



國立臺灣大學社會科學院政治學系

博士論文

Department of Political Science

College of Social Sciences

National Taiwan University

Doctoral Dissertation

公共治理與人工智慧的交引纏繞—

行動者網絡理論分析途徑

The Entanglement of Public Governance and

Artificial Intelligence :

An Actor-Network Theory Approach

丁玉珍

Yu-Jen Ting

指導教授：林子倫 博士

Advisor: Tze-Luen Lin, Ph.D.

中華民國 113 年 1 月

January 2024

國立臺灣大學博士學位論文
口試委員會審定書

公共治理與人工智慧的交引纏繞—
行動者網絡理論的分析途徑

**The Entanglement of Public Governance and Artificial
Intelligence: An Actor-Network Theory Approach**

本論文係 丁玉珍 君（學號：D06322003）在國立臺灣大學政治
學系完成之博士學位論文，於民國 112 年 11 月 17 日承下列考試
委員審查通過及口試及格，特此證明
口試委員：

丁玉珍

（簽名）

（指導教授）

張文配

葉俊榮

陳致遠

黃心怡

謝辭



衷心感謝子倫老師的悉心指導與師生情誼，子倫老師的鼓勵，我才有勇氣展開學術研究的旅程，開啟新視野，享受學術研究的樂趣。感謝行政院人事行政總處蘇俊榮人事長、財政部財政資訊中心張文熙主任、國立政治大學公共行政學系陳敦源教授、荷蘭萊頓大學政府治理與全球事務學院公共行政系黃心怡教授協助指導，給我諸多實質幫助及啟發，順利完成學位。感謝勞動部許銘春部長的支持。

感謝父親丁聞琴，我完成了您的期望。感謝我親愛的家人，媽媽林秀琴，坤木、奕如、奕心的支持。感謝論文寫作過程中太多人的協助，無盡的感激。

中文摘要



人工智慧(AI)將全面性重塑當代社會，包括政府機關、公共治理理論。然而，基於對 AI 風險的疑慮，公部門使用 AI 的執行速度和案例成熟度不足。

公部門使用 AI 必須比私部門更加謹慎。然而，現有公共治理理論無法有效解釋公部門使用 AI 的現象；雖然新興 AI 治理研究社群提出避免 AI 風險的理論、框架、策略和指南，除了太抽象、也與公共治理理論幾乎無關；臺灣亦缺乏公部門使用 AI 的實證案例研究。有鑑於 AI 對政府機關的影響將相當巨大，公部門應有理論及適切的參考指引，使公部門能負責任地使用 AI。

本研究使用行動者網絡理論(ANT)的分析途徑，實證研究臺灣公部門使用 AI 的 5 個案例。以往文獻有關公共治理領域的 AI 理論研究，集中於探索性的概念框架，理論化不足，也缺少證據的支持。本研究則以循證為基礎(evidence-based)，推演公部門使用 AI 的公共治理理論。研究個案包括「財政部智能稅務服務計畫」、「環境部智慧判煙系統」、「臺北市政府 AI 智慧交控管理」、「勞動部 1955 多元智能電服中心」、「勞動部 AI 人機協作分文」。另為使本研究所提出的理論具有強固性，再利用文獻分析法分析國外公部門使用 AI 的 7 個案例佐證本研究的推論成果。

本研究範圍聚焦於政府是使用者，研究方法使用多重研究方法，包括文獻分析法、個案研究法，以 ANT 為分析途徑。ANT 不同於公共治理理論以人為中心的方法論，主張「廣義對稱性原則」、「物體也有能動性 (agency)」、「交纏的社會世界觀」；該理論的重要性是我們可以把非人類(例如 AI)當成分析單位，能有效詮釋公部門使用 AI 的現象。接續利用本研究梳理文獻後所提出公部門使用 AI 的挑戰：1.公共行政改革；2.AI 官僚的合法性；3.AI 系統的品質(資料治理與演算法治理)；4.在公共治理領域中人機協作的管理；5.勞動力影響(勞

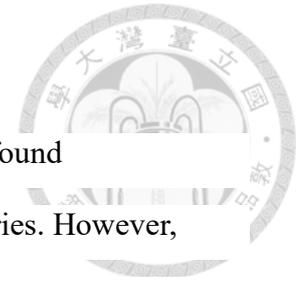
動力替代與轉型)。並從本實證研究成果歸納公部門的 AI 風險緩解機制。

AI 治理爭議與公共治理理論喪失解釋力，主因是以人為本、以人為分析單位的治理典範，長期忽視技術。本研究以此為始，藉由臺灣、國外合計 12 個案例的研究成果提出解決之道。公共治理理論層次，提出典範轉移路徑，包括：1. AI 應該是分析單位；2. AI 鑲嵌的公共政策過程可能呈現循環交錯的現象；3. AI 能動性(agency)透過人機協作對人類官僚的認知能力產生影響；4. 「AI 官僚」也能行使行政裁量權。在公共治理分析層次，當公部門使用 AI 時，ANT 分析途徑顯現出優異的公共政策分析能力。本研究補充了以人為中心的公共治理實踐層次的新觀點，包括：1. 策略性運用 AI 風險緩解機制，提升公共治理效能及控制風險；2. 強化文官人力資源的策略：AI 賦能的新官僚；3. 緩解 AI 治理爭議的方法：以循證為基礎。藉此強化公共治理理論的解釋力，並同時擴展至 AI 應用。

本研究主要貢獻是將研究進程從探索性研究往前推進至修正公共治理理論層次，包括：1. 將公部門使用 AI 的決策及行為理論化，提出公共治理理論的分析單位應包括 AI，並關注 AI 是行動者，有自適應性和能動性；2. 提出「AI 風險緩解機制」的意涵與具體實踐做法；3. 提出修正人機協作的理論，包括認知層面的人機協作。克服文獻對新興 AI 治理框架、策略和指南無法具體實踐的批判。

關鍵詞：人工智慧、公共治理、行動者網絡理論(ANT)、AI 能動性、AI 自適應性、AI 風險緩解機制、人機協作

Abstract



Artificial intelligence (AI) is anticipated to bring about a profound transformation, impacting government and public governance theories. However, concerns about AI risks necessitate a more cautious deployment in the public sector compared to the private sector. The public sector faces challenges not only in the speed of AI implementation but also in the maturity of application. Additionally, existing public governance theories struggle to incorporate AI effectively. While emerging AI governance research communities propose guidelines to mitigate risks, these guidelines often remain overly abstract and lack direct relevance to established public governance theories. Additionally, Taiwan lacks empirical case studies on the use of AI in the public sector. Recognizing the significant impact of AI, there is a pressing need for both theoretical and practical guidelines to facilitate the responsible use of AI in the public sector.

This study employs the Actor-Network Theory (ANT) approach to empirically investigate five cases of AI utilization within the Taiwanese public sector. Previous literature on AI theories in the field of public governance has predominantly focused on exploratory conceptual frameworks, lacking theoretical depth and evidence-based support. However, this study is evidence-based and deduces theories of public governance regarding the use of AI within the public sector. The researched cases include the Ministry of Finance's "Smart Tax Services project", Ministry of Environment's the "Intelligent Smoke Plume Identification System", the Taipei City Government's "AI Smart Traffic Control Management", the "1955 intelligent online customer services" implemented by the Ministry of Labor, and the "AI e-document distribution system (AI-eDDS)" also by the Ministry of Labor. Furthermore, to

strengthen the robustness of the theories proposed in this study, a literature analysis method is used to analyze seven cases of AI utilization in foreign public sectors, corroborating the conclusions drawn in this research.

The research focuses on the government as a user and employs literature analysis, case study, and ANT approach. The study argues that the controversy surrounding AI governance and the diminishing explanatory power of public governance theory are primarily attributed to its longstanding anthropocentric paradigm, which neglects technology. ANT, which treats nonhumans (such as AI artifacts) as the analytical unit and advocates for “the strong program principle of symmetry”, posits that “objects too have agency” and views “the social world as an entanglement of interactions”, differing from the anthropocentric methodology of public governance theory. Therefore, ANT can effectively interpret the phenomenon of the public sector using AI by considering AI artifacts as the analytical unit. Following the literature review conducted in this study, the challenges identified in the use of AI within the public sector are as follows: (1). public administrative reform; (2). legitimacy of AI in bureaucracy; (3). quality of AI-based systems (data governance and algorithm governance); (4). management of human-AI collaboration in public governance; and (5). the impact of AI on labor (labor replacement and transformation). The study then summarizes the risk mitigation mechanisms for AI in the public sector derived from the findings of this empirical study.

This study aims to propose solutions based on the research findings from a total of 12 cases in Taiwan and abroad, thereby suggesting a paradigm shift in public governance theory: (1). AI should be considered as an analytical unit within public governance; (2). The public policy process of AI embedding may exhibit cyclical



interaction phenomena; (3). AI agency will affect the cognitive abilities of human bureaucrats through human-AI collaboration ; and (4). Acknowledging that 'AI bureaucrats' can exercise administrative discretion. In terms of public governance analysis levels, when the public sector utilizes AI, the ANT approach demonstrates outstanding capabilities in analyzing public policy. This study introduces novel perspectives to the domain of human-centric practices in public governance: (1). Strategic application of AI risk mitigation mechanisms can enhance public governance effectiveness and control risks; (2). The AI-empowered bureaucracy can serve as a strategy to strengthen civil service resources; (3). An evidence-based approach can effectively mitigate AI governance disputes. Through these measures, the explanatory power of public governance theory is reinforced, concurrently extending its application to AI scenarios.

The primary contribution of this study is the refinement of public governance theory, covering the following key aspects: (1). Theorizing the decision-making and behavior of the public sector in using AI, the analytical unit of public governance theory should encompass AI, emphasizing AI as an actor with adaptive and agency; (2). Suggesting the definition and practical approaches to the AI risk mitigation mechanism for the public sector; (3). Proposing a theory to modify human-AI collaboration, incorporating cognitive aspects. The study presents the potential for implementing a tangible AI governance framework in the public sector.

Keywords: Artificial Intelligence, Public Governance, Actor-Network Theory, AI Agency, AI Adaptive, AI Risk Mitigation Mechanism, Human-AI Collaboration

目次



第一章、前言.....	1
第一節 研究動機.....	5
第二節 研究問題與目的.....	17
第三節 研究限制.....	22
第二章、AI 與公共治理的對話.....	25
第一節 AI 的定義與範圍概述.....	25
第二節 公共治理理論.....	55
第三節 公部門應用 AI 的潛力與挑戰.....	76
第四節 公共治理導入 AI 的研究現況.....	108
第三章、AI 與公共治理的研究設計.....	120
第一節 方法論—ANT.....	121
第二節 研究架構.....	134
第三節 研究方法.....	146
第四章、臺灣公部門應用 AI 的個案分析.....	153
第一節 財政部「智能稅務服務計畫」.....	153
第二節 環境部「智慧判煙系統」.....	185
第三節 臺北市政府「AI 智慧交控管理」.....	208
第四節 勞動部「1955 多元智能電服中心」、「AI 人機協作分文」.....	225
第五章、AI 與公共治理的交引纏繞.....	241
第一節 AI 與公部門行動者的新關係.....	241
第二節 公共治理理論的典範轉移路徑.....	253
第三節 與文獻對話：公部門的 AI 挑戰.....	261
第四節 公部門 AI 風險緩解機制.....	270
第六章、結論.....	297
第一節 研究發現.....	298
第二節 研究貢獻.....	318
第三節 政策建議.....	324
第四節 未來研究方向.....	338
參考文獻.....	341



圖次

圖 1-1 研究概念.....	4
圖 2- 1 AI 風險的管制強度.....	31
圖 2- 2 AI 與公部門行動者的關係.....	119
圖 3- 1 人機協作的理論.....	135
圖 3- 2 AI 的公共治理研究架構圖.....	137
圖 3- 3 案例選擇架構.....	142
圖 3- 4 研究案例.....	145
圖 3- 5 AI 和公共治理理論分析流程.....	151
圖 3- 6 AI 與公部門行動者互動分析流程.....	152
圖 5- 1 人機協作魔力象限.....	252
圖 5- 2 公部門 AI 風險緩解機制概念圖.....	273
圖 6- 1 公部門應用 AI 的決策及行為理論化概念圖.....	320



表次

表 2- 1 Russell & Norvig 對 AI 的分類	27
表 3- 1 ANT 的分析框架與敘事重點	131
表 3- 2 AI 的公共治理研究內容	138
表 5- 1 AI 與人的互動與感知彙整表	244
表 5- 2 智能稅務服務計畫的 AI 挑戰	262
表 5- 3 智慧判煙系統的 AI 挑戰	264
表 5- 4 AI 智慧交控管理的 AI 挑戰	265
表 5- 5 1955 多元智能電服中心的 AI 挑戰	267
表 5- 6 AI 人機協作分文的 AI 挑戰	268
表 5- 7 公共行政方法之 AI 風險緩解機制彙整表	275
表 5- 8 臺灣公部門導入 AI 案例	282
表 5- 9 資訊科技風險管理方法之 AI 風險緩解機制彙整表	293
表 6- 1 人類官僚和 AI 官僚的管理方式對照表	308
表 6- 2 公共治理理論缺口與補充觀點	311
表 6- 3 公部門可信任 AI 參考指引(草案).....	333



第一章、前言

AI 在特定領域的表現優於人類，且 AI 的許多重要用途涉及增強人類，而不是取代人類(Malone et al., 2023)，成為解決政府人力不足，提升公共政策決策品質、公民服務品質、政府績效等公共治理問題的重要手段之一。然而，AI 的自動決策(automated decision)或自主決策(autonomous decision)¹，以及認知²技術(Cognitive technology)可能對人類的認知權利(Epistemic Rights)(Kitchner, 1983; Watson, 2018)和後續的具體作為產生影響，使以往**公共治理理論「以人為中心」的治理典範，主要關注人的能動性(human agency)**，以及在公共行政研究中，技術僅從功能或工具的角度來促進或限制人類行為(Nisar & Masood, 2018)，**可能無法解釋 AI 主動性(或自適應性)及能動性所產生的現象**。舉例來說，傳統官僚理論中基層官僚具有自由裁量權和相當程度自主權，AI 自動化決策影響基層官僚自由裁量權(de Boer & Raaphorst, 2021)，甚至變成實質上的決策者；又例如透過演算法的搜索引擎排名功能，操縱公民的認知和行為，進而影響選舉結果，使演算法被認定為具有「政府權力(governmental power)」和「把關功能(gatekeeping function)」的新型態參與者(Napoli, 2014)，甚至是強大的執行者。

此外，有關因果關係推理問題(causal inference problem)，以往公共治理理論強調因果關係的成立，必須包括相關性、時間序列正確性、邏輯合理性等，需要領域知識的理論、嚴格統計技術的政策分析、人類判斷等，以精確理解潛在的因

¹ 自動決策(automated decision)是從已知選項評估最佳者；自主決策(autonomous decision)是由系統自主決定。

² 「認知」是我們獲取知識、理解世界並解決問題的過程；人類一直在努力設法推展自己的認知邊界，尤其是借助科技發展的協助(蘇經天, 2023)。



果機制。儘管 AI 使用訓練數據，利用「關聯學習(learn relationships)」和「機率推理(probabilistic reasoning)」等相關性(correlations)生成的預測模型，在多項任務上的表現已經並將繼續超越人類的能力，但**演算型決策(algorithmic decision making, ADM)**只能說明相關性、可能性，不代表因果關係³，使得 AI 在確定因果關係方面不如人類熟練(Jarrahi, Lutz, & Newlands, 2022)。

再者，AI 缺乏人類智慧的本質，無法理解人類的經歷，無法透過資料蒐集歸納社會成員共同生活的基礎價值，在人類的意義脈絡下，更深層度的網路和更大量的數據不太可能解開 AI 的「意義障礙(barrier of meaning)」(Mitchell, 2019)。演算法輸出所呈現的數據關係，對人類而言是不直觀(nonintuitive)，無法直接對應到一般人常識性理解。若 AI 取代官僚機構的裁量權，在理論或邏輯上即與現行公共治理理論產生差異(Obermeyer, Powers, Vogeli, & Mullainathan, 2019)。舉例來說，使用 AI 的「相關性」，但是做為與「公民權利高度相關」的裁決，將產生是基於發生可能性，而不是事實必然性的問題，顛覆公共治理理論長久以來的因果關係邏輯。

因此，以往公共政策模型所欲達成權衡不同人類價值觀，公共政策過程的次序性、可控制性，公共政策執行時官僚行使裁量權的獨立決策自由度的詮釋邏輯，可能無法回應當政府使用 AI 處理公共事務時所產生的現象，有學者指出 AI 科技對政府機關所帶來的影響，或許已達孔恩(Thomas Kuhn)所言之科學典範轉移(paradigm shifts)的關鍵時刻(黃心怡、陳敦源, 2023)，未來幾年內可能產生新的全球治理典範(Barcevičius et al., 2019; Mergel, Kattel, Lember, & McBride, 2018)。因此，公部門⁴應用 AI 亟需適當的理論或分析框架，以最大化 AI 的正面影響及降

³ 可能原因包括：(1).技術進程：目前是弱 AI，有侷限性。(2).機器學習所產生的偏誤(bias)。(3).資料偏誤(bias)：樣本不具代表性、資料偏見、極端值等。

⁴ 公部門(public sector)使用 AI 系統可能比私部門更多或更早，通常是軍事戰略考



低負面影響。

行動者網絡理論(Actor-Network Theory, ANT)提供一個新的視角，主張給予技術物廣義對稱性原則，強調非人類能動性的重要性，並將「非人類」行動者整合到我們對社會世界的理解，該理論的重要性是我們可以把非人類(例如 AI)當成分析單位，可以擺脫公共治理理論長久以來對人的專門分析；交引纏繞的社會世界觀，可以詮釋人與 AI 的動態互動關係、AI 能動性如何影響人類決策、如何形塑新的治理形式等複雜情況。

因此，本研究利用 ANT 提供 AI 技術物(AI artifacts)是分析單位的論述基礎，解決公共治理理論缺乏對技術關注的困境，並使用 ANT 的研究方法進行案例研究，探索公共治理理論典範轉移方向，期能提高公共治理理論的解釋力，研究概念詳如圖 1-1。

量。例如圖靈(Alan Mathison Turing)二次世界大戰期間負責德國密碼分析、911 後預防犯罪和反恐的資料探勘(Bignami, 2022)，物聯網最初也是為軍事用途而設計。再者，公部門比一般行政機關範圍大，為利以簡潔方式陳述，本研究所指公部門意指一般行政機關。

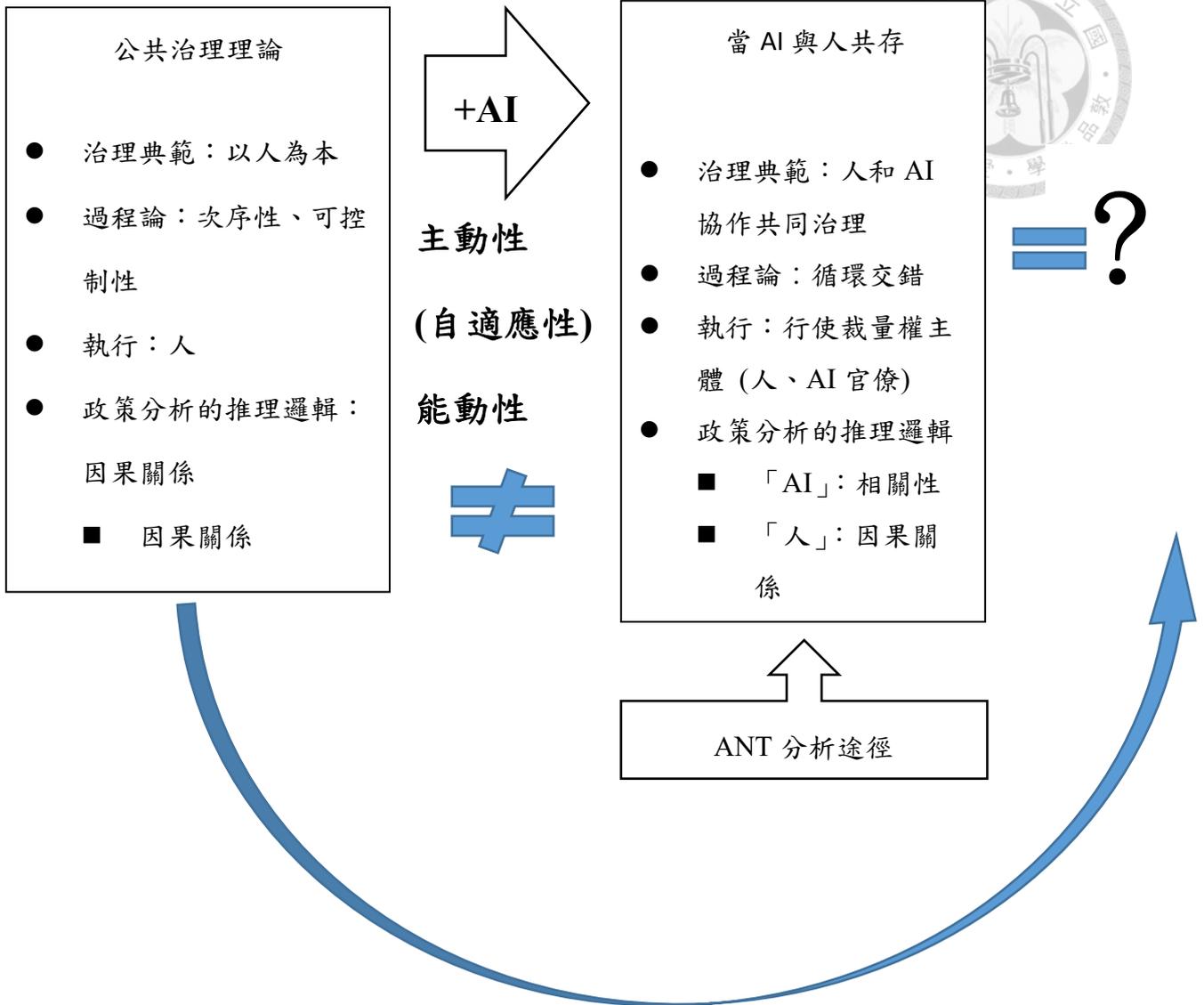


圖 1-1 研究概念

資料來源：本研究設計

第一節 研究動機



AI 是第四代工業革命的關鍵推力，其影響力更勝於以往的工業革命⁵和數位革命⁶，AI 能夠使人類認知增強(Cognitive Augmentation)和生理增強 (Physical Augmentation)，人類和 AI 共同努力可以提高績效(Burke et al., 2019; Malone et al., 2023)，AI 也被認為可以替代、補充和增強人類目前執行的所有任務(Casares, 2018; Makridakis, 2017; Sun & Medaglia, 2019)。因此，AI 所產生的作用和效果，預期將全面性重塑當代社會(Calo, 2017)；甚至有文獻指出，我們往往會高估一項技術的短期效果，而低估長期效果，AI 將推動社會乃至人類的演化(Pitt, Paschen, Kietzmann, Pitt, & Pala, 2023)，使得近幾十年來 AI 幾乎是學術界唯一研究的議題 (Jimenez-Gomez, Cano-Carrillo, & Lanas, 2020)。

因此，AI 對社會將產生重大影響已經不再是問題，當前辯論取而代之是：「將產生多大的正、負面影響，對於誰、什麼面向、什麼地方、什麼時間點？」(Floridi et al., 2018)。另一方面，當世界各地的組織和政府正在以快速增長的速度採用和實施 AI 系統(Merhi, 2023)，美國、中國、歐盟等先進數位國家的政府組織也已經啟動公部門採用 AI 的計劃(Engstrom, Ho, Sharkey, & Cuéllar, 2020; Misuraca

⁵ 機器提供替代，補充和增強人類完成人工工作(manual work)方式克服人類力量 (human force)的限制。

⁶ 機器提供替代、補充和增強人類執行日常智力工作(mental routine)方式，克服人類認知能力(human cognitive capacity)限制的「認知機器時代(Age of Cognitive Machines)」。

& van Noordt, 2020; Molinari F., 2021)^{7 8 9 10}。然而，公部門導入 AI 技術應用的執行速度和案例成熟度不足，且相關公共政策學術研究也不多，與政府大張旗鼓的導入策略形成強烈的對比。



壹、公共治理困境的成因

當今公共治理困境主要是「預算赤字」、「績效赤字」及「信任赤字」等(孫本初, 2003)，各先進工業化民主國家所推動的政府再造運動，也改變過去對「過程」(process)負責的心態，變成對「結果」(result)負責(陳敦源、簡鈺肆, 2018)。儘管各國政府都不斷努力提高公共治理效能，公共治理的發展概念與指標也不斷與時俱進，然而，時至今日，公共治理面臨的挑戰仍不斷增加，甚至更形嚴峻，不管是內部自我要求或外力要求，已非執政者所能掌握(陳金貴, 2013)。

我國亦有治理績效亟待提升的問題(蘇彩足, 2013)。依據 2016 年臺灣公共治理指標調查結果顯示，在五大面向：「法治化程度」、「政府效能」、「透明與課責」、「廉潔程度」、「公共參與程度」中，「政府效能」得分最低；而效能有 2 個方面，1.管理機構設置合理，管理程序科學化，管理活動彈性；2.最大限度地降低管理成本；並建議持續強化公共參與，藉以提升課責透明與政府效能(蘇彩足, 莊文忠, & 郭銘峰, 2017)。因此，提升政府效能成為提高公共治理績效當務之急，而在國家的作用和能力都是有限的情況下(蘇彩足 et al., 2017)，增強政府治

⁷ 檢自 <https://www.brookings.edu/blog/techtank/2022/09/22/understanding-artificial-intelligence-spending-by-the-u-s-federal-government>

⁸ <http://cpc.people.com.cn/BIG5/n1/2023/0714/c64387-40035514.html>

⁹ 檢自

<https://indsr.org.tw/respublicationcon?uid=12&resid=1950&pid=3939&typeid=3>

¹⁰ 檢自 <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>



理能力最重要的方法之一就是善用科技改變政府做事的方法。

貳、公共治理運用 AI 的潛力與挑戰

公部門導入 AI 對未來政府工作、決策、以及公共服務帶來決定性的變革(曾冠球、廖洲棚、黃心怡, 2019)，有潛力解決長久以來公共治理所面臨的挑戰。歐盟 2020 年 AI 白皮書即指出利用 AI 可以改善公民生活品質、降低公共服務及聯合國永續發展目標的成本(Nikolinakos, 2023)，未來新數據戰略是 AI 的開發和採用，並視為未來幾十年經濟成長關鍵技術(Misuraca & van Noordt, 2020)。

然而，AI 通用技術(General purpose technologies)¹¹、雙重用途(dual use)¹²(Floridi, 2018)、自適應性 (adaptive)、能動性(agency)或作用、政治性、偏差或偏誤(bias)等特性，可以根本上改變權力本質、結構、分配，並使權力差距擴大，甚至可能會出現新的強大參與者或行動者(Cath, 2018; Dafoe, 2018)，使 AI 的應用和挑戰是彼此密切相關且可能產生協同效應(synergy effects)。

參、公共治理理論假設與缺口

當前公共治理理論，包括公共政策模型、公共政策過程、倫理、官僚組織、行政裁量權等，「政策利害關係」¹³是公共政策理念的基礎，「政策利害關係」

¹¹ 經濟學家將通用技術描述為「經濟增長的引擎(Engines of Growth)」(Bresnahan & Trajtenberg, 1995)，適用於大量提高生產率的應用，與其他限制性技術相比，規模和持續時間更有影響力(Brundage, 2019)。

¹² 在社會科學、電力工程領域是指同一種工具有雙重用途的特質，如何使用受目的及對社會和個人的影響決定(Djeffal, 2020)，例如科技目的通常是為了生產或戰爭，AI 可以容易地用於促進公平或造成損害(Young, Bullock, & Lecy, 2019)。

¹³ 政策利害關係人的定義是受到公共政策影響，直接或間接、有形或無形影響公



的政策意涵是以「人」為主要討論對象。然而，機器人、計算機化和 AI 的快速發展，智人(intelligent human)和非人類行動者(non-human actors)共同組成的世界，人類與 AI 技術結合在一起使人類整體增強的「H+」¹⁴等等，已經超越公共治理理論以人為主的治理典範。舉例來說，AI 在決策過程中的權重過大，可能導致人類被邊緣化(蘇經天, 2023)，甚至 AI 在公共政策中成為實際的決策者，成為「AI 官僚」(Unver, 2018)；或是 AI 自適應性(Adaptive AI)可能超過人類可以預測和控制的範圍；AI 能動性(agency) (Mele, Spena, Kaartemo, & Marzullo, 2021; Thaler & Sunstein, 2008; Yeung, 2018)使 AI 技術能具體對人類行動者的行為產生影響，甚至對行為者的作為(actors' practices)具有很強的顛覆性潛力(Mele et al., 2021)，已超越我們今天人類中心世界觀，迫使我們不得不從本質上重新考慮我們如何理解智慧(intelligence)、人際關係或社會存在(Bloom, 2020)。

因此，若要從現有公共治理理論機制與假設出發，公共治理的研究必須重新思考這些理論的假設(Ruppert, Law, & Savage, 2013)、AI 的認識論、現有研究方法的規範和倫理假設(normative and ethical assumptions) (Waldherr, Geise, & Katzenbach, 2019)、AI 技術和人之間的關係。

從實務觀察，AI 對公共治理理論的影響為何？在傳統官僚機構中，基層官僚具有自由裁量權和相當的自主權，運用自己的判斷力決定如何執行政策，同時在複雜的現實生活中處理時間等資源稀缺問題，AI 改變了基層官僚自由裁量權(de Boer & Raaphorst, 2021)。舉例來說，2012 年波蘭勞動和社會政策部(Polish Ministry of Labour and Social Policy, MLSP)建立具有成本效益且個人化服務的「失業評分系統(unemployment scoring system)」，最初構想是建議工具，同時保留諮

共政策的人，涉及到個人的行為、互動的對象，以及整體的社會脈絡等(Dunn, 2015; 丘昌泰, 2008; 吳定, 2017)

¹⁴ H+ (a common abbreviation) means the enhancement of human beings as a whole.



