性(governmentality)浪潮下⁶⁵，卻可成為對人及其相關事

63 關於difference maker的中文翻譯，我採取王鵬翔與張永健的用法譯為「差異製造者」，以有別於「差異製造事實」。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205、234(2015年)。「差異製造者」就是造成某個結果發生的原因。

64 「介入」是Judea Pearl在思考因果問題所提出三個層次中的第二層次。「差異製造者」是使某件事情(結果)發生的原因；掌握「差異製造者」才能藉由人為介入(intervention)來進行因果控制。JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT 27-36 (2018).

65 See, e.g., Fleur Johns, *Governance by Data*, 17 ANN. REV. L. & SOC. SCI. 53, 53-71 (2021).

物的各種「剖繪」(profiling)，進而再轉化為可據以進行某種預測性控制(介入)的基礎。究竟在資料驅動的治理性下，人們是如何將不具有因果關係的剖繪或相關性資訊，轉化為可以採行預測性控制(介入)行動的基礎？

上述問題可以藉由「差異製造」的概念，重新刻畫如下：在A與B僅存在相關性的世界裡，A和B並非彼此發生的「差異製造者」，則機器學習發掘出A與B間存在相關性的這個事實，確實並未直接指出「介入」A會對B結果的發生與否或發生機率造成差異，也因此A與B間存在相關性的這個事實本身並不是「差異製造事實」，無法直接成為採取特定行動的「理由」⁶⁶。不過，A與B間存在相關性的這個事實，卻可轉化為某個新的以機率為基礎的「差異製造事實」得以成立的基礎事實。讓我們以購買家具保護貼與償還銀行貸款的相關性為例來加以說明：

- (1)購買保護貼並非償還貸款的差異製造者（因為人為介入買或不買保護貼，並不會改變償還貸款的可能性）。
- (2)「購買保護貼與償還貸款存在相關性」的事實本身，並未指出「介入」買或不買保護貼，對於償還貸款結果的發生或發生機率造成差異，因此不是一個「差異製造事實」，也因此不是促成人們去購買保護貼，以便提高償還貸款可能性的行動理由。
- (3)但「購買保護貼與償還貸款存在相關性」的事實卻是「不貸款予未曾購買保護貼的人則可避免銀行呆帳風險」，以及「貸款給曾購買保護貼的人可增加銀行信貸獲益可能」這些以機率為基礎的新的差異製造事實得以成立的基礎⁶⁷。

66 將理由視為差異製造事實的主張，可稱為「基於差異製造的理由論」(difference-making-based theory of reasons)。王鵬翔受王一奇的啟發，發展出成果豐碩的法理學論述。請參見王鵬翔(註44)，頁325。

67 在此，「存在相關性的事實」應當屬於「給予理由的事實」(reason-giving

(4)前述(3)的差異製造事實在銀行追求提高收益可能與減少呆帳風險的目的下（目的論證），即成為應該要「貸款給曾購買保護貼的人」以及「不貸款予未曾購買保護貼的人」的理由。

我們也可以再以 A_n 特徵與再犯風險之間的相關性為例，來進行另一種結果論證模式的說明：

- (1)設若 A_n 特徵與再犯之間並無因果關係，二者僅單純存在相關性，因此人為事後介入調整 A_n 特徵並不會改變再犯可能性（ A_n 特徵不是再犯與否的差異製造者）。
- (2)「 A_n 特徵與再犯風險之間存在相關性」的事實本身，並不代表介入調整 A_n 特徵對於再犯結果的發生或發生機率造成差異，因此該事實不是一個「差異製造事實」，不構成藉由移除 A_n 特徵以便降低再犯可能性的行動理由。
- (3)但「 A_n 特徵與再犯風險之間存在相關性」的事實卻是「若於審前先行羈押有 A_n 特徵之人」、「若將有 A_n 特徵之人處以較長刑期」、「若拒絕有 A_n 特徵之人提前假釋」則可產生「降低整體社會犯罪率」結果的這些以機率為基礎的新的差異製造事實得以成立的基礎。
- (4)由於降低整體社會犯罪率（機率）的結果是好的（結果論證），前述(3)的差異製造事實就成為應該「審前羈押有 A_n 特徵之人」、「將有 A_n 特徵之人處以較長刑期」以及「拒絕有 A_n 特徵之人提前假釋」的理由。

fact），是可以使作為理由的差異製造事實得以成立的事實。關於「理由」與「給予理由的事實」的區分，請參見王鵬翔、張永健（註63），頁235-236；王一奇，理由與提供理由的事實，收於：謝世民編，理由轉向：規範性之哲學研究，頁105-140（2015年）。

機器學習的相關性發掘能力即是透過上述轉化，從其所發掘出的相關性資訊中，「創造」出採取特定預測性控制（介入）行動即可以提高達成特定目的或結果之可能性（減少呆帳、降低再犯）的行動「理由」，再以此為據，由機器做出自動化的預測性控制決定。

二、超越既有法治節制的新挑戰

機器學習的相關性發掘能力可經由轉化，從所發掘的相關性中「創造」出採取特定預測性控制行動以達成特定目的或結果（例如減少呆帳風險、降低再犯率）的新「理由」，並由人工智慧的自動化決策系統據以做出預測性控制的決策行動。相關性發掘型人工智慧系統的此種運作模式，顯然有別於以模仿為基礎的法律人工智慧系統，也因此對法治帶來截然不同的挑戰。基於以下三個理由，本文認為「相關性發掘型」人工智慧系統的上述「特性」，不僅與法治理念的部分要求不相容，更已脫逸出法治在節制權力恣意與濫用上能發揮作用的最大疆界。

（一）欠缺（外部）可解釋性所導致的法治困境

機器學習的相關性發掘能力最終雖可經轉化而「創造」出採取特定預測性控制行動以達成特定目的之新「理由」，但除了少數例外，機器學習所發掘的複雜相關性資訊，多半不在人類現有的認知範圍，也不被現有因果知識所掌握，使得機器學習所發掘的相關性資訊通常難以被一般人所理解。然而，「相關性發掘型」人工智慧系統雖然與前述「模仿型」法律人工智慧系統一樣，都無法就其內部演算經過提出可為一般人理解的說明，導致欠缺「內部可解釋性」。但「相關性發掘型」對法治帶來真正挑戰之處在於，「相關性發掘型」並無法藉由模仿法律人的法律論證，針對相關性衍生的決策行動提出三段論式的「得心證理由」以履行其說理義務。例如，難以針對「因當事人未曾購買保護貼，所以拒絕其貸款的決策行動」，提出三段論式的說理以證成行動的合法性。我將此稱為欠缺