詢人員對失業分類最終決定權。但基層官員信任 AI 提供的信息，不會進一步搜索更多資訊，或他們開始害怕出現不一致的意見，只對案例做微小更改，結果是諮詢人員超過 99% 依據 AI 系統的推薦分類，導致 AI 系統變成實質決策者 (Misuraca & van Noordt, 2020)；此時，AI 是「工具」還是「公務員」？公共治理理論長久以來所討論街頭官僚機構(street-level bureaucracies)的行政裁量權，可能無法有效解釋螢幕官僚機構(screen-level bureaucracies) (Bovens & Zouridis, 2002; de Boer & Raaphorst, 2021)的現象。

因此，機器人成為新一代官僚—「AI 官僚」(Unver, 2018)，不只讓我們有機會看到 AI 在人類社會中的角色，也可能對人類作為主體的觀點提出挑戰，將顛覆以「人」為思考的公共治理理論，引發學界、政府重新關注 AI 在全球政治、公共治理中的作用。有研究從社會科學角度，將 AI 視為「非人社會行動者」(Schwartz, 1989; 劉育成, 2020)。有從公民權角度授予 AI 機器人「Sophia」國籍與公民權¹⁵。有從法制角度建議賦予 AI 「電子人格」、「法人格」，以迴避世界科學知識與技術倫理委員會(The World Commission on the Ethics of Scientific Knowledge and Technology Commission)認為將 AI 稱為「人」是高度違反直覺的，只要他們不具備通常與人相關的某些特質，例如自由意志或道德能力等，就不能被認為具有人格(王一旅, 2021)。

儘管人們普遍認為 AI 將從根本上挑戰公共治理理論，目前研究人員主要關注 AI 應用的技術，研究具體技術問題和解決方案，例如 2000 年至 2019 年之間發表的 AI 文章幾乎全部集中在私部門的技術發展和應用，只有極少數(1438 個中的 59 個)討論公部門的 AI (Desouza, Dawson, & Chenok, 2020)，討論內容也主要為技術面向，例如優化人工神經網絡(Artificial neural networks, ANN)的演算法提高正確率等(de Sousa, de Melo, Bermejo, Farias, & Gomes, 2019)，藉由提高 AI 的技術

¹⁵ 檢自 <https://www.thenewslens.com/article/82000>



來提升公共治理效能。

除了高度技術性的研究領域，另一個面向則以社會學、政治學或公共政策的
角度，探討 AI 科技應用伴隨的風險，例如如何控制 AI 伴隨的倫理道德風險(黃
心怡、曾冠球、廖洲棚、陳敦源, 2021)，包括：

1. 機器人取代公務人員工作。
2. 法規賦予公務員行政裁量權變成是演算型決策的課責性(Barth & Arnold, 1999)。
3. 公共領域的人臉辨識是否會過度侵害隱私權(Whittaker et al., 2018)。

然而，過度強調 AI 治理的倫理和公平會遏制技術創新¹⁶，歐盟建議應在學術
上應更加關注 AI 在「政府事務(business of government)」的應用(Misuraca & van
Noordt, 2020; Misuraca & Viscusi, 2015)。舉例來說，實務上 AI 系統旨在促進人與
機器之間的協作(Mills et al., 2021)，或 AI 已持續增強和取代官僚機構的自由裁量
權，而當前相關的研究，仍側重於對行政自由裁量權潛在和觀察得到的影響
(Young et al., 2019)，極少有研究採用理論的觀點來探討 AI (Sun & Medaglia,
2019)，也沒有健全的公共管理模型來描述對行政國家治理的影響(Reis, Santo, &
Melão, 2019; Sharma, Yadav, & Chopra, 2020)，例如公部門採用 AI 對公共治理、政
策、影響的研究，人與機器(human-to-machine, H2M)的互動(Wirtz, Weyerer, &
Geyer, 2019)關係，亟需合適的理論和概念框架來理解(Duan, Edwards, & Dwivedi,
2019)。

最後，對未來如何因應的挑戰也缺乏共識(Zuiderwijk, Chen, & Salem, 2021)，
例如，有實證研究指出，參與 AI 技術採用的利益相關者認為，大多數挑戰都不

¹⁶ 檢自 <https://www.brookings.edu/blog/techtank/2021/05/13/analyzing-artificial-intelligence-plans-in-34-countries/>



是技術或數據(Sun & Medaglia, 2019)，而是如何在官僚環境中應用 AI (Höchtel, Parycek, & Schöllhammer, 2016)，創新過程中不能與公共行政改革脫鉤(Misuraca & van Noordt, 2020; Misuraca & Viscusi, 2015)，關鍵在於 AI 的組織或管理問題，公部門場域的 AI 技術設計、如何將 AI 技術整合到公共行政流程(Salganik et al., 2020)。有論者指出公部門採用 AI 最大的挑戰即是公共治理的回應：未被充分利用(underused)產生的機會成本，或過度使用和濫用(overused and misused)產生的風險(Floridi et al., 2018)；或是「如何最大化從 AI 獲益同時最小化對人類的傷害」(Dignam, 2020)。也有論者指出倫理、法律和社會挑戰並非 AI 所獨有，而是 AI 所加劇的(Henman, 2020)。

因此，各國政府、國際組織、非營利組((NGO)和公司另起爐灶，發布許多相似的倫理準則(ethics codes)、原則(principles)、指南(guidelines)、框架/frameworks)和政策策略(policy strategies)等(Misuraca, 2020; Schiff, Biddle, Borenstein, & Laas, 2020)，或許可以稱之為新興 AI 治理。但除此之外，在發展 AI 的同時消除不必要的障礙，以確保社會利益能納入考慮並保護的面向，幾乎沒有任何作為 (L. Zhang, 2020)；更重要的是，在大多數情況下，這些單純從技術角度出發的新興 AI 治理，所提出的指導方針忽略現有的治理機制和機構，就好像現有的治理結構在「AI 時代(the age of AI)」完全不匹配一樣。歐盟經常發出警告，若單純從技術角度出發的治理框架，容易忽略對現有治理結構的影響，建議應該從現有機制設計創新、修正以提高可行性 (Misuraca, 2020)，這種研究是不同於以往單純對技術、工作和組織的傳統研究，而是整合技術與公共治理的研究(Bailey & Barley, 2020)。

肆、公部門應用 AI 的困境

雖然私部門已經有很多應用 AI 的經驗，然而，公部門使用 AI 必須比私部

門更加謹慎，使得私部門的 AI 作為和數位轉型策略無法直接複製到公部門 (Fatima, Desouza, & Dawson, 2020)。公部門和私部門使用 AI 為何有差異呢？原因包括：

1. 公部門有統治權：公部門具有行使統治權，能片面決定改變相對機關或相對個人權利義務的權力。再者，AI 通用技術、雙重用途 (Floridi, 2018)、能動性 (agency) 或作用、政治性、偏差或偏誤 (bias) 等特性，可以根本上改變權力本質、結構、分配，並使權力差距擴大。
2. 公民無法選擇退出 (opt-out)：公民可以選擇不使用特定私人公司產物或服務，但公民不能選擇不使用公共服務 (UK, 2020)，政府必須在符合民主價值 (democratic values)、權利保護 (rights protection)、法治 (the rule of law)、安全 (safety) 和保障 (security) 的最高指導原則下，為公民提供足夠的公共服務保障 (Jimenez-Gomez et al., 2020)。
3. 公部門使用大數據發展 AI 仍受到許多限制¹⁷：人們每天產生超過 2.5 兆位元組的資料 (2.5 quintillion bytes of data)，企業可以利用 AI 為商品和服務增加數兆美元的價值 (Margetts & Dorobantu, 2019)。理論上，政府所蒐集可用的大量公民數據大於企業，但各國政府在應用 AI 來改善政策和公民服務方面進展緩慢。政府收集大量公民數據所遭遇的困境，包括個人資料保護法、透明度、倫理等議題，例如我國 2022 年憲判字第 13 號 [健保資料庫案]¹⁸、英國「國家數據選擇退出」(ND opt-out) 制度 (Meszaros, Ho, & Corrales Compagnucci, 2020)。因此，政府如果事先不使用參與的方式促進社會的辯論，政府將無法擴大採用 AI 技術 (de Sousa et al., 2019; Mikhaylov, Esteve, &

¹⁷ Rethink government with AI. 檢自 <https://www.nature.com/articles/d41586-019-01099-5>

¹⁸ 檢自 <https://cons.judicial.gov.tw/docdata.aspx?fid=38&id=309956>



Campion, 2018)。

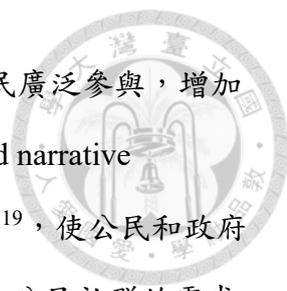
4. 公部門需要最大化公共價值：政府與企業對於目的的追求，在相當程度上，企業幾乎可以無所限制的追求市場的效率與利潤，而政府經常必須權衡 (trade off) 效率與公平，以及其他公共價值之間的平衡(蔡允棟, 2002)。

公部門有統治權、公民無法選擇退出、公部門使用大數據受到許多限制、公部門需要最大化公共價值，使公部門使用 AI 必須比私部門更加謹慎。

然而，過度謹慎可能會低估 AI 的潛在效益，以及錯失 AI 的準確性 (accuracy)、規模(scale)和效率(efficiency)、有機會做出更好的判斷的優勢(Young et al., 2019)。因此，亟需如何因應公部門使用 AI 的理論基礎，以及如何在公共治理面向設計緩解機制避免社會反彈，什麼條件下 AI 技術可以發揮最大效用，影響 AI 採用的成功或失敗因素，導入 AI 的直接和間接影響(H. Chen, 2019)等的實證研究。

鑑於 AI 在公共治理領域內逐漸加速擴大實施，與在政府中應用 AI 影響的實證研究進展緩慢形成鮮明對比；更甚者，技術社群追求 AI 系統的效能和正確率的推波助瀾之下，AI 系統變得越來越複雜且難以預測(J Hernández-Orallo, 2014; José Hernández-Orallo, 2017)，在缺乏具體 AI 治理理論的情況下，大多數政府對在公部門使用 AI 的影響理解有限，甚至不清楚會如何影響公共治理(Zuiderwijk et al., 2021)。

舉例來說，如何在工作場所採用和執行 AI 之前，讓技術、工作和組織的學者參與新智慧技術設計，以避免技術性失業及技術可能帶來的倫理問題(Bailey & Barley, 2020)。在政策規劃階段，為避免傳統「官僚性政府」所具有相當強的菁



英決策色彩(江明修, 1997), 如何設計「公民參與」機制, 使公民廣泛參與, 增加公民的信任(林子倫, 2017), 利用 AI 驅動的敘事構建(AI-powered narrative building)提高公民參與人數、範圍、品質、利益相關者的可近性¹⁹, 使公民和政府能夠在政策設計方面進行更有意義的對話和合作, 政策規劃符合公民社群的需求與價值, 加快和優化決策過程(Carrascosa & Banjac, 2021; Marmolejo-Ramos et al., 2022), 減少對技術官僚專家的依賴, 增加議程制定和政策制定的合法性。如何進行資料治理及演算法治理, 避免數據限制演算法的學習能力(Desouza, 2018), 以及人與非人的社會技術組合(socio-technical assemblages)所構成的演算法模型(algorithmic model)的價值觀受到專業人士世界觀所影響(Kolkman, 2020); 以及社交媒體使用者不具有人口統計學代表性(Höchtel et al., 2016), 必須考慮最容易數位落差的弱勢族群(Pencheva, Esteve, & Mikhaylov, 2020)。在政策執行過程中如何將 AI 技術整合到實際行政流程中, 增加使用者(公民、公務員)的科技接受度; 如何使公務員合理運用 AI 工具, 保護公民避免受到演算法潛在傷害(UK, 2020), 或法律裁量權中沒有適當考慮個人的處境(Henman, 2020)。在政策評估過程中如何評估人機協作的績效, 避免無法課責的問題等(Davenport & Kirby, 2016; House, 2019)。

從公部門使用 AI 的現況分析, 波士頓諮詢集團 2021 年的調查顯示公部門 AI 在「負責任 AI 的建置成熟度」(a mature implementation of a responsible artificial intelligence, RAI) 排名倒數第 2 (Mills et al., 2021), 顯示目前公部門應用 AI 的成熟度相較私部門明顯不足。另依據 2022 臺灣 AI 國力調查, 臺灣應用 AI 的現況顯示, 成熟度較為落後的也是公共部門, 該調查另指出臺灣企業的模式部署率(即

¹⁹ CitizenLab 平台, Youth4Climate(Cuau, 2019) 組織在該平台上從世界各地的青年蒐集大量應對氣候變遷的想法, 之後該平台利用其自動分析功能將這些想法轉化為有意義的建議和行動; 義大利使用 CITBot 聊天機器人來促進公民參與。



開發出的模型用於實際生產的比例)在 50-75%，與 2021 年的 20-50%相比明顯提高，且高於全球平均(部署率約 20%)，推估是因為臺灣有大量經驗豐富的 AI 技術團隊和具有領域專長的合作夥伴，故善於挑選、實驗和測試模型，有利於提高使用模型的成功率(葉哲良、李育杰, 2022)。

因此，歐盟建議各會員國未來應該針對 AI 實施成功、失敗的案例做更深入的研究，了解及分析成功和失敗的原因，透過大規模的試驗性措施(the Large-Scale Pilots initiative)，累積各會員國深度質化案例研究成果來獲得不同脈絡的 AI 治理問題，作為未來提供給公部門使用 AI 的理論基礎、政策設計框架、成功與失敗的分析框架、實施路線圖、關鍵成功因素等建議(Duan et al., 2019; Misuraca & van Noordt, 2020; Molinari F., 2021)，以克服公部門應用 AI 的困境，提高應用 AI 的成熟度。

伍、小結

鑒於公共治理領域中越來越廣泛使用 AI 作為提高政府治理能力以及達成永續治理目標的手段，並且預期政府將成為 AI 的最大採用者之一(Dwivedi et al., 2019)。而傳統公共治理理論以人為本的治理典範，當公部門應用 AI 時產生解釋力不足的現象，Dafoe(2018)建議此類研究可跨領域尋找現有的理論，而不需重新發明研究方法(Dafoe, 2018)。

本文嘗試於在科學科技與社會(Science, technology and society, STS)尋找適合理論，由於行動者網絡理論(Actor-Network Theory, ANT)凸顯技術在社會建構中的角色，強調非人能動性，有利詮釋人與 AI 之間權力關係的解構與重構的複雜情況，與本研究方向契合，故選擇 ANT 分析途徑為本研究的方法論，研究方法與重點如下：

一、使用 ANT 及其研究方法，將 AI 技術物理化：將公共治理理論中「AI 技

術物視為理所當然的事情」轉移到「具有獨特文化和計算能力的特定技術」(Höchtel et al., 2016; Orlikowski & Iacono, 2001)，解決政策科學中缺乏對技術關注的問題，探索公共治理理論典範轉移路徑。

- 二、新增探索人與 AI 如何互動相關理論：分析 AI 對人的影響，AI 與人如何互動、公共治理領域人機協作的理論基礎等，補充現有公共治理理論的不足。
- 三、有別於目前 AI 治理研究中占主導地位的倫理準則、戰略、策略、政策和管制等法律領域規範的研究，著重於可近性及務實性。



第二節 研究問題與目的

本文基於前述研究動機提出三大研究方向：

一、公共治理理論層次

公共治理理論長久以來承襲經濟學、社會學²⁰的方法論是以人為本的治理典範，忽略 AI 技術物(AI artifacts)，例如公務員的行政裁量權在「人」、「AI」之間移轉的現象。另從 AI 領域角度，涉及「蒙特利爾宣言(The Montreal Declaration)」：「需要在以人為主導的決策與以機器為主導的決策之間取得平衡」所探討「計算機系統的自治權(the autonomy of computer systems)」(Floridi et al., 2018)。再者，目前 AI 屬於弱 AI 階段，有侷限性，有研究建議採用「人機協作」的方式。

在公共治理理論層次的研究問題與目的：

1. 公共治理理論的分析單位，除了人，是否應該包括 AI？原本公共治理理論可能無法解釋公部門應用 AI 的現象，而 ANT 的廣義對稱性原則，是否可以提供 AI 成為分析單位的論述基礎。
2. 公共治理理論典範轉移路徑，以及應補充的理論基礎？公共政策模型、過程論、因果關係推理問題等的詮釋邏輯是否需要因應 AI 改變。
3. 公共治理領域中人機協作的理論基礎？探討影響 AI 與公務員互動關係的變數，以及應如何管理 AI 與公務員的互動關係等。

²⁰ 社會學者涂爾幹在思考社會的構成時，仍然只有考慮到人，而沒有納入技術(techniques) (雷祥麟, 2004)。



二、公共治理分析層次(政府使用 AI 的政策分析方法)

公共政策分析具有著重於解釋而不在於開處方、致力於追求公共政策的前因與後果，評估政策的何去何從(林水波、張世賢, 2006)。然而，當公部門使用 AI 時，原本公共政策的分析方法，無法有效解釋公部門使用 AI 的現象，例如強調因果推理的詮釋邏輯變成相關性時，需要在公共治理領域找到能生成操作定義 (operational definitions)、萃取意義 (extract meanings)、解釋結果 (explain outcomes) 的經驗研究方法 (empirical methods)。Zuiderwijk 等人 (2021) 建議重新審視 AI 出現以後的理論，採跨學科合作 (multidisciplinary collaboration) 方式，研究方法應轉向方法多樣性，以提高公共治理理論的嚴謹性 (Zuiderwijk et al., 2021)。

由於 ANT 不同於公共治理理論以人為中心的方法論，主張「廣義對稱性原則」、「物體也有能動性 (agency)」、「交引纏繞的社會世界觀」，可以把非人類 (例如 AI) 當成分析單位，彌補公共治理理論長期忽視技術的理論缺口。再者，AI 是行動者，有自適應性 (adaptive) 和能動性 (agency)，然而，人類受限於經驗主義客觀現實，不易提出人和 AI 共存的因果推理假設；而 ANT 發展概念是從經驗研究 (empirical studies) 中構建出來的，ANT 對描述而非解釋的關注 (Latour, 2005; Waldherr et al., 2019)，可以彌補公共治理理論關注因果關係推論的分析方法。因此，本研究認為，ANT 分析途徑有機會成為公共治理領域公部門使用 AI 的政策分析方法。

在公共治理分析層次的研究問題與目的：

1. ANT 的分析途徑是否可以成為公部門使用 AI 的公共政策分析方法？
2. 驗證 ANT 及其分析途徑在公共治理領域的可用性？

三、公共治理實踐層次

雖然 AI 是第四代工業革命的關鍵推力，但理論和實作間可能產生落差，技術

的意義要在使用中發現，唯有使用經驗研究可揭露技術隱匿面向(王志弘、高郁婷譯, 2023)，公部門的數位轉型也並非線性或可預測地進行(Molinari F., 2021)，有研究指出 87%的 AI 和大數據專案會失敗並且從未部署(Merhi, 2023)，甚至公部門應如何使用 AI 仍有許多懸而未決的難題與困境：

1. 公部門 AI 科技創新原則：如何在管理風險/關切(managing risks/concerns)與促進創新之間找到合理的組合，關鍵的論述是，任何規範當前數位轉型的嘗試都會扼殺創新，AI 錯誤甚至是必須的，因為 AI 需要從錯誤中學習和改進以不斷優化其決策(王一旅, 2021)。相反的觀點是，面對不確定性，可以基於預防原則採用強有力的監管方法(Cohen, 2019; Misuraca, Barcevicus, & Codagnone, 2019)。因此，政策制定者與管制者常面臨兩難：「預防原則(precautionary principle)」與「不需許可的創新 (permissionless innovation)」(Thierer, Castillo O'Sullivan, & Russell, 2017)，「預防原則」要求推動創新發明與服務者的產物不超出現有規範，以免未來對公眾構成危害，但不容易產生政策創新、妨礙由下而上的政策解決方案，嚴重的威脅技術進步；「不需許可的創新」是為避免應用限制，而是以循證為基礎的政策(evidence-based policy)，自下而上彈性解決方案，而非自上而下剛性的控制(Thierer et al., 2017)。

2. 人類監督強度：Rahwan(2018) 認為將 AI 應用到整個社會，影響領域更廣泛，社會需要有代表所有利益相關人的規範方法，提出維持社會的管制模型，從「將人類置於循環中監督，以增強人類的決策能力(human-in-the-loop, HITL)」模型調整為「將社會置於循環中監督，以增強社會的決策能力(society-in-the-loop, SITL)」模型，因為在 SITL 模型中，整個社會首先必須權衡不同人類價值觀(例如隱私和安全)，並且必須確保 AI 技術的收益和成本在不同的利益相關者之間合理分配(Rahwan, 2018)，然而，SITL 雖可以降低風險，但勢必將增加

公部門導入 AI 的困難度。



- 3.AI 系統的品質：AI 系統的效能/正確率(performance/accuracy)和可解釋性(explainability)之間存在反比關係的情況下，需要進行權衡(Dwivedi et al., 2019)，應如何權衡？何時應該注重效能/正確率，何時應該注重可解釋性。
- 4.AI 在政府職能的應用領域：在啟動 AI 計劃之前，公部門必須了解哪種技術可以執行哪種類型的任務，以及每種技術的優勢和侷限性(Duan et al., 2019)。
- Mehr(2017)提出避免將 AI 決策用在對人民權利和義務產生重大影響領域(Mehr, Ash, & Fellow, 2017)；Bullock(2019)則提出複雜性和不確定性較低的任務是 AI 處理，而複雜性和不確定性較高的任務則應由人類自行決定(Bullock, 2019)；有研究甚至以程序正當性的理由主張核心公共機構，例如負責刑事司法、醫療保健、福利和教育的機構等高風險領域，不應再使用 AI (Agarwal, 2018)。究竟哪些政府職能或公共服務應用領域適合導入 AI？(Busch & Henriksen, 2018; Kuziemski & Misuraca, 2020) 2023 年 6 月 14 日歐洲議會表決通過「AI 法(A.I. Act)(草案)」，將禁止警察單位利用過往犯罪紀錄來發展犯罪預測系統²¹。再者，各國對於人臉辨識技術(facial recognition technologies, FRTs)的應用領域也存在著許多爭議及法規不確定性，例如使用人臉辨識技術識別和定位人口販運的受害者²²？歐盟「AI 法(A.I. Act)(草案)」禁止基於執法目的在公共空間使用即時人臉識識系統(real-time facial recognition systems)，但允許用於執法以外目的(例如邊境管制)。

²¹ 檢自 <https://www.artificial-intelligence-act.com/>

²² 檢自 <https://aws.amazon.com/tw/rekognition/the-facts-on-facial-recognition-with-artificial-intelligence/>



5.價值選擇：這是最困難的，「科技優勢 VS 法律原則」²³，涉及「能力可行性及規範合理性」的「倫理辯證」，例如「AI 法官」的價值衝突。例如法官需要足夠的事實知識、理論、實踐和系統知識以及靈活的思維，只有人才能做出公平合理的權衡決定(FRA, 2020)，或是演算法抵觸憲法第 80 條²⁴，或是法院使用機器學習確定是否還押或釋放被控犯罪的個人，以及根據過去案件的數據，預測個人是否會重返監獄的特徵，評估尋求假釋囚犯的再犯風險等(Kleinberg, Lakkaraju, Leskovec, Ludwig, & Mullainathan, 2018)，是否侵犯公民權利？有論者提出「AI 法官」必須在特定的條件、脈絡與案型的觀點；然而，這樣的觀點仍然陷入另一層次的論辯²⁵，有法官認為簡易的案件交由 AI 處理，讓法官把心力留給需要裁量的「疑案」，效率與品質都能得到提升；但也有法官認為 AI 學習舊判決，有極大的可能會複製社會結構的錯誤；甚至可能忽略 AI 也具有生成能力的事實，使 AI 變成「有罪判決產生器」²⁶。

以上所述需要權衡之處亦需亟需進一步釐清。因此，在公共治理實踐層次的研究問題與目的：

1. 如何提升公共治理效能及控制風險，提高 AI 專案的成功率？
2. 公部門使用 AI 時，應如何權衡 AI 爭議？(1).科技創新原則、(2).人類監督強度、(3). AI 系統的品質、(4). AI 在政府職能的應用領域、(5).價值選擇等。
3. 提出臺灣公部門可信任 AI 參考指引(草案)？

²³ 檢自 <https://research.sinica.edu.tw/ai-judge-justice-chien-liang-lee/>

²⁴ 憲法第 80 條：「法官須超出黨派以外，依據法律獨立審判，不受任何干涉。」

²⁵ 檢自 <https://castnet.nctu.edu.tw/castnet/article/14771?issueID=788>

²⁶ 檢自 <https://www.chinatimes.com/realtimenews/20231005002011-260407?chdtv>



第三節 研究限制

AI 和公共治理的範圍都很廣泛，本研究為利聚焦討論設定研究範圍，同時也形成研究限制，說明如下：

壹、AI 定義範圍的限制

一、AI 領域的快速發展、案例的激增、應用的環境及解釋隨著時間不斷更新、涵蓋許多不同的專業領域(例如生成式 AI 與傳統 AI 明顯的區隔)，使得 AI 沒有普遍接受的定義(Bignami, 2022; Wang, 2019)。但如果沒有對 AI 定義，很難評估 AI 對公共治理理論的影響，為利研究能夠聚焦，本研究使用 AI 的範疇(scope)，同時使用兩個定義〔AI Russell & Norvig(2002)、歐盟 2023 年 6 月 14 號表決通過的 AI 法(A.I. Act)(草案)〕互相補充。

二、AI 技術發展程度分成 3 個層次：包括弱 AI(the label artificial narrow intelligence, ANI)、通用 AI(artificial general intelligence, AGI)、超級 AI(artificial super intelligence, ASI)，目前是弱 AI，研究案例亦屬弱 AI 之範疇。

貳、政府角色範圍的限制

政府可以身兼使用者、推動者、促進者和管制者的多重角色(Molinari F., 2021)，本研究聚焦於政府是使用者。

參、研究方法的限制

本研究採取多重研究方法，包括文獻分析法、個案研究法。茲就二種研究方法可能造成研究限制分別說明如下：



一、文獻分析法：缺乏政府是 AI 使用者的相關文獻

政府作為 AI 「使用者(user)」的研究遠沒有「管制者(regulator)」角色受到關注(Zuiderwijk et al., 2021)。

二、個案研究法：樣本數不夠充分

行動者網絡理論的研究方法是「跟隨行動者(Follow the actors)」(Latour, 2005)，基於案例資料的可取得性，案例選擇聚焦於臺灣公部門應用 AI 的案例，國外案例做為補充或佐證，不做實證上研究。然而，臺灣公部門使用 AI 的實證案例不多，部分政府職能亦未導入 AI，樣本選擇受限，例如政府職能樣本分布不均、稀缺等，可能無法選擇最典型案例，使立意抽樣(purposive sampling)的判斷性樣本(judged sampling)或代表性(representative sample)產生偏差，高估樣本代表性，使得利用研究結果推論母群體時形成推論的限制：

1.大部分案例屬於試驗或部分部署階段

臺灣應用 AI 的進度與美國、歐盟類似，分布在規劃(planning)、試驗或部分部署(piloting or partially deployed)、完全部署(fully deployed)三階段，大部分仍屬於試驗或部分部署(piloting or partially deployed)，成熟的案例不足，仍需時間觀察。對一個尚在發展理論的領域來說仍是未論，或太早下定論恐有失偏頗，尚待累積更多實證經驗提高理論的強韌性。舉例來說，荷蘭政府 2014 年開發 SyRi，2020 年取消 SyRi，歷時約 6 年；波蘭 2012 年開發 AI 失業評分系統(unemployment scoring system)，2019 年正式停止，歷時約 7 年；愛沙尼亞 2018 年開發 SATIKAS32 系統，目前仍在持續開發並擴展其特性和功能；比利時 2014 年佛蘭芒兒童和家庭事務局(Flemish Agency for Child and Family)開發的 AI 預測系統(Predictive System)，目前仍在使用中。進一步分析

案例，荷蘭和波蘭的案例在 AI 導入過程中即爭議不斷，ANT 的分析途徑，對一個發展中的理論而言仍有其優勢。



2. 管制性政策的案例比重不高

本研究立意選樣挑選案例包括不同政府職能或任務領域、組織層級及組織型態、計畫成熟度、可比性(comparability)、應用普遍性或後續擴散潛力等多樣化覆蓋範圍，使研究成果獲得綜觀性理解。然而，國家權力涉入程度低的象徵和勸告型，例如政府資訊的聊天機器人、提高公民服務品質的智慧交通號誌等政府案例較普遍，成熟度較高。但國家權力涉入程度高的管制型，AI 自動化演算結果作為管制或干預人民行動的依據或參考案例較少，成熟案例稀缺，使案例選擇性相當侷限。究其原因，可能是管制型政策容易產生爭議，例如「澳大利亞聯邦政府 Robodebt」、波蘭「失業評分系統」、荷蘭 SyRi 系統，主要爭議原因在於與現有治理機制扞格，包括

- (1). 法律授權範圍觀點：AI 行使公權力的效果已經超過法律對行政裁量權的授權範圍(Henman, 2017)。
- (2). 技術觀點：使用 AI 的「相關性」，但是做為與「公民權利高度相關」且須「因果關係」的裁決，是基於可能性而不是事實的法律問題等。



第二章、AI 與公共治理的對話

本研究採跨領域的研究方法，涉及領域包括 AI、公共治理理論、行動者網絡理論，不同領域人員對同一名詞可能會有不同的理解，故本章首先提出本研究所採用跨領域術語或專有名詞的定義、概念或真正意義，以利後續討論。接續論述公部門應用 AI 的機會與挑戰、AI 與公共治理的研究現況，以利研究規劃與設計。

第一節 AI 的定義與範圍概述

「人工智慧(Artificial Intelligence, AI)」領域的快速發展、案例的激增、應用的環境及解釋隨著時間不斷更新、涵蓋許多不同的專業領域(例如生成式 AI 與傳統 AI 明顯的區隔)，使得 AI 沒有普遍接受的定義(Bignami, 2022; Wang, 2019)。再者，不同領域對 AI 的理解、名詞與應用也未盡一致(Bhatnagar et al., 2018; Wang, 2019)。

以 AI artifacts 為例，在資訊科學領域，AI artifacts 是軟體代理人(software agent)的建置(implementation)或是具體化(instantiation)，是所選用的軟體或硬體具備智慧的特性，因此，AI artifacts 指「AI 演算法(AI Algorithm)」、「AI 程式(Artificial intelligence programming) 或是代理人程式」、「AI 產品或 AI 產物(AI production)」、「內含 AI 功能的工具及產物」等。另一方面，在科學科技與社會(STS)、行動者網絡理論(the Actor-Network Theory, ANT)的領域，以「AI 技術物」來表達。

為利跨領域理解與討論能夠聚焦，本文後續視上下文脈絡而交互使用 AI、

AI artifacts、AI 產品或 AI 產物(AI production)、AI 演算法(AI Algorithm)，或 AI 程式(Artificial intelligence programming)、AI 系統、AI 技術等，期使不同領域都能理解本研究所指。



壹、AI 的定義與本研究使用的 AI 範疇

梳理文獻 AI 定義(Gellers, 2020; Samoili et al., 2020; Zuiderwijk et al., 2021)，包括：

1. *Russell & Norvig*：所著「人工智慧——一種現代方法(*Artificial Intelligence: A Modern Approach*)」²⁷的定義「AI 是智慧代理人或軟體機器人(*intelligent agents*)的概念，從環境中獲得感知並採取行動的代理人(*agents*)，每個代理人(*agents*)都能夠連續感知並採取行動。²⁸」(*S. Russell & Norvig, 2002*)。

其中，*agents*：在資訊科學、公共政策領域，主要翻譯為「代理人」。在科學科技與社會(STS)領域大多翻譯成「能動者(*agents*)」；*Latour* 定義 *agents* 是「能對其他能動者的行動產生影響」(*Latour, 2005; Sayes, 2014*)；*Datnow*、*Hubbard* 和 *Mehan*(2002)對能動者的定義是「有能力改變現狀的人」(黃琳、霍秉坤, 2016)。為利跨學科理解，本文後續將依據不同脈絡採「代理人(*agents*)」、「能動者(*agents*)」方式交替使用。

Russell & Norvig(根據一些 AI 定義，利用 2 個維度整理分成 4 種類型：「像人一樣思考的系統、像人一樣行動的系統、理性地思考的系統、理性地行動的系統」，圖上半部所定義的是思惟過程和推理，而下半部的定義則強調

²⁷是 AI 領域排名第一的教科書

²⁸ 原文：The main unifying theme is the idea of an intelligent agent. We define AI as the study of agents that receive percepts from the environment and perform actions. Each such agent implements a function that maps percept sequences to actions.



行為，左側的定義根據對人類技能保真度來衡量其成功與否，而右側的定義則根據所謂理性的理想效能測量來進行衡量。一個系統如果能夠在他所知的範圍內做「正確的事」，它就是理性的。

表 2- 1 Russell & Norvig 對 AI 的分類

像人一樣思考的系統	理性地思考的系統
像人一樣行動的系統	理性地行動的系統

資料來源：Russell & Norvig (2003)、(歐崇明、時文中、陳龍編譯, 2011)

因此，Russell & Norvig 認為智慧主要與理性行動相關，理想上，智慧代理人(intelligent agents)要採取在一個狀況下最好的可能行動，而 AI 是要研究如何建造在這個意義上具備智慧代理人(agent)。然而，「要達到完美的理性—即總能做正確的事情—在複雜的環境下是不可行的，這對運算能力的要求實在太高了。」，因此，智慧代理人(intelligent agents)是在時間不足以完成全部計算時，如何採取「有限理性」的適當行動，AI 所強調「正確的推論」的「思維法則」所產生的推論並不總是全然理性的(S. J. Russell, 2010; 歐崇明、時文中、陳龍編譯, 2011)。

2.HLEG：「AI 系統是由人類設計的軟體(可能還有硬體)系統，先給定一個複雜的目標，其會通過數據蒐集來感知週遭環境，以此來解釋所蒐集的結構化數據(structured data)或非結構化數據(unstructured data)，從這些數據中推論知識或處理資訊，並決定為實現該給定目標而採取的最佳行動；AI 系統既可以使用符號規則(symbolic rules)，也可以學習數值模型(numeric model)，還可以通過分析



環境如何受到先前行動的影響來調整其行為。²⁹

AI 作為一門學科，包括多種方法和技術，例如機器學習(machine learning)下有深度學習和強化學習等分支、機器推理(machine reasoning)包括最佳化規劃、知識表示暨推理和搜索；機器人學(robotics)包括控制、感知、傳感器和執行器，以及將前述所有其他技術合併整合到虛實整合系統(cyber-physical systems)中。」(HLEG, 2019; Samoili et al., 2020)

3. Wirtz 等人(2019)：在綜各種定義後認為，AI 是以機器為基礎的系統(machine-based systems)和類似人類智慧行為(human-like intelligent behavior)的特徵，嘗試更有效率地去解決人類的問題，是僅以技術的角度定義，因此嘗試以公共管理脈絡提出 AI 的定義：「AI 是指計算系統具有某些類似人核心能力(包括感知、理解、行動和學習)的智慧行為³⁰。」，AI 應用是「將 AI 的技術整合到人與計算、人與資料互動的應用領域³¹。」(Wirtz et al., 2019)。

²⁹ 原文："Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans* that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions." (* Humans design AI systems directly, but they may also use AI techniques to optimise their design.)(HLEG, 2019)

³⁰ 原文：AI refers to the capability of a computer system to show humanlike intelligent behavior characterized by certain core competencies, including perception, understanding, action, and learning.

³¹ 原文：the integration of AI technology into a computer application field with



4. OECD：AI 為「一臺以機器為基礎的系統，對於人類所給定義的目標，可以做出影響真實或虛擬環境的預測、推薦或決策。AI 系統可以設計成具有不同程度自動化的機器。³²」(OECD, 2020)。
5. 歐盟 2020 AI 白皮書(EU White Paper on AI)：將 AI 定義為「結合數據、演算法和計算能力的技術組合。³³」(Nikolinakos, 2023)。
6. 美國 2020 年「國家 AI 倡議法(National Artificial Intelligence Initiative Act of 2020)」(DIVISION E, SEC. 5001)的定義³⁴：「AI 是指一種基於機器的系統，它可以針對人類定義目標做出影響真實或虛擬環境的預測、建議或決策。AI 系統使用機器和人類的輸入來 1.感知真實和虛擬環境；2.提取此類感知自動分析為模型；3.使用模型推論來制定資訊或行動的選項。」；「AI 通常涵蓋的主題包括機器學習、電腦視覺、自然語言處理和理解、智慧決策支持系統和自主系統，以及這些技術在各個領域的新穎應用。」³⁵。

human– computer interaction and data interaction.

³² 原文：An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

³³ 原文：a collection of technologies that combine data, algorithms and computing power.

³⁴ 檢自 <https://www.ai.gov/about/>

³⁵ 原文：The term ‘artificial intelligence’ means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and

7. 歐盟 2023 年 6 月 14 號表決通過的 AI 法(草案)，定義 AI：「...使用[特定]技術和方法[在附件 1 列出]開發的軟體，可以針對人類所設定的一系列目標，產生影響與其互動環境的輸出，例如內容、預測、建議或決策等。因此，『AI 系統 (AI system) 』的概念將指一系列以軟體為基礎的技術，包括『機器學習 (machine learning) 』、『基於邏輯和知識(logic and knowledge-based) 』的系統，以及『統計(statistical) 』方法。」³⁶，AI 系統可以獨立使用或作為產物的一部分使用³⁷。此外，2023 年 6 月 14 日議會通過的法案中特別針對通用 AI 的基礎模型和生成式 AI 新增規範。「生成式 AI」基本定義³⁸：

(1). 「基礎模型(Foundation models)」：從廣泛的數據來源和大數據上以監督學習、強化學習等方式進行訓練，為了通用性和多功能性之目的基於演算法而開發，以完成廣泛的下游任務，基礎模型可以是單模態或多模態，可以

human-based inputs to – (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.” Topics typically encompassed by the term “AI” include machine learning, computer vision, natural language processing and understanding, intelligent decision support systems, and autonomous systems, as well as the novel application of these techniques to various domains.

³⁶ 原文：...software that is developed with [specific] techniques and approaches [listed in Annex 1] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. & Annex 1 of the proposal lays out a list of techniques and approaches that are used today to develop AI. Accordingly, the notion of 'AI system' would refer to a range of software-based technologies that encompasses 'machine learning', 'logic and knowledge-based' systems, and 'statistical' approaches.

³⁷ 檢自 <https://artificialintelligenceact.eu/the-act/>

³⁸ 檢自歐盟人工智能法案對「生成式 AI」的規範 from https://www.saint-island.com.tw/TW/News/News_Info.aspx?IT=News_1&CID=266&ID=62704



利用開源(open source)或者應用程式(app)做為實現特定用途的AI系統，或者做為通用AI系統。

(2). 「生成式AI (Generative AI)」：基礎模型中用於生成具有不同自主程度之文章、圖像、音頻或視頻等內容的通用AI系統。

此外，AI法(草案)對AI採取技術中立的定義(a technology-neutral definition of AI systems)，但仍採取基於風險的方法(a risk-based approach)，訂定AI風險程度，包括：(1).不可接受的風險(unacceptable risk)；(2).高風險(high risk)；(3).有限風險(limited risk)；和(4).低風險或極小風險 (low or minimal risk)；並依據AI風險的程度制訂相對應的法律干預規則(如圖 2-1)，包括使用AI系統之前設計和開發的通用強制性要求，以及事後控制。

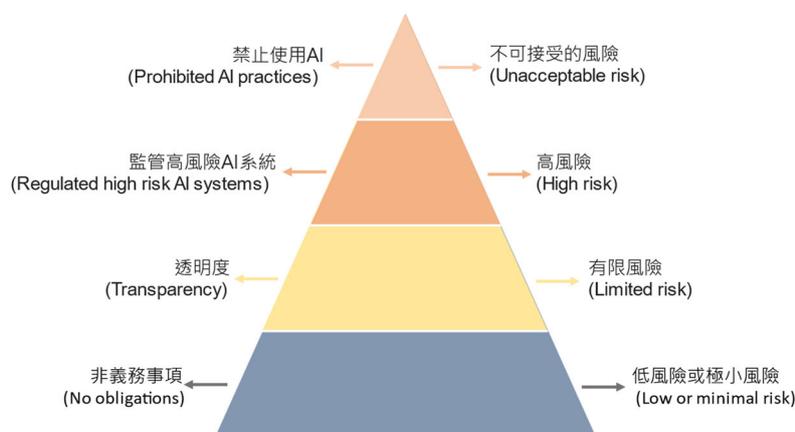


圖 2-1 AI 風險的管制強度

資料來源：EU³⁹

³⁹ 檢自 <https://artificialintelligenceact.eu/the-act/>



8. 中國的 AI 法規，其中最具體、最有影響力的法規及定義如下：

- (1).2021 年第一個針對演算法具有約束性的法規—演算法推薦(recommendation algorithms)的法規：「互聯網信息服務算法推薦管理規定」，其中「演算法推薦技術」的定義為「第 2 條：利用生成合成類、個性化推薦類、排序精選類、檢索過濾類、調度決策類等演算法技術向用戶提供信息。」。
- (2).2022 年深度合成演算法的法規：「互聯網信息服務深度合成管理規定」，其中「深度合成技術」，即深度偽造(deepfakes)，定義為「第 23 條：利用深度、虛擬現實等生成合成類算法製作文本、圖像、音頻、視頻、虛擬場景等網絡信息的技術。」。
- (3).2023 年「生成式人工智能服務管理暫行辦法」^{40 41}：其中「生成式 AI (Generative AI)」的定義為「第 2 條：基於算法、模型、規則生成文本、圖片、聲音、視頻、代碼等內容的技術。」，其涵蓋的內容與深度合成法規幾乎完全相同，但細緻化深度合成的管理規定，是中國對生成式 AI 產業首部監管的法律。生成式 AI 包括演算法推薦技術與深度合成技術，例如合成式生成圖像(synthetically generated images)、ChatGPT 模型的聊天機器人^{42 43}。

由於各種 AI 定義的目的不太一樣，*如果沒有對 AI 定義，政策制定者很難評估 AI 在不久的將來能夠做什麼，以及該領域如何實現這一目標*(Bhatnagar et al., 2018; Wang, 2019)。雖然歐盟委員會聯合研究中心 (Joint Research Centre ,JRC) 與

⁴⁰ 檢自 <http://cpc.people.com.cn/BIG5/n1/2023/0714/c64387-40035514.html>

⁴¹ 檢自 <https://www.tipo.gov.tw/tw/cp-885-927264-397c7-1.html>

⁴² 檢自

<https://indsr.org.tw/respublicationcon?uid=12&resid=1950&pid=3939&typeid=3>

⁴³ 檢自 <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>



通信網絡、內容和技術總局 (the Directorate General for Communications Networks, Content and Technology, DG CONNECT) 合作開發 AI Watch 監測歐洲 AI 的發展、採用和影響，梳理 1955 年至 2019 年間相關文獻 AI 的定義及概念主要特徵，認為歐盟 AI 高級專家組(High-Level Expert Group on Artificial Intelligence, AI HLEG) 的定義是全面的、技術性強且詳細。Zuiderwijk 等人(2021)在探討 AI 在公共治理的應用時亦採用 AI HLEG 的定義(Zuiderwijk et al., 2021)。然而，有研究選取過去五年內發表的作品(2014 年至 2020 年間發表的 AI 學習介入措施的研究論文)，最受認可的人工智慧學習系統(AI-enabled learning systems)是自適應性學習系統(Adaptive Learning Systems) (Kabudi, Pappas, & Olsen, 2021)。自適應性系統(Adaptive systems)是在情境感知機制(context-aware mechanisms)的驅動下自動改變其行為⁴⁴(Fischer, 2023)；而成功的自適應性學習環境需要兩種類型的適應：(1). 在解決問題過程中提供高度具體、即時和有效反饋的能力；(2).根據每個使用者的個人技能熟練程度構建自適應性學習內容的能力(Minn, 2022)等。再者，Gartner 發布 2023 年十大戰略科技趨勢報告，提出 10 項當前關鍵科技，包括自適應性人工智慧(Adaptive AI)，並定義自適應性 AI 系統(Adaptive AI systems)是「支援以更快決策為中心的決策框架，同時在出現問題時保持靈活調整。這些系統旨在運行時根據新數據不斷學習，以更快地適應現實環境的變化。人工智慧工程框架可以幫助編排和優化應用程式以適應、抵抗或緩解干擾，從而促進自適應性系統的管理」⁴⁵。

⁴⁴ 原文：Adaptive systems change their behavior automatically, driven by context-aware mechanisms including models of their users and of specific tasks.

⁴⁵ 原文：Adaptive AI systems support a decision-making framework centered around making faster decisions while remaining flexible to adjust as issues arise. These systems aim to continuously learn based on new data at runtime to adapt more quickly to changes in real-world circumstances. The AI engineering framework can help



因此，本研究認為 Russell & Norvig 的定義：[「AI 是智慧代理人或軟體機器人(intelligent agents)的概念，從環境中獲得感知並採取行動的代理人(agents)，每個代理人(agents)都能夠連續感知並採取行動。」(S. Russell & Norvig, 2002)]不僅簡潔易理解，且與 AI 當前關鍵科技接近。再者，歐盟 AI 法(草案)的定義細述使用的 AI 技術，偏向法律要規範的角度定義(歐盟內部還在討論)，具有使用的名詞外行人比較熟悉的好處。為利跨領域整合，本研究有關 AI 的定義：

本研究使用 AI 的範疇(scope)，包括技術觀點的「Russell & Norvig(2002)」，和風險控制觀點的「歐盟 2023 年 6 月 14 號表決通過的 AI 法(A.I. Act)(草案)」互相補充。

綜上分析，本研究歸納 AI 不同於以往科技的重要特徵：*「自適應性 AI (Adaptive AI)」*：

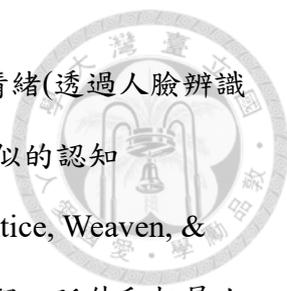
自適應性 AI (Adaptive AI)：需學習(有資料、能回饋)、可調適(能進步、有彈性)、有限制(有偏誤、會犯錯)等特質。

貳、AI 技術進程

1950 年代開始開發 AI，演化出許多理論和技術。根據 AI 功能，AI 可分為分析型 AI (Analytical AI)、受人類啟發的 AI (human-inspired AI)和(即將推出)的人性化 AI (humanized AI)。分析型 AI 表現為認知智慧和預測(forecasting and

orchestrate and optimize applications to adapt to, resist or absorb disruptions, facilitating the management of adaptive systems. From:

<https://www.gartner.com/en/information-technology/glossary/adaptive-ai>



predicting)學習。受人啟發的 AI 旨在識別和理解決策中的人類情緒(透過人臉辨識工具了解人類的情緒)。人性化的人工智慧將擁有與人類非常相似的認知(cognitive)、情感(emotional)和社交智慧(social intelligence) (Prentice, Weaven, & Wong, 2020)。若從人類的角度分析,「弱 AI (weak AI)」只能模擬、延伸和拓展人類的低端智慧(low-end human intelligence),即人類的感受(feelings)、感知(perceptions)以及常規的(conventional)、程序化的邏輯推理(programmed logical reasoning)。「強人工智慧(strong AI)」階段能模擬、延伸、擴展人類的高端創造性智慧(high-end creative intelligence)、想像力、直覺、潛力,以及只能透過行為表達的隱性知識、經驗、技能等非常規、非程序化的個人化智慧(Dong, Hou, Zhang, & Zhang, 2020)。

根據 AI 技術發展程度及在政府應用可分成 3 個層次：

第一代 AI—「弱 AI(the label artificial narrow intelligence, ANI)」

目前可用的 AI 系統被認為是弱 AI(narrows/weak)(Ang, Choolani, See, & Poh, 2023),現在幾乎無處不在,是數位計算機至少能夠表現出「擁有並表現出像人的智慧」或「能夠理性地解決問題以達成目標」⁴⁶(Turing, 2009; Young et al., 2019),為一項或多項特定任務而設計的系統,亦即「專用型 AI (Special-purpose AI, Narrow AI, Weak AI),並且不能轉移到未經訓練的未知和不確定環境的其他領域(McLean et al., 2023)。例如 AlphaGo 就是專用型 AI 系統的里程碑,Facebook 能夠人臉辨識並標記用戶,Siri 能夠理解聲音並採取行動,特斯拉自動駕駛汽車等,都是專用型 AI 系統。目前專用型 AI 在特定領域甚至比人類還厲害。

⁴⁶ 原文：A more technically precise definition narrows AI applications to digital computers that either (1) possess and exhibit human-like intelligence or (2) are capable of rationally solving problems to achieve a goal.

第一代 AI 的公共服務主要利用知識庫(Knowledge base)回應民眾查詢和解決基本問題，例如聊天機器人(Chatbots)。弱 AI 只能說明相關性、可能性，不代表因果關係，因果關係是由人類決定，所以有侷限性。



第二代 AI—「通用 AI(artificial general intelligence, AGI)」

目前的 ANI 系統通常被用作支持人類行為的工具，但通用型 AI (Artificial General Intelligence, AGI；General Purpose AI, GPAI；Strong AI) 將擴展當今廣泛使用的 AI 或弱 AI 系統，是一個**自主能動者或自主代理人(autonomous agent)**，AGI 能夠針對從未設計過的任務自主地推理、計劃和解決問題，而無需任何人工幫助，或無人監督的方式學習(McLean et al., 2023)；能夠使用廣泛的認知能力，而不僅僅是為單一任務複製和模擬人類智慧，預計在各方面超越人類智慧。

AlphaGo Zero 使用類神經網絡的深度學習(Deep Learning)、強化學習(Reinforcement Learning)、自主學習，從感測器收集而來的資料、動作、重要因子與獎賞(reward)來調整能動者(agent)的策略以及一系列的決策，不再需要人類棋手的棋譜數據；因此，AlphaGo Zero 正處於 ANI 和 AGI 之間的過渡階段。目前絕大多數開發者希望藉由開發「專用型 AI」的經驗和成果找出開發「通用型 AI」的橋樑。有研究認為 AGI 仍然是一種理論結構，目前還不存在，有研究則認為預計 2040 年左右發揮作用(Pitt, Paschen, Kietzmann, Pitt, & Pala, 2023)；也有研究認為將在本世紀某個時候出現(McLean et al., 2023)；2023 年 9 月 25 日吳恩達(Andrew Ng)則表示還需要 30 到 50 年才有可能實現⁴⁷；OpenAI CEO Altman 認為 AGI 時代「即將來臨 (reasonably close-ish future)」⁴⁸。

⁴⁷ 檢自 <https://futurecity.cw.com.tw/article/3227>

⁴⁸ 檢自 <https://the-decoder.com/openai-ceo-sam-altman-says-gpt-4-is-just-a-preview-of-whats-to-come/>



第二代公共服務領域將了解民眾在做什麼或需求，並會主動提供服務，例如民眾換住址時，AI 會主動通知民眾要不要換駕駛執照等；公路公共運輸車輛導入公共運輸車聯網 (Connected vehicles, CV)，結合先進駕駛輔助系統(Advanced driver assistant systems, ADAS)⁴⁹，提升公路公共運輸車輛與行人交通安全。

第三代 AI—「超級 AI(artificial super intelligence, ASI)」

牛津大學哲學家尼克·博斯特羅姆(Nick Bostrom)將 ASI 定義為「在幾乎所有感興趣的領域大大超過人類認知表現的智慧」⁵⁰(Zohuri, Zadfathollah, Balgehshiri, & Paydar, 2023)。另有學者認為 ASI 是指機器具有自我意識並超越人類智慧和能力的計算機世界⁵¹，並將 ASI 將定義一個有感知的計算機能夠訓練其他計算機的世界⁵²；亦即 ASI 能超越人類的認知能力、自我升級，學習新任務並以難以理解的速度進步，執行以前難以想像的任務，且具有高度投機的特性。因此，當演算法能「學會如何學習(learn how to learn)」，並超越人類最初設定的程式碼範圍的情況下，AI 真正的全方位超越人類個體和整體的時候，甚至達到所謂的技術奇點

⁴⁹ 應用車聯網與先進駕駛輔助系統於我國公路公共運輸行車安全 檢自
<https://www.iot.gov.tw/cp-2273-201899-18e71-1.html>

⁵⁰ 原文：Artificial Super Intelligence (ASI) is the ultimate level of Artificial Intelligence beyond the capabilities of the human brain! This is the “Third and last phase” of designing and applying Artificial Intelligence to our world. Oxford philosopher Nick Bostrom defines Super Intelligence as “Any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest”.

⁵¹ 原文：Artificial Super Intelligence (ASI) refers to the type of computer world in which machines become self-aware and surpass the capacity of human intelligence and ability.

⁵² 原文：ASI will define a world in which sentient computers will be able to train other computers.



(singularity)，也就是 AI 達到人類無法控制，AI 發展不可控制、無法逆轉與預測⁵³(Pitt et al., 2023)，人類的命運將端視我們到那時候之前，是否已達全面準備(蘇經天, 2023)。

而對於 AGI 到來的猜測各不相同，從「AGI 之後不久」到「不會在可預見的未來」；也有學者認為 ASI 是 AI 超越人腦能力的終極水平，我們的世界設計和應用 AI 的「第三個也是最後一個階段」。

第三代的公共服務不僅可以查詢，還可以基於非查詢、直接資料處理(無需公民採取行動而有新的駕駛執照)、從鍵盤轉向語音和視聽的輸入。又例如在自動駕駛，汽車能夠從其環境中學習做出預測或決策，以及執行，而無需提供各種交通情況明確的程式碼(Mehr et al., 2017; 蘇經天, 2023)。

參、資料治理的重要性：資料科學和 AI 發展關連性

每個 AI 專案都從「資料(data)」開始，資料分析及 AI 彼此相輔相成，正確和適合的數據是 AI 系統的基本驅動力(Thierer et al., 2017; Wirtz et al., 2019)，成功使用 AI 的前提，使得世界上最有意義的資源不再是石油，而是數據⁵⁴(Economist, 2017)。未來數據擁有者將擁有權力，利用 AI 機器提高組織生產力，降低勞動成

⁵³ 原文：Artificial Super Intelligence (ASI) refers to the type of computer world in which machines become self-aware and surpass the capacity of human intelligence and ability. While this is a massive step up from AGI, speculations about its arrival vary from “soon after AGI” to “not in the foreseeable future.” ASI will define a world in which sentient computers will be able to train other computers. Under this scenario, we will approach a technological singularity where, AI growth becomes uncontrollable and irreversible, resulting in unpredictable and possibly undesirable changes to human civilization.

⁵⁴ 原文：the world’s most valuable resource is no longer oil, but data

本的人可能會變得更強大，可能會引發社會、組織和人類之間的重大權力轉移，導致勝者全拿 (the winner takes it all) 所產生社會差距擴大的風險(Seeber et al., 2020)，以及可能隨之而來極端的治理挑戰(Dafoe, 2018)。

因此，政府必須將資料視為發展 AI 的核心要素，組織內必須要有資料科學家等職業，政府在利用 AI 技術之前能夠獲得高品質、具代表性、有價值、公正的、足夠的可用性(availability)數據，有助於提升 AI 正確率、降低 AI 偏誤(bias) (Jimenez-Gomez et al., 2020; Misuraca et al., 2019)，發揮 AI 預期效用。更甚者，AI 系統功能與效率，尤其是基於機器學習(machine learning, ML)的系統，受到數據資料訓練的影響比系統本身設計邏輯還大(Lee & Rich, 2021)。因此，計算機科學常用：「垃圾進，垃圾出(Garbage in, garbage out)」(Desouza, 2018)來彰顯資料的重要性。

資料治理的定義是「組織及其人員定義、應用和監控規則和權限的模式，以指導組織內部和跨組織的數據和演算法的整個生命週期的正常運行並確保其問責制。」，因此，資料治理是對數據管理的權力和控制的行使。(Janssen, Brous, Estevez, Barbosa, & Janowski, 2020)。數據管理包括數據取得(data acquisition)、準備(preparation)、標註(annotation)、整合(integration)和保存(preservation)，為利不同資料庫之間數據的可操作性，必須建立標準化的數據格式(standardization of data formats)、文檔(documentation)、工具(tools)、分析結果(results of analyses)以及其他與這些數據相關的資源(House, 2019)，並滿足「FAIR 原則」：F—可搜尋 (Findable)、A—可近性 (Accessible)、I—互操作性 (Interoperable)、R—可重複使用 (Re-usable)等四大核心原則(Wilkinson et al., 2016)。因此，許多組織正在將資料治理作為一種手段來控制數據品質，以及管理資料相關風險，目的是增加資料的價值並最小化與資料相關的成本和風險，使資料治理符合相關法律和倫理規範，以保證做出值得信賴的決策(Abraham, Schneider, & Vom Brocke, 2019)。

公部門雖然擁有公共和私人提供的巨量且多型態數據資料(multimodal digital data)(Mergel, Rethemeyer, & Isett, 2016)，然而公部門使用大數據發展 AI 仍受到許多限制，包括：

- 1.數據相關性不是因果關係(correlation is not causation)：限制了公共數據在政府運作和公共事務研究中的使用(Mergel et al., 2016)。
- 2.資料缺乏互通性使可用資料庫規模不足：公共部門資訊科技 (IT)系統的孤立性質、缺乏資料標準來控制資料收集、收集方式、儲存格式，限制了政府跨系統整合多個資料庫處理關鍵主題問題的潛力(Campion, Gasco-Hernandez, Jankin Mikhaylov, & Esteve, 2022)。
- 3.可能造成權力差距擴大的疑慮：學術界認為數據可能導致監督、猜疑、對文化的依賴以及複製歷史上的結構不平等，從而加劇了社會不平等，並有可能分裂社會(Henman, 2020)。
- 4.隱私和安全問題：公民對隱私的喪失感到不安，即使政府的開放數據源(open data sources)被去標識化(de-identified)，但已有的研究表明可以很容易地利用多個數據源來重新識別個人資料，使公部門使用數據資料集(digital data sets)產生隱私和數據匿名(digital anonymity)的問題。因此，為保護參與者個人資料之隱私，歐盟於 2016 公布「個人資料保護規則(The General Data Protection Regulation, GDPR)」訂定個人資料去識別化規範，包括匿名化(anonymous)、假名化(Pseudonymisation)⁵⁵。我國則有「個人資料保護法」、「人體生物資料庫管理條例」等，對政府在公務機關依職權蒐集當事人與衛生福利相關之資料進行相當程度之規範。

⁵⁵ 去識別用語不一致(吳全峰、許慧瑩, 2018)。



基於以上種種限制，使得儘管資料治理通常被認為是非技術問題⁵⁶，但它卻成為是否能夠順利採用 AI 的關鍵因素(O'Leary, 2014)。

肆、AI 風險的重要來源之一：AI 的偏誤(bias)

本研究係屬跨領域，在理論或概念上，社會科學領域 AI 技術物的偏差(bias)、人類偏見(bias)，與在資訊科學 AI 產品的偏誤(bias)⁵⁷不同，為利跨學科理解，本文後續將依據不同脈絡採「偏差(bias)」、「偏誤(bias)」，以及「偏見(bias)」方式交替使用。

在社會科學領域，「偏差(bias)」的概念不含主觀價值判斷，而是敘述性(descriptive)詞彙，表示一種涉及相關行為者間的特定利益分配方式。任何的法律條文、法律慣例、意識型態、政治過程、社會價值或組織制度的安排，都會隱含特定的偏差(bias)，理性行為者必然權衡各種利害關係，在聯合和衝突的過程中各種相關偏差(bias)的比較分析和抵換(trade-off)，選擇較有利的偏差(bias)或「偏差的管理」(蕭全政, 1990, 1991, 1992)。

在資訊科學領域，機器學習有 5 種偏誤(bias)，包括演算法偏誤(algorithmic bias)、樣本偏誤(sample bias)、偏見偏誤(prejudice bias)、測量偏誤(measurement bias)、排除偏誤(exclusion bias)(TDS, 2021)等。AI 演算型決策不可避免的產生使用者不易了解的偏誤(bias)，本研究以演算法偏誤(algorithmic bias)為例，進一步說明偏誤(bias)可能產生的原因。

演算法是程式執行、資料處理和其他的計算⁵⁸，在不同的脈絡中執行最佳任

⁵⁶ 資料治理非技術問題，但實務上數據仍然需要技術工具來協治理。

⁵⁷ 資訊科學，預測值跟實際值的差距是 bias，實驗中觀察值跟真實值的測量差距是 error。

⁵⁸ 原文：the algorithm is the thing that programs implement, the thing that gets data



務(R. K. Hill, 2016)，具有識別模式(patterns)、生成知識、產生洞察力、客觀性和準確性，不會有人類的偏見(the bias of humans) (Klinger & Svensson, 2018)。那麼，為何會產生演算法偏誤(algorithmic bias)？

一、演算法本質的瑕疵：

技術限制(technological constraints)、錯誤(errors)或設計決定(design decisions)的技術偏誤(bias)(Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016)，包括：

- 1.演算法立論偏好：同一個問題的解法，用立論基礎不一樣的演算法，例如決策樹、貝式定理等。
- 2.演算法門檻值：多數演算法都設有門檻值，特徵選擇(the choice of features)、權重(weights)和目標函數(objective functions)等的設定或選擇(Kordzadeh & Ghasemaghaei, 2022)影響答案。決策問題是選擇演算結果，並非選擇門檻值。

二、演算法自主學習的偏好(preferences)：

機器學習賦予演算法具有某種程度的自主性，可以通過以往的經驗或資料，自主定義或修改決策規則(Mittelstadt et al., 2016)，自動改進電腦的演算法，使得AI具有主動性(active)。再者，任何系統的行為，無論是人類還是非人類，只要足夠複雜以保證它的存在，就可以在某種不同的意義上被視為有意的(intentional) (Johnston, 2001)，而自主(autonomy)或自由程度(freedom)取決於系統的複雜性或目標，並因情況而異(Boychenko, 2019)。因此，AI系統可能超過初始程式設計者或政策規畫者的預期，產生有意(intentional)、無意(unintentional)或無法預期(unexceptional)的結果。

processing and other computation done.



三、模型設計和訓練的認知壓縮：

為節省儲存空間和算力，使用各種壓縮技巧，例如權重共享、權重量化、知識蒸餾等，來減少模型的大小和計算需求。認知壓縮主要是訊息的篩選與處理，必然無法保證資訊的完整性，以及造成資訊的漏失和扭曲(蘇經天, 2023)。

從資訊科學領域學者或程式設計者的觀點，演算法設計過程中每一步的選擇都是當前狀態下最好或最佳(即最有利)的選擇，對整體而言雖然未必是最佳的解決方案，因此常常認為機器學習的偏誤(bias)是技術本質，並不帶任何價值判斷的偏見。

然而，從社會科學領域學者的觀點，由於演算法(algorithm) 沒有正式統一定義，Hill 將的演算法定義為「一個數學結構與『一個有限的、抽象的、有效的、複合的控制結構，在特定的規則(provisions)下完成特定的目的(purpose)』」⁵⁹。Hill 進一步論述，演算法包括數學結構、執行(技術、程式)和構想(應用程式)⁶⁰(R. K. Hill, 2016)。因此，演算法依據「規則(provisions)」和「目的(purpose)」實施和執行將抽象的數學結構融入技術中去執行特定的任務，並認為操作參數由開發人員指定的，構想是由使用者考慮到某些價值觀和利益優先於其他價值和利益所提出的，演算法不可避免的具有價值觀(value-laden) (Mittelstadt et al., 2016)。因此，「AI 演算法決策從來都不是中立的⁶¹」(Klinger & Svensson, 2018)，或是「AI 所產生的權力效果對公民而言都不是中立的」(Kuziemski & Misuraca, 2020)，意旨

⁵⁹ 原文：a mathematical construct with “a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions.”

⁶⁰ 原文：algorithms as mathematical constructs, implementations (technologies, programs) and configurations (applications)

⁶¹ ADM is never neutral. From 檢自 <https://algorithmwatch.org/en/what-we-do/>

AI 技術在政治或價值觀方面永遠不會是中立的(Waldherr et al., 2019)

綜上分析，在資訊科學領域，演算法常常是中立的；在社會科學領域，演算法具有價值觀(value-laden)，雖然「理念」與「權力」之爭，才是政治社群的本質(林子倫、陳亮宇, 2009)，與其陷入不同領域的觀點之爭，不如回歸到行政實務上的考量，將技術限制與行政實務相結合，控制 AI 可能產生的風險，降低 AI 可能產生偏誤的效果。

伍、AI 能動性(agency)：對人的行政裁量權產生約束和促進作用

在科學科技與社會(STS)領域的 AI 技術物及其能動性(agency)，與在資訊科學領域 AI 是智慧代理人(intelligent agents)及具有代理關係(agency relations)的理解不同。在 STS 領域，「能動性(agency)」被理解為對事件的影響或修正，從這個意義上說，能動性是有影響力並且可以產生影響的一切事物⁶²，也就是說能動性(agency)並不意味著「意圖、自主或責任」，而是事物通過對他人產生影響來「行動(act)」：「事物可能會授權、容許、提供、鼓勵、允許、建議、影響、阻止、使成為可能、禁止等等」⁶³。因此，是否有能動性(agency)，要問任何能動者(agent)的問題如下：是否對其他能動者的行動(agent's action)產生影響？⁶⁴(Latour,

⁶² 原文：Agencies are always presented in an account as doing something, that is, making some difference to a state of affairs, transforming some As into Bs through trials with Cs.

⁶³ 原文：Agency in terms of ANT does not imply intentionality, autonomy, or responsibility. Things “act” by having an impact on others: “Things might authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, forbid, and so on”.

⁶⁴ 原文：Thus, the questions to ask about any agent are simply the following: Does it make a difference in the course of some other agent's action or not? Is there some trial that allows someone to detect this difference? The rather common sense answer should

2005; Sayes, 2014)。因此，在公共治理領域，*AI 能動性(agency)* 可以被理解為 *AI 演算型決策的作用(agency)* 對人類行為所產生的影響，例如 *AI 能動性(agency)* 對人的行政裁量權產生約束和促進作用。

在資訊科學領域，智慧代理人是用一個軟體來執行一個代理人的功能，主體是人，代理人就是仲介的功能；代理關係是指一個或一個以上的委託人(principal)，雇用並授權給另一個或一個以上之代理人(agent)，代其行使某些特定的行為，彼此間所存在的契約關係(Jensen & Meckling, 1976; 柯于璋, 2013)。本文後續將依據不同脈絡採 AI「能動性(agency)」、AI「代理(agency)」、AI「作用(agency)」方式交替使用。

技術與社會理論(Technology and social theory)以技術的理論化(theorizing technology)為背景，認為技術不單純是權力關係的承載者，技術本身就是強大的行動者，技術協助創造出權力關係，應該視技術為權力本身，因此，有必要了解非人類能動性的可能。接續在 STS，ANT 主張非人類也有能動性(agency)，視技術物為「承擔責任、表述承諾，並且在人類關係的疆域負起能動者角色」(王志弘、高郁婷譯, 2023)，將「非人類」⁶⁵行動者整合到我們對社會世界的理解(Waldherr et al., 2019)，我們與物質能動者一起塑造了集體，也藉由且通過(with

be a resounding 'yes'.