「外部可解釋性」問題，以別於前述欠缺「內部可解釋性」問題。而造成「相關性發掘型」人工智慧系統的決策行動欠缺「外部可解釋性」的原因則有以下二者：

第一，因策略性隱蔽相關性資訊而導致決策行動欠缺外部可解釋性。機器學習所發掘出的隱藏相關性，通常並不在人類現有認知範圍內，例如，前述購買家具保護貼與準時償還銀行貸款存在相關性一事，並非眾所周知的事實。超越人類現有因果知識範圍的相關性資訊不僅對發掘出此等相關性的機器學習系統開發者或領域專家（field experts）而言具有「非直覺性」（non-intuitiveness），對一般人而言更具有難以意想其存在的「隱蔽性」（opaqueness）⁶⁸。相關性資訊的隱蔽性使據此進一步創造出的「行動理由」（決策行動所依據的「差異製造事實」），在未刻意予以揭露時，因隱蔽性遞移而不為人知；「行動理由」的隱蔽性則使建立於其上的人工智慧自動化決策系統，因未能就其「決策行動」提供說理，最終導致決策行動欠缺「外部可解釋性」。欲打破「相關性發掘型」人工智慧系統所創造之「行動理由」的隱蔽性，讓人工智慧系統的自動化決策行動具有「外部可解釋性」，就必須先打破機器學習所發掘之相關性資訊的隱蔽性，使利用該等相關性資訊而創造人工智慧系統的「行動理由」，被理解與說明。然而，一旦對外揭露機器學習所發掘之相關性及據此創造之「行動理由」，將使行動者可利用該等相關性資訊進行策略性反制，導致人工智慧系統面臨行動者操弄相關性資訊而造成「博弈風險」（gaming risk）⁶⁹。例如，一旦對外公開揭露「相關性發掘型」人工智慧系統利用「購買保護貼與償還貸款存在

68 學者W. Nicholson Price II及Arti K. Rai將機器學習的隱蔽性進一步區分為（模型）複雜性（complexity）、（結果）非直覺性（non-intuitiveness）與（系統）祕密性（secrecy）三者。W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 784-88 (2021).

69 關於博弈風險的一般性論述請參見Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 10 (2018).

相關性」的資訊，創造出是否貸款的「行動理由」(「貸款申請人是否購買保護貼」是「銀行貸款後是否能獲利或出現呆帳」的差異製造事實)，可能會使申請貸款者得以事前刻意購買保護貼的方式來矇騙取得銀行貸款，最終導致「相關性發掘型」人工智慧系統欲達成的預測控制目的落空。反之，若為了避免「相關性發掘型」人工智慧系統的預測控制功能遭「博弈」利用而瓦解，就必須策略性地保持相關性資訊的隱蔽性，但也因此導致必須策略性地隱蔽人工智慧系統自動化決策所仰賴的「差異製造事實」，使其決策理由欠缺「外部可解釋性」而無法滿足法治對說理義務的要求⁷⁰。值得一提的是，「相關性發掘型」人工智慧系統面臨「維持預測控制目的」與「履行說理義務之法治要求」的兩難困境，並不會發生在「模仿型」人工智慧系統；後者並不因為模仿法律人提供「得心證之理由」的法律推論說理，就使系統面臨「博弈風險」，也不會有為了避免「博弈風險」而對行動理由保持「策略隱蔽性」的需要。這是一般論者在未清楚區分「模仿型」與「相關性發掘型」人工智慧系統下，較容易忽略的差異。

第二，機器學習所發掘的相關性欠缺驗證可能，導致決策行動欠缺外部可解釋性。由於機器學習的模型參數與訓練資料經常被當作「秘密」而未予公開⁷¹，因此除非能仰賴機器學習以外的其他工具或方法，取得額外資訊（例如進行隨機控制實驗所獲得的實驗結

70 學者討論機器學習之「黑箱」時雖較常使用「秘密」(secret)一詞來描述外界無從得知演算法內在運作邏輯的現象，例如Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 657-60 (2017); Price & Rai, *supra* note 68, at 784，但並未如本文詳細區分「決策行動理由的隱蔽」與「驗證決策行動理由為真的隱蔽」(見下述)二者。本文刻意使用「策略隱蔽性」(strategic opacity)一詞用以指稱為避免博弈風險而產生的「決策行動理由隱蔽」需求，以有別於多半基於營業競爭考量以致阻礙驗證行動理由為真的「秘密」。

71 在此之秘密包含Frank Pasquale所稱之「實際上的秘密」(real secrecy)與「法律上秘密」(legal secrecy)兩種情形。PASQUALE, *supra* note 54, at 6-8.

果)以針對機器學習所發掘之相關性進一步完成因果推論⁷²，否則外界通常難以驗證機器學習透過模式辨認所發掘的相關性本身是否為真。例如，我們缺少必要的資訊驗證「購買家具保護貼」與「準時償還銀行貸款」二者的相關性真的存在。機器學習發掘的相關性「欠缺驗證可能」的問題，係獨立於前述「策略隱蔽性」問題而存在；換言之，即使公開揭露人工智慧系統自動化決策行動所依據的「相關性資訊」，也不當然解決相關性本身欠缺驗證可能的問題。

由以上分析可知，相關性資訊的「策略隱蔽性」使人工智慧系統的「決策行動理由」(決策行動所依據的「差異製造事實」)無法被公開，而相關性「欠缺驗證可能」則使「決策行動理由」無法被驗證為真。這兩個因素是相關性發掘型人工智慧系統的決策行動欠缺「外部可解釋性」的成因。我們也可以說，「外部可解釋性」是由「決策理由的公開」與「公開之決策理由可被驗證為真」這兩個因素所構成。而相關性發掘型人工智慧系統的決策行動之所以欠缺「外部可解釋性」正是因為系統開發者與部署者想要在保有系統開發秘密的情況下(決策理由無法被驗證)，利用系統進行預測控制(決策理由不公開)。那麼，在法律上賦予當事人「解釋權」(right to explanation)或課予系統開發者與部署者「解釋義務」等透明化策略，是否就能克服相關性發掘型人工智慧系統之決策行動欠缺「外部可解釋性」的困境⁷³？

72 關於如何透過實證研究來進行因果推論，請參見王鵬翔、張永健(註63)，頁246-264。

73 針對當事人是否就自動化決策享有「解釋權」、解釋權之實定法依據以及其內涵為何等問題，國內外法學界已有相當豐碩的論著，例如黃詩淳，AI可解釋性的法學意義及其實踐，臺大法學論叢，52卷特刊，頁931-972(2023年)；see Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189 (2019); Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 79-90 (2017); Kiel Brennan-Marquez, "Plausible Cause": *Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249 (2017); Andrew D. Selbst & Julia Powles, *Meaningful*

我們不難發現，系統開發者或部署者隱匿系統開發資訊以致無法驗證「決策行動理由」是否為真的問題，並非不能由立法者在權衡秘密所欲維護之利益（例如營業秘密欲維護秘密擁有者的財產上利益）與驗證決策行動理由所追求法治價值二者之輕重後，透過法律規定在一定條件下解除秘密，以供驗證「決策行動理由」。例如，制定法律要求系統開發者或部署者必須提供系統開發之資訊予管制者或第三方驗證機構⁷⁴，以便對人工智慧系統決策行動所依據的「差異製造事實」（相關性資訊）進行事前（*ex ante*）或事後（*ex post*）的公共驗證⁷⁵。此種為確保公共課責性（accountability）的管制模式已可見於新藥上市許可程序，強制開發新藥之藥廠必須提供性質上屬營業秘密的臨床試驗資訊，以供主管機關能據以審查新藥的安全性與有效性。公部門部署相關性發掘型人工智慧系統用於公共決策時，更有義務提供「可驗證其決策理由」的資訊。在此意義下，由法律賦予當事人「解釋權」或課予系統開發者與部署者的「解釋義務」，即能解決機器學習所發掘的相關性欠缺驗證可能的問題，部分緩解決策行動欠缺「外部可解釋性」的困境。

Information and the Right to Explanation, 7 INT'L DATA PRIVACY L. 233 (2017); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841 (2018); Boris Babic & I. Glenn Cohen, *The Algorithmic Explainability "Bait and Switch"*, 108 MINN. L. REV. 857 (2023). 針對「解釋權」之相關議題進行完整的探討與分析，超越本文自我設定的任務範圍。本文之主要目的係分析「解釋權」的法制策略，對解決「驗證決策行動理由為真的隱蔽」問題雖具有一定功用，但對於「決策行動理由的隱蔽」問題卻存在侷限性。

⁷⁴ 歐盟數位服務法（Digital Services Act）即兼採第40條第3項由主管機關掌握主動監理權限，與第37條課予部署者自行委託第三方進行獨立稽核（independent audit）義務的雙重管制模式。Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1.