⁶⁵ 依據 ANT 的術語，「非人類(nonhuman)」作為一個總稱，用於涵蓋範圍廣泛但最終範圍有限的實體(entities)，例如事物(things)、物體(objects)、動物(animals)、自然現象(natural phenomena)、工具和技術人工製品(tools and technical artifacts)、文本(texts)和經濟商品(economic goods)等，被排除在這個術語之外的是人類(humans)、本質上完全是象徵性的實體(entities that are entirely symbolic in nature)、超自然的實體(entities that are supernatural)，以及人類和非人類組成的實體(entities that exist at such a scale that they are literally composed of humans and nonhumans)(Sayes, 2014)。



and through)技術而彼此關聯(王志弘、高郁婷譯, 2023)。

以實際發展現況觀察，AI 成功執行複雜任務範圍增加，特別是在圖像識別、語音識別、自主駕駛系統、GPT 生成式語言模型等特定知識領域。AI 自主學習(self-learning)、自主開發系統(self-developing systems)功能越強與正確率越高，人類更加依賴 AI，AI 對人的影響越大，特別是在演算法文化(Karppi & Crawford, 2016)、社交媒體平臺政治和倫理(Gillespie, 2010)以及社交機器人干預(Guilbeault, 2016; Klinger & Svensson, 2018)的實證研究顯示，AI 對人產生約束和促進作用，從而衍生出「削減論點(curtailment thesis)或「賦能論點(enablement thesis)」(Buffat, 2015; Busch & Henriksen, 2018)。

因此，若只關注人類能動性(agency)，在分析非人類和人類的互動和依賴關係就幾乎無法獲得全面性的理解，越來越多的研究人員開始認真對待非人類，並開始將非人類能動性(nonhuman agency)理論化，因此，在概念上區分非人類和人類行動者的能動性(agency)就變得越重要(Waldherr et al., 2019)。

雖然有研究認為演算法缺乏反思能力(reflexive capacity)，不能處理意外情況，進而反駁 AI 有能動性(agency)，並主張 AI 無法超越程式設計師/軟體工程師所期望的結構和方向。例如 AlphaGo 演算法在某些方面能夠預測(anticipate)和計畫(project)，但它仍然無法超越自身，或超越設計者預期的圍棋結構或圍棋方向。AI 演算法是人類和機構的能動性(agency)所塑造，根據人類意圖進行設計，雖然人類無法預測其設計和行為的所有連鎖反應，但很大程度上仍然是人類行為的結果，爭論演算法本身有能動性不僅會掩蓋人類、結構和關係背後的權力和利益，更可能使人類逃脫課責，與其擔心機器不如擔心機器背後的人(Klinger & Svensson, 2018)。

然而，ANT 提醒我們，只關注人類的能動性幾乎無法完整理解人與 AI 的權力互動關係，雖然我們確實應該檢查影響 AI 背後的人類的條件和結構，但從

ANT 的角度來看，觀察與理解 AI 的行為模式(behavioral patterns)和規範標準(normative standards)同樣重要(Waldherr et al., 2019)，如此才能對行動者網絡內人和 AI 的關聯變化進行批判性分析。因此，本研究將依循 ANT 對非人類(nonhuman)和能動性(agency)的觀點：

AI 也有能動性(agency)。

AI 能動性(agency)意旨：「AI 技術物(AI artifacts)對人類行動產生影響」。

依據權力的觀點，一般的說法是「一組加在其他行動上的行動(a set of actions upon other action)」(李承嘉、廖本全、戴政新, 2010)，例如高權力者使低權力者做出其原本不會做出的行徑。而權力的產生，不是依繫行動者既有的權力，而是依繫行動者之間建立起來的關係，只要轉譯完成及網絡穩固，權力就成為結果並浮現出來(Latour, 1984; 李承嘉, 2005; 官美吟、彭立沛, 2018)。因此，權力是根據關係來定義的、是一段關係中雙方對彼此影響力的作用結果(Dahl, 1957; 李怡青, 2012)、是通過一個網絡來運作、是通過網狀的組織而施展和運作、是一系列權力(power)、規範(normative)和溝通(communicative)關係的結果(Johnston, 2001)，權力唯有通過具體配置(agencement concrets,)才有可能(王志弘、高郁婷譯, 2023)發生。行動者和網絡被認為是可逆：「網絡由其行動者和關聯(associations)所定義，行動者完全由其網絡定義，這裡的網絡一詞聚焦於行動(action)和能動性(agency)的再分配。」(Latour, 2011)。一旦技術到位，黑箱(black-boxed)產生，演算法和人類協作之間潛在的權力結構就會隱藏起來⁶⁶(Ramati &

⁶⁶ 原文：the underlying power structure of algorithmic and human collaboration that becomes hidden once a technology is in place—thus, black-boxed.

Pinchevski, 2018; Waldherr et al., 2019)，在具體脈絡中黑箱化(black-boxed)⁶⁷成特定客體(Latour, 1987; 林文源, 2017)。

由於公共治理的範圍極為廣泛，AI 能動性(agency)對人類行動產生影響的面向也很多，本研究聚焦探討人與 AI 的互動關係，亦即：

AI 能動性(agency)對人的行政裁量權或決策產生影響。

陸、AI 的政治性與對公民權利義務的影響

馬克思是社會理論家中最早觀察人造物的政治性(王志弘、高郁婷譯, 2023)，馬庫色(Herbert Marcuse)定義技術是一種控制與支配的工具(王志弘、高郁婷譯, 2023)—對自然和對人的支配，將支配的目的和利益透過技術投射進技術裝置本身的構造中(Habermas, 1968; Klinger & Svensson, 2018)。美國技術哲學家蘭登·溫納(Langdon Winner) 1980 年的論文「技術物有政治性嗎？」，是科學與技術研究(STS)領域最著名的論文之一，以公部門建築師羅伯特·摩西 (Robert Moses)在長島(Long Island)公路沿線的建橋計畫「摩西低橋」為案例，提出技術物具有政治特性這一概念，說明摩西利用技術限制特定社會利益，刻意降低的橋梁高度，反映摩西的社會階級偏見和種族偏見。

羅伯特·摩西 (Robert Moses)是 1920 年代至 1970 年代紐約的道路、公園、橋樑和其他公共工程的主要建造者，儘管摩西宣稱瓊斯海灘是一座公眾海灘，然而長島上的兩百多座低橋是有意設計的，目的是為了達到特定的社會效果的技術：

⁶⁷ Latour 所指黑箱與資訊科學領域所理解的黑箱概念不同，對資訊科學領域的人來說，黑箱的概念未必代表被隱藏或無法解釋，例如黑箱預測器(A black box predictor)是一種資料探勘(data-mining)和機器學習的模糊模型(machine-learning obscure model)，其內部結構要麼觀察者未知，要麼人類已知但無法解釋(Guidotti et al., 2018)。

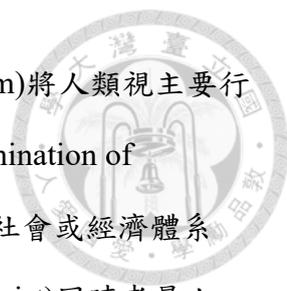


12 英尺高的公共汽車無法通過低橋，使用公共交通工具的窮人和黑人被禁止上路，據以限制少數族裔和低收入群體進入瓊斯海灘。兩百多座天橋成為固化的有形基礎設施，這些巨大的混凝土和鋼結構，看似無害的橋樑，突然變成物質化種族歧視意圖的承載者，產生最具壓迫性的立法行動及系統性的社會不平等 (Winner, 2017; 林崇熙譯, 2004)。

溫納摩西低橋案例提出一個非常重要且具有挑戰性的主題：技術物真的是中立的嗎？溫納主張現代物質文化的機器、結構和系統，除了可以提升效率和生產力、對環境產生正面和負面影響外，技術甚至還可以展現特定形式的權力(power)和權威(authority) (Winner, 2017)。摩西低橋案例衝擊著技術中立性的觀點，同時也引起「技術的社會決定論(social determination of technology)」的抨擊。

從技術中立性的觀點認為：

1. 技術工具論(technological instrumentalism)：認為技術是中立的，社會文化因素對技術的影響是次要的，因為工具只作為手段服務於特定目的，與價值無涉 (賴曉黎, 2012)；或是「技術過程和程序是中性的……它們，或者更準確地說，它們所服務的目的，需要『道德審查』」 (Garvey, 2007; Grant, 1995)；
2. 技術決定論(technological determinism)：技術作為首要的行動者，技術是自主的，具有決定性力量，技術塑造社會關係，並造成社會變遷，人類僅是次要的(王志弘、高郁婷譯, 2023)。因此，主張技術發展獨立於社會、經濟或政治力量(Wyatt, 2008)，技術遵循內部功能邏輯自主發展(Feenberg, 1992)，或是「社會被迫適應技術的轉變，進而限制歷史的軌跡」 (Sismondo, 2010)，或是技術具有自主性或所謂的科技動量(technological momentum)(Hughes, 1993; 賴曉黎, 2012)，只有當技術被引進社會時，政治才進場。



面對技術中立性觀點挑戰，社會建構論(social constructionism)將人類視主要行動者(王志弘、高郁婷譯, 2023)；技術的社會決定論(social determination of technology)則認為：「重要的不是技術本身，而是技術所嵌入的社會或經濟體系⁶⁸」。比起偏袒任一種形式的決定論，後人類主義學派(anti-humanist)同時考量人、技術、同伴物種、非人能動者等，同時接受了技術和人類的能動性，其所倡議分散的能動性(distributed agency)，能動性分布於所有被視為行動者的對象之間(王志弘、高郁婷譯, 2023)。

摩西低橋案例很有說服力的顯示堅持技術中立的立場非常天真，充分論證了技術能動性、人造物的政治性(王志弘、高郁婷譯, 2023)。溫納為他的觀點提出論辯：使用技術之前，技術的設計、製造與安排的重要性，技術的安排將從根本上改變了權力的行使和公民的體驗；技術可以以增強某些人對權力、權威和特權的使用；技術變革是人類的各種動機的華美服飾，絕對不只是某些人希望支配其他人；技術有時會完全超越了「有意(intended)」和「無意(unintended)」的後果，技術有可能遠遠超出其直接使用的目的；「技術」的東西是在我們的世界中建立秩序的方式，而非世界中的單純客體，並存在著傅柯所謂的「權力的技術(technologies of power)」(王志弘、高郁婷譯, 2023)，只有當某個特別的儀器、系統、或技術剛被引進時，人們還有最大的自由度，由於自由度是與某種有形設備、經濟投資、與社會習慣等相依相隨的，因此，決定採用某種技術後，原有可能的使用彈性也就隨之消失，在此意義下，技術創新與新法案或新政黨的成立相類似，都為社會秩序建立起一套延續許多世代的架構。因此，各種技術系統在現代政治條件下深深地交織在一起。而溫納所說的「政治」是指「人類關聯中權力

⁶⁸ What matters is not technology itself, but the social or economic system in which it is embedded.

和權威的安排以及在這些安排中發生的活動」⁶⁹。

溫納(1980)進一步說明技術物可以包含政治屬性的兩種方式。首先是特定技術設備或系統的發明、設計或安排成為解決特定問題的一種方式。其次是可以稱為「天生帶有政治意味的技術(inherently political technologies)」(林崇熙譯, 2004)的技術物，即似乎需要或與特定類型的政治關係強烈兼容的人造系統，這也是技術物最具政治性之處(Winner, 2017)。

而對於另一面向「技術的社會決定論(social determination of technology)」學派的抨擊，溫納則回應，與其堅持將一切歸結為社會力量的相互作用，不如建議我們關注技術對象的特徵以及這些特徵的含義，作為對技術的社會決定理論(theories of the social determination of technology)的**必要補充，而不是替代**，這種觀點將某些技術本身無法將技術自社會脈絡抽離(王志弘、高郁婷譯, 2023)，而將技術確定為政治現象(Winner, 2017)。

另一面向，哲學家詹姆斯·加維(James Garvey)也認為工具和手段也必須有價值載荷，不存在絕對中立的情形，任何手段都必須符合道德，牽涉價值關聯，「技術是手段」不能等價於技術中立性。因為技術物是獨特的創造物，旨在以特定和有限的方式發揮作用，技術物可能有多種用途這一事實並不會削弱技術物具有主要或一組主要用途的信念，以「槍」來說，「準備好放在手邊」，即使沒有實質上的使用，仍然有恫嚇效果，因此主張技術具有政治性，以及技術在某種程度上是自主的，不受我們的控制(Garvey, 2007)。綜上分析，本研究主張：

AI 技術物(AI artifacts)具有政治性。

⁶⁹ By "politics," I mean arrangements of power and authority in human associations as well as the activities that take place within those arrangements.



AI 技術物(AI artifacts)具有政治性(Wong, 2020)，可以舉出很多案例。2016年美國總統大選時，劍橋分析(Cambridge Analytica)⁷⁰事件：劍橋分析蒐集美國選民的數據，利用 AI 認知技術(Cognitive technology)誘導人們以不同的方式思考或採取不同的行動(Yeung, 2018)，亦即不是透過理性說服(Reijula & Hertwig, 2022)形塑偏好來影響決策和行為，而是利用認知技術形成所謂的「數位推力(Digital nudging)」、「智慧推力(Smart nudging)」(Mele et al., 2021)，操弄輿論時對症下藥，讓川普從不被看好到打贏選戰。因此，AI 以不會造成任何明顯傷害的方式，利用人類的慾望、需要和激情，或使我們的注意力偏離真正的議題，或提供誤導性證據，讓人們以新的、意想不到的方式理解和概念化世界，進而對政治產生影響(Mittelstadt et al., 2016)。

另一案例，美國最高法院 Rucho v. Common Cause 案，利用 AI 判斷黨派「政黨傑利蠟蝟(PARTISAN GERRYMANDERING)」的現象。黨派立法者以 AI 繪製不公平的選區來操縱選舉結果，儘管選民仍然能夠投票，但他們卻無法影響選舉，選民被剝奪了選舉權，顯示 AI 技術已侵蝕人的領域。而本案例最後，法院也是利用 AI 提供司法上可管理的標準來判斷選區地圖是否過於偏黨派因素，開始啟動去除黨派或政客控制的因素(Tam Cho, 2019)，這案例不僅展現 AI 技術物政治性的特徵，更充分展現 AI 技術物雙重用途(dual use)。

英國脫歐公投(Brexit referendum)社交機器人(Social bots)則是另一個經典案例，AI 所驅動的社交媒體影響公民的情緒(affect)、價值(appreciation)和判斷(judgment)，從而擴大留歐者(Remainer)與脫歐者(Brexiters)的政治對立(political polarization) (Bouko & Garcia, 2020)。英國脫歐派(the Leave)利用殭屍網絡(Botnet)和使用者生成超黨派新聞(User-Generated Hyperpartisan News)，經過社交媒體上模

⁷⁰ 檢自 <https://www.inside.com.tw/feature/documentary-and-digital-world/23967-the-great-hack-review>



仿人類活動的政治運動社交機器人(Social bots in Political Campaigns)自動生成、轉推的功能，機器人成為有效資訊傳播者(Bastos & Mercea, 2019)。雖然 AI 所驅動的社交媒體究竟是否會影響公投結果，有研究指出如果不是最後一刻 AggregateIQ (AIQ)等「AI 政治代理人(agents of AI politicking)」的 AI 負面廣告政治方法，結果可能會有所不同(Moss, Robinson, & Watts, 2020; Wogu et al., 2020)。再者，根據國際人權組織「自由之家」(Freedom House)的 2023 年的報告，去年至少有 16 個國家利用生成式 AI 來散播懷疑、抹黑對手或影響公眾辯論。如何避免 AI 製造假新聞干擾選舉？因此，有研究主張，付費網路政治訊息並非新聞，乃是在新介面呈現的政治廣告，應強制揭露及公開資金與資訊流足跡，提升選民識別能力(陳柏良, 2021)。

2017 年牛津大學研究員維亞切斯拉夫·波隆斯基(Vyacheslav Polonski)在世界經濟論壇的文章斷言，AI「默默地接管了民主(silently took over democracy)」，AI 所驅動的數位廣告、社交媒體平台的力量和大眾傳播破壞者(機器人和巨魔)，從針對性政治廣告到面部識別等，利用自動化和演算法系統本身或副作用，以線上和離線的方式影響政治系統、選舉、決策和公民(citizenship)的意願(will)和意識(awareness)(Unver, 2018)。因此，AI 已對民主治理有效性(effectiveness)、合法性(legitimacy)和社會正義(social justice)的價值(Fung, 2015)產生影響。

柒、結語：在公共治理領域中將 AI 技術物(AI artifacts)理論化

從溫納提出「技術政治理論(The theory of technological politics)」；肯尼迪大法官(Justice Kennedy)撰寫：「技術既是威脅，也是希望」⁷¹；克蘭茲伯格第一定律(Kranzberg's first law)：「技術沒有好壞之分，也不是中立的(Technology is neither

⁷¹ 原文：Technology is both a threat and a promise. From：檢自

<https://www.law.cornell.edu/supct/html/02-1580.ZC.html>



good or bad; nor is it neutral)」等等均顯示，科技的影響不僅是正面或負面的，而且是矛盾的，其對人類和社會的影響往往遠遠超出科技的預期目的(Kranzberg, 1991; Pitt et al., 2023)，目的是提醒大家，認真對待技術物的觀點：**使用技術之前，技術的設計、製造與安排的重要性**。技術的最終目的，應該是充滿人類價值的生命(王志弘、高郁婷譯, 2023)，透過源頭管理，避免 AI 技術物的偏誤(bias)、能動性(agency)、通用技術、雙重用途、政治性的特徵，對政治文化和權力分配濫用與操控造成權力差距擴大(Cath, 2018)。

雖然歐盟 2023 年 6 月 14 號表決通過的 AI 法(草案)，對 AI 採取技術中立的定義(a technology-neutral definition of AI systems)，但仍採取基於風險的方法(a risk-based approach)，訂定 AI 風險程度，因此，本研究將以「技術物有政治性」，以及技術不單純是權力關係的承載者，技術能創造出權力關係，技術為權力自身；以及思考技術時，應注意所有權(ownership)、控制權(control)、近用權(access)、使用後果(use)、非意圖後果(unintended consequence)(王志弘、高郁婷譯, 2023) 等觀點，將 AI 技術物確定為具有權力、產生政治現象，關注 AI 的特徵與含義、可能產生的影響，作為對公共治理理論的補充。



第二節 公共治理理論

當今公共治理理論以人為本、以人為中心的治理典範或分析框架，都以理性而自利的個人作為立論的起點與基礎，因而都有偏好、理性選擇等假設；然而，現行公共治理理論無法有效解釋 AI 與人共存的現象，似乎隱含另一次典範轉移的可能性。

「典範(paradigm)」代表一套基本信仰系統，其植基於本體論、知識論與方法論之預設。一個典範提供了問題框架、探究線索、研究方法、重要範圍、以及最終意義。因此，任何學術典範都隱含特定的假設、概念、命題、理論、模型、分析方法與所能處理的問題。孔恩(Kuhn)主張知識並非以線性漸進方式進展，而是通過一連串改變場域的革命而開展，當特定典範無法處理重大問題時即可能發生危機，而由另一個能處理該重大問題的典範所取而代之，他稱這些激烈變化為「典範轉移(paradigm shift)」(王志弘、高郁婷譯, 2023; 潘慧玲, 2003; 蕭全政, 1990)。

為解決公共治理理論長久以來對技術的忽視，必須從方法論上徹底檢討才能瞭解問題與限制。本節的目的即希望透過相關概念的討論與界定，消除在公共治理理論與技術分立的現況，建立更具包容性的理論與概念框架。

壹、公共治理理論的概念和定義

公共治理的概念和定義受到不同學科部門、研究目的、研究人員的地理和文化背景、行政傳統和改革趨勢影響而具有高度複雜性，因此，公共治理仍然是一個不精確和模糊的模型(model)，很難找到一個共同的和被廣泛接受的定義(Cepiku, 2013)，經梳理文獻各種公共治理定義和概念，包括：



1.世界銀行(1992)將治理定義為「行使權力管理一個國家的經濟和社會資源以促進發展⁷²。」，主張「善治(Good governance)體現在可預測的、公開的和開明的決策(即透明的過程)；充滿專業精神的官僚機構；對其行為負責的政府執行機構；以及參與公共事務的強大民間社會；所有人都在法治之下行事⁷³。」(Kulshreshtha, 2008)。

2.Kooiman and van Vliet (1993)區分公共治理三個分析領域：微觀(單一公共行政)、中觀(相互關聯的行政系統)和宏觀(社會經濟環境)，在這個多層次的框架中，公共行政部門負責必須確保能滿足需求的績效水平(Cepiku, 2013)。

3.Rhodes 提出治理是一個比政府更廣泛的概念，是網絡(networks)扮演核心作用，國家空心化(hollowing out of the state)的過程(Cepiku, 2013)。Rhodes (1996)進一步提出治理的六種用途：最小國家(the minimal state)、合作治理(corporate governance)、新公共管理(New Public Management, NPM)、善治(good governance)、社會控制系統(socio-cybernetic system)和自組織網絡(self-organizing networks)(Rhodes, 1996)。

4.Stoker (1998) 指出治理理論源自於制度經濟學(institutional economics)、國際關係、組織研究、發展研究、政治學、公共管理和受福柯啟發的理論家

⁷² 原文：the manner in which power is exercised in the management of a country's economic and social resources for development.

⁷³ 原文：Good governance is epitomized by predictable, open, and enlightened policymaking (that is, transparent processes); a bureaucracy imbued with a professional ethos; an executive arm of government accountable for its actions; and a strong civil society participating in public affairs; and all behaving under the rule of law.



(Foucauldian-inspired theorists)(Stoker, 1998)。

5.Kettl (2000) 提出包括國家的「空心(hollowed-out)」觀點和網絡的擴散，並指出傳統的、縱向的、等級(hierarchical)的關係與新的橫向關係並存(Kettl, 2000)，並認為「*治理是政府把事情做好的方法*⁷⁴。」(Kettl, 2006)。

6.Bovaird & Löffler 指出：*公共治理的目的不僅僅是嘗試確保內部的良好管理，更關注那些受到外部利益相關者重視的公共政策成果*⁷⁵(Bovaird & Löffler, 2001; Bovaird & Löffler, 2002)。

7.Schedler (2003)將公共治理定義為「確保民主決策和可靠公共行政的結構和過程」。

8.Lynn 等人指出治理是縱向和橫向工具的混合體，*協調和指導不同行為者能夠實現共同目標；治理包括「公眾支持的體制、法律、法規、司法決定、行政作為*⁷⁶*的限制、規定和促成。」*(Cepiku, 2013; Lynn Jr, 2006; Lynn Jr, Heinrich, & Hill, 2001)，等級治理(hierarchical governance)仍然是主要的治理手段(C. J. Hill & Lynn, 2004)。

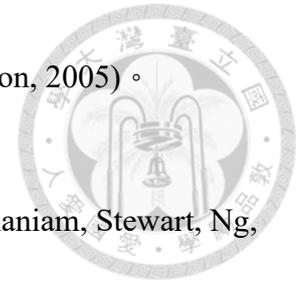
9.Frederickson (2005)表示治理體制的成果受到環境因素、委託人(client)特徵、

⁷⁴ 原文：Governance is the way government gets the job done.

⁷⁵ 原文：Governance is the way government gets the job done.

⁷⁶ 原文：regimes, laws, rules, judicial decisions and administrative practices that constrain, prescribe and enable the provision of publicly supported goals and services.

組織的核心流程、結構、管理角色和行動的影響(Frederickson, 2005)。



10. 政府在維繫公共利益方面仍然是最重要的角色(Subramaniam, Stewart, Ng, & Shulman, 2013)。

11. 蘇彩足等人(2017)指出公共治理理論的前提是國家的作用和能力都是有**限的**(蘇彩足 et al., 2017)。

12. Alqooti (2020)指出公共治理，係指政府使用經過驗證的管理方法，利用行政改革，提供可靠和靈活性的服務以實現最大效率和創新(Alqooti, 2020)。

13. Zuiderwijk et al. (2021)提出的概念：治理被定義為「制定和執行規則，並提供服務」；公共性為「公共產物和服務的生產和遞交」；公共治理的主要行動者參與集體決策，以實現公共治理的目標，這些行動者包括公共、私營和非營利部門的個人、公民、組織和組織系統。最後將公共治理定義為「與公共政策和服務相關的所有規則和行動」⁷⁷(Zuiderwijk et al., 2021)。

⁷⁷ 原文：Governance is defined as the activity to “make and enforce rules, and to deliver services”. Publicness is defined by the object of governance as “production and delivery of publicly supported goods and services”. The main actors involved in public governance encompass individuals, citizens, organizations, and systems of organizations in public, private, and nonprofit sectors. These actors engage in collective decision-making that is constrained, prescribed, and enabled by laws, rules, and practice to achieve the object of public governance. Building on these existing definitions, we define public governance as all the rules and actions related to public policy and services.



貳、本研究使用公共治理理論的範疇(scope)

很難想像任何涉及政府、政治或行政管理的東西都不是治理，但也正因為如此，也增加了公共治理概念上的模糊性和解釋的靈活性(Cepiku, 2013)，因此，**本研究將以政府是使用者為主要分析範圍**，並參考世界銀行(1992)、Bovaird & Löffler(2002)、Zuiderwijk et al. (2021) 和蘇彩足 et al. (2017)的定義，提出本研究所採用公共治理意旨：

「公共治理理論的前提是國家的作用和能力都是有限，以及公部門管理國家時行使權力的方式，以及公部門的制度、流程、環境、系統的公共管理原則，可以確保內部的良好管理、提升公共服務效能、最大化利益相關者價值。」

公共治理效能包括：

- 1.微觀：個人效率；
- 2.宏觀：組織管理、組織流程、組織效能、組織效率、公平性、可管理性、合法性和政治可行性等。

參、AI 使公共治理理論解釋力不足？

公共治理真的忽略了技術嗎？ A. Meijer & Löfgren(2015) 為檢驗這一假設，將技術與政府政策之間的關係分為 3 大類：

- 1.技術作為自變量，即技術決定論(technodeterminism)。
- 2.技術作為因變量，即社會決定論(socio-determinism)。
- 3.技術與政策之間的相互依存關係。

對政策科學領域的主要期刊進行了文獻回顧，證實技術作為變量通常被忽



略，即便討論技術，大部分是將技術概念化為自變量、或因變量，亦即技術決定論的觀點或社會決定論的觀點，技術與政府政策之間的相互依賴關係較少，且這些研究大多來自科學科技與社會(STS)的研究，這凸顯了在政策理論中需要對技術與政策之間的關係採取更複雜的觀點，在政策變革中將技術的作用理論化(Meijer & Löfgren, 2015)。

在政策科學領域，主要仍以研究制度結構(institutional structures)、偏好(preferences)、互動模式(interaction patterns)、社會經濟發展(socioeconomic developments)和理念(ideas)，以了解政策選擇，技術很少受到關注。如果考慮技術，它被概念化為一種工具(Meijer & Löfgren, 2015)，技術被視為對現有組織和行政作為的補充，幾乎不被視為能夠引起政策方向變化、組織結構變化(Pollitt, 2011)的因素。然而，實務上，技術往往是政策變遷的主要推動力，不僅限於對現有服務和程序的改進(Bannister & Connolly, 2014)。因此，在理論的層次上，究竟AI技術對公共治理理論產生了什麼影響？

一、公共政策模型

公共治理理論探討公共政策的制定是由誰來決定？從而衍生了不同的公共政策模型，不論是描述性模型，包括菁英主義(elitism)、多元主義(pluralism)、統合主義(corporation)、制度主義(institutionalism)等模型；或是規範性模型，包括廣博理性模型、滿意決策模型、漸進主義、綜合掃描模型、規範最佳模型等，主要是根據個人、利益團體之價值與偏好事實取向的表現結果，或是根據人類對完美理性價值追求的價值取向的表現結果(丘昌泰, 2008)，各種模型的行動者都是以「人」為基礎，並未考慮技術所產生的影響。然而，「澳大利亞聯邦政府Robodebt」—澳洲史上最大型集體訴訟案及政策慘敗(policy fiasco)案例顯示，公共政策模型的理論受到「AI官僚」的影響。

澳大利亞聯邦政府社會服務部(Services Australia)使用的 Robodebt 具有判斷



民眾是否溢領社會福利金、與民眾確認收入、發出債務通知，以及隨後債務追繳過程等自動化功能，因此，AI 演算型決策是直接對對標的團體(target group)管制或干預。然而，Robodebt 演算法未能依照法律規定執行、違反行政程序相關規定(Whiteford, 2021)，且行使公權力的效果超過法律對行政裁量權的授權範圍，凸顯演算法不能簡單地視為行政過程自動化，反而可能會從根本上改變政府制定的行政原則和公民權利(Henman, 2017)。回到本段一開始的提問，公共政策的制定是由誰來決定？現有公共政策模型，可能無法針對 AI 對公共政策的制定所產生的影響提出合理解釋。

二、公共政策過程

公共政策過程的研究途徑主要以階段論、反階段論為主，其中階段論受理性決策模型的影響，是公共政策過程理論的主流。階段論將政策過程的階段劃分為政策規劃、合法化、執行、評估四個步驟分明的階段，每一個階段都有其先後次序性。因此，政策過程具有可控制性、步驟發展的必然性與各階段的次序性(丘昌泰, 2008)。

然而，AI 系統自動決策產生即時動態數據的特徵，無論是新形式的電子參與(林子倫、蔡甫昌, 2021)、社交媒體的輿情分析、立法方案的分析、政策執行的數據，或不同政策間複雜的影響評估等，在階段論中的各個過程均可以提供實證分析數據，使得評估成為整個政策週期中不可或缺的一部分，而不是在決策過程結束時的單獨步驟(Höchtel et al., 2016)，持續評估的概念縮短傳統的政策週期，改變政府治理的過程。

美國紐約市大量挖掘和連結約 60 個不同城市機構現有數據庫，從中找到關連性，藉此重新調整工作計畫的重點(Desouza & Jacob, 2017)。美國麻薩諸塞州埃塞克斯郡議會(Essex County Council)預測 14 歲年輕人到 18 歲時會成為「不在教育、就業或培訓中」(not in education, employment or training, NEET)的風險，再



與學校合作制定早期干預措施(Mikhaylov et al., 2018)。德國聯邦勞動和社會事務部(Bundesagentur für Arbeit) 利用大數據分析定義和評估失業者與部分工時者的特徵，提供更有效的安置和諮詢服務，並能夠更準確地分析其安置計畫的結果，找出效率相對較低的計畫並加以改善或終止(Manyika, 2011)。

上述案例，使用 AI 即時行為模式分析及時發現問題(Mergel et al., 2016)、識別高風險民眾，以循證的方式改善計畫、開發特定對象的干預措施和提升行政效能(Mikhaylov et al., 2018)等，皆是公部門內部管理運用 AI 優化工作、工作自動化、衡量員工績效、進行內部控制和監督等(Maciejewski, 2017)，儘管公共政策過程本質上仍是一個政治過程，AI 已經改變公共政策過程階段論中次序性的概念。

三、公共政策執行

公共治理理論探討科層體制的控制過程、基層官僚(street-level bureaucrats)行政裁量權，然而，AI 的出現使得官僚的工作發生了根本性的變化(Giest & Kliavink, 2022)，引起是誰(人或機器人)在行使行政裁量權？以及所衍生課責性的論辯。舉例來說，目前公部門最常看到、最典型的「AI 官僚」或「隱藏官僚」(hidden bureaucrat)－聊天機器人，而真正的人類公務員成為調解人(mediators)而不是決策者(decision-makers)，產生行政裁量權可能被濫用，以及缺乏問責制的問題(Aoki, 2020; Janssen et al., 2020; Wihlborg, Larsson, & Hedström, 2016; Zuiderwijk et al., 2021)。

按照 Max Weber 的定義：「官僚機構是一種以許多規則、標準化流程、程序和規定、辦公桌的數量、細緻的分工和責任、清晰的層級和專業的、員工之間的互動幾乎沒有人情味為特徵的組織結構。」，是最適合 AI 化，進而產生 AI 官僚的組織(Unver, 2018)。

公共治理理論探討基層官僚的行政裁量權時，依據赫伯特·西蒙(Herbert



Simon)所言，「決策是行政管理的核心，而智慧是決策的基本能力，行政裁量權的特徵就是官僚參與決策(存在不確定性)時，如何最佳化的提供政府服務、利益和懲罰的複雜任務。」(Bullock, 2019; Simon, 1947)，行政裁量權主要研究重點為基層官僚利用自己的判斷力實現政府政策、公民服務並解決問題時，在複雜且不確定的問題中的獨立決策自由度、回應能力、判斷力和課責、動機和態度，包括公共服務動機(Perry, 1996; Perry & Wise, 1990)，外部動機(Bullock, Hansen, & Houston, 2018)和組織承諾等(Bullock, 2019)。

Barth & Arnold 1999 年提出一個非常經典的問題：「AI 技術的出現，行政自由裁量權的論點將變得毫無意義？」，於是引發 AI 對行政自由裁量權影響的論辯，一派觀點擔憂 AI 的行政決策減少了基層官僚的自由裁量權，行政決策可能無法適當考慮個人處境(Henman, 2020)。另一派論點可以從赫伯特·西蒙(Herbert Simon)的觀點開始談起，他將計算機視為「一種非政治性技術(an apolitical technology)，可以提高決策的合理性和組織運營效率」，具有將事實與價值分離能力的技術，反觀人類不是理性自動機(rational automatons)，行政自由裁量權的一個眾所皆知的問題是人為偏見(human bias)或認知限制(cognitive limitations)，這些限制會阻止代理人(agent)做出正確的決定(Bullock, 2019; Kahneman, 2011)。雖然 AI 也可能產生偏見(prejudices)，但是可以把 AI 程式碼列印出來檢查程式設計師是否有偏見，不受個人利益、偏見或價值觀的影響，但是人類的社會態度(social attitudes)卻不能這樣做；AI 機器具有整合諸如價值、動機和目標等人類特質的能力，可以開發符合公共政策目標的理想 AI 系統，該系統可以近似理想的理性人(ideal rational man)(Barth & Arnold, 1999)，因此，AI 雖然減少了行政自由裁量權，也同時消除人為干預。再者，AI 機器具有學習、合理化和處理指令的能力，能夠檢查基本假設和價值、識別專業基層官僚所獨有的推理過程、解決在許多經典的公共政策案例中人類脆弱性、人類決策缺陷進而影響判斷問題的優勢。Barth



& Arnold 為他自己的問題提出結論：「只要僅將 AI 系統視為工具或輔助工具，就可以消除這種恐懼，AI 的出現絕對沒有使行政自由裁量權論證毫無意義」(Barth & Arnold, 1999)。然而，Barth & Arnold (1999)論證將 AI 系統視為工具或輔助工具的論證卻無法得到實證上的支持。

第一個案例是 2012 年波蘭勞動和社會政策部(Polish Ministry of Labour and Social Policy, MLSP)的「失業評分系統(unemployment scoring system)」。最初構想，是利用 AI 失業評分系統提供失業者分類建議，是半自動化的建議工具，同時保留諮詢人員對失業分類結果的最終決定權。實際情況，工作人員不願質疑計算機系統作的決定，超過 99%依據失業評分系統的推薦分類，幾乎放棄行政裁量權，即使他們的專業判斷表明並非如此，使得失業評分系統變成實質上的決策者(Misuraca & van Noordt, 2020)。本案例中，顯然 AI 系統不是工具或輔助工具，是人類依賴資料和演算法預測結果做出決策，侵害社會最弱勢群體的公民權利，擴大了權力差距。

第二個案例是臺北市智慧號誌系統，透過動態的交通控制，取代原本以人力調整的時制計畫，或單一或多個路口的定時號誌，提升重點壅塞區域號誌控制的效率。在本案例中，AI 系統不是工具或輔助工具，AI 合理分配通行權的預測結果，優於經驗豐富人類專家的判斷，大幅提升交通通行能力(江伊嵐、陳奕廷、葉梓銓, 2020)。

從組織管理的觀點，在人機協作、人與 AI 之間的互動，公部門的職權或國家權力何時可以直接使用 AI 預測 (機器決策)，何時必須有人為介入(人類決策)? 回到本段一開始的提問，公共政策的執行是由誰來執行? 現有公共政策執行理論，若不考慮技術的因素，可能無法針對 AI 對公共政策執行產生的影響提出合理的解釋。

此外，AI 課責性的論辯，英國政府將立法免除人類司機在自動駕駛運作時



的事故責任⁷⁸。然而，這樣的觀點要在公部門實施相當困難。從代理人理論觀點，民主國家的人民為「委託人(principal)」，而公部門組織為「代理人」，公共部門的組織運作，都可能涉及權力授予(delegation of power)，擁有權力的人委託他人代理執行某些任務、功能，代理人依照委託人的目標付諸行動、完成所託(陳敦源、徐仁輝, 1999)。公民委託政府，公民可能無法接受人類公務員可以免除課責，且會使公共治理所面臨「信任赤字」的困境更加嚴峻。

綜上分析，AI 是對民主治理(即負責任的決策或行使裁量權)的一種改進，前提是要理解應用 AI 背後的理論基礎及假設，並嘗試予以控制，否則，我們不會比目前的情況更好，甚至會更糟(Barth & Arnold, 1999)。因此，公共治理理論必須重新思考「AI 是行動者(actor)」、「AI 能動性(agency)」對人類行動產生影響，以及人與 AI 之間互相構成(constitute)的過程中所隱藏「演算法和人類協作潛在的權力結構(the underlying power structure of algorithmic and human collaboration)」(Ramati & Pinchevski, 2018; Waldherr et al., 2019)。由於公共治理理論解釋力不足，研究社群開始轉向科技決定論，或是科學科技與社會互動的理論去尋找更大的解釋力。

肆、新興 AI 治理研究社群興起

當 AI 治理的議題成為主流，而在公共治理理論領域對 AI 技術關注很少，加上現有理論、政策和法律無法有效因應 AI 治理需求，**技術研究社群開始慢慢發展 AI 治理的議題，也是目前研究的主流**，廣泛性的關切 AI 對人類各層面的影響，並以全球 AI 治理的角度，發展各項 AI 倫理準則，以避免 AI 風險，是普遍應

⁷⁸ 檢自 <https://mashdigi.com/uk-government-to-legislate-to-exempt-human-drivers-from-liability-for-accidents-while-autonomous-driving-is-in-operation/>



用性的觀點。

由於 AI 包括跨越多個部門的許多技術，不同利益相關者對適當的治理模式和程度也存在分歧，但仍有共通性，可歸納包括：「以人為本」的 AI 設計、識別和解決 AI 的不確定性和風險(Winfield & Jirotko, 2018)以及所衍生的倫理與辯論(Morley, Floridi, Kinsey, & Elhalal, 2020)等，例如從 2016 年至 2019 年已經發布 80 多份的倫理準則、原則、指南、框架和政策策略等(Misuraca, 2020; Schiff et al., 2020)，包括美國、中國、歐盟和印度等主要地緣政治大國的政府和私人公司都在制定 AI 政策和倫理聲明。

在國際組織層面，2019 年歐盟執委會(European Commission) AI 高級專家小組(High-Level Expert Group on Artificial Intelligence, AI HLEG) 提出「可信任 AI 的倫理準則(Ethics Guidelines for Trustworthy AI)」，強調以人為本(human-centric)並具備 3 大要件：合法(lawful)、符合倫理(ethical)及健全(robust)。至於實際上如何開發一個「合法、合乎倫理、健全而值得被信賴的 AI」，歐盟提出下列 7 項要求(requirement) 作為實踐之參考，其中第一項即人類代理與監督(human agency and oversight)：AI 系統應作為人類自主性(human autonomy) 與人類決策之輔助，尊重並保護人的基本權利，且需實現人類重要的目標或價值，並由人類加以監督。2020 年發布「AI 白皮書」及「歐盟資料策略」。歐洲理事會(European Council)於 2022 年 12 月 6 日通過了關於 AI 法案(Artificial Intelligence Act)的一般作法(General Approach)，確保在歐盟使用的 AI 系統是安全的，並尊重現行法律的基本權利(fundamental rights)和歐盟價值(Union values)⁷⁹。

2018 年第 40 屆國際數據保護和隱私委員會國際會議(The International Conference of Data Protection and Privacy Commissioners, ICDPPC)發布「AI 中的

⁷⁹ 檢自 <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>

倫理與數據保護宣言」；OECD 2019 年發布 AI 倫理準則；WEF 在 2019 年發布 AI 治理主題的白皮書。聯合國教科文組織(UNESCO)2021 年通過第一份全球性「AI 倫理問題建議書」，禁止利用 AI 實施社會信用評分制度或監控大眾等。

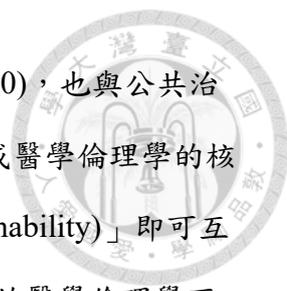
在非營利組織層面，未來生活研究所(The Future of Life Institute) 2017 年發布 23 項 Asilomar AI 原則，降低 AI 技術的可能風險。電機電子工程師學會(Institute of Electrical and Electronics Engineers, IEEE)2018 年發布「倫理合規設計(Ethically Aligned Design)」。OpenAI，一家具有營利性和非營利性機構的混合組織，2018 年發布了「OpenAI 章程(OpenAI Charter)」⁸⁰，確保通用 AI(AGI)能使人類整體受益。

在行政組織層面，美國國家標準暨技術研究院(The National Institute of Standards and Technology, NIST)依川普「行政命令」，訂定可靠、健全和可信任的 AI 系統的技術標準。在企業層面，美國/起源於美國的多家跨國技術公司都發布 AI 倫理聲明，尤其是「Microsoft AI Principles」(Daly et al., 2019)。

此外，亦有從軟體設計思維層面提出的 AI 治理理論，包括敏捷式治理(agile governance)理論、負責任的研究與創新(Responsible Research and Innovation, RRI)理論、價值敏感設計(Value Sensitive Design, VSD)等。例如負責任機器人基金會(Foundation for Responsible Robotics)將 AI 的技術創新從「對風險的治理」轉化為「對創新本身的治理」(Winfield & Jirotko, 2018)。價值敏感設計指出應控制整個設計 AI 過程必須考慮人類價值(Friedman, Kahn, Borning, & Huldtgren, 2013)。

以上所述的理論、框架、策略和指南，目的是期待 AI 技術的研究與創新者必須負責保證有利於社會利益，並且確保利益相關者可以參與到技術創新的過程，進一步積極監督研究創新過程。然而，除了美國行政命令，幾乎所有 AI 倫

⁸⁰ 檢自 <https://openai.com/charter/>



理和治理文件都不具有約束力(Daly et al., 2019; Morley et al., 2020)，也與公共治理理論幾乎無關。此外，這些倫理準則大量借鑒了生物倫理學或醫學倫理學的核心原則，有研究認為只要再加一個「AI 可解釋性原則(AI explainability)」即可互通(Floridi et al., 2018)，然而，AI 與以促進患者健康為共同目標的醫學倫理學不同，AI 的發展不是一個共同的目標，而是多個目標，而且這些 AI 倫理原則中有許多相互矛盾，例如，如何在確保公平(fair)和平等(equal) 對待受演算法影響的人，同時提高演算法預測的正確率(accuracy)? (B. Zhang & Dafoe, 2020)；再者，這些框架和原則往往是缺乏系統性(unsystematic)、太高層次(too high-level)、高度抽象(a highly abstract level)，無法具體實踐(Henman, 2020; Whittlestone, Nyrupe, Alexandrova, Dihal, & Cave, 2019)，因而無法有效解決目前許多公共治理上懸而未決、爭議權衡的難題，例如如何權衡 (1).科技創新原則、(2).人類監督強度、(3). AI 系統的品質、(4). AI 在政府職能的應用領域、(5).價值選擇等問題。

另一個研究領域，是以科學科技與社會(STS) 為主，STS 開發了精密的方法來處理社會系統與技術之間的關係，例如「技術的社會塑造(the Social Shaping of Technology)」(Williams & Edge, 1996)、「結構化方法(the structuration approach)」(DeSanctis & Poole, 1994; Orlikowski, 1992)、「技術制定框架(the technology enactment framework)」(J. E. Fountain, 2004; Garson, 2003)、「技術的社會建構(the social construction of technology, SCOT)」(Bijker, Hughes, & Trevor, 1987; Jackson, Poole, & Kuhn, 2002; Meijer & Löfgren, 2015)、「行動者網絡理論(the Actor-Network Theory, ANT)」(Callon, 1984; Crawford, 2020; Latour, 1984, 1996, 2005, 2011)等。

以 STS 理論為基礎的電子化政府理論(e-government theories)為例，電子化政府理論試圖將技術視為自變量或因變量的論文之間架起一座橋樑，亦即在技術與政府政策之間的關係概念定位為「技術與政策之間的相互依存關係」，強調技術是決策的關鍵組成、技術並不能決定政策的制定、但技術為政策制定創造了條

件。雖然成功的技術是最能解決社會問題的技術，但技術的過程不是一個線性的、理性的、進化的和解決問題的過程；人與任何技術的互動都形成了「實踐中的技術(technology-in-practice)」(Orlikowski, 2000, 2002)。

SCOT 理論建立者 Bijker 提出「技術是社會和政治構建的，社會(包括政治)是技術構建的。」(Bijker, 2010)，人與技術的互動過程中塑造(shapes)了技術，技術既能促進也能限制人類行為；技術是政策的產物，政策是由技術塑造的。

Latour 的行動者網絡理論(the Actor-Network Theory, ANT)，強調人類和非人類混合體的積極角色，Latour 主張無論是「人」或「科技物」應都具有能動性(agency)，進而對彼此發生作用、造成變化。

殊為可惜的是，雖然 STS 提出將「人」與「物」給予對稱化的社會建構論點，然而這些分析似乎與政策科學無關(Meijer & Löfgren, 2015)。

伍、公共治理理論與新興 AI 治理的路徑爭議

現今 AI 治理，無論是技術決定論，或是 STS 的理論，均忽略了現有的治理機制和機構，並未整合到有關公共治理的理論中(Meijer & Löfgren, 2015)，這不僅是一個學術問題，而是延伸到公共治理具體實踐的問題。舉例來說，當 AI 的決策攸關到人命、司法判決、權利義務時，當無法解釋 AI 決策的過程，將會考驗人類對於 AI 的信任程度。根據 KPMG 2023 年調查，61%的受訪者表示對信任 AI 系統持謹慎態度，67%對 AI 的接受度較低至中等，AI 在人力資源領域的使用是最不值得信任和接受的，而在醫療保健領域的使用是最值得信任和接受的⁸¹。

AI 必須被解釋(explanation)的原因(reason)在於必須理解黑箱子(black box)本

⁸¹ 檢自 <https://kpmg.com/xx/en/home/insights/2023/09/trust-in-artificial-intelligence.html>，檢自 <https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf>



身是如何運作的，決策的數據是如何產生的(Guidotti et al., 2018)。然而，學者們對於可解釋性(interpretability)或可解釋性(explainability)的機器學習，以及如何衡量和評估仍然沒有達成共識。Escalante 等人(2018)將「可解釋性(interpretability)」定義為「模型結構解釋什麼功能？」，將「可解釋性(explainability)」定義為「做出決策背後的理由？」來區分這兩個術語(Escalante et al., 2018)。此外，de Bruijn 等人 (2022) 將「可解釋的 AI (explainable AI, XAI)」定義為「大眾對 AI 結果洞察力的程度」，這意味著可以向社會解釋做出決策的基本原理(de Bruijn, Warnier, & Janssen, 2022)。這兩個概念通常被視為可以互換，但有細微差別，兩者都旨在提高演算法的公平性(fairness)、透明度(transparency)和課責(accountability) (Cath, 2018)。然而，讓人類更信任 AI 技術的「可解釋的 AI(Explainable Artificial Intelligence, XAI)」⁸²，從技術的角度和社會科學角度，就有不同觀點。

從技術角度，就透明度(transparency)、黑箱的可解釋性而言，可以是全局可解釋性(Global Interpretability)模型，可能是完全可解釋的(completely interpretable)，我們能夠理解模型的整個邏輯，並遵循導致所有不同可能結果的整個推理；也可以是局部可解釋性(local interpretability)，可以只理解特定決策的原因的情況：只有單一預測/決策是可解釋的(Guidotti et al., 2018)。在提高 AI 系統軟體品質及功能的層面，從技術角度，大多數的科學社群學者、資料科學家在解決預測問題時，往往傾向關注「均方根誤差〔root-mean-square deviation

⁸² 檢自 <https://www.bnnext.com.tw/article/53335/fujitsu-explainable-ai-with-knowledge-graph>

(RMSD)、root-mean-square error(RMSE)]⁸³、正確率(accuracy)⁸⁴、AI「召回率(Recall)」⁸⁵和「精確率(Precision)」⁸⁶等指標(Dignam, 2020; Ekelund, 2017)。例如應用 AI 技術作為坡地崩塌預測分析，提供防災等計畫研擬之策略，研究關注重點是經由混淆矩陣比對預測成果和歷史發生崩塌斜坡單元比較，成果以優於 80%正確率證實其預測分析技術之可行性(林彥廷、顏筱穎、張乃軒、林宏明、韓仁毓、楊國鑫、陳俊杉、鄭宏達、徐若堯, 2021)。

從社會科學角度，認為 AI 演算法系統的內部運作是一個「黑箱(black box)」

⁸³ 測量數值之間差異的量度，其數值常為模型預測的量或是被觀察到的估計量。均方根偏移代表預測的值和觀察到的值之差的樣本標準差(sample standard deviation)。

⁸⁴ 正確率：在所有情況中，正確判斷真假的比例 $TP+TN / (TP +TN+ FP +FN)$

TP：True positive – number of persons with disease who have a positive test result with the assay in question.

TN：True negative – number of persons without disease who have a negative test result with the assay in question.

FP：False positive – number of persons without disease who have a positive test result with the assay in question.

FN：False negative – number of persons with disease who have a negative test result with the assay in question.

⁸⁵ 召回率是在所有正樣本當中，能夠預測多少正樣本的比例：Fraction of persons with disease who get a positive test result with the assay in question. Calculated as $TP / (TP + FN)$ ；

⁸⁶ 精確率為在所有預測為正樣本中，有多少為正樣本：Precision =PPV：Positive predictive value – fraction of persons with a positive test result who do have the disease. Calculated as $TP / (TP + FP)$.

87：給定一個問題，系統會提供答案，但無法解釋如何產生答案，且該系統缺乏自我解釋的能力，因此在高風險的應用中無法信任，特別是因為機器學習系統偶爾會發生災難性的錯誤(House, 2019)，因此，AI的可解釋性是相當必須的，以及這些關鍵決策是誰決定的。舉例來說，立法院司法及法制委員會審查 2024 度中央政府總預算案，民進黨立委林淑芬質詢司法院：「民眾一定想知道判決書哪段是 AI 寫的？如果交給 AI，屆時誤判如何究責？」，並表示發展 AI 的前提是要能課責，將 AI 用在執法與法律屬高風險，建議應有使用、參考「指引」⁸⁸。

行政程序法有關透明原則的內容係指資訊必須具備「公開性(openness)」與「可理解性(comprehensibility)」兩種基本特質，最主要的目的在於使一般大眾了解行政機關作成決策的過程，一方面利益相關者得因此選擇最適合其自身利益的因應策略，降低其法遵成本(compliance costs)，同時達到維護個人權益與提高行政效能的目的。另一方面，藉由公開決策程序，強化課責(accountability)監督機制(陳敦源, 2009; 葉俊榮, 1999)。因此，公民可以知道國家做出決定的理由，對決定提出質疑並向法院提出上訴或向國家提起訴願，法官或政府也可以依據該決定的理由進行裁決(Dignam, 2020)。

然而，政府應用的 AI 涵蓋 AI 的全部技術範圍，從傳統的機器學習到使用自然語言和圖像數據的「深度學習」，複雜的模型通常可達到更出色的效能表現(林守德, 2021)，雖然增加 AI 系統軟體品質及功能、預測的能力，但卻可能難以理解演算法的決策方式，也失去解釋如何得出結果的能力。然而，若使用不太複雜的 AI 技術或工具可能更難提高正確率(accuracy)和效率(efficiency) (Engstrom et al., 2020)。

從社會科學角度，Oswald(2018)主張公部門利用演算型決策時必須受到程序

⁸⁷ 對外行人是黑箱的部分，對內行人未必是黑箱。

⁸⁸ 檢自 <https://udn.com/news/story/7321/7486194>



和知識的控制，演算法不可能用較低的標準來提供藉口或辯護，並認為我們對演算法要求更高標準的期望需要被調整。換言之，公部門決策者在應用機器學習演算法時，必須要能夠對該機器的預測結果進行解釋(Oswald, 2018; 陳敦源、廖洲棚、黃心怡、張濱璿、王千文, 2022)。另一方面，從技術角度，可解釋性可能只會以效率為代價(Engstrom et al., 2020)。畢竟最終重要的是決策的結果，而不是決策背後的邏輯，如果 AI 能夠提高公共價值，從而導致公平和健全的決策，公民對公共服務滿意，基本邏輯將不那麼重要(Busch & Henriksen, 2018)。人力資源諮詢公司 OC Tanner 的副總裁兼首席技術官 Niel Nickolaisen 表示：「我同意演算法必須具有一定的透明度和可解釋性，但是這樣做會削弱『機器學習』測試不同模型的能力，以及建立因果關係的能力；是否需要這種折衷方案？」(Dwivedi et al., 2019)

在 AI 系統的效能/正確率(performance/accuracy)和可解釋性(explainability)之間存在反比關係的情況下(Dwivedi et al., 2019)，如何忠實地解釋模型的決策與人類易於理解的方式之間進行權衡(Schmidt & Biessmann, 2020)？如何在提高公共治理效能與保護公民避免受到演算法潛在傷害之間的權衡？在正確率和可解釋性之間的權衡？從不同的分析路徑產生了不同的觀點與治理方式。

有論者認為在醫療保健中正確率比可解釋性更重要；在行政裁量面向，有研究認為正確率非常重要(Henman, 2020)，但也有研究認為高影響力的決定(例如可能影響公民權利或補助金的決定)，可解釋性非常重要；低影響力的決策(例如為提高管理效率而做出的決策)，不太需要解釋；檢測違規行為演算法的決策，在有合法正當理由的情況下，可以選擇不解釋(UK, 2020)。



陸、搭起 AI 與公共治理理論的橋梁：AI 的理論化與人機協作

有研究從文獻提出成功實施 AI 的關鍵因素為 4 類：組織、技術、流程和環境，關鍵因素是技術(Merhi, 2023)。因此，如何在公共治理領域以人為主的社會系統中，處理公共治理理論與 AI 技術之間的關係？如何將 AI 技術的作用在公共治理中理論化？Schoemaker 和 Tetlock (2017)建議採用混合方法，將人和機器相結合，以達到所謂的「更好的判斷(superior judgements)」，去解決 AI 系統缺乏人類更廣泛和脈絡化智慧的問題(Androutsopoulou, Karacapilidis, Loukis, & Charalabidis, 2019; Schoemaker & Tetlock, 2017)。Stead(2018)斷言人與 AI 建立夥伴關係的重要性，AI 機器將進行計算和/或預測(calculate and/or predict)，而人類對行動進行解釋和決定(explain and decide)，亦即人類可能會專注於 AI 流程處理和輸出所需要的設計、分析和因果關係的闡釋(interpretation)等更高價值的活動。未來的組織可能會透過整合人類和 AI 協作的方式來創造價值(Dwivedi et al., 2019; Stead, 2018)。例如人類增強 AI (Human-augmented AI)(由人類訓練並根據人類輸入不斷提高其性能的 AI 系統)、AI 增強人類智慧(AI Augmented human intelligence)(AI 系統傾向於透過提供預測分析等支援系統來擴展或增強人類的能力，從而增強人類智慧)、人與 AI 互動(Human-AI interactions) (人類與 AI 有效互動產生人類增強人工智慧和增強人類智慧)。有鑑於 AI 系統是一種社會技術系統，只有人類和 AI 透過相互理解相互作用並相互補充，才能在效能上取得進展；也就是說，機器必須更好地理解人類如何推理和操作，而人類必須更好地了解機器的決策邏輯(Jarrahi et al., 2022)。因此，Miller(2018)主張一種新的人機合作互利共生(human-machine symbiosis)的必要性，並呼籲人們重新思考「人與機器為什麼必須合作互利共生地工作來擴增(augment)和強化(enhance)彼此的能力」⁸⁹(Miller, 2018)。

⁸⁹ 原文：how humans and machines need to work symbiotically to augment and enhance each other's capabilities.

當人類無可避免的將與機器並肩工作，人與機器如何協作成為公共治理理論與實務應用的新挑戰。雖然 Schoemaker 和 Tetlock、Stead 和 Miller 米勒的主張應該要有進一步有研究和證據，但已展開技術作為變量的解釋潛力，將 AI 技術物視為行動者，在政策利益相關者的觀點中給予廣義對稱性的對話空間。

第三節 公部門應用 AI 的潛力與挑戰



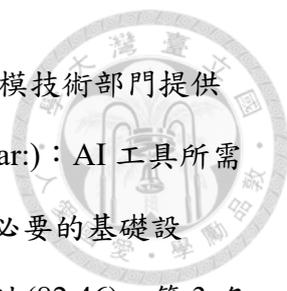
世界各地的組織和政府正在以快速成長的速度採用和實施 AI 系統，PWC 的研究，約 86% 的參與者表示，AI 在 2021 年成為其公司的「主流技術(mainstream technology)」；AI Journal 的調查，參與研究的高階主管中有 72% 對 AI 在未來的角色持正面態度，顯示 AI 的採用將繼續擴大，其對組織、社會和個人的影響也將增加(Merhi, 2023)。然而，儘管全世界大部分的政府組織都已啟動在公部門採用 AI 的計劃，但政府採用 AI 的整體應用潛力、挑戰的相關研究非常少，使 AI 技術在公部門的實施舉步維艱(Boyd & Wilson, 2017; Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018; Wirtz et al., 2019)。本節對公部門使用 AI 的潛力與挑戰作概括式的論述，以利進一步探究如何提升公共治理效能及控制風險的理論。

壹、臺灣與其他國家、地區或國際組織的 AI 發展現況

有關臺灣 AI 的競爭力，2021 年 Tortoise 研究全球 62 個國家或地區的 AI 競爭力，發布全球人工智慧指數(Global AI Index)，以 7 個角度分析：人才、基礎設施、運營環境、研究、發展、政府策略和商業，臺灣排第 24 名⁹⁰。2021 年英國牛津洞析研究組織(Oxford Insights)與國際發展研究中心(International Development Research Centre, IDRC)發布的「政府 AI 準備指數(Government Artificial Intelligence Readiness Index 2021)」，評估「政府利用 AI 提供公共服務的準備程度⁹¹」，評分項目包括 3 個基準(pillar)：「1. 政府基準(The Government pillar:)：政府對於如何開發與管理 AI 有戰略眼光，並輔以適當的管制和對倫理的關注；2. 技術部

⁹⁰ 檢自 <https://www.tortoisemedia.com/intelligence/global-ai/>

⁹¹ 原文：How ready is a given government to implement AI in the delivery of public services to their citizens?



門基準(Technology Sector pillar)：政府依賴該國具有競爭力和規模技術部門提供大量 AI 工具；3.數據和基礎設施基準(Data and Infrastructure pillar)：AI 工具所需大量、高品質、具代表性數據。最後，AI 工具提供公民服務時必要的基礎設施。」，針對 160 個國家排名結果：第 1 美國(88.16)、第 2 新加坡(82.46)、第 3 名英國(81.25)，第 4 芬蘭(79.23)，第 5 名荷蘭(78.51)；臺灣第 18 名(71.98) (Insights, 2021)。

從上面的分析數據顯示，臺灣 AI 的能力在全球評比均屬於前段班，臺灣在推動 AI 的重要的國家發展戰略，包括：

- 1.2017 年為臺灣 AI 元年並於 2018 年啟動四年期的「臺灣 AI 行動計畫」，打造臺灣的 AI 生態系統(行政院, 2018)。
- 2.2023 年臺灣 AI 行動計畫 2.0 (2023-2026 年) (行政院, 2023)，持續培育人才、重視 AI 倫理法制、建構落實可信任 AI 的發展環境，並且要把 AI 應用於解決臺灣面臨的勞動力短缺、高齡化社會及淨零碳排等重大挑戰，讓全民皆能受益於 AI，並將推動與國際介接的規範與標準、積極參與國際相關組織及與 AI 領先國家建立合作關係。
- 3.行政院 2023 年 8 月 31 日通過國家科學及技術委員會(國科會)擬定的「行政院及所屬機關(構)使用 AI 生成式參考指引」，利基於強調運用生成式 AI 產出資訊時應適當揭露、須由人進行客觀專業的最終判斷。
- 4.國科會為臺灣打造生成式 AI 對話引擎⁹²。

在國外部份，美國擁有相對較多的技術部門、最多的科技獨角獸和高價值大型科技公司位居榜首。新加坡總理辦公室內設置國家 AI 辦公室，已有 20 個部會

⁹² 國產可信任 AI 對話引擎 TAIDE 開放測試申請，初期聚焦辦公室 4 大應用。檢自 <https://www.ithome.com.tw/news/158607>



使用 AI 計劃。在 2021 年政府 AI 準備指數排名的 160 個國家中，近 40% 國家已經發布或正在起草國家 AI 戰略，表示 AI 正在迅速成為世界各國領導人最關心的問題(Insights, 2020, 2021)。

美國 AI 創新處於世界領先地位，歐巴馬政府在 2014 年及 2015 年成立科技政策辦公室(Office of Science and Technology Policy)，並發布兩份關於大數據報告。川普政府 2016 年再度針對大數據與演算法發佈相關報告，強調「大數據是客觀的」這個假說前提是錯誤的，並提醒勿過於依賴演算法，忽略其所附帶歧視風險，原因包括：

1. 演算法資料選取偏誤、不正確或不完整；
2. 演算法設計不佳、個人化的推薦服務限縮使用者自主性、決策系統的錯誤假設導致歧視、資料蒐集不足以代表特定人口。

美國 2019 年宣布美國國家 AI 戰略：美國 AI 倡議(the American AI Initiative)，提出 5 個關鍵領域：研發、釋放 AI 資源、建立 AI 治理標準、建立 AI 勞動力，以及國際協作與保護；8 項國家 AI 研發戰略，其中一項研究重點是「了解並解決 AI 的倫理、法律和社會影響」，依照 2 個階段進行研究：「設計、開發、展開和使用」、「監視和測試」，並藉由研究如何解決公平性(fairness)、透明性(transparency)、隱私性(privacy)、可信度(reliability)、韌性(resilience)、演算法偏誤(algorithmic bias)、課責性(accountability)和可解釋性(explainability)等問題，瞭解 AI 系統運作方式、潛在挑戰、隨著技術進步這些挑戰如何演化、如何緩解這些挑戰、革命性技術和設計的解決方案所造成的社會影響等；並提出在聯邦政府中設計、開發、獲取和使用 AI 時，各機構應遵守以下原則(House, 2019; 陳敦源、廖洲棚、黃心怡、張濱璿、王千文, 2022)：

1. 合法並尊重國家價值觀(lawful and respectful of our nation's values)：尤其重視隱



私、人權與人民的自由。

- 2.有目的性和績效導向(purposeful and performance-driven)：利益必須大於風險。
- 3.AI 應用必須「正確、可靠和有效(accurate, reliable, and effective)」。
4. AI 應用必須「安全、保全和韌性(safe, secure, and resilient)」。
- 5.可理解(understandable)：AI 應用的操作和結果必須可以理解。
- 6.負責任和可追溯(responsible and traceable)。
- 7.定期監測(regularly monitored)。
- 8.透明(transparent)：相關資訊必須揭露。
- 9.可課責的(accountable)：對於 AI 應用的正確使用和運作。

2019 年發布「維持美國 AI 領導地位的行政命令(Executive Order on Maintaining American Leadership in Artificial Intelligence)」，指出 AI 將影響幾乎所有的行政機關(構)的任務，為確保美國在 AI 的領先地位，行政機關(構)應追求六個戰略目標(Trump, 2019)：

- 1.與產業、學術界、國際合作夥伴和聯盟、以及其他非聯邦實體(Non-Federal entity)合作，持續投資 AI 的研發，並將研究成果貢獻於經濟和社會發展。
- 2.強化高品質且可完全追溯的數據資料、模型和計算資源，以增加數據對 AI 研發的價值，同時使數據能符合法律和政策對於安全(safety)、保全(security)、隱私性(privacy)和機密性(confidentiality)保護的要求。
- 3.減少 AI 技術創新應用障礙。
- 4.確保技術標準兼顧降低惡意行為者的攻擊和 AI 技術系統的創新，以維持公民信任。
- 5.培訓 AI 的研究人員、使用者、聯邦工作者；培訓重點為科學、技術、工程和數學(Science, Technology, Engineering, and Mathematics, STEM)教育，以

確保能夠充分利用 AI。

6. 制定並實施行動計劃，以保護美國在 AI 和技術的策略優勢。



有關公部門如何導入 AI，美國的建議如下：可以參考以前政府轉型、私部門導入 AI 的執行經驗，以及六種策略(Mehr et al., 2017)：

1. AI 為目標的一部分、以公民為本 (make AI a part of a goals-based, citizen-centric program)
2. 公民參與(get citizen input)
3. 利用現有資源(build upon existing resources)
4. 做好數據準備工作並妥善處理隱私問題(be data-prepared and tread carefully with privacy)
5. 減輕倫理風險並避免 AI 決策(mitigate ethical risks and avoid AI decision making)
6. 增強員工能力，而不是替換他們(augment employees, do not replace them)

此外，美國白宮科學與技術政策辦公室(The White House's Office of Science and Technology Policy , OSTP) 2020 年發布「AI 應用管制指引」(Guidance for Regulation of Artificial Intelligence Application)，為美國政府機關起草 AI 規範並進行管制時提供指引⁹³。在公共管理實際應用的研究議題上，美國國家公共行政學院(National Academy of Public Administration)現正朝向下列相關議題研究(NAPA, 2020)：

1. 如何使用 AI 改善服務遞交(service delivery)。
2. 如何發展具有 AI 技能的勞動力 (AI-ready workforce)。

⁹³ 檢自 <https://stli.iii.org.tw/article-detail.aspx?no=67&tp=5&d=8444>



- 3.如何將 AI 融入公共管理課程。
- 4.如何提高解決 AI 相關問題的倫理意識。
- 5.如何制定適當的多層次治理計劃(multi-level governance schemes)，以防止無意的偏差(unintended bias)。
- 6.如何確保所有人都能享受 AI 的好處。
- 7.如何處理 AI 政府間和跨部門層面的問題。