⁷⁵ See also David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REG. 800, 828-54 (2020); Shea Brown et al., *The Algorithm Audit: Scoring the Algorithms That Score Us*, 8(1) BIG DATA & SOC'Y 1, 1 (2021); Ellen P. Goodman & Julia Trehu, *Algorithmic Auditing: Chasing AI Accountability*, 39 SANTA CLARA HIGH TECH. L.J. 289, 306-14 (2023).

然而，造成相關性發掘型人工智慧系統之決策行動欠缺「外部可解釋性」的「策略隱蔽性」問題，則與法治存在更為直接的緊張關係，難以藉由法律賦予「解釋權」或課予「解釋義務」而解決⁷⁶。不可諱言的是，「策略隱蔽性」問題在一些情況下有被擊敗的可能，從而使「解釋權」或「解釋義務」在表面上協助解決了部分的「策略隱蔽性」問題：當相關性所連結的因素是對外彰顯且難以改變的個人特徵（immutable characteristics，例如種族、膚色、性別等）或其他無法為個人於事後輕易操控的外在因素時（例如個人學歷、天候狀況等），公開揭露（說明與解釋）人工智慧系統的預測控制行動係仰賴此等無法為個人於事後輕易操弄的特徵或因素作成其行動決策，並不會使系統的決策行動面臨「博弈風險」；因此公開機器學習所發掘的相關性資訊，使人工智慧系統的決策行動理由（「差異製造事實」）可被說明與解釋，不致影響人工智慧系統所欲達成的「預測控制目的」。不過，上述情況實際並非由「解釋權」或「解釋義務」獨自克服「策略隱蔽性」造成決策行動欠缺「外部可解釋性」的問題，毋寧是「策略隱蔽性」問題在上述情況下的自我解消，創造了「解釋」與「預測控制目的」得以並存的空間。因此，除了「策略隱蔽性」在事實上可被擊敗的上述情況外，相關性發掘型人工智慧系統仍會在「履行說理義務的法治要求」與「維持預測控制目的」間，面臨真正的兩難困境，而無法藉由賦予當事人「解釋權」或課以開發者與部署者「解釋義務」加以解消。換言之，「解釋權」或「解釋義務」雖可解決相關性發掘型人工智慧系統因相關性欠缺驗證可能所產生的問題，卻無力調解「預測控制目的」與法治要求「說理義務」間的衝突關係。

76 此處之「解釋權」係提供論者所稱針對「特定決策的事後解釋」（*ex post explanations of specific decisions*）。See Wachter, Mittelstadt & Floridi, *supra* note 73, at 78.

(二) 機率決策的行動合理性與難以節制的控制慾望

相關性發掘型人工智慧系統對法治理念的另一個挑戰則是系統的決策行動以「機率」而非確定的因果關係作為其「決策行動理由」。「機率」作為決策行動理由對法治帶來的挑戰雖與前述關於決策行動欠缺「外部可解釋性」有關，但二者不盡相同。

機器學習所發掘的隱藏相關性建立在機率或蓋然性的關係之上，多半不具有因果性⁷⁷。若從「禁止恣意」的法治必要條件或更為直接的合理性或比例原則加以觀察，則相關性所蘊含的機率資訊雖有助於預測性決策達成預測控制目的，或可符合適當性原則，但顯然並不適於在「必須以較高證明度所確認之事實作為個人責任基礎」的案件中作為決策依據。例如，是否構成犯罪的各種事實認定，包含客觀與主觀構成要件及罪責，均不應利用相關性發掘型人工智慧系統的機率預測能力來完成，否則將違反「犯罪事實應依據證據認定之，無證據不得推定其犯罪事實」的無罪推定原則。而即使具體危險犯（致生○○危險）或適性犯（足生XX侵害）的認定涉及個案中系爭行為所造成危害風險的機率判斷，但系爭行為是否是犯罪構成要件所禁止的危險，一般認為仍以該行為對危險結果的發生具有決定論式的「因果力」（causal power）或「差異製造能力」為必要⁷⁸，方能在確保行為人對行為結果的客觀預見可能性與結果避免可能性的前提下論斷其罪責，因此若直接由未經因果驗證的相關性發掘型人工智慧系統的機率預測，作為危險犯行認定的基礎，同樣也欠缺合理性。

77 從相關性到因果性，需要仰賴機器學習以外的其他工具與方法以取得額外資訊（例如控制隨機實驗的結果）來進行因果推論，因而超出人工智慧系統本身所具有的「相關性發掘型」能力。關於相關性與因果性的區別，另請參見PEARL & MACKENZIE, *supra* note 64, at 5-6, 82-84.

78 關於「決定論的因果概念」與「可能性的因果概念」在刑法上的不同意義，請參見許恒達，食品安全衛生管理法刑事責任之檢討——以強冠劣油案為中心，收於：倪貴榮編，食品安全與司法救濟，頁149-206（2020年）。

相對地，在預測性決策上利用機器學習所發掘的相關性，對特定結果發生的可能進行機率預測，從而做出法律規定在該結果可能發生時的相應裁量決策，則具有初步的（*prima facie*）合理性。預測性決策普遍存在於管制與秩序行政當中，例如，警察職權行使法第6條第1項第1款規定，警察於公共場所或合法進入之場所，合理懷疑有犯罪之嫌疑或有犯罪之虞者，得查證其身分；入出國及移民法第18條第1項第13款規定，有危害我國利益、公共安全或公共秩序之虞者，移民署得禁止外國人入國；證券交易法第156條規定，主管機關對於已在證券交易所上市之有價證券，發生重大情事而有影響市場秩序或損害公益之虞者，得命令停止其一部或全部之買賣；石油管理法第15條第2項規定，國內石油市場因突發事故，致油品供需失調或有失調之虞時，中央主管機關得限制石油輸出業者輸出石油等。即使在刑事法領域，保安或保全處分亦常有以預測未來某種風險發生可能性為據而作成決策者，例如刑法第89條第1項規定，因酗酒而犯罪，足認其已酗酒成癮並有再犯之虞者，於刑之執行前，令入相當處所，施以禁戒；刑事訴訟法第101條第1項第3款規定，所犯為最輕本刑五年以上有期徒刑之重罪，有相當理由認為有逃亡、湮滅、偽造、變造證據或勾串共犯或證人之虞，非予羈押，顯難進行追訴、審判或執行者，得羈押之；刑事訴訟法第101條之1規定，特定犯罪之被告經法官訊問後，認為其嫌疑重大，有事實足認為有反覆實行同一犯罪之虞，而有羈押之必要者，得羈押之；刑法第77條第1項及第2項第3款規定，受徒刑之執行而有悛悔實據者，經執行一定刑期後，得許假釋出獄，但犯妨害性自主、妨害風化及強盜、海盜等罪而接受強制治療後，經鑑定評估其再犯危險未顯著降低者，則不得假釋⁷⁹。由於此等預測性決策係立足於現在向未來達成特定的預警性（*precautionary*）管制、保全等控制與

79 以「之虞」為檢索字詞即可在法源資料庫查詢到2,741筆法律條文及中央機關之行政命令條文。

介入目的，而非針對過去已發生事態對個人的究責，因此機器學習所發掘之相關性似乎可以被「合理」地轉化為以機率為基礎的「差異製造事實」，而成為超前部署與控制的行動理由。前述「購買保護貼與償還貸款」的相關性也是在預警性地控制銀行借貸風險的目的下，被「合理」地轉化為「不貸款予未曾購買保護貼的人則可避免銀行呆帳風險」以及「貸款給曾購買保護貼的人可增加銀行信貸獲益可能」的機率差異製造事實，而成為決定是否貸款的標準。

然而正是此等在法治原則下具備初步合理性的預測控制能力，對法治帶來顛覆性的挑戰。機器學習的相關性發掘能力為人類進行各種超前部署的社會控制慾望，創造源源不絕的新差異製造事實，並可以具體展現在三個應用層面上：

首先，機器學習所發掘的隱藏相關性資訊可為現有法規中既存的預測性決策（例如前述各種管制與秩序行政法律規定），創造出個案中的「決策行動理由」，進而使人工智慧系統可利用決策行動理由所刻畫的標準認定「裁決事實」（*adjudicative facts*）⁸⁰，於個案中據以自動化判斷「預測性構成要件事實」是否該當。荷蘭社福主管機關偵測社福詐欺的SyRI系統、歐盟機場的iBorderCtrl等也都為此類應用之事例。而美國威斯康辛州的量刑規定要求法官量刑前應考量犯罪的嚴重性（*the gravity of the offense*）、犯罪行為人本身的特質（*the character of the offender*），以及保護公眾安全的需要（*the need to protect the public*）⁸¹，也因此當法院利用COMPAS軟體預測被告「再犯可能性」以作為法官量刑參考依據時⁸²，雖非由

80 關於「裁決事實」的說明請參見Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402 (1942); 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、244（2008年）。

81 *State v. Rodgers*, 203 Wis. 2d 83, 552 N.W.2d 123 (Wis. Ct. App. 1996); WIS. STAT. § 973.013 (2023).

82 關於COMPAS之進一步說明及威斯康辛州案*State v. Loomis*的評論，請參見Liu et al., *supra* note 62, at 126-33.

COMPAS直接作成量刑的自動化決策，仍可歸類為藉由相關性預測結果進行個案裁決事實認定（認定犯罪行為人特質與保護公眾安全需要等量刑要件）的利用模式⁸³。

其次，機器學習所新發掘出的相關性資訊也可轉而用於「創造新的預測性決策的立法理由」，也就是透過創造新「立法事實」（legislative facts）⁸⁴作為立法者制定特定法規範的經驗基礎，從而給予應該進行特定立法的理由。例如，倘若機器學習的相關性發掘能力可以發現乘坐大眾捷運前具某些個人特徵與反社會暴力傾向間具有關聯性，此等相關性資訊就可能轉而成為制定「有暴力傾向之虞者，大眾捷運營運機構應禁止其搭乘大眾捷運」之預測性決策立法的理由，並於立法後交由此等相關性發掘型人工智慧系統在捷運閘門前作成預測性自動化決策，以阻擋被系統演算為有暴力傾向之虞者搭乘大眾捷運，或進行其他的預警性防範作為。機器學習的相關性發掘能力為既有的預測性決策創造決策行動理由時，尚必須以既有的預測性決策為其應用範圍；相對地，當機器學習的相關性發掘能力被用於「創造新的預測性決策立法理由」時，則僅受限於機器學習本身的相關性發掘能力，因此只要新的相關性繼續被發掘，其創造立法事實、給予立法理由的能力也就應運而生⁸⁵。

第三，機器學習的相關性發掘能力除了能用於創造一般性立法理由之外，也可進一步將相關性所掌握之「差異製造事實」當中的關聯特徵，以類似刑法抽象危險犯的立法模式，直接規定為啟動預測性控制介入行動的構成要件。抽象危險犯的立法模式是立法者依其生活經驗認為某一類型行為對於特定法益帶有一般高度危險性，

83 美國威斯康辛州的量刑規定，於應報主義之外暗示應考量犯罪行為人未來再犯風險，該等規定是否符合行為主義的罪刑相當原則，雖不無疑義，但不在本文檢驗之範圍。

84 關於「立法理由」概念的中文說明，請參見邱文聰（註80），頁239-257。

85 相似的觀點請參考Frank Fagan, *Law's Computational Paradox*, 26 VA. J.L. & TECH. 1, 18-28 (2022).

因此透過「立法上擬制之危險」，將符合構成要件所描述之事實視為已對法益造成侵害危險，無待個案中進行具體認定。刑法處罰抽象危險犯係對個人行為之究責，其構成要件所禁止的抽象危險行為，固然應對結果的發生具有決定論式的「因果力」或「差異製造能力」，才能在行為人對行為結果具有客觀預見可能性與結果避免可能性的情況下，達成刑法規制個人行為之目的，已如前述，因此也不應直接利用機器學習所發掘的相關性來制定抽象危險犯罪的構成要件。但預測性管制立法既然是立足於現在向未來採取預警性管制、保全等控制與介入手段，利用機器學習的相關性發掘能力，將其產生的各種對人及其相關事物的「剖繪」特徵，明文規定為啟動預測性管制之構成要件，並非不能「合理」地利用其所蘊含的機率資訊達成預警性控制目的。而此種利用機器學習之相關性發掘能力進行預測控制的方式，事實上是前述第二種「給予立法理由」之應用方式的一種特殊型態。不過，若為了避免因法制化的明確性要求而公開相關性資訊後產生被控制者策略性反制的「博弈風險」，則預測控制目的下的「剖繪構成要件化」也必須同時符合「無法為個人於事後輕易操控」的條件。

上述藉由機器學習的相關性發掘能力進行公權力預測控制的三種法制應用模式，經常遭遇以下違反平等原則的質疑：機器學習所發掘的相關性資訊可能存在各種偏誤，或在表面中立的相關性特徵下隱藏平等原則所禁止的歧視而背離法治⁸⁶。我們可將此等違反平等原則的質疑進一步分為以下兩個主要論點：

第一個論點質疑利用相關性資訊進行預測之「手段正確性」，並指出不正確的預測違反「恣意禁止」的平等要求。不可諱言，機

⁸⁶ See, e.g., Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 AM. CRIM. L. REV. 1553, 1553-77 (2019).