2020 年「AI 政府法案(the AI in Government Act of 2020)」，在美國總務管理局(General Services Administration)設立「AI 卓越中心(AI Center of Excellence)」提高聯邦政府對 AI 的採用(Bignami, 2022)。美國 2021 年 1 月 1 日通過「國家 AI 倡議法案(National AI Initiative Act)」，法規規範整個聯邦政府加速 AI 研究和應用計劃，確保公部門和私部門開發和使用值得信賴的 AI、為現在和未來的美國勞動力做準備、促進國家經濟繁榮和國家安全。依據該法案在 OSTP 下成立「國家 AI 倡議辦公室(the National Artificial Intelligence Initiative Office, NAIIO)」⁹⁴，在美國政府的積極推動下，截至 2019 年，已經有 45%的聯邦政府機構嘗試使用 AI(Engstrom et al., 2020)。

有關公部門如何導入 AI，英國政府要求公部門導入 AI 必須強制性作影響評估，理由包括：(UK, 2020)

- 1.公部門可以瞭解 AI 系統可能帶來的風險等級及時因應。
- 2.公部門 AI 經驗很少，可能會忽略某些風險，例如數據偏誤(bias)。
- 3.發揮課責功能。
- 4.確保公民利益和權力。

⁹⁴ 檢自 <https://www.ai.gov/naiio/>



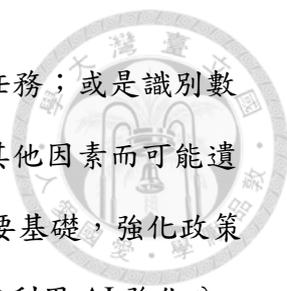
歐盟建議作法如下(Misuraca & van Noordt, 2020)：

1. 激發 AI 的意識和潛力(Stimulating awareness and potential of AI)：讓公務員認識 AI，以利開發 AI 可以產生工作價值的領域。
2. 改善 AI 的數據管理：改善公部門數據的品質(quality)、可用性(availability)和可近性(accessibility)。
3. 建立 AI 內部能力：對所有公務員進行 AI 培訓、使公務員瞭解如何將工作與 AI 結合、建立對 AI 的信任；建立新的職位或機構，例如首席數據官(Chief Data Officers)或專門的 AI 團隊，以刺激公部門內 AI 的發展。
4. 做中學(Learning by doing)：公部門對 AI 新技術的開發和應用方式知之甚少，因此，許多國家將執行經驗和成效作為機構間知識共享及後續修正 AI 策略的參考。
5. 開發倫理和法律的 AI 框架(Developing ethical and legal AI frameworks)：AI 的開發和使用涉及許多倫理問題，公共服務面向尤為重要，因此建立公部門使用 AI 倫理框架和相關準則，可以幫助公務員和公民之間建立信任。
6. 分配資金和採購：提供足夠資金或機制讓公部門導入 AI，例如直接提供財務支援進行 AI 實驗性方案或計畫、以公私協力的方式刺激 AI 或 GovTech 新創公司為公部門量身定制 AI 解決方案、修改採購法規等。

貳、政府使用 AI 的應用潛力

一、公共政策理論面向分析

AI 具有卓越的數據處理能力，透過數據分析、模式識別和預測，提供即時、更快、更全面的決策，有利提升我們對各種現象與問題的理解，突破傳統決策模式的侷限性(林子倫, 2017)。例如 AI 可以在幾秒鐘內分析高品質、完整、即



時的大量數據，自動執行人類需要數天或更長時間才能完成的任務；或是識別數據中的模式和趨勢，找出人類分析師可能受限規模、複雜性或其他因素而可能遺漏的資料關連與趨勢(許淑梅, 2019)，作為政策規劃與評估的重要基礎，強化政策推估預測能力(OECD, 2014)。再者，在政策規劃過程中，也可以利用 AI 強化公民參與的機制(Seeber et al., 2020)，以包容和開放的方式納入利益相關者，達成更公平、循證的政府決策(evidence-informed policy-making)，可以挽救信任赤字(trust deficit)。在政策執行過程，利用 AI 解決組織中簡單重複的任務，現有人力可將更多精力放在複雜的工作，提升專業工作水平及更有效地利用批判性思維能力；例如基層官僚逐案審查非常耗時，若能利用 AI 部分自動化，或整個決策過程自動化來行使自由裁量權，提高決策效率、節省大量時間和金錢，有更多時間解決公民需求及建立更好的關係(Busch & Henriksen, 2018)。

再者，繁文縟節(red tape)經常被視為降低公共服務效率、影響組織及個人績效表現的原因(陳敦源、黃建勳, 2019)，更甚者影響公共服務動機、造成公務人員工作壓力、職業倦怠(陳敦源、吳舜文、陳揚中、王光旭, 2019)，然而，有論者認為繁文縟節是政府運作的必要之惡，某人眼中的繁文縟節，在另一個人眼裡卻可能視之為安全屏障(陳敦源、簡鈺肆, 2018)。而 AI 技術的導入可以提高記錄和管理等繁文縟節的效率(Mehr et al., 2017)，AI「事半功倍」(do more with less)(McQuillan, 2020)的關鍵解決方案，可以很大程度上解決傳統官僚機構運作(例如高成本、缺乏靈活性、超負荷、違法者追蹤等)的問題，無需考慮權衡取捨(Peters & Zittoun, 2016)。換言之，藉由 AI 降低繁文縟節，可以跳脫繁文縟節的論辯，同時提高公共治理效能。



二、政府使用 AI 的應用領域分析

早期 AI 在公部門的應用類別以技術分類為主，使用 8 種不同的 AI 技術：人工神經網絡(Artificial neural networks, ANN)、模糊邏輯(Fuzzy logic, FL)、機器學習(Machine learning, ML)、多重代理人系統(multi-agent systems, MAS)、自然語言處理(Natural language processing, NLP)、基於案例的推理(case-based reasoning, CBR)、認知圖(cognitive mapping, CM)、基因算法(genetic algorithms, GA)，最常出現的技術是人工神經網絡(Artificial neural networks, ANN)(de Sousa et al., 2019)

近期歐盟(Misuraca & van Noordt, 2020)、Wirtz 等人(2019)提出以機器學習演算法應用領域做 10 個分類(Wirtz et al., 2019)，以避免 AI 技術不斷發展，技術分類過時。從應用普遍性分析，主要以聊天機器人(Chatbots)、智慧助理(Intelligent Digital Assistants)、虛擬代理人和推薦系統(Virtual Agents and Recommendation Systems)為主，其次是預測分析(Predictive Analytics)、模型和數據視覺化(Pattern Recognition Simulation and Data Visualisation)，第 3 是身分識別(Computer Vision and Identity Recognition)和專家決策系統(Expert and Rule-based Systems, Algorithmic Decision Making) (Misuraca & van Noordt, 2020)

然而，以機器學習演算法應用領域分類仍不脫以技術層次的角度，若依據聯合國政府職能分類(Classification of the functions of government, COFOG)⁹⁵(UN, 1980)，最常出現的是「公共服務」、「經濟事務」、「環境保護」；唯一未提及的是娛樂、文化和宗教，可能政府將 AI 集中在具有戰略意義的領域(de Sousa et al., 2019)。

⁹⁵ 依據聯合國發布政府職能分類 (COFOG) 分為 10 類：「F1：一般公共服務；F2：公共秩序與安全；F3：國防；F4：經濟事務；F5：環保；F6：住房和社區設施；F7：健康、F8：娛樂、文化和宗教；F9：教育；F10：社會保障。」。



AI 在政府中的三個核心應用包括：(1).機器人和認知自動化，透過機器人流程自動化等技術將人類勞動力轉移到高價值工作，(2).透過更好的預測來實現認知洞察能力；(3).透過回答公民詢問進行認知參與。AI 在公民服務方面至少有五種新興應用：(1).回答問題；(2).填寫和搜尋文件，(3).發送請求，(4).翻譯，(5).起草文件(Ojo, 2019)。

Zuiderwijk 等人(2021)則將 AI 在政府中的潛在效益分為九類：1.提升效率和績效；2.風險識別和監控效益；3.經濟效益；4.數據和資訊處理效益；5.公共服務效益；6.整個社會效益；7.決策效益；8.參與和互動效益；以及 9.可持續性效益 (Zuiderwijk et al., 2021)。

美國聯邦行政會議(The Administrative Conference of the United States, ACUS) 針對演算法治理範圍中的政府運作，提出包括強化執行(enforcement)、監理分析與監控(regulatory research, analysis, and monitoring)、裁決(adjudication)、公共服務(public services and engagement)以及內部管理(internal management)(Engstrom et al., 2020)。而已應用 AI 工具強化治理的任務分類⁹⁶則包括：

⁹⁶ 原文：AI tools are already improving agency operations across the full range of governance tasks, including:

- *Enforcing* regulatory mandates centered on market efficiency, workplace safety, health care, and environmental protection;
- *Adjudicating* government benefits and privileges, from disability benefits to intellectual property rights;
- *Monitoring and analyzing* risks to public health and safety;
- *Extracting* useable information from the government’s massive data streams, from consumer complaints to weather patterns; and
- *Communicating* with the public about its rights and obligations as welfare beneficiaries, taxpayers, asylum seekers, and business owners.



- 1.執行以市場效率、工作場所安全、醫療保健和環境保護為中心的監管要求。
- 2.裁決政府福利和權益，從身心障礙福利到知識產權。
- 3.監測和分析公共健康和環境安全風險。
- 4.萃取政府巨量資料中有用的資訊，從消費者投訴到天氣模式。
- 5.與大眾溝通福利受益者、納稅人、尋求庇護者和企業主等的權利和義務。

然而，要實現政府這些新的公共治理能力，取決於是否擁有有利的公共管理和行政環境(Ojo, 2019)。

機器學習在美國聯邦政府中非常普遍，2019 年 1 月至 8 月期間進行的一項研究發現，在接受調查的 142 個政府機構中，近一半(64 個機構，即 45%)進行過實驗利用機器學習技術來執行核心功能。其中許多機構將機器學習用於多個目的，導致 64 個機構產生 157 個案例。這 157 個案例，53%-33%已完全部署，而其他案例處於「規劃」階段或「試點或部分部署」。在聯邦層面，執法部門是機器學習的最大使用者；其他較突顯領域包括健康、金融監管、社會福利和商業。

參、政府使用 AI 的應用挑戰

社會對 AI 的接受和信任，取決於是否成功解決可能產生的負面影響，必須妥為因應以避免阻礙 AI 的採用(House, 2019)。有關 AI 在政府治理應用與挑戰的研究，Wirtz 等人(2019)已研究歸納出以下四類：(Wirtz et al., 2019)

1. AI 技術執行上有 AI 安全、系統/資料的品質和整合、財務可行性、缺乏專業知識和專家的問題。
2. AI 法律與法規的控制和治理需注意自主智慧系統治理、責任和課責、隱私/安全性的問題。



3. AI 的倫理問題，如機器與人類價值的包容性、道德困境 (moral dilemmas)、AI 的歧視。

4. AI 對社會影響，如勞動力替代和轉型、社會信任、人與機器互動問題。

該研究係透過分析文獻提出潛在應用領域和挑戰的初步研究，Wirtz 等人 (2019) 在研究限制時表示仍有必要進行實證檢驗。

Zuiderwijk 等人 (2021) 綜合文獻提出政府使用 AI 挑戰分為八類：1. 數據挑戰；2. 組織和管理挑戰；3. 技能挑戰；4. 可解釋性 (interpretation) 挑戰；5. 倫理和合法性挑戰；6. 政治、法律和策略挑戰；7. 社會個人和社會整體結構運作挑戰 (social and societal challenges)；以及 8. 經濟挑戰 (Zuiderwijk et al., 2021)。

Gartner 2021 年調查在美國、德國和英國已部署 AI 或打算在三年內部署 AI 的組織發現，平均而言，從專案開始計算，85% 會失敗，AI 專案可以從試辦走向生產約 54%⁹⁷ ⁹⁸。阻礙 AI 實施的主要障礙，包括 1. 無法/難以衡量價值；2. 缺乏對 AI 優勢和用途的理解；3. 數據可近性等。

另有研究從文獻整理出 19 個影響 AI 實施至關重要的因素：包括 1. 高層支持度、2. 戰略願景模糊、3. 組織文化、4. 組織架構、5. 缺乏可見的好處、6. 強而有力的專案管理者、7. 員工抗拒變化、8. 缺乏 AI 技術專長、9. 倫理議題、10. 責任與課責 (Responsibility and accountability)、11. AI 系統與現有系統、資料庫進行整合的複雜度 (Integration complexity)、12. 低數據品質、13. 資料量不足、14. 缺乏 IT 基礎設施、15. 數據的安全和保密 (Security and confidentiality)、16. 數據治理議題 (與

⁹⁷ Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence. from: 檢自 <https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence>

⁹⁸ 檢自 <https://www.gartner.com/en/newsroom/press-releases/2022-08-22-gartner-survey-reveals-80-percent-of-executives-think-automation-can-be-applied-to-any-business-decision>



資料和演算法相關的法律、法規和政策)、17.可擴展且靈活的系統(Scalable and flexible system)、18.廠商的選擇(廠商採用一個可以與目前系統整合的系統)、19.AI 成本高等，並指出 87%的 AI 和大數據專案會失敗並且從未部署(Merhi, 2023)。另有研究針對新技術和 AI 的採用和開發提出常見的障礙，包括：沒有明確的 AI 戰略、高層缺乏對 AI 承諾、缺乏適當的 AI 能力、組織文化和能力(H. Zhang, Zhang, & Song, 2021)。

綜上分析，有關 AI 在政府治理的挑戰與因應，本研究主要參考 Wirtz 等人(2019)、Zuiderwijk 等人(2021)文獻，並梳理公共治理相關理論，聚焦政府是使用者，整合分類如下：

1. 公共行政改革
2. AI 官僚的合法性
3. AI 系統的品質(資料治理與演算法治理)
4. 在公共治理領域中人機協作的管理
5. 勞動力影響(勞動力替代與轉型)

茲分述如下：

一、公共行政改革

從公部門應用 ICT 的經驗顯示，開發人員想要運用 AI 實現的目標和實際運作之間可能會有很大差異，非科技因素往往才是推動成效不確定性的來源(陳敦源、朱斌好、蕭乃沂、黃東益、廖洲棚、曾憲立, 2020)。例如公部門不願改變其做法(Busch & Henriksen, 2018)、組織分散、制度化的不信任和財務激勵措施失調等。因此，資訊技術的最終效果不是僅取決於技術能力，還取決於技術如何的被使用，個人、組織和政策制定者如何做好準備及回應(House, 2019)。有研究即指出成功實施 AI 的主要障礙並不是技術，在私營和公部門都是如此，演算法僅佔



專案成功的 10%，技術和工程環境(technology and engineering environment)佔 20%，剩下的 70%取決於人員和流程(people and processes)(Patel et al., 2021)。有關公共行政改革可以分幾個面向探討：

(一)、政策制定者的創新困境

當政策制定者必須在創新與常規之間做出選擇時，常面臨創新的困境，新方法可能帶來更大的未知危險以及更大的潛在優勢(Ben-Haim, Osteen, & Moffitt, 2013)。根據財星(FORTUNE)雜誌 2022 年的報導，AI 專案的失敗率在 83-92%之間⁹⁹。當使用 AI 的成功率不高、可能需要長時間的試誤過程，又沒有可供政策學習(policy learning)、政策移植(policy transfer)的對象時，常令政策制定者卻步。因此，無論是國內、外政府公部門(本研究意指「一般行政機關」)對 AI 的廣泛採用遠比私部門來得緩慢，大多數 AI 應用仍處於測試或初期採用階段。再者，有些風險都是猜測的，並沒有實證研究來證實這些猜測(Valle-Cruz, Criado, Sandoval-Almazán, & Ruvalcaba-Gomez, 2020)。

綜上分析，公部門導入 AI 的主要挑戰是似乎仍是執行速度和能力，而不是處理 AI 所產生的風險(Misuraca & van Noordt, 2020)，政府導入創新科技 AI 的思考應該從「創新者的困境(innovator's dilemma)」轉變為「政策制定者的創新困境(Policy-maker's innovation dilemma)」。從方法論的角度來看，若能提供一個系統性視角、具有理論基礎的分析框架，分析可能帶來的各種挑戰(Misuraca & Viscusi, 2014)，讓政策制定者得以在創新與常規之間評估，可以適度舒緩政策制定者的創新困境。

⁹⁹ Want your company's A.I. project to succeed? Don't hand it to the data scientists, says this CEO 檢自 <https://fortune.com/2022/07/26/a-i-success-business-sense-aible-sengupta/>



(二)、終端使用者的科技接受度

由於公務員和公民對 AI 技術的認知是公共服務和政策中可持續使用和實施 AI 的關鍵要素(Misuraca & van Noordt, 2020)。研究顯示，對於沒有直接參與演算法系統技術開發的人來說，這系統像一個「黑箱」，它接受輸入、執行一些被認為機密或專有的演算法，再提供無法解釋的輸出，自然降低終端使用者的科技接受度(Lee & Rich, 2021)。特別是，如果公務員擔心工作會被 AI 取代，例如印尼總統佐科威 (Joko Widodo) 希望將政府機構從目前的四級縮減為兩級，並以 AI 取代兩級的公務員職位，以減少阻礙投資的繁文縟節^{100 101}，當涉及工作權問題，所產生的抗拒會使公部門採用 AI 更加棘手¹⁰²。

因此，有關公部門導入 AI 所產生技術影響的討論，逐漸從 AI 技術人員、專家、倫理學家等，轉移到受影響最大的公務員和公民，亦即所謂的基層專家 (Low level experts) (Misuraca & van Noordt, 2020; Veale, 2020)。再者，在新興研究中，人類增強(human augmentation)或使用 AI 的人類決策者(human decision-makers using AI)被認為是提高終端使用者對 AI 信任的一種方式(Lee & Rich, 2021)。

從理論的觀點，欲增加終端使用者的科技接受度，Davis 在 1985 年提出科技接受模式(Technology Acceptance Model, TAM)，瞭解和解釋使用者接受或拒絕資訊系統的原因，以及分析使用者接受新資訊科技的各項影響因素。

在 TAM 的構面中，「知覺有用性 (Perceived Usefulness, PU)」(相信使用特定系統可以增進工作績效的程度)以及「知覺易用性(Perceived Ease of Use, PEOU)」

¹⁰⁰ 檢自 <https://www.businessweekly.com.tw/international/blog/3000878>

¹⁰¹ 檢自 <https://www.aljazeera.com/economy/2019/11/28/indonesia-will-replace-some-civil-servants-with-ai-says-jokowi>

¹⁰² 檢自 <https://www.npf.org.tw/1/18403>



(相信使用特定系統不需體力或智力付出的程度)是主要兩信念變數，其中「有用性」比「易用性」影響程度更大。有用性及易用性影響「使用態度」(Attitude Toward Using)、「行為意圖」(Behavioral Intention to use)、「實際使用」(Actual System use)，使用態度是「愉快或非愉快地對物體、人、事件、機構或其他人的世界中可分辨的層面的特質，強調正負向、喜好不喜好的評價」，行為意圖衡量「使用者在進行特定行為的意願強度」，實際使用是「使用者滿意度」(user satisfaction)和「使用系統」(system usage)」，而實際使用成為資訊技術接受指標(洪新原、梁定澎、張嘉銘, 2005)。

(三)、經費來源

政府推動任何計畫，首要之務就是確認是否有經費來源，欠缺經費常是政府應用 AI 的推動障礙。美國愛達荷州 2012 年的 Medicaid 計畫使用 AI 預測受補助對象，導致 Medicaid 計畫削減 4000 身心障礙者的補助款，美國公民自由聯盟(American Civil Liberties Union, ACLU)提起訴訟成功，最後州政府 Medicaid 計畫恢復付款，ACLU 隨後指出測試這些系統要花很多錢，州也沒有足夠的動力去花錢提高 AI 正確率(Dignam, 2020)。然而，最近的研究顯示，電腦硬體和資料儲存成本的下降，與十年前相比，現在實施 AI 的成本要低得多(Merhi, 2023)。

(四)、採購制度

採購制度也是政府應用 AI 的推動障礙，主要是數據科學(Data Science)/人工智慧(AI)/機器學習(Machine Learning)專案的概念與傳統的 IT 或軟體開發專案不同，主要的區別包括¹⁰³：

¹⁰³ How AI/Machine Learning projects are different from traditional IT/Software Development . from : 檢自 <https://www.linkedin.com/pulse/how-aimachine-learning->



- 1.專案的焦點：傳統 IT 聚焦於應用發展；AI 聚焦於資料探索和洞悉。
- 2.專案的需求：傳統 IT 聚焦於業務的規則和功能；AI 在於針對業務問題找答案。
- 3.業務的知識：傳統 IT 是將業務規則告訴程式設計師；AI 是提供業務資料去發現規則。
- 4.流程：傳統 IT 流程是可管理的；AI 很難固定流程。
- 5.執行流程：傳統 IT 是迭代有形(tangible)的產出(deliverable)；AI 是多重迭代的試誤。
- 6.方法：傳統 IT 是被結構化和被證明的方法論；AI 是實驗、發現、試誤、多重迭代。
- 7.進程管理：傳統 IT 是漸進式的有形產出；AI 是 60-70%時間花在數據工程(data engineering)，很難估計進度。
- 8.風險：傳統 IT 是改變消費者需求；AI 是資料品質、或資料中沒有發現價值。
- 9.失敗標準：傳統 IT 是客觀的標準，系統無法依照設計執行；AI 是主觀的標準，所產生的洞見是沒有業務價值或沒有結果的行動。
- 10.失敗風險：傳統 IT 是低到中；AI 是中到高。
- 11.專案負責：傳統 IT 是 IT 部門；AI 是 IT 或業務部門。

綜上分析，現行的政府採購制度，公部門需求與技術規格可以在公開招標文件中清晰完整地描述，開發成效容易預期，但 AI 採購不確定因素甚廣。例如完全組態演算法(A fully configured algorithm)，將抽象的數學結構具體化到特定領域的任務中，對於特定案例或問題的操作必須進一步調整，以利在給定情況下採取

projects-different-from-itsoftware-mehta

最佳行動、或對數據的最佳解釋等(Mittelstadt et al., 2016)。廠商必須提供為公部門開發量身定制、創新解決方案的研發服務，研發成本的不確定性直接影響招標金額；公部門資料治理良莠，所提供的數據資料數量與品質直接影響開發成效，導致機器學習執行的任務很難事先預測(如何處理新輸入)或事後解釋(如何做出特定決策)。廠商缺乏數據的知識(例如與數據的範圍、來源和品質)(Mittelstadt et al., 2016)，不確定性將阻礙廠商投標意願。最後，公務員必須願意採用 AI 的推薦數據、同時提供領域知識偕同廠商來訓練機器人等。

因此，政府應用 AI 的過程，**政府機關與廠商之間是「共同設計、共同創造、共同生產」的關係，才能達成預期開發成效，與現行採購作業政府機關與廠商之間是「需求端、供給端」關係的概念截然不同**，在採購作業前期與招標時可能無法以足夠精確的方式確定其需求與技術規範，已顛覆過往政府機關採購的邏輯，因此，必須因應 AI 開發創新採購制度(Molinari F., 2021)。

二、AI 官僚的合法性

美國北卡羅來納州政府使用聊天機器人來處理將近 90% 民眾的諮詢電話 (Mehr et al., 2017)，愛沙尼亞(Estonian)計畫使用 AI 處理小型案件(例如賠償金額不超 6400 歐元)的行政裁量¹⁰⁴(FRA, 2020)，澳大利亞聯邦政府社會服務部 (Services Australia) 「自動債務追繳系統(Automated Debt Raising and Recovery System)」追繳社會福利金補助等，表面上看似應用 AI 技術解決方案，但實質上

¹⁰⁴ AI 法官來了！愛沙尼亞計畫使用人工智慧審理小型訴訟案件 檢自
<https://tomorrowsci.com/technology/ai-%E6%B3%95%E5%AE%98-%E6%84%9B%E6%B2%99%E5%B0%BC%E4%BA%9E-%E4%BA%BA%E5%B7%A5%E6%99%BA%E6%85%A7-%E8%A8%B4%E8%A8%9F%E6%A1%88%E4%BB%B6%E5%AF%A9%E7%90%86/>



已然產生法律效果，使原來的核心法律原則遭到破壞(Busch & Henriksen, 2018; Kuziemski & Misuraca, 2020)。德國在聯邦公務員法(the Federal Civil Servants Act)第 4 條規範，「公務員的決定，不得僅由 AI 自動系統做出決定¹⁰⁵。」(Djeffal, 2020)。美國密西根州 2017 年立法規定失業保險局(Unemployment Insurance Agency,UIA)使用「密西根整合數據自動化系統 (Michigan’s integrated data automated system, MiDAS)」時，必須由人類做出欺詐的判定，並禁止 MiDAS 系統直接做決定(Giest & Klievink, 2022)。美國聯邦法律，憲法規定的正當程序權包括聯邦政府中由人類決策者的權利，而 AI 系統產生的決定都可能因違反行政程序法(Administrative Procedure Act, APA)的任意和反覆無常的標準而受到質疑(Bignami, 2022)。

由於現行的一般法律原則是規範人類而非機器的行為，建議需要有新的法律規範，將演算法決策等同於人類決策(Henman, 2020)。例如德國議會 2015 年修改「行政程序法」第 35a 條¹⁰⁶規定，明確表明法律上可以使用 AI 演算型決策，但前提是必須滿足兩個法律要求：1.法律必須允許 AI 決策，2.該決策沒有裁量權(discretion)，也不能有立法裁量(margin of appreciation)。因此，「code is law(程式碼就是法律)」，AI 系統的自動化決策也必須要遵守法治(the rule of law) (Djeffal,

¹⁰⁵ 原文：According to this provision, decisions relating to civil servants may not be taken by automatic systems exclusively.

¹⁰⁶ 「行政程序法」第 35a 條：「一項行政行為可以完全由自動設備採用，但前提是法律規定許可，並且沒有裁量權(discretion)、立法裁量 (margin of appreciation)(原文：An administrative act may be adopted entirely by automatic devices, provided that this is permitted by a legal provision and that there is neither discretion nor margin of appreciation)。」。德國行政法術語，裁量權(discretion)是指議會授權行政機構決定是否採取行動以及採取何種措施的情況。與此相反，立法裁量 (margin of appreciation)表示專家機構有權審查行政裁量之合法性。

2020)。



三、AI 系統的品質(資料治理與演算法治理)

(一)、資料治理

資料治理技術上的做法本章第一節已有論述，本段著重於在公共行政上的作法。由於機器學習依據數據及演算法做出決策，低品質或不受信任的數據對政府組織來說是一項重大挑戰 (Thierer et al., 2017; Wirtz et al., 2019)。如果數據品質不良(bad)，例如不具代表性、或包含內在和外在的種族、性別或意識形式偏誤(bias)，或演算法偏好(bias)，可能使機器學習所產生的預測模型產生偏誤而不具代表性，甚至可能更加深化或放大歷史偏見，所造成的歧視不亞於公務員的歧視。甚至有文獻指出，機器學習演算法沒有偏差的想法是錯誤的，因為注入模型的資料是無偏差的假設是錯誤的，更具體地說，預測模型實際上可能存在固有偏差，因為它學習並保留歷史偏差(Pessach & Shmueli, 2023)。

舉例來說，為避免政策分析時無法得到接近民意的調查結果，美國加州 2018 年立法禁止機器人(Bot)帳戶假裝為真人，機器人與他人互動必須表明其為機器人之原則(劉彥伯, 2019)。另一個案例，紐西蘭移民署(Immigration New Zealand)的試驗計畫(pilot programme)，建立一個傷害模型(harm model)演算法，利用國家簽證申請程序所蒐集移民的年齡、性別、種族、犯罪紀錄、醫療費用欠繳紀錄等資料，預測居住在紐西蘭的移民，如果續留是否從事犯罪行為或帶來更多醫療支出。一旦模型顯示某些移民可能對國家造成治安疑慮或增加醫療負擔，這些移民將無法重新申請簽證，更甚者，將被直接驅逐出境。一些律師與人權團體因為演算法模型納入種族資料，強烈抨擊移民署的計畫背後隱藏著種族歧視(Daly et al., 2019)。

因此，行政上的做法，是確保 AI 團隊的多樣性，使公務員能使用沒有歧視和



偏見的證據(evidence)，能公平、公正地採取行動和做出決定(UK, 2020)。

(二)、演算法治理

美國聯邦行政會議將演算法治理定義為「使用基於 AI 的工具來支持政府決策、執行和互動。」¹⁰⁷(Engstrom et al., 2020)，顯見演算法治理的重要性。

AI 與傳統 IT 不同，傳統 IT 或早期的 AI，以專家系統為例，是依賴領域專家輸入邏輯規則條件(if-then)來產出自動化決策，因為使用相對較少的解釋變量建立模型，模型相對簡單。現在的 AI，以機器學習為例，是給予電腦判斷、預測、分群的能力，而不需要把所有詳細規則、程式碼寫出來，有些規則是電腦自己利用統計、演算法從資料中自動歸納學習得來的結果。因此，AI 系統的開發人員或操作員可能無法控制或預測其後續行為(Johnson, 2015; Wirtz et al., 2019)。

由於演算型模型是資料、參數、加演算法。演算法歧視可能有不同的來源：

- 1.演算法決策的輸入資料可能權重不佳，導致不同的影響。
- 2.使用演算法本身的決定可能會產生歧視。
- 3.演算法可能會因在不同環境下誤用某些模型而導致歧視。
- 4.以回饋循環的形式，有偏差的訓練資料既可以用作演算法使用的證據，也可以用作演算法有效性的證明(Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018)。

因此，演算法並不會有政治偏好，但演算法必存在方法的偏好，例如是歸納型或演繹型會影響建置及計算效率。政治態度是偏好，設計者也許可以藉由調整權重、參數來形塑偏好，使演算型模型所產出的結果具有政治態度、特定價值觀或特定利益分配方式。因此，若演算型決策應用在對公民生活重大影響領域，例

¹⁰⁷ 原文：The use of AI-based tools to support government decisionmaking, implementation, and interaction—what could be called “algorithmic governance”



如從銀行貸款、司法假釋決定、住房、就業、教育、社會福利、自由、隱私和尊嚴等，所產生的新型態歧視，將有別於傳統歧視(Dignam, 2020)。

然而，受限於研究人員對 AI 神經網絡如何做出決策的理解還處於早期階段(OECD, 2019)，美國國防高等研究計劃署 (Defense Advanced Research Projects Agency, DARPA) 已致力於「可解釋 AI (Explainable Artificial Intelligence, XAI)」的研究。「可解釋性人工智慧(XAI)」是「新的或經過改進的機器學習技術，結合有效的解釋技術可產生可解釋的模型，可使最終使用者理解、信任並有效管理新興的 AI 系統」(De Spiegeleire, Maas, & Sweijjs, 2017)。Adadi 和 Berrada(2018)指出需要可解釋性的 4 個原因：證明(justify)、控制(control)、改善(improve)和發現(discover) (Adadi & Berrada, 2018; Dwivedi et al., 2019)。簡言之，XAI 指開發能夠解釋與詮釋學習模型的 AI 演算法，可解釋性是解釋特定決策，分類或預測背後的原因的能力。Derek Doran 等人指出了二種不同的解釋途徑：意會系統(comprehensible system)、邏輯系統(interpretable system)，邏輯系統顯然是最為透明且直接的理解途徑，但意會系統也已經讓黑箱不再完全無法捉摸(賈文字, 2020)。另亦有技術研究根據數據特徵在模型中的重要性對數據特徵進行排名(ranking)、排序(sorting)和評分(scoring)，或通過使用視覺化技術或文本論證來顯示模型的決策，從而使機器學習模型更具可解釋性。雖然如此，技術挑戰依然存在，尤其是對於更複雜的演算法模型(Engstrom et al., 2020)。

因此，在 AI 技術目前尚無進一步重大突破的情況下，有研究提出現階段建議採用「人機協作」的方式，但目前還沒有廣泛使用且具有廣泛可比性的測試來評估人機協作相對於單獨人類、單獨 AI 性能或品質的方法(Malone et al., 2023)，也尚未充分分析人類與 AI 共存和協同工作的最佳條件(Seeber et al., 2020)，再者，人機協作也有課責問題。另一方面，亦有從控制的角度出發，例如英國建議在 AI 流程的每個階段，從設計、採購、部署等，都應該有一個人為控制的元素



(UK, 2020)。

(三)、AI 系統的品質

AI 系統功能的增強與人類依賴它們的趨勢之間有著直接的關係(Bullock, 2019)，從 AI 系統功能觀點，雖然 AI 系統的功能逐漸增強，但仍有侷限性：

陳昇瑋和溫怡玲(2019)使用「情境相關性」和「資料可得性」等兩個軸度分類。AI 技術最擅長的領域是解決和情境無關且樣本量多的問題，而最不擅長的領域則是解決高度情境相關且樣本量少的問題(陳昇瑋、溫怡玲, 2019; 黃心怡、曾冠球、廖洲棚、陳敦源, 2021)。

Bullock(2019) 根據公共政策問題的「複雜性(complexity)」和「不確定性(uncertainty)」兩個關鍵構面來判斷適合人類主導還是機器主導，複雜性高(與規範的偏差較大)和不確定性高(難以分析)的任務很可能仍然是由人類自行完成；複雜性較低(與規範的偏差較小)和不確定性較低(易於分析)的任務更有可能由機器完成，而不確定性高的任務通常相對應的是「可自由裁量的程度」(Bullock, 2019)。