器學習過程確實可能因為訓練資料欠缺代表性、資料清理不足、模型選擇錯誤、過度擬合等「統計或運算上的偏誤」(statistical or computational bias)問題，或者因為開發者本身在認知上的「人為偏誤」(human bias)問題⁸⁷，導致利用相關性資訊在真實世界進行預測控制可能產生不正確的結果。不正確的預測不僅違反「恣意禁止」的平等要求，也根本無助於預測控制目的之達成而欠缺工具理性。因此如何在人工智慧系統開發的測試、評估、確認與驗證(test, evaluation, validation, and verification, TEVV)階段，去除偏誤以提高並確保預測的正確性⁸⁸，避免預測結果的偽陽性與偽陰性二者的錯誤比例在不同對象間存在明顯差異而存在恣意⁸⁹，以滿足工具理性的最基本要求，應當是相關性發掘型人工智慧系統能符合法治原則的必要條件。

第二個論點則以「禁止可疑分類」為中心，質疑相關性資訊所指認的新差異製造事實，無論其預測是否正確，可能將平等保障原則或反歧視法禁止對「可疑分類」群體（例如種族、性別）的故意「差別對待」(disparate treatment)歧視，刻意包藏或無意間反映在表面中立而無涉可疑分類的「剖繪」特徵當中⁹⁰。在「禁止可疑分

87 「統計或運算上的偏誤」與「人為偏誤」是美國國家標準暨技術研究院(National Institute of Standards and Technology)對人工智慧系統中可能存在「偏誤」(bias)的三種分類當中的兩種。Reva Schwartz et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence* 6-9 (NIST Special Publication 1270, 2022), <https://doi.org/10.6028/NIST.SP.1270>.

88 *Id.* at 20-31.

89 學者Deborah Hellman認為，以偽陽性與偽陰性二者的比例作為標準，可用以偵測某一風險評估預測方法適用在不同族群時，是否存在明顯不平等的偏誤，Hellman稱為「錯誤率等價性」(error rate parity)，see Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 834-46 (2020).

90 See Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 694-714 (2016); Michael Selmi, *Algorithms, Discrimination and the Law*, 82 OHIO ST. L.J. 611, 619-32 (2021); Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 117-19 (2018).

類」的平等法制框架下，一旦預測控制所仰賴的剖繪特徵使用「可疑分類」作為變項，即使該剖繪所提供的預測正確無誤，仍構成違法的故意「差別對待」歧視。前述使用「無法為個人於事後輕易操弄」之特徵進行「剖繪構成要件化」的應用模式，若涉及傳統的「可疑分類」，確實將構成「差別對待」歧視。但「無法為個人於事後輕易操弄」的特徵並不必然需以「可疑分類」為基礎，也不必然均涉及與生俱來「無法改變的個人特徵」(immutable personal traits)。機器學習駕馭高維資料的能力，已能輕易地在可疑分類或與生俱來特徵之外，發掘其他表面中立而無法為個人事後輕易操弄的相關性資訊(例如過去工作經歷、駕駛行為模式等)，作為預測「剖繪」的基礎。而利用表面中立的剖繪進行預測控制，無疑就增加平等法制在「禁止差別對待」模式下的法遵檢驗難度；對於僅將憲法平等保障對公權力的制約，侷限在「主觀歧視故意」類型的憲法解釋實務而言⁹¹，機器學習產出表面中立之相關性資訊的能力，也製造此類公權力預測控制脫逸出憲法平等保障原則射程範圍的危機⁹²。

誠然，當「可疑分類」群體相對於其他「非可疑分類」群體，在表面中立的特定預測控制措施下若遭受不合比例的不利影響時，「禁止可疑分類」的平等法制仍可在無從證明主觀歧視故意的情況下，透過客觀上「差別影響」(disparate impact)的檢驗，認定該措

91 例如美國聯邦最高法院自1976年起的實務見解即認為，欲主張憲法平等保障者必須能證明立法者或其他政府的作為具有「歧視目的或意圖」(discriminatory purpose or intent)，否則僅適用合理性審查基準。Washington v. Davis, 426 U.S. 229, 242 (1976); McCleskey v. Kemp, 481 U.S. 279 (1987)。

92 憲法平等保障原則對公權力的限制是否應僅限於具備「主觀歧視目的」之類型，牽涉平等理論中形式平等觀與實質平等觀，以及客觀理論與主觀理論的爭辯。相關討論請參見黃昭元，從平等理論的演進檢討實質平等觀在憲法適用上的難題，收於：李建良編，憲法解釋之理論與實務（九），頁286-293（2017年）；黃昭元，論差別影響歧視與差別對待歧視的關係——評美國最高法院 Ricci v. DeStefano (2009)判決，中研院法學期刊，11期，頁43-45（2012年）（以下簡稱：黃昭元，差別影響與差別對待）。

施違反平等保障⁹³。不過，在憲法層次的平等保障是否將純粹客觀存在的「差別影響」視為當然違憲歧視仍有待論辯之前⁹⁴，欲循此路徑將平等保障原則的憲法誠命，加諸在仰賴相關性資訊進行預測控制的公權力之上，將會是一個前景不明的策略。就算撇開憲法層次的平等保障是否包含「禁止差別影響歧視」的爭議，一般亦僅需證明手段與目的具備合理關連性，以及欠缺其他有效但影響較小之手段，即可將客觀上對可疑分類群體的「差別影響」正當化為合理的差別待遇⁹⁵。在此理解下，「禁止差別影響歧視」事實上乃是「反非理性差別待遇」平等觀的一種具體操作標準⁹⁶；利用機器學習的相關性發掘能力進行表面中立的三種法制上的預測控制應用，只要預測手段本身能滿足正當的超前部署控制需要（預測正確而能達到控制目的），就不至於是一種恣意的非理性差別待遇⁹⁷。

在上述平等原則檢驗下，相關性發掘型人工智慧系統在公權力預測控制上的應用，除了受到「預測正確性」的最基本拘束外，並不容易在「禁止可疑分類」或「反非理性差別待遇」的平等原則標準下受到節制。特別是，當機器學習進行相關性發掘所仰賴的資料

93 關於美國法上「差別對待歧視」與「差別影響歧視」概念的一般性說明另可參見黃昭元，差別影響與差別對待（註92），頁22-33。

94 關於我國大法官釋憲實務上是否承認不需證明「歧視意圖」的「差別影響歧視」（間接歧視），相關討論請參見湯德宗，大法官平等權解釋會通釋義，收於：司法院行政訴訟及懲戒廳編，行政訴訟新制二十週年研討會會議實錄，頁155-169（2022年）。

95 美國民權法第七編關於「差別影響」的就業歧視相關法制，即規定若雇主能證明表面中立之措施與「工作相關且業務上必要」（job related and consistent with business necessity），又無其他可滿足雇主需要而影響較小的替代方案（alternative practice that would serve the defendant's interests while reducing adverse impact），即可正當化具有差別影響的措施。See *Selmi*, *supra* note 90, at 636-44.

96 在欠缺歧視之主觀故意的情況下，對產生差別影響之表面中立措施的平等原則（平等權）違憲審查標準，以合理審查基準為預設，只是為了「追求『合法的公共利益』，且其差別待遇之手段與所欲追求目之達成間具有『合理關聯』（rationally related）者，即應認定為合憲。」，湯德宗（註94），頁88、200。

97 至於是否存在其他有效而影響較小的手段能滿足超前部署的控制目的，則只能就具體情境個案判斷。

本身若存在制度或歷史因素造成的「結構性偏誤」(systemic bias)時⁹⁸，機器學習所發掘的相關性不免沿襲此等偏誤，使預測的「正確性」陷入「自證預言」(self-fulfilling prophecy)的陷阱。舉例而言，倘若有色人種在現實世界中因根深蒂固的文化偏見而被恣意攔查與逮捕的比例較高，從而後續訴追與定罪比例也必然較高，一旦直接以現實中法院的再次定罪資料，作為機器學習「再犯率」預測的學習訓練對象，將得出有色人種與再犯率間存在相關性的客觀發現，使利用此等相關性資訊預測有色人種有較高再犯風險獲得表面上的合理性。此外，利用相關性資訊進行預測控制也可能造成社會過度分化(balkanization)而破壞社會團結(social cohesion)的不利結果⁹⁹。

然而，挑戰表面中立之預測控制背後存在「結構性偏誤」的任務，並不在「禁止差別影響歧視」原則當前的守備範圍；而在長期遭受歧視的傳統「可疑分類」群體之外，防止相關性資訊不斷產生具破壞社會團結可能性的「新」分類，也超越現行憲法平等保障原則的射程。在資訊驅動之治理理性浪潮下，相關性發掘型人工智慧系統已開啟難以透過法治原則的「合理性」加以節制的控制慾望。與其說機器學習的相關性發掘能力滿足了既有的社會控制需求，不如說，相關性發掘能力不斷開創並餵養了無盡的控制慾望。事實上，當法治理念的諸原則原是用以節制國家權力的不合理施展與運作時，能以各種差異製造事實證成社會控制之合理性的「相關性發掘型」人工智慧系統，顯然已非當前法治的諸原則所能有效駕馭。

98 關於「結構性偏誤」之概念，請參見Schwartz et al., *supra* note 87, at 5-7.

99 「反巴爾幹化」是憲法追求平等保障的一種可能理由，請參見Reva B. Siegel, *From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 YALE L.J. 1278, 1330-46 (2011).

（三）秘密立法的合法性與民主正當性爭議

機器學習的相關性發掘能力經由轉化所創造出的差異製造事實，被應用在個案中作為「決策行動理由」時，依此而打造的人工智慧系統即可利用差異製造事實所刻畫的標準於個案中據以自動化地認定「裁決事實」，例如在前述貸款、審前羈押、假釋等程序中作為自動化認定個案還款可能性、再犯可能性等的標準。然而新的差異製造事實一旦成為可普遍化的決策行動理由、成為個案裁決事實的認定標準，就形同在三段論的法律適用過程中被悄悄地安置成為額外的規範大前提（「凡未曾購買保護貼者不予貸款」、「凡具有A_n特徵者不予假釋」），也相當於實質上的規則制定（rule-making）或法制定（law-making）行為¹⁰⁰。此外，機器學習在相關性發掘過程中，也必須在樣本範圍、界定損失函數等後設科學標準（metascientific standards）的價值問題上做出決定¹⁰¹。凡此都使「相關性的發掘」與衍生的「新差異製造事實的創造」，成為法治社會中原本應該透過民主程序作成或至少應具備民主正當性的準立法活動。少了民主程序的控管與民主正當性的加持，相關性發掘型人工智慧系統以「依演算法而治」（rule of algorithms）取代「依法而治」¹⁰²，將違背法治原則要求國家權力應按照透明可問責之民主立法程序制定之法律以行治理的合法性原則¹⁰³。

¹⁰⁰ Suksi, *supra* note 2, at 71.

¹⁰¹ Yoan Hermstrüwer, *Artificial Intelligence and Administrative Decisions under Uncertainty*, in REGULATING ARTIFICIAL INTELLIGENCE 199, 207 (Thomas Wischmeyer & Timo Rademacher eds., 2020).

¹⁰² See, e.g., Meneceur, *supra* note 36, at 139-42; 陳弘儒，依法而治或依演算法而治？，收於：張麗卿編，人工智慧與法律衝擊，頁261-268（2020年）。

¹⁰³ See, e.g., Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 45 (2019). 其他關於人工智慧對民主制度之挑戰的一般性討論請參見，MARK COECKELBERGH, THE POLITICAL PHILOSOPHY OF AI: AN INTRODUCTION 62-92 (2022).

三、小結

相關性發掘型人工智慧系統之共通特徵係藉由機器學習所發現的相關性資訊，經轉化而「創造」出採取特定預測性控制行動即可提高達成特定目的或結果之可能性的「差異製造事實」，從而成為預測性控制的決策行動的「理由」，並由人工智慧系統據以做成自動化決策。然而相關性發掘型人工智慧系統並無法比照模仿型人工智慧系統，針對如何從相關性演算出決策行動提出三段論式的得心證理由，以履行其「說理義務」，因而欠缺「外部可解釋性」。相關性發掘型人工智慧系統欠缺外部可解釋性的問題，雖可藉由有條件保護秘密的公共驗證，克服因隱匿系統開發資訊以致「無法驗證決策行動理由是否為真」的困境，但卻無法在要求「全面公開決策理由」的情況下，同時維持相關性發掘型人工智慧系統所欲達成的「預測控制目的」。只有在相關性所創造的差異製造事實係以「無法為個人於事後輕易操弄」的因素（例如種族、膚色、性別等）作為相關性之「剖繪」條件時，似乎才可能在滿足外部可解釋性的要求下，發揮相關性發掘型人工智慧系統的預測控制功能。除此之外，「必須以較高證明度所確認之事實作為個人責任基礎」的案件必須滿足行為人預見可能性與結果避免可能性等額外的法治原則，也不適合以相關性之「機率」作為決策行動理由。但立足於現在向未來達成特定的預警性管制、保全等控制與介入目的之「預測性決策」，則似乎可合理地利用以機率為基礎的差異製造事實，創造各種超前部署與控制的行動理由。此一具備合理性的能力無疑餵養了難以為法治原則所節制的無盡控制慾望。最後，藉由相關性所創造的差異製造事實在個案中據以自動化認定「裁決事實」的應用方式，雖滿足合理性的要求，但卻存在秘密立法的問題，而與法治的合法性及民主正當性存在緊張關係（參見表1）。

表1 相關性發掘型人工智慧系統與法治原則的相容性

	與法治相容的應用模式	與法治不相容的應用模式
外部可解釋性	以公共驗證方式，解決「決策行動理由是否為真」的驗證問題	隱蔽系統開發資訊使「決策行動理由是否為真」欠缺驗證可能
	相關性之「剖繪」使用「無法為個人於事後輕易操弄」的因素進行刻畫，可在「揭露決策行動理由」的前提下，仍維持系統的「預測控制目的」	為避免博弈風險致無法達成預測控制目的，而策略性隱蔽決策行動理由，導致欠缺外部可解釋性
以機率為基礎的決策行動合理性	立足於現在向未來達成特定的預警性管制、保全等控制與介入目的之「預測性決策」： 個案裁決事實認定 創造新立法理由 剖繪的構成要件化	利用機率資訊做成原本須以較高證明度所確認之事實作為個人責任基礎的決策而欠缺合理性
平等原則	表面中立、結果正確且與目的存在合理關聯的預測手段	預測手段欠缺正確性（形成恣意） 預測直接使用可疑分類（違法差別對待） 表面中立之預測形成差別影響但手段目的欠缺合理關聯（違法差別影響）
通案立法的合法性與民主正當性	利用相關性創造差異製造事實的過程及後設科學標準的決定，藉由民主參與強化依機器而治的合法性與正當性	個案裁決事實認定過程隱形的規則制定或法制定行為
總體評價	表面上與法治相容，卻可能餵養無法為法治所駕馭的無盡控制慾望	明顯與法治不相容而應予禁止