Young 等人(2019)針對公部門的行政任務提出「人工智慧裁量(artificial discretion, AD)」的概念，AD 的定義是使用 AI 來「增強(augment)」或「自動(automate)」行使行政裁量權。Young 等人(2019)提出「任務特異性(task specificity)」和「環境複雜性(environmental complexity)」二個軸度的理論框架，幫助公共管理人員如何判斷 AI 的決策：「人工智慧裁量(artificial discretion, AD)」或「人類裁量(human discretion, HD)」。「任務特異性」取決於任務的複雜性、數據的品質和可用性、技術要求、AI 系統的限制、與任務相關的風險和不確定性，以及使用非人類解決方案的可行性。任務特異性大致上可以用完成任務所需的自由裁量程度(the degree of discretion)來描述，AD 適合越低自由度裁量權、任務需要較低判斷力、不同人判斷品質差異很大等；HD 適合任務所需越高自由裁量權。「環境複雜性」是脈絡分析的程度(Contextual Level of Analysis)，無爭議和



非競爭性的公共產物(例如安排救護車通過)、直接的、狹義的、相對受限的任務(例如開停車罰單)、有限數量的利益相關者等，適合 AD；HD 適合越高自由裁量程度、涉及對特定利益群體或人群產生不同影響的權衡取捨、政策或任務定義不明確、利益相關者很多等。簡言之，AD 的關鍵任務維度：所需行政裁量權程度(the level of discretion required)和分析的程度(the level of analysis)(Busch & Henriksen, 2018; Young et al., 2019)。

從公務員行使行政裁量權的觀點，當公部門遇到政策目標模糊、高度的情境相關性、可用資源不足、管理控制困難、當事人的結構性弱勢以及複雜人性特質時(Buffat, 2015)，政策制定者被迫在政策內容中，以多元但是模糊的目標來賦予執行者更多的裁量權(葉俊榮, 2005)，正是 AI 系統可能不擅長的領域。

若欲在 AI 可能不擅長領域使用 AI，Young et al.(2019)即指出以「人工智慧裁量(artificial discretion)」取代「人類裁量(human discretion)」時，應仔細考慮政府職能品質的適當閾值(appropriate threshold) (Young et al., 2019)，意旨不同政策工具或政府職能，可以容許的 AI 偏誤範圍不同。然而，如何評估 AI 系統的適當閾值或偏誤範圍？

歐洲委員會(EC)發布的可信任 AI (trustworthy AI)的倫理準則(Ethics guidelines)增加質量評估項目(Kuwajima & Ishikawa, 2019)。然而，如何進行 AI 系統品質的評估？由於現有資訊系統的原理和方法均無法對 AI 軟體有效進行質量評估(Quality Assessment)，如何對 AI 軟體功能進行質量驗證(quality validation for AI software function features)已成為學術界和工業界的關注重點和研究課題(Kuwajima & Ishikawa, 2019)，例如 AI 軟體推薦服務的精準度(precision)是多少？如何執行測試和品質驗證？再者，AI 軟體設計通常具有情境相關性，不同於傳統軟體，可能發生的錯誤涵蓋功能或程式錯誤、訓練數據的錯誤或缺陷，必須增加功能測試品質評估(function test quality evaluationis)，如果評估結果未被利害關係



人接受，則必須再次進行建模(modeling)以進行新的測試迭代(testing iteration)，如何對 AI 系統進行品質驗證是一大挑戰(Tao, Gao, & Wang, 2019)。因此，AI 系統的質量評估的方法，有研究認為應該包括 AI 系統軟體評估及功能評估(Tao et al., 2019)，有研究建議以傳統評估軟體品質框架或模型的 SQuaRE (System and Software Quality Requirements and Evaluation)為基礎，再以機器學習特性(ML nature)為基礎的 AI 系統增刪質量評估項目。

最後，從 AI 系統品質需求的標準討論，有文獻指出，AI 使用在和公民權利義務相關的領域時，技術正確率(Technical accuracy)非常重要(Henman, 2020)。也有研究以美國社會安全局身心障礙者申請補助的正式裁決的案例，AI 的決策品質超過正確率(accuracy)超過 80%，就可以減少公務員案件處理時間 12%、降低 7.5%的訴願案件。另一個案例，美國食品和藥物管理利用 AI 預測嚴重的藥物不良反應真陽性率(true positive rate) 91%、假陽性率(false positive rate) 4.9%，由於不正確(inaccurate)的預測可能會導致生死攸關的後果，即便是具有精確度(precision)和專業知識的 AI 模型，FDA 至今沒有完全成功部署(Engstrom et al., 2020)。

四、公共治理領域人機協作的管理

人機協作可以使公務員認知增強(Cognitive Augmentation)和生理增強(Physical Augmentation)，公務員和 AI 像一個團隊共同努力提高績效(Burke et al., 2019)、實現共同的目標(Semeraro, Griffiths, & Cangelosi, 2023)，有機會提高公共治理能力：

1. 認知增強：我們的大腦認知能力有限，難以處理大量資訊(Kolkman, 2020; Todd, 2007)，利用 AI 與人腦相互關聯、相互支撐前進，AI 有效地延伸了人類大腦，拓展人類的認知範圍(Dong et al., 2020)，強化人類思考和做出更好決策

的能力。由於認知是人類通過感官輸入、生活經驗、學習、思考來獲取知識的過程，AI 使人類在無須接受密集又昂貴培訓課程取得專業技術或知識。

2. 生理增強：包括感覺增強(sensory augmentation)、附肢和生物功能增強(appendage and biological function augmentation)、大腦增強(brain augmentation)、基因擴增(genetic augmentation)等。

人類判斷和演算法結果結合可以提高判斷力或正確率。例如深度學習技術預測癌症，演算法錯誤與人為病理學家的錯誤不相關，演算法的預測結果與病理學家診斷相結合，有助於將人為錯誤率從 3% 減少到不足 1%(Desouza, 2018)。

然而，AI 技術能賦能人類，同時也限制和制約人類(Bueger & Stockbruegger, 2017)，使得人與機器互動的結果，AI 系統可以用來支持/協助(support/assist)人類決策者，也可以用來代替(replace)人類決策者。例如依規定街頭官僚仍有裁量權，可以接受或拒絕計算機提出的決定，由於演算法的原理對人是難以理解，使得決策的正當性(legitimacy)難以挑戰，產生不同的人在決策時可能會採取不同的態度和行動(Duan et al., 2019; Mittelstadt et al., 2016)，包括演算法趨避和自動化偏誤：

(一)、演算法趨避(algorithmic aversion or algorithmic avoidance)

許多研究使用人類決策作為衡量人們對 AI 感知的基準，結果顯示人們對 AI 決策的信任遠低於人類決策，特別是對於主觀和/或被認為需要人類能力的決策，即使已明確實證研究證明，演算法判斷的績效優於人類判斷(Awad et al., 2018; Bonnefon, Shariff, & Rahwan, 2016)。在某些案例中，組織雖然採用 AI 技術，儘管 AI 模型優於人類直覺(Longoni, Bonezzi, & Morewedge, 2019)，人們還是依靠自己的直覺或他人的直覺做出決策，而不是 AI；或是 AI 挑戰人類專家的決策，衍生刻意忽視 AI 的建議(Seeber et al., 2020)。再者，人們對演算型決策(Algorithmic



Decision making, ADM)的寬容似乎遠不如對人類的寬容：演算法犯了一個錯誤(即使其總體性能優於人類)，人們仍傾向選擇人類的決策者(Araujo, Helberger, Kruikeimeier, & De Vreese, 2020)。有研究針對企業管理者(經理)調查「人機協作」偏好的組合，是人類在決策過程中佔 70%的權重，機器佔 30%的權重(Haesevoets, De Cremer, Dierckx, & Van Hiel, 2021)，在人機協作的過程中，人類似乎傾向於低估 AI 潛在的效益(Bobadilla-Suarez, Sunstein, & Sharot, 2017)。

因此，公務人員可能以 AI 設計開發人員或管理人員無法想像的方式使用這些技術，例如以懷疑態度對待 AI 或完全忽略它的建議(Bailey & Barley, 2020; Misuraca & van Noordt, 2020)，或是公務員不信任他們「AI『同事』(AI“colleagues”)」而仔細重複檢查(double-check)所有機器的工作成果，反而可能會降低生產力(productivity)而不是增強(enhancing)能力。

(二)、自動化偏誤(autonomous bias)

有學者指出縱然決策者知道 AI 演算結果有可能不正確、不完全或有偏誤，但對機器產出之(參考)結果，往往不會站在中立立場，而是會給予積極的評價(鄭明政, 2020)。有些基層官僚傾向贊同機器所產生的決策，而忽略與機器決策相反的資料，或相衝突的人類決策，通常是直接接受計算機決策(Skitka, Mosier, & Burdick, 1999)，導致機器決策的誤用。例如：AI 輔助偵測系統決策牽涉太多細節，以致難以覺察時，在快速檢查過程中，人類過度信賴、監控不足、盲目同意等(王自雄、張腕純, 2020)。舉例來說，美國普渡大學(Purdue University)的研究，ChatGPT 有 52%回答不正確，只有 48%正確，但 39.34%的題目會讓參與者忽略 ChatGPT 的錯誤信息(Kabir, Udo-Imeh, Kou, & Zhang, 2023)。過於依賴這些 AI 機器的危險，遠比機器本身自主學習產生的危險更甚(Barth & Arnold, 1999)，因此，在人機協作的新工作模式中，公務人員亟需主動介入，而非毫無質疑地將權

力移轉給 AI (黃心怡、曾冠球、廖洲棚、陳敦源, 2021)。再者, AI 技術與以往的技術不同, AI 系統具有自主學習能力, 從而破壞治理的合法性(legitimacy of governance), 此時, 包括系統設計師、法律政策官員, 特別是 AI 專家又被視為以前街頭官僚的新成員(Bovens & Zouridis, 2002; Unver, 2018)。

然而, 在人機協作的模型極為複雜, 一項關於人機協作的調查, **大部分的研究主要關注生理層面人機協作(physical human-robot collaboration)**, 從感知環境推斷, 例如協作策略(collaborative strategies)、學習方法(learning methods)、感測器(sensors)和執行器(actuators)等層面, 完成實際的物理任務(actual physical task), 包括協同組裝(collaborative assembly)、物件移交(object handover)、物件搬運(object handling) 和協作製造(collaborative manufacturing); **缺少關注認知層面的人機協作(cognitive human-robot collaboration)**, 例如人和機器互動過程中, 決定要執行下一個動作時, 主要考慮因素包括人的意圖(human intention)、人類對機器人的信任、機器對場景(the scene)的認知等, 亦即當前互動狀態的資訊評估要採取的行動, 而不是從另一項行動中所獲得的資訊評估下一個動作(Semeraro et al., 2023)。

Gartner 發布 2023 年十大戰略科技趨勢報告, 提出 10 項當前關鍵科技, 包括對 AI 的信任、風險和安全管理 (AI trust, risk and security management, AI TRiSM)、自適應性人工智慧(Adaptive AI)等¹⁰⁸。其中與人機協作相關的是自適應性 AI (Adaptive AI), 是一種能夠根據不同情境、使用者的反饋(例如人類偏好、領域知識修正)等數據來源, 自我學習、自動調整其模型和演算法的 AI 系統, 以提高其效率和準確性; 且不同於以往的「訓練 → 部署」的直線流程, 是具有高度時間依賴性。因此, 在公共治理中若欲提高 AI 的應用潛力, 同時避免數位裁量權和數位官僚主義進一步擴張, 公部門應妥善管理人與 AI 互動關係。例如探

¹⁰⁸ 檢自 https://www.profetai.com/en/post/gartner_2023_aia_recap

究在人機協作過程中，「人的認知」和「AI 的認知」所產生的交互影響，以及不同公共任務與人機協作的細部關係，包括「是否」適合及「如何」發展出人機協作的可能性，以及相對應的完全交由機器處理、部分交由機器處理，以及維持傳統人工處理的模式(黃心怡、曾冠球、廖洲棚、陳敦源, 2021)。

五、勞動力影響(勞動力替代與轉型)

有研究認為 AI 和機器人技術不斷發展將意味著人類的工作終結，然而，許多經濟學家認為，過去經驗顯示技術突破最終導致勞動力和工資需求，沒有理由擔心這次會有所不同(Acemoglu & Restrepo, 2019)。再者，過去的經驗顯示生產力和工資之間一直存在很強的相關性，要利用 AI 同時達成提高生產力和提高工資，不僅取決於技術變革的性質，更重要的是政策和機構如何協助工作者為 AI 做好準備，以因應勞動市場的衝擊(Holdren & Smith, 2016)。

數位發展部長唐鳳把 AI 理解成「Assistive Intelligence」¹⁰⁹，輔助人類的智慧，人類才是主體，AI 取代的是人類不適合做、不想去做的事情，人類可以從事更有創意的工作(蘇經天, 2023)。因此，為因應未來工作性質可能發生變化(Lund et al., 2019)，勞動力工作技能必須轉換，或是工作需要職務流程再設計(黃彥鈞, 2019)，例如勞動力的角色可能會轉向需要更多監督的功能及 AI 特定技能(Wirtz et al., 2019)的數據科學家或機器學習工程師；或是利用 AI 和人類協作，使工作者認知增強，從而擴展工作內容(Desouza, 2018)。為達成上述目標，雇主利用 AI 提高生產力和員工擴增能力應該同時進行，以達成雇主與員工共享共榮的雙贏目標，否則所得分配不均的政治效應可能會使新技術減緩甚至完全停止採用和發展(Acemoglu & Restrepo, 2019)。美國密西根州 2014 年使用「密西根整合數據自動化系統 (Michigan's integrated data automated system, MiDAS)」，降低行政成本並

¹⁰⁹ 檢自 <https://finance.ettoday.net/news/2307456>



打擊詐領失業保險給付，然而 AI 系統導致失業保險局(Unemployment Insurance Agency, UIA)第一線服務人員 1/3 被解雇(260 人減少到 184 人)，另外聘僱人員從事系統監督及資料蒐集的工作(Giest & Klievink, 2022)。因此，為避免結構性失業，公部門在 AI 導入的過程應協助員工為 AI 做好準備，緩解技術落差。可能的作法如下：

1. 「以人為本」設計 AI 系統

AI 設計師的背景是計算機科學(computer science)、資訊科學(information science)和智慧系統(intelligent systems)，技術優先於社會的文化，使技術設計的思考是提高效率和降低勞動力成本，而不是新技術對社會產生的影響，一旦一項新技術完成並開始使用後，要再修改技術設計的機會將會縮小。另一方面，經濟學家和政策制定者傾向分別在經濟和社會層面評估 AI 對工作的影響，政策制定者再制訂緩解社會影響或勞動力衝擊的政策。然而，已有實證研究指出，若能在 AI 技術設計階段，組織跨領域學者、受技術變革影響利害關係人研究團隊，透過協作，產出能兼顧新技術、協助工作者「增強智慧(augmented intelligence)」的 AI 技術，可以避免技術性失業的問題，換句話說，相同的 AI 技術仍有可能對不同的組織產生不同的結果(Bailey & Barley, 2020)。

因此，AI 技術「以使用者為中心(user-centered)」和「價值敏感(value-sensitive)」的設計，遵循「在設計過程的早期階段專注於使用者和任務性質」、「使用經驗評估來評估設計的效果」、「不斷進行使用者新資訊和經驗評估的迭代設計(design iteratively)」等原則，或利害關係人與 AI 系統設計者共同參與資料治理，AI 模型協作(collaboration)的「參與式設計(participatory design)」，包括共同創造(co-creation)、共同設計(co-design)、共同調校(Berditchevskaia, Malliaraki, & Peach, 2021)，可以增加使用者的接受度、避免技



術可能帶來的倫理問題。良好設計的方法包括脈絡調查(contextual inquiry)、任務分析(task analysis)、觀察(observation)、焦點團體(focus groups)、使用者訪談(user interviews)、使用者體驗(user personas)、調查(surveys)和使用者實地測試(user testing in situ)等(Bailey & Barley, 2020)。

2.強化培訓協助員工數位轉型

政府 AI 技術能力建構的策略，通常分為「製造(make)」或「購買(buy)」決策。理論上，私部門擁有更多的專業知識，可以以更低的成本生產；在實踐中，不涉及核心職能或行政裁量權的任務，政府通常利用私部門的專業知識和效率；任務涉及核心職能或較多裁量權，通常用「製造(make)」的方式(Engstrom et al., 2020)。

雖然可運用公私協力等解決政府 AI 技能落差的問題，但公務員仍然必須要具有一定水平的技術知識判斷 AI 系統的可解釋性、與政策是否相符、精確操作 AI 系統、確保輸入正確數據、識別錯誤、識別演算法模型中可能的偏見、避免機器學習成果失真，謹慎使用 AI 決策等。高階領導者必需給予員工足夠的訓練、必須確保數據品質經過適當的檢查、避免公務員隨意更改 AI 做決定的基礎，以確保對 AI 系統造成的任何風險進行適當的治理(UK, 2020)；如果未對 AI 系統的操作人員進行適當的培訓，要求他們對接受或拒絕 AI 決策負責是不合理的(Wirtz et al., 2019)。因此，AI 系統的最終責任在於決定採用和執行 AI 系統的高階領導者(UK, 2020)。此外，AI 技術訓練應該在發展 AI 系統之前(OECD, 2019)，以專案需求為導向的訓練(Project-Based Learning, PBL)(陳敦源、朱斌好、蕭乃沂、黃東益、廖洲棚、曾憲立, 2020)，並且是整個 AI 系統生命週期的持續過程，而不是一次性事件(UK, 2020)。

美國 2019 年「維持美國 AI 領導地位的行政命令」要求行政機關(構)應培訓



AI 的研究人員、使用者、聯邦工作者(Trump, 2019)。政府 AI 準備指數前幾名的國家對公務員 AI 能力的培訓不遺餘力，英國 2019、2020 年政府 AI 準備指數是第 2 名，2014 年在工作與退休金部(Department for Work and Pensions, DWP)建置數位學院(Digital Academy)；另為使英國具有領先世界、最精通數位技術的公務員，2017 年發布「政府轉型策略 2017-2020(Government Transformation Strategy 2017 to 2020)」，其中「公務員計劃(Civil Service Workforce Plan)」培訓公務員數位、數據和技術(digital, data and technology, DDaT)專業，包括服務管理(service management)、內容設計(content design)和資料科學(data science)等，以確保能將數位工具和技術鑲嵌入公務人員原有專業領域中，以因應政府數位轉型的變化(G. UK, 2017)。新加坡 2019 年政府 AI 準備指數是第 1 名，新加坡於 2018 年 6 月公布「數位政府藍圖」(Digital Government Blueprint)，預計 2023 年前，所有公務人員都要有基本的數位能力(basic digital literacy)、2 萬名公務員培訓數據分析(data analytics)和資料科學(data science)能力¹¹⁰。2020 年政府 AI 準備指數是第 8 名的丹麥政府將建立一個中央政府內部學院，一般人員提供 AI 通識教育課程，資訊人員培訓處理大數據和 AI 發展(Misuraca & van Noordt, 2020)。

¹¹⁰ Digital Government Blueprint。檢自 <https://www.tech.gov.sg/digital-government-blueprint/>



第四節 公共治理導入 AI 的研究現況

公共治理使用 AI 的研究進程，目前集中於概念框架(conceptual framework)或框架分析¹¹¹，理論化不足(Valle-Cruz, Alejandro Ruvalcaba-Gomez, Sandoval-Almazan, & Ignacio Criado, 2019)。本節嘗試整理既有研究方法、理論、分析框架等，歸納主要因素(the main factors)、結構(constructs)或變量(variables)，以利進一步研究設計與探究可能的理論基礎。

壹、公共治理導入 AI 的研究方法

有關在公共服務中使用 AI 的研究方法，依序為定量(56%)、定性(28%)、定性和定量相結合(17%)。定量方法較多的原因是針對實際公共政策問題提出技術解決方法，包括新型 AI 技術方法、比較不同演算法技術或模型提高 AI 決策品質，改善公部門的績效、提高政府職能(the functions of government)的公共價值，例如城市道路二氧化氮污染值的預測模型比較(Cabaneros, Calautit, & Hughes, 2017; de Sousa et al., 2019)；或是統計和數學方法分析公共服務應用領域。

在學術研究上，採用定量方法的困境包括 AI 應用過程的數據密集性(data-

¹¹¹ 概念框架與理論、模型或理論框架(theoretical framework)是不同，當處理非常複雜的現象時，概念框架通常是第一步(Misuraca et al., 2019)。框架以圖形或敘述的形式解釋要研究的主要因素、結構或變量，以及它們之間的建議相互關係。這樣的框架不僅促進複雜研究主題的抽象和概念化 (Goffman, 1974; Sun & Medaglia, 2019)，且有助提供新興研究領域研究方向(Kankanhalli, Charalabidis, & Mellouli, 2019)。



heavy nature)和社會嵌入性(social embeddedness)，所產生數據存取(data access)、蒐集、管理和匿名化方法技術(anonymization methods)的障礙，以及在政府中實施 AI 的本質，例如人臉辨識、追蹤和監督、自主代理、管制性政策查核方式所產生隱私及安全性的問題。

有關 AI 與公共治理之間關係，現有研究大多數都採用定性方法，但未使用特定的定性方法，且**很大程度上是探索性研究**，探索性研究集中在於文獻綜述、案例研究、研究框架(research framework)、政策和倫理影響、機會與挑戰、公共服務應用領域潛力等，**探索性研究缺乏理論框架和嚴謹性**(Zuiderwijk et al., 2021)。另一類研究方法則主要依賴研究(或借鑒)私部門的實施成果應用於公部門環境(Valle-Cruz et al., 2020)。

因此，Zuiderwijk 等人(2021)呼籲公共治理研究應該超越探索性研究，並認為解釋性研究設計(explanatory research designs)可以應對這些研究挑戰，進一步了解如何(how)、為什麼(why)、以及因果關係(causal research)，包括公部門採用 AI 應如何治理、如何解決潛在挑戰、各種公共價值(例如透明度和 AI 系統效能)之間如何權衡的公共治理理論或通用框架(Aoki, 2020; Cath, 2018; Dwivedi et al., 2019; Kuziemski & Misuraca, 2020; Zuiderwijk et al., 2021)。

貳、公共治理導入 AI 的理論

Zuiderwijk 等人(2021)在探討 AI 在公共治理的應用，利用系統性文獻回顧方法搜索以英語撰寫並於 2010-2020 年發表的期刊文章和會議記錄，目前公共治理研究使用的特定理論，包括：

1. Androutsopoulou 等人(2019)，使用媒體豐富性理論(media richness theory)和頻道擴展理論(channel expansion theory)論述 AI 聊天機器人改善政府與公民的溝通。



2. Sun 和 Medaglia(2019)使用框架(framing)理論來收集利益相關者的觀點。
3. Ojo et al. (2019)使用技術採用理論(technology adoption theory)來論述研究背景及其結論。
4. Wirtz et al. (2020)使用管制理論(regulation theory)作為 AI 治理框架的基礎。

從本質上講，大多數的研究方法是實用的，並且側重於概念框架，但所有研究都沒有進一步測試或擴展理論(Androutsopoulou et al., 2019; Ojo, Mellouli, & Ahmadi Zeleti, 2019; Sun & Medaglia, 2019; Wirtz, Weyerer, & Sturm, 2020; Zuiderwijk et al., 2021)。

另一方面，Kankanhalli 等人(2019)研究智慧政府或智慧城市的各種 IoT 和 AI 框架，提出物聯網賦能 AI 系統(IoT-enabled AI Systems)的研究框架，另整出多種理論觀點(Chatfield & Reddick, 2019; El-Haddadeh, Weerakkody, Osmani, Thakker, & Kapoor, 2019; Kankanhalli et al., 2019; Tang & Ho, 2019)，例如：

1. El-Haddadeh 等人(2019)技術採用(technology adoption)。
2. Sun & Medaglia(2019)框架。
3. Tang & Ho(2019)路徑依賴(path dependence)。
4. Androutsopoulou 等人(2019)媒體豐富性理論。
5. Chatfield & Reddick(2019)動態能力觀點(dynamic capabilities perspective)等。

此外，本研究另整理：

1. Gasser, U., & Almeida, V. A. (2017)倫理與法律理論(Gasser & Almeida, 2017)。
2. Rahwan, I. (2018) 社會契約理論(Rahwan, 2018)。
3. Wright, S.A. & Schultz, A. (2018) 整合利益相關者理論與社會契約理論(Wright & Schultz, 2018)。

- 
4. Almeida, P., Santos, C., & Farias, J. S. (2020) 彙整 15 個獨特的框架和幾種不同的理論(倫理、法律與技術的互動式治理模型、管制理論、社會契約理論、利益相關者理論、機器人原理、敏捷式治理等) (Almeida, Santos, & Farias, 2020)。
 5. Valle-Cruz, D., Alejandro Ruvalcaba-Gomez, E., Sandoval-Almazan, R., & Ignacio Criado, J. (2019)政策週期理論(Valle-Cruz et al., 2019)。
 6. Winfield, A. F., & Jirotko, M. (2018) 採用敏捷式治理理論(agile governance)、負責任的研究與創新理論(Responsible Research and Innovation, RRI)、公民參與(public engagement) (Winfield & Jirotko, 2018)。
 7. Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013) 控制理論 (Friedman et al., 2013)。

分析現有使用理論，聚焦於社會決定論或技術決定論領域，缺少在現有公共治理論中將公部門應用 AI 的決策及行為理論化的觀點，也少有採用科學科技與社會互動的理論：將人和技術物同時納入為行動者。

參、公共治理導入 AI 的框架

本研究整理文獻，與公共治理較為相關概念框架、分析框架以及理論框架大致可以分為從技術社群角度、從社會科學角度二大類分析：

一、從技術社群的角度

目前研究大部分框架是由技術研究社群所提出(Almeida et al., 2020; de Sousa et al., 2019; Dwivedi et al., 2019; Friedman et al., 2013; Gasser & Almeida, 2017; Kuziemski & Misuraca, 2020; Mok & Hyysalo, 2018; Molinari F., 2021)，框架目的是降低 AI 風險，主要變量是 AI 技術、倫理準則和原則、法律、管制、公民參與



(或利益相關者)、應用領域、互動式治理等。舉例來說：

1. Kuziemski, M., & Misuraca, G.(2020) 提出目標(goals)、驅動因素(drivers)、障礙(barriers)和風險(risks)來評估在公部門使用 AI 的潛在影響，以及 AI 的使用如何加劇現有的權力不對稱性(power asymmetries)的觀點。
2. Gasser, U., & Almeida, V. A. (2017) 提出 AI 治理的分層模型，消除開發人員與公民、決策者之間的資訊不對稱。
3. Dwivedi 等人(2019) 提出跨領域、多學科觀點的研究方法，並提出未來人類與 AI 之間的關係將朝人機協作的方向發展，人類和 AI 協作共同創造價值；而本篇所提出公共政策框架：TAM-DEF 框架(TAM-DEF framework)，主要是指導公部門在使用任何 AI 系統之前必須提出的問題、客觀測試，決定是否使用特定的 AI 系統，然後再啟動提供公眾使用。
4. Molinari F.等人(2021) 提出 AI 影響評估分析框架包括四個分析維度：公部門 AI 解決方案的設計(design)、採用(adoption)、實施(implementation)和使用(use)，強調政府與利益相關者共同設計(co-design)、共同創造(co-created)、共同生產(co-production)、共同遞交(co-delivering)或共同評估(co-evaluating)，可以使公民及企業獲得公共服務的關鍵利益同時避免風險。本篇研究另提出公部門的數位轉型並非線性或可預測地進行，公部門採用 AI 的成功關鍵因素，不是只有與演算法、數據和計算能力有關，更與人有關，不只是為人而做(only done “for” the people)，而是人需要與 AI 共存(needs to be done together “with” them.)¹¹²。本篇研究考慮 AI 技術物的特性，給予廣義對稱性的思考邏輯，頗值得參考，但未敘明如何與公共治理理論結合。

¹¹² 原文：AI is not only about algorithms, data and computing power, but also about people. Therefore, its implementation may never be completely successful is only done “for” the people, but needs to be done together “with” them.