資料來源：作者製表。

伍、結論

機器學習型人工智慧帶來由資料驅動決策的數位化治理革命。本文以全自動化決策在公部門的應用為課題，探討此一席捲全球的趨勢對法治理念的影響與衝擊。本文認為，目前為止探究此等問題的許多努力，因為誤解機器學習型人工智慧系統的真正能力（與侷限），而錯置了批判的重點。本文認為要在人工智慧懷疑論與福音派間，找到正確評價其潛力與真正威脅的方法，必須從區分「模仿型」與「相關性發掘型」人工智慧系統二者開始。唯有如此，方能在理解不同人工智慧系統的不同限制下，找到可與法治相容的人工智慧應用模式，並遠離可破壞民主基本價值的真正威脅。

本文認為，模仿型法律人工智慧系統不僅因其所具有的默會知識學習能力，並非不可能超越目的解釋屏障，而在法律適用的內部證立上模仿法律人的目的解釋與其他涵攝推論，並且只要權限移轉獲法律授權、同等履行獲得擔保，並保留由人類進行最終救濟決定的可能性，以檢視是否存在系統性偏誤、是否需打破成規時，就能有條件地與法治原則相容。

相對地，相關性發掘型人工智慧系統則是藉由機器學習所發現的相關性，經轉化而「創造」出採取特定預測性控制行動可提高達成特定目的或結果之可能性的「差異製造事實」，從而成為預測性控制的決策行動的「理由」，並由人工智慧系統據以做成自動化決策。運用相關性發掘型人工智慧系統進行自動化決策時，若因系統開發者或部署者隱蔽系統開發資訊而導致「決策行動理由是否為真」欠缺驗證可能，或為了避免博弈風險致無法達成預測控制目的而策略性隱蔽「決策行動理由」時，都無法滿足提供說理的「外部可解釋性」要求。此外，若運用相關性發掘型人工智慧系統做成原本須以較高證明度所確認之事實作為個人責任基礎的決策，則因為無法滿足行為人預見可能性與結果避免可能性等額外的法治要求，

而欠缺合理性。至於運用機器學習所發掘的隱藏相關性資訊，為現行法規中既存的預測性決策創造個案中可普遍化的「決策行動理由」，則形同迴避民主程序的秘密立法，違反法治的合法性及民主正當性要求。該等運用相關性發掘型人工智慧系統的方式既與法治原則無法相容，則應予以禁止。

另一方面，若能以公共驗證方式，解決「決策行動理由是否為真」的驗證問題，同時僅使用「無法為個人於事後輕易操弄」的因素進行相關性「剖繪」之刻畫，即可在「揭露決策行動理由」的前提下，維持相關性發掘型人工智慧系統所欲達成的「預測控制目的」。而運用機器學習發掘的機率資訊做成各種「預測性決策」，則符合立足於現在向未來達成特定預警性管制、保全等控制與介入之目的，而具有法治原則所要求的合理性。然而，真正的挑戰在於，在資料驅動之治理理性浪潮下，相關性發掘型人工智慧系統將逐步餵養難以透過法治加以節制的「合理」控制慾望，因而對民主體制帶來最深刻的威脅。

參考文獻

1. 中文部分

- 王一奇（2015），理由與提供理由的事實，收於：謝世民編，理由轉向：規範性之哲學研究，頁105-140，臺北：臺大出版中心。
- 王鵬翔（2015），規則的規範性，收於：謝世民編，理由轉向：規範性之哲學研究，頁325-356，臺北：臺大出版中心。
- 王鵬翔、張永健（2015），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），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。
- （2020），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，收於：李建良編，法律思維與制度的智慧轉型，頁135-166，臺北：元照。
- （2022），亦步亦趨的模仿還是超前部署的控制？——AI的兩種能力和它們帶來的挑戰，收於：李建良、林文源編，人文社會的跨領域AI探索，頁285-299，新竹：國立清華大學出版社。
- 張乃文（2020），基於統計準則式方法偵測生醫文獻中的生物關聯，國立臺灣大學生醫電子與資訊學研究所博士論文。
- 陳弘儒（2020），依法而治或依演算法而治？，收於：張麗卿編，人工智慧與法律衝擊，頁261-268，臺北：元照。
- （2020），初探目的解釋在法律人工智慧系統之運用可能，收於：李建良編，法律思維與制度的智慧轉型，頁225-299，臺北：元照。

- (2022), 法律人工智慧的法哲學反省：判斷權限、執法機制與法治理念，歐美研究，52卷2期，頁175-245。
- 許恒達 (2020)，食品安全衛生管理法刑事責任之檢討——以強冠劣油案為中心，收於：倪貴榮編，食品安全與司法救濟，頁149-206，臺北：元照。
- 湯德宗 (2022)，大法官平等權解釋會通釋義，收於：司法院行政訴訟及懲戒廳編，行政訴訟新制二十週年研討會會議實錄，頁63-248，臺北：司法院。
- 黃居正 (2018)，與人工智慧相關的國際法議題——從國際人道法到生命體法，收於：劉靜怡編，人工智慧相關法律議題芻議，頁215-241，臺北：元照。
- 黃昭元 (2012)，論差別影響歧視與差別對待歧視的關係——評美國最高法院 *Ricci v. DeStefano* (2009) 判決，中研院法學期刊，11期，頁1-63。
- (2017)，從平等理論的演進檢討實質平等觀在憲法適用上的難題，收於：李建良編，憲法解釋之理論與實務（九），頁271-312，臺北：中央研究院法律學研究所。
- 黃詩淳 (2023)，AI可解釋性的法學意義及其實踐，臺大法學論叢，52卷特刊，頁931-972。
- 謝碩駿 (2023)，論全自動作成之行政處分，收於：黃丞儀編，2017行政管制與行政爭訟：行政程序法2.0，頁213-298，臺北：中央研究院法律學研究所。

2. 外文部分

- Appel, Steven M., and Cary Coglianese. 2020. Algorithmic Governance and Administrative Law. Pp. 162-181 in *The Cambridge Handbook of the Law of Algorithms*, edited by Woodrow Barfield. Cambridge: Cambridge University Press.

- Babic, Boris, and I. Glenn Cohen. 2023. The Algorithmic Explainability “Bait and Switch”. *Minnesota Law Review* 108:857-909.
- Bambauer, Jane, and Tal Zarsky. 2018. The Algorithm Game. *Notre Dame Law Review* 94:1-48.
- Barocas, Solon, and Andrew D. Selbst. 2016. Big Data’s Disparate Impact. *California Law Review* 104:671-732.
- Brennan-Marquez, Kiel. 2017. “Plausible Cause”: Explanatory Standards in the Age of Powerful Machines. *Vanderbilt Law Review* 70:1249-1301.
- Brown, Shea, Jovana Davidovic, and Ali Hasan. 2021. The Algorithm Audit: Scoring the Algorithms That Score Us. *Big Data and Society* 8(1):1-8.
- Brown, Tom B., Mann Benjamin, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. Unpublished manuscript. On file with Cornell University, arXiv. Available at <https://arxiv.org/pdf/2005.14165.pdf>.
- Choi, Jonathan H., Kristin E. Hickman, Amy B. Monahan, and Daniel Schwarcz. 2022. ChatGPT Goes to Law School. *Journal of Legal Education* 71:387-400.
- Coeckelbergh, Mark. 2022. *The Political Philosophy of AI: An Introduction*. Cambridge: Polity Press.
- Coglianese, Cary, and David Lehr. 2019. Transparency and Algorithmic

- Governance. *Administrative Law Review* 71:1-56.
- Collins, Harry M. 1990. *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: The MIT Press.
- Davis, Kenneth Culp. 1942. An Approach to Problems of Evidence in the Administrative Process. *Harvard Law Review* 55:364-425.
- Denvir, Catrina, Tristan Fletcher, Jonathan Hay, and Pascoe Pleasence. 2019. The Devil in the Detail: Mitigating the Constitutional & Rule of Law Risks Associated with the Use of Artificial Intelligence in the Legal Domain. *Florida State University Law Review* 47:29-98.
- Dreyfus, Hubert L. 1992. Response to Collins, Artificial Experts. *Social Studies of Science* 22:717-726.
- Dworkin, Ronald. 1986. *Law's Empire*. Cambridge, MA: Harvard University Press.
- Engstrom, David Freeman, and Daniel E. Ho. 2020. Algorithmic Accountability in the Administrative State. *Yale Journal on Regulation* 37:800-854.
- Engstrom, David Freeman, Daniel E. Ho, Catherine M. Sharkey, and Mariano-Florentino Cuéllar. 2020. Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. NYU School of Law, Public Law Research Paper No. 20-54. Available at <http://dx.doi.org/10.2139/ssrn.3551505>.
- Enqvist, Lena, and Markus Naarttijärvi. 2023. Discretion, Automation, and Proportionality. Pp. 147-177 in *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*, edited by Markku Suksi. Cham: Springer.
- Fagan, Frank. 2022. Law's Computational Paradox. *Virginia Journal of Law and Technology* 26:1-72.
- Fallon, Richard H. Jr. 1997. "The Rule of Law" as a Concept in Constitutional Discourse. *Columbia Law Review* 97:1-56.

- Goodman, Ellen P., and Julia Trehu. 2023. Algorithmic Auditing: Chasing AI Accountability. *Santa Clara High Technology Law Journal* 39:289-338.
- Hamilton, Melissa. 2019. The Biased Algorithm: Evidence of Disparate Impact on Hispanics. *American Criminal Law Review* 56:1553-1577.
- Hellman, Deborah. 2020. Measuring Algorithmic Fairness. *Virginia Law Review* 106:811-866.
- Hermstrüwer, Yoan. 2020. Artificial Intelligence and Administrative Decisions under Uncertainty. Pp. 199-223 in *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher. Cham: Springer.
- Hildebrandt, Mireille. 2018. Algorithmic Regulation and the Rule of Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376(2128):20170355:1-11.
- Johns, Fleur. 2021. Governance by Data. *Annual Review of Law and Social Science* 17:53-71.
- Kaminski, Margot E. 2019. The Right to Explanation, Explained. *Berkeley Technology Law Journal* 34:189-218.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2018. Discrimination in the Age of Algorithms. *Journal of Legal Analysis* 10:113-174.
- Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable Algorithms. *University of Pennsylvania Law Review* 165:633-705.
- Liu, Han-Wei, Ching-Fu Lin, and Yu-Jie Chen. 2019. Beyond *State v Loomis*: Artificial Intelligence, Government Algorithmization and Accountability. *International Journal of Law and Information Technology* 27:122-141.
- Liu, Pengfei, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi,

- and Graham Neubig. 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys* 55(9):1-35.
- Meierhenrich, Jens. 2021. *Rechtsstaat* versus the Rule of Law. Pp. 39-67 in *The Cambridge Companion to the Rule of Law*, edited by Jens Meierhenrich and Martin Loughlin. Cambridge: Cambridge University Press.
- Meneceur, Yannick. 2023. Artificial Intelligence, Public Administration, and the Rule of Law. Pp. 117-145 in *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*, edited by Markku Suksi. Cham: Springer.
- Naarttijärvi, Markus. 2023. Situating the Rule of Law in the Context of Automated Decision-Making. Pp. 15-31 in *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*, edited by Markku Suksi. Cham: Springer.
- Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- . 2019. A Rule of Persons, Not Machines: The Limits of Legal Automation. *The George Washington Law Review* 87:1-55.
- Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books.
- Pöysti, Tuomas. 2023. Legislating for Legal Certainty, with a Right to a Human Face, in an Automated Public Administration. Pp. 33-63 in *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*, edited by Markku Suksi. Cham: Springer.
- Price, W. Nicholson II, and Arti K. Rai. 2021. Clearing Opacity Through Machine Learning. *Iowa Law Review* 106:775-812.

- Radavoi, Ciprian N. 2020. The Impact of Artificial Intelligence on Freedom, Rationality, Rule of Law and Democracy: Should We Not Be Debating It?. *Texas Journal on Civil Liberties and Civil Rights* 25:105-127.
- Sahoo, Pranab, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. Unpublished manuscript. On file with Cornell University, arXiv. Available at <https://doi.org/10.48550/arXiv.2402.07927>.
- Schwartz, Reva, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, and Patrick Hall. 2022. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. NIST Special Publication 1270. Available at <https://doi.org/10.6028/NIST.SP.1270>.
- Selbst, Andrew D., and Julia Powles. 2017. Meaningful Information and the Right to Explanation. *International Data Privacy Law* 7:233-242.
- Selmi, Michael. 2021. Algorithms, Discrimination and the Law. *Ohio State Law Journal* 82:611-651.
- Siegel, Reva B. 2011. From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases. *Yale Law Journal* 120:1278-1366.
- Suksi, Markku. 2023. Formal, Procedural, and Material Requirements of the Rule of Law in the Context of Automated Decision-Making. Pp. 65-95 in *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*, edited by Markku Suksi. Cham: Springer.
- Tamanaha, Brian Z. 2009. A Concise Guide to the Rule of Law. Pp. 3-16 in *Relocating the Rule of Law*, edited by Gianluigi Palombella and Neil Walker. London: Bloomsbury Publishing.

- Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law and Technology* 31:841-887.
- Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law* 7:76-99.
- Zalnieriute, Monika, Lyria Bennett Moses, and George Williams. 2019. The Rule of Law and Automation of Government Decision-Making. *The Modern Law Review* 82:425-455.

Algorithmic Governmentality and Its Challenges to Rule of Law:

A Study of “Imitative” and “Correlation-Discovery”
AI Systems

*Wen-Tsong Chiou**

Abstract

This article explores the issues of the application of fully automated decision-making in the public sector, with a particular focus on the data-driven decision-making revolution brought about by machine learning-based artificial intelligence (AI) and its impact on the principles of the rule of law. The article argues that understanding the potential and threats of AI requires distinguishing between “imitation-based” and “correlation discovery-based” AI systems. This distinction helps to recognize the limitations of different AI systems and to identify applications compatible with the rule of law, thereby avoiding genuine threats to fundamental democratic values. The article asserts that imitation-based machine learning systems, after sufficient training, can overcome the barriers of teleological interpretation in legal application and possess the ability to mimic various forms of legal reasoning. The use of such imitation-based AI in fully automated legal decision-making may be conditionally compatible with rule of law principles. On the other hand, correlation discovery-based AI, by uncovering hidden patterns and continuously generating new reasons for intervention, could lead to an expansion of power in the public sector that exceeds the control afforded by current rule of law principles.

* Research Professor, Institutum Iurisprudentiae, Academia Sinica.

KEYWORDS: fully automated decision-making, artificial intelligence, rule of law, teleological interpretation, explainability, correlation, causation, profiling, difference-making facts, reasons for action.