二、從社會科學社群的角度

目前研究框架是從不同類別的期刊或學者背景提出，包括公共政策、商學、人類學等(Rahwan, 2018; Sharma et al., 2020; Valle-Cruz et al., 2019; Wirtz et al., 2019; Wright & Schultz, 2018)；框架目的是確保 AI 技術的收益和成本在不同的利益相關者之間合理分配；主要變量包括 AI 技術、演算法的擬人化¹¹³、公共政策週期、管制、社會契約、協作治理、利益相關者、倫理、應用領域。舉例來說：

1. Wirtz 等人(2019)主張將受影響的利益相關者、公共和私人利益集團的代表均納入 AI 系統的協作治理。
2. Valle-Cruz 等人(2019) 提出 AI 和大數據將改變公共治理理論中公共政策過程階段論的概念，並提出動態公共政策週期來擴展和補充公共治理理論。
3. Rahwan (2018) 提出維持社會的管制模型，將「將人類置於循環中監督，以增強人類的決策能力(human-in-the-loop, HITL)」模型調整為「將社會置於循環中監督，以增強社會的決策能力(society-in- the-loop, SITL)」模型，權衡不同人類價值觀，HITL AI 是將個體或群體的判斷嵌入具有微弱影響的 AI 系統，以優化 AI 系統；而 SITL AI 則是將社會價值作為一個整體，嵌入對社會具有廣泛影響的演算法治理中。本篇研究主張因為將 AI 應用到整個社會，影響領域更廣泛，在社會需要一種代表所有利益相關者規範方法，所提出框架探討 AI 系統中社會價值重要性，頗值得參考，但缺乏面對社會價值衝突時如何權衡方法，且這樣的做法勢必增加公部門應用 AI 的困難度。

肆、公共治理導入 AI 的理論缺口

分析上述目前理論、框架的研究進度，不論是技術社群角度、社會科學角度，都同時指出 AI 應作為主要變量，其他包括利益相關者、演算法的擬人化、

¹¹³ 原文：any kind of living being

公民參與、倫理、協作、互動、應用領域也常出現；再其次是依據不同學術領域而有不同變量，例如法律、管制、公共政策週期、社會契約等。



經分析目前研究進度後，目前理論、框架仍有以下不足之處：

1. 缺少 AI 是行動者，且具自適應性及能動性(agency)的觀點：

既有文獻所提出的概念框架或框架分析大多數忽略 AI 是行動者及有能動性(agency)，甚至 AI 是智慧代理人 (intelligent agents)及具有代理關係(agency)，因而無法有效解釋 AI 官僚、AI 自主決策(autonomous decision)、AI 與人共存，原本屬於人類官僚的行政裁量權發生權力移轉的現象。

2. 缺少人類監督程度、AI 自主、人機協作(AI 與人的關係)的觀點

AI 系統旨在促進人與機器之間的協作，然而既有理論或框架主要側重以人為主的社會決定論，或以技術為主的技術決定論，缺少人與機器必須相互依存、合作互利共生的詮釋邏輯，例如人與 AI 互動關係的理論或框架。

3. 缺少國家權力涉入程度、國家官僚化(bureaucratization)的威權(authority)、AI 與國家權力關聯的觀點：

國家理論(theories of the state)的中心論點是國家的行政、立法和司法機構對公共政策的性質具有重要影響，其中官僚更發揮舉足輕重的作用(Hall, 1993)。當官僚和 AI 結合，能夠轉變國家職能(the transformation of state functions)或擴大官僚威權(authority)影響的範圍(McQuillan, 2020)；例如擁有更多關於公民的資訊並控制資訊(Busch & Henriksen, 2018)，強化國家統治能力。然而，既有 AI 治理理論或框架偏向普遍應用性概念，缺少政府部門具有行使統治權作用之行為。因此，當公部門採用 AI，國家權力與 AI 結合，在各種難以預料的因素(Barcevičius et al., 2019)產生相加效應(additive effect)、協同效應



(synergism)、增強作用(potentialiation)等，可能擴大權力差距，也可能縮小權力差距。

綜上分析，在理論層次上，必須先探究 AI、人與 AI 共存與互動關係、相關變數，才能在意義上釐清 AI 對公共治理理論的影響。

伍、AI 與人的互動關係

在公共治理中的人與 AI 互動關係如何呈現？有研究提出超越自由個人主義(liberal individualism)和技術理性(technical rationality)控制模式(modality of control)：「機器+人類(Machine + human)」的認知組合(cognitive assemblages)(Crosset & Dupont, 2022)。另有研究使用「賽博格¹¹⁴官僚(Cyborg bureaucracy)」，cyborg 意指「人機混合體(human-machine hybrids)」，使人們能夠理解人類和技術在公共服務中的整合(Breit, Egeland, & Løberg, 2019)，以及技術如何約束官僚可用的自由裁量權程度(the degree of discretion)(Nisar & Masood, 2018)。

本研究從公共治理、資訊科學、ANT 三個觀點分析 AI 與人的互動關係：

一、公共治理觀點

使用 AI 作為決策支持工具，利用機器預測(Machine Predictions) 來增強人類決策(Human Decisions)，以遠遠超過人類認知速度的方式，增加人類可用資訊的範圍和品質；如同 Kleinberg 等人(2018)指出必須經過「數據→預測→決策(Data→Prediction→Decision)」的過程(Kleinberg et al., 2018)。因此，人與 AI 的互動關係是分工(division of labor)的關係，而不是代工的關係，AI 是協助而不是替

¹¹⁴ 英文「Cyborg」是「Cybernetic Organism」的結合，又稱人機整合，是控制論有機體(Cybernetic Organism)的簡稱，表示任何混合有機體(Organic)與電子機器的機電一體化(Biomechatronic)生物。from：檢自 <https://zh.wikipedia.org/zh-tw/%E8%B3%BD%E5%8D%9A%E6%A0%BC>



代者，最終判斷權責仍必須歸屬於人。

AI 系統決策結果是否可以直接使用，或是予以調整，必須是「由人決定」。舉例來說，在參考 AI 系統的預測結果後，人類基於「演算技術的完備性、投入資料的不偏性、解釋動機的中立性」，以及領域知識等，權衡 AI 系統運用的經濟邏輯，以最小行政成本將 AI 工具整合到「明確預測和決策之間聯結(link)」的經濟框架中，以建構一個無偏誤決策(unbiased decision)。

綜上分析，依循公共行政觀點，人與 AI 之間的關係，可能是權力授予(delegation of power)的「代理關係(agency relations)」¹¹⁵，也可能是協作(collaboration)或合作(cooperation)等現象共存(coexistence)的「雙重關係(dual relations)」。

二、資訊科學、工程、控制等領域觀點

自動化是指完全或部分替換先前由人類執行的功能，這意味著自動化不是全部或沒有，而是可以在連續的水平上變化，從最低水平的全手動性能到最高水平的全自動化。Sheridan 為解釋計算機的自動化程度，計算機如何接管人類執行的任務，開發了一個 10 階層的量表來表示自動化程度，等級越高代表計算機對人類行為的自主性越強，例如自動化階層(Levels of Automation) 1 為：「計算機不提供任何協助，人類必須作所有的決策和行動。」，自動化階層 4 為：「計算機建議一個決策替代方案，但人類保留執行該替代方案或選擇另一個替代方案的權

¹¹⁵ 權力授予(delegation of power)：「有權力的人委託他人代理執行某些任務、功能，代理人依照委託人的目標付諸行動、完成所託。」。代理關係：「代理關係是指一個或一個以上的委託人(principal)，雇用並授權給另一個或一個以上之代理人(agent)，代其行使某些特定的行為，彼此間所存在的契約關係。」(Jensen & Meckling, 1976; 柯于璋, 2013)



力」，自動化階層 10 為：「計算機決定所有的工作，自主行動、忽略人的干預」(Busch & Henriksen, 2018; Parasuraman, Sheridan, & Wickens, 2000; Sheridan, 1992)。

綜上分析，人與 AI 之間的關係，依 AI 的自動化程度，或是計算機系統的自治權(the autonomy of computer systems)」(Floridi et al., 2018)觀點，可能是階級關係(hierarchy)、協作關係、機器自主(automony)等。

三、ANT 觀點

人和 AI 都有能動性(agency)，人與 AI 能彼此互相影響，Latour 指出最重要的是找出特定型態的能力是如何在不同種操控類型的「交轉過程」(cross over)中產生(雷祥麟, 2004)，如何發生替代(displacements)、轉換(transformations)、協商(negotiations)、調整(adjustments)。因為一連串不可預知的替代，所有的過程都可以被描述為一種轉譯(translation)，結果是某些實體(entities)控制其他實體的情況，形成新的權力關係(Callon, 1986)。

舉例來說，當公務人員行使職權時，在法律許可的範圍內得為自由之判斷，法律授權的主體是人，權力移轉途徑為「國家威權→公務員」。然而，當人與 AI 互動時，彼此的能動性(agency)交互影響，AI 的權力本質在網絡中被促動(enacted)出來(林文源, 2007)時，此時，人是「有限理性」，可能過度信任 AI 系統，或喪失自信心，放下領域知識與行政裁量權，導致 AI 系統原本用來支持/協助(support/assist)人的行政裁量權，可能成為用來代替(replace)人的行政裁量權，使得授權主體發生改變，權力移轉路徑可能變成「國家威權→AI」。

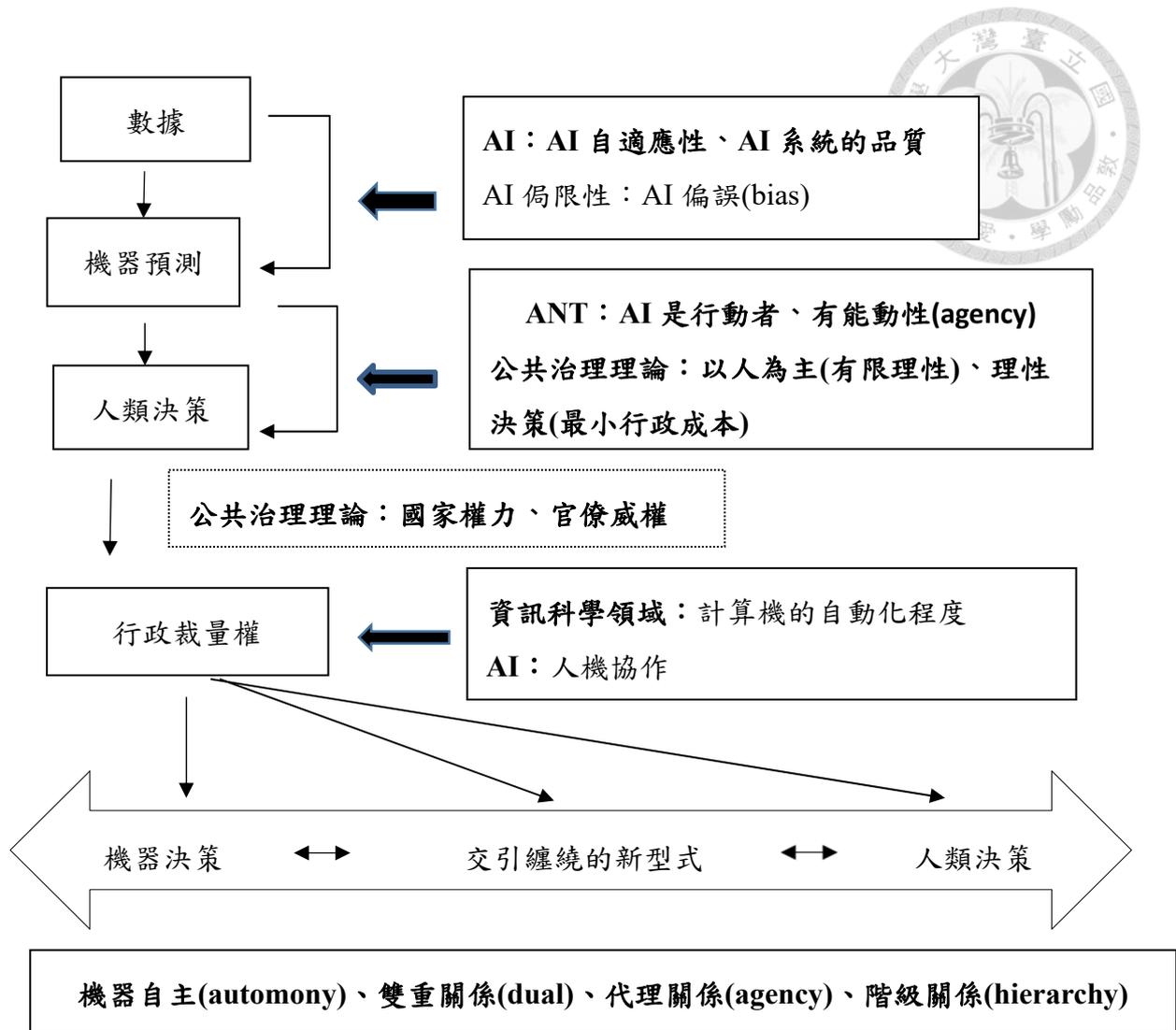
本研究結合 AI、公共治理理論、ANT、資訊科學等領域觀點，人與 AI 對彼此影響力作用的結果，或是權力結構再分配的結果，可能偏向人類決策、也可能偏向機器決策。人與 AI 之間建立起來的權力關係，在各種難以預料的因素、循

環交錯的過程，可能產生不同態樣的關係，包括：階級關係(hierarchy)、代理關係(agency relations)、機器自主(autonomy)、雙重關係(dual)等。而影響人與 AI 互動關係的主要影響因素，包括：

1. 國家威權：政府部門具有行使統治權作用之行為。
2. AI 能動性：人與 AI 互動後改變現狀。
3. AI 系統的品質：AI 系統功能的增強與人類依賴之間有著直接的關係。

後續將以實證研究案例進一步驗證。

另為利跨領域理解，補充說明如下：圖 2-2 所指決策與資訊科學在 AI 領域決策指「決策變數的選擇」不同。在 AI 領域，做完決策變數的選擇就結束，跟制定公共政策的決策不一樣且距離有點遠。公共政策通常是將 AI 系統視為政策分析的工具，將 AI 系統的預測結果轉換成政策資訊，並進行政策分析，決策結果作為政府選擇作為或不作為、或是政府對社會價值從事權威性分配的行為。



- 機器決策：AI 自主決策(autonomous decision)的行使行政裁量權
- 人類決策：AI 系統「賦能(empower)」人類，人類行使裁量權

圖 2-2 AI 與公部門行動者的關係

資料來源：本研究修改調整自(Busch & Henriksen, 2018; Callon, 1986; Kleinberg et al., 2018; Latour, 2005; Parasuraman et al., 2000; Sheridan, 1992)等文獻



第三章、AI 與公共治理的研究設計

第二章第四節的分析，大部分的 AI 治理概念框架或分析框架是由技術研究社群所提出，也發布在科學技術相關期刊、會議報告，是以技術角度的 AI 倫理、敏捷式治理、價值敏感設計等，與公共治理理論幾乎無相關。另一方面，社會科學期刊關於 AI 治理不僅數量非常有限，且研究取向側重於「人」或「以人為本」原則(UK, 2020)，不脫社會學或社會解釋(social explanation)的範疇，對技術主要的觀點仍著重於參與技術發展的各個社會群體利益協調的社會建構過程，或都由社會或利益團體所充實，技術不過是社會(如衝突、妥協、競爭、合作等)的再現，技術物本身並未被賦予能動性地位(張國暉, 2017)，並且停留在技術變革的描述和解釋中(Choi, Yeo, & Won, 2018)。為結束一般社會科學與技術研究領域之間的隔離，已經有研究嘗試將公共治理和數據科學相結合的跨學科方法，再納入技術以強化理論基礎的研究方法(Höchtel et al., 2016)。

行政學是個「借用的領域(borrowing field)」(江明修, 1997)，常見的做法是源自於心理學、經濟學、社會學等領域；再者，在跨學科(interdisciplinary)和應用研究(applied research)中常見的做法是從其他研究領域借用理論(Berkovich, 2020; 陳金貴, 2013)。在 AI 時代，對公共行政而言，尤其當我們一時沒有更好的解決方案時，借用不同的理論進行研究，成為本研究的思考重點。因此，本章將從技術應作為變量開始，說明本研究為何跨學科尋找科學科技與社會(Science, technology and society, STS)領域、為何選擇 ANT 作為研究方法，本研究設計方式如何能夠達到本研究目的，研究結果具有說服力。

第一節 方法論—ANT



史諾(C.P. Snow)主張有兩種文化區隔了現代社會：科學與人文。科學關注的是沒有人的物質世界，例如技術決定論者卡爾·馬克思(Karl Marx)視技術為社會和經濟關係的指標，他的政治計畫是要駕馭技術，令其真正為人類所用。另一方面，人文則聚焦沒有物質性的人類世界，例如社會決定論徹底忽略物質技術物，以及技術物的政治性(politics of artifacts)，一切都聚焦在社會，技術僅是社會建構，沒有自己的能動性，這有助於解釋社會理論為何不願意處理物體(王志弘、高郁婷譯, 2023)。因此，對於應如何探索「科技與社會之間如何互動」及更歸根究底地探索科技的本質為何等問題，長久以來不易脫出科技決定及社會建構分立的兩個途徑，進而逐漸變成關鍵學術難題。

行動者網絡理論(Actor-Network Theory, ANT) 嘗試利用科學知識社會學(sociology of scientific knowledge, SSK) 將「人」與「物」給予對稱化的社會建構論點，將技術與社會兩端持續交織起來，突破以人為中心的敘事型態，凸顯技術在社會建構中的角色；強調非人能動性，賦予技術「存有論的尊嚴」，避免以技術或社會任何一端化約掉另一端的狀況，任何成功的建構都同時仰賴人類和非人類客體(王志弘、高郁婷譯, 2023)來消解這個分立難題：「科學和技術是社會建構的，技術也構建了社會」；不是「社會塑造 socially shaped」的技術，而是技術提供社會聯結增強性和持久性(Latour, 2005)；並如 John Law 所說不是彼此「用以解釋的資源」(Latour & Woolgar, 2013; 林文源, 2014; 張國暉, 2017; 楊弘任, 2017)，而透過「人與非人的異質聯結」(the heterogeneous association of human and nonhuman)才讓「某些技術物」與「某種社會」同時萌生出來(楊弘任, 2012)，在此，非人類通常扮演領導角色，這樣的觀點可以對政策科學中缺乏對技術關注的現象提出可行的解釋路徑。



壹、行動者網絡理論

ANT 是 1980 年代由主要從事科學和技術社會學研究的社會學者卡隆(Michel Callon)、拉圖(Bruno Latour)及羅(John Law)等人，從經驗研究(empirical studies)中開發出如何進行科學分析的一種方法論，認為科學技術產物強大的推進力量使「社會(society)」發生非常激進的轉型(transformation)，因此將物質(materiality)賦予與人類同地位，即廣義對稱性原則(generalized symmetry)，並設計了一個異構網絡，其中人類和自然行動者的問題是相互關聯，他們的利益也相互交叉；技術賦能人類，但同時也限制和制約人類(Bueger & Stockbruegger, 2017)。Latour 爭論科技物應被視為具有權利(rights)，就像是可組成國會(parliament of things) 一般，會產生社會行為(規約、限制、要求等)的作用，與「人」形成一連串的網絡互動，以此為 ANT 基礎構想。

ANT 與 SCOT 雖然同源於社會建構論，在將技術放回理論議程上貢獻卓著，ANT 與 SCOT 最大的差異在於前者並不認為從「社會性(the social)」途徑解釋科技發展具有特權(privilege)，Latour 主張人(humans) 與非人(nonhumans) 的對稱，同樣賦予能動性(人是 actors，非人是 actons)(張國暉, 2017)，有影響力並且可以產生影響的一切事務，均可為對方給予設定(norms) 或附加義務(obligations)，進而對彼此發生作用、造成變化(Latour, 2005)。再者，由於 AI 的演算法具有學習能力，可能導致原本預設的權衡(tradeoffs)，超出程式設計的預期，甚至可能與設計者的初衷完全不同，這種影響通常被描述為「始料不及後果定律 (the law of unintended consequences)」(Bannister & Connolly, 2014)，使演算法產生經濟學家所說的負面外部性(negative externalities)——第三方未參與決策的成本 (Rahwan, 2018)，加以公部門採用 AI 對整個社會影響的領域更廣泛，在社會中需要一種代表所有利益相關者的規範方法(Rahwan, 2018)，而 ANT 闡明唯有解決每一個行動者的問題，皆能達成渠等利益賦予、徵召、動員，才能提高導入 AI 成功的機



會，使得 ANT 具有較佳的解釋潛力。

Latour(2005) 認為社會學作為「共同生活的科學(science of living together)」，應該能夠履行以下三個職責：應該能夠部署可能的關聯以及這些關聯的所有爭議，應該能夠提出解決爭議的方案並維持這些解決方案，通過研究物體來定義集體組成的正確程序；簡言之，就是依序 3 種職責：「部署(deployment)」、「穩定(stabilization)」和「組合(composition)」；爭議的部署、這些爭議的穩定以及尋求政治影響力，因此適合研究發展中科技與社會之間如何互動。在具體的實踐上，ANT 如何可以緊密扣合公共治理理論的精隨進一步闡述如下：

一、強綱領的廣義對稱性原則(the strong program principle of symmetry)

許多社會科學的「以人為本」源於自然與社會之間的二元論，並清楚劃分「自然」與「社會」、「人類」(humans) 與「非人類」(nonhumans)，不斷將人類定位為唯一重要的行為者，以人為中心的觀點(human-centred perspective)，無法充分考慮構成我們世界的各種非人類(Murdoch, 1997; 洪榮志、蔡志豪, 2011)的現象，例如一種由多重的人和機器組成的社會賽博格(Sociocyborg)(王志弘、高郁婷譯, 2023)。

ANT 理論的視角放棄社會和技術「維度(dimensions)」之間的人為劃分，主張任何可以改變事務狀態的東西都是行動者(an actor)；或者，如果它還沒有輪廓(figuration)，那就是行動體(an actant)¹¹⁶。因此試圖重新定義行動者，以某種方式影響或擾亂技術社會系統活動的任何實體，包括人類或非人類(Crawford, 2020; Latour, 2005)。Callon 對聖布里厄灣扇貝(the scallops of St Brieuc Bay)的研究顯示，扇貝是有能力擾亂、抵制和破壞人類作為的「行動者(actors)」，它們不能輕易被控制和馴化(domesticated)(Bueger & Stockbruegger, 2017)，行動者不必是社會

¹¹⁶ any thing that does modify a state of affairs by making a difference is an actor—or, if it has no figuration yet, an actant.

的，但「可以是任何東西，只要它被授予作為行動的來源。」¹¹⁷(Latour, 1996)。這種觀點一旦建立，人類無法孤立於非人類世界之外，必須對物質性的問題寄予嚴肅關注(王志弘、高郁婷譯, 2023)，物質就會在社會互動中真正採取行動，在其中扮演平等和對稱的角色，例如在 Web 研究非人類與日俱增的相關性時，研究人員可以將 AI 視為參與社會型塑的行動者(Waldherr et al., 2019)，使 AI 技術被概念化為一個主權行動者(a sovereign actor)，甚至是強大的執行者。

簡言之，ANT 主張強綱領的廣義對稱性原則，擺脫以人為主體的思考模式，將 AI 納入為研究對象，能具體回應「AI 官僚」的現象。

二、物體也有能動性(Objects too Have Agency)

ANT 主張任何事物都具有潛在的能動性(agency)，包括自然(nature)和物質對象(material objects) (Bueger & Stockbruegger, 2017)，特定網絡中的行動體(An actant)通過網絡並與其他行動體(actants)關聯來實現能動性。換句話說，網絡決定(或影響)行動體(actant)的形式並將其轉變為具體行動者(actor) (Bueger & Stockbruegger, 2017)。因此，AI 能動性(agency)對 AI 與人類行動體(actant)互動的詮釋方法，可以展現人類成為科技的一個組成部分，科技成為人類的一個組成部分，沒有一個(例如科技)支配另一個(例如人類)，理想情況下人類和科技都不會主導對方(Pitt et al., 2023)，而是透過網絡中的關聯，行動者的身份、互動的可能性和調動的幅度得到協商和界定(Callon, 1984; Choi et al., 2018)。

簡言之，ANT 主張物體也有能動性(Latour, 2005)，在公共治理理論探討人與 AI 的互動時，藉由交引纏繞新形式(new shapes)的概念：「互動的可能性和調動的幅度得到協商和界定」，觀察人與機器(human-to-machine, H2M)互動(Wirtz et al.,

¹¹⁷ 原文：Actors do not have to be social but “can literally be anything provided it is granted to be the source of an action “.

2019)時權力控制幅度(the margins of manoeuvre)的變化，理解 AI 自動化決策是如何改變基層官僚的自主行政裁量權，進而產生權力移轉的現象。



三、交引纏繞 (an entanglement of interactions)的社會世界

ANT 假設社會是一種瞬時關聯的典型(a type of momentary association)，物體成為行動過程的參與者，並參與重新關聯(re-association)和重新組合(reassembling)的運動，所有這些異質元素可能會在某些給定的事態中重新組合¹¹⁸，其特徵在於它聚集在一起形成新形式(new shapes)，有可能產生一個被理解為交引纏繞的社會世界¹¹⁹(Latour, 2005)。這些網絡中的實體(無論是人類還是非人類)，彼此互相影響，彼此穿透混合而形成一個共同網絡的集體(collective)，通過連接它們的數量、廣泛性和穩定性獲得權力，而不是通過其他任何方式(Couldry, 2008)，由較少的「交引纏繞」發展至大規模且非常深入的「交引纏繞」(雷祥麟, 2004)。因此，ANT 的基本思想是整個世界都是由網絡組成的，每個組織、過程或實踐都可以用網絡術語來描述，並為避免線性和單邊效應的概念，強調社會和技術之間是動態(dynamic)、循環(circular)和交互(reciprocal)、共同進化過程(Bueger & Stockbruegger, 2017; Latour, 2005; Waldherr et al., 2019)。

因此，ANT 對社會世界交引纏繞(an entanglement of interactions)的假設，有利詮釋 AI 與人類互動式機器學習(interactive machine learning)的過程與現象，例如 AI 在真實環境中與人類進行互動，經由觀察和分析人類的行為和反應，理解人類的思維模式和行為模式，以及 AI 的自我評估，學習人類行為數據和社會互

¹¹⁸ 原文：those heterogeneous elements might be assembled anew in some given state of affairs.

¹¹⁹ 原文：it's possible to generate a social world understood as an entanglement of interactions.



動數據。而 AI 系統整合使用者(人類行動者)的反饋，包括人類使用者的評價、專家的評論等，能幫助 AI 機器更快或更有效地學習；AI 模型再進一步據以調整學習策略和理解模型等，逐步提高對人類的理解。而在人類與 AI 的互動過程中，AI 同時也逐步協助人類提高認知的邊界(蘇經天, 2023)。

綜上分析，ANT 提供了一種理解非人類(AI)和人類行動者之間關係的方法，非人類(AI)和人類都是能動者(agents)，都有能動性(agency)，都能交互影響，可以從概念上解釋社會技術世界(socio-technological world)的複雜性、人和 AI 關係深度交織的本質。非人類中特別是 AI，有別於其他文獻技術物(artifacts)或產物(artifacts)是沉默(silent)的行動者，AI 在社會關係中是具有主動(active)性的積極行動者，更有機會成為行動者網絡中的關鍵、有力的行動者，能進一步結合(amalgamation)成一個社會網絡(social network)，透過感知、回應、互相學習、溝通、情感交流而改變彼此的認知、定位、功能、決定、作為和行為意義歸屬，從而使人類不再被視為權威控制者(authoritative controller)，並有能力在一個情境行動中(situated action)起更積極的作用，例如在一個資訊系統中負擔大部分的認知功能，並與人類之間形成某種形式的協同作用(a form of co-agency)，避免只考慮人的能動性(human agency)，產生對結果無法解釋的困境。再者，已有文獻利用 ANT 研究人與非人類行動者之間的關係，或是建議 ANT 可以當成研究人與 IT/IS 互動關係的研究方法(Boychenko, 2019; Boychenko, Teixeira, Kuzovkova, & Boychenko, 2020; Ghazinoory & Hajishirzi, 2012; Johnston, 2001)。

因此，使用 ANT 的分析途徑研究 AI 與人的權力關係，可以從方法論上解決公共治理理論長久以來中對技術的忽視，將技術物理論化，消除在公共治理理論與技術分立的現況，促使本研究使用 ANT 及其分析方法作為案例研究的分析方法。



貳、ANT 的分析途徑(analytic approach)

ANT 是研究權力的分析框架(analytical framework)，權力論點主要依據是轉譯，包括不同階段轉譯的四個「時刻」(four 'moments' of translation)，四個可以重疊的時刻組成，轉譯的過程就是從物質(material)替代(displacement)成為資訊(information)，再轉成行動(action)的形式(form) (Callon, 1984; Choi et al., 2018)。轉譯包括在網絡內一連串的協商，在這個交互影響的動態過程中，透過轉譯，每一主體或行動者在網絡中的主體性(identity)、行動者的能力、特性、角色、功能、地位和利益都重新界定、分配、安排和賦予(林文源, 2007; 邱大昕, 2011)，權力關係亦重新築構。因此，「人」或「非人」的能力或特性都不具先驗的本質，行動者的本質(nature)不但可變，且會受到其他行動者的影響，因此可以在「共同的(communal)、權變的(contingent)、情境相關性的目標與利益(contextual goals and interests)」下，進行交引纏繞 (Entanglement)(Latour, 1987)，並連結成網絡 (周信輝、蔡志豪, 2013)。

轉譯的過程也會發生替代(displacements)¹²⁰：目標、利益、策略(device)、人類、非人類(AI)的替代(Callon, 1986)，因為一連串不可預知的替代，行動者隨時處於極端不確定的狀態(radical indeterminacy)，這些替代使行動者遠離強加於他們的「強制通行點(Obligatory Passage Point, OPP)」¹²¹。新的替代取代了以前的替代，轉譯繼續，但平衡已經修改，某些實體(entities)控制其他實體，變成「有權

¹²⁰ 替代(displacements)指過程中這些行動者存在狀態發生變化(林文源, 2014)。

¹²¹ 強制通行點係指網絡中的主要行動者經由明確的路徑建構必要的問題，這一問題足以促使其他行動者的利益與主要行動者的目標一致(Callon, 1986)。在 Callon (1986) 海扇貝的例子裡，其強制通行點為網絡中的主體必須知道下列問題：海扇貝如何附著？而且各個主體認定他們的結盟在此一問題上都可以獲利(李承嘉、廖本全、戴政新, 2010)。



力者的解構(deconstructing of the powerful)」(Murdoch, 2000; 李承嘉、廖本全、戴政新, 2010)。有權力者是否會發生解構的關鍵在於轉譯能力的大小，若一個行動者網絡的強制通行點(Callon, 1986)能夠引起越多行動者的注意，且行動者在行動者網絡中被賦予的利益越確定，就能徵召更多的行動者(李承嘉、廖本全、戴政新, 2010)，能夠徵召及動員越多異質的主體，此一網絡就越能對網絡進行成功地轉譯。

一、轉譯的過程

(一)、問題化(problematisation)

研究人員試圖定義網絡中行動者的性質和問題，建議行動者就強制通行點與其他行動者進行談判，這些問題將得到解決。強制通行點為實現參與者的目標而必須提出的一組共同的成功條件，在這個階段，一個認識到問題的行動者出現，指出問題並提出解決問題的新想法、解決問題的網絡、確定參與網絡的主要行動者，並定義其他參與者的角色和他們「想要(want)」的東西。

(二)、利益賦予(interessement)

研究人員試圖為其他行動者定義角色的過程，行動者的屬性和身份在利益賦予過程中被合併和/或重新定義，利益賦予建立在對尚未徵召的行動者是什麼、想要什麼，以及這些行動者與哪些實體相關聯的某些解釋。因此，利益賦予的策略創造了一種有利的權力平衡，這是行動者參與新網絡必備的強大動力，並使其他行動者打破他們現有的網絡。

(三)、徵召(enrolment)

研究人員試圖定義和關聯其他行動者角色的策略，徵召就是多邊談判、實力測試，角色的定義和分配是多邊談判的結果。在此階段，行動者的身份被確

定和測試，行動者積極參與網絡，認清自己的角色並努力擴大新建立的網絡。



(四)、動員(mobilization)

研究人員創建代表網絡行動者的代言人(spokesman)，確保各種相關集體的假設代言人能夠適當地代表這些團體，而不會被背叛所使用的方法。在此階段，行動者的漸進替代過程依次進行，從而產生網絡的聯合代表。當我們進入這個系統穩定的階段時，我們看到「黑箱(black box)」已經完成，這意味著網絡已經穩定；如果動員成功，在四個時刻結束時就已經建立了一個有約束力的關係網絡。

轉譯¹²²是一個過程，沒有時間順序，可以重疊進行，一旦轉譯成功完成，就建立一個有權力的網絡。但是這種共識及聯盟隨時可能受到質疑，主要是代言人的代表性受到質疑、討論、談判、拒絕等，因此轉譯也可能失敗(Choi et al., 2018)。換句話說，當特定創新發明者嘗試以「正向方案」(program)進行拉長化的

¹²² 「轉譯」概念不易理解，另一種解釋方式：問題化(problematization)(知道你的行動體，針對它們的問題來提出解答)、利害與共(interressement)(令新徵募者脫離其先前網絡，將它們結盟到你的網絡)、納編(enrolment)(通過壓力、呼籲或同意，以及分配角色給其他人，來鞏固新網絡的認同)，以及動員 (mobilization) (確保發言人代表他們的集體，並避免背叛)(王志弘、高郁婷譯, 2023)。拉圖(Latour)針對轉譯技巧提出了具細緻的說明：關注其他人的利益、提供阻礙者其他選頭、提供更快的成功之路、轉換利益和目標、讓你自己變得不可或缺。同樣的，他列出了各式各樣的戰術，包括創造出新的問題、目標和群體，以及暗中繞路(Latour, 1987; 王志弘、高郁婷譯, 2023)。

網絡結盟時，其他人或非人的行動者可能也在進行「反向方案」(anti-program)的動員，所謂一項成功或失敗的技術創新發明過程，也就是「正向方案」與「反向方案」各自網絡結盟角力之後的結果(楊弘任, 2012)。



二、分析方法

ANT 發展概念是從經驗研究(empirical studies)中構建出來的，並明確建議「技術／社會現象」的最佳研究方法，即是「跟隨行動者」(Latour, 2005)，ANT 家置身田野，在社會聯結中緊緊跟隨科學家或工程技術者(楊弘任, 2012)，專注於當前的經驗故事，不會在主體和客體之間做出明確區別，通過詳細追蹤案例的價值和利益，特別是衝突和爭議的部分，可以揭示行動者網絡中隱藏的社會秩序、審查權力和等級結構、解決行動者間價值競爭和意義的問題，並分別識別和分析人和 AI 等行動者間複雜的能動性(agency) (Waldherr et al., 2019; 王志弘、高郁婷譯, 2023)。

因此，透過 ANT 實際案例的敘事方法，可以理解政治、權力、技術和轉型(transformation)發展的複雜情況(Bueger & Stockbruegger, 2017)，案例研究敘事重點如下：

1. 行動者：行動者身份(identity of actors)。
2. 能動性：互動可能性(the possibility of interaction)。
3. 強制通行點(obligatory passage point, OPP)：實現參與者的目標而必須提出的一組共同的成功條件。
4. 交引纏繞：控制幅度(the margins of manoeuvre)的協商和劃定過程。
5. 代言人：代言人是否具有代表性。
6. 轉譯成功及失敗的原因：分析技術成功和失敗的原因。



表 3-1 ANT 的分析框架與敘事重點

轉譯 過程 行動	問題化	利益賦予	徵召	動員
行動者	1. 一個認識到問題的行動者出現 2. 定義行動者的問題 3. 定義行動者身份和他們「想要(want)」的東西 4. 找出強制通行點	1. 定義角色的過程：尚未徵召的行動者是什麼、想要什麼 2. 行動者的屬性和身份在利益賦予過程中被合併和/或重新定義	行動者的身份被確定和測試	代言人：創建代表此類網絡的行動者，產生網絡的聯合代表
能動性 (互動可能性)	確定參與網絡的主要行動者，並定義其他參與者的角色	這些行動者與哪些實體相關聯	多邊談判、實力測試	假定代言人能夠適當地代表這些團體，而不會被背叛所使用的方法
交引纏繞 (控制幅度的協商和劃定過程)	行動者就強制通行點與其他行動者進行談判	創造了一種有利的權力平衡	角色的定義和分配是多邊談判的結果	找出轉譯成功及失敗的原因

資料來源：本研究整理自(Callon, 1984; Choi et al., 2018; Latour, 2005)



參、ANT 的應用範圍

由於 ANT 內容龐雜，不同的研究者可能以不同的方式使用 ANT，**本研究是利用 ANT 研究 AI 與人、AI 與公共治理理論的關係；而非利用 ANT 設計新的分析方法**。因此，本研究使用 ANT 的範圍如下：

一、案例研究分析方法的理論基礎：

ANT 分析框架(analytical framework)可以描述行動者如何透過行動，達成各自的目的進而相互權謀的策略；從凝聚共同問題、因而引發興趣追求目標，到無法獨力解決問題而設法吸引其他行動者的呼應而相互連結、搭架，從而獲取特定定位與利益，並形成可以解決問題的網絡關係(Callon, 1984; 周信輝、蔡志豪, 2013); ANT 也可以描述技術失敗的原因，是各種行動者和利益相交但無法保持一致(Bueger & Stockbruegger, 2017)等。因此，ANT 對描述而非解釋的關注(Latour, 2005; Waldherr et al., 2019)，適合研究發展中科技與社會之間如何互動，以及科學和技術在構建權力關係中所起作用的分析框架。

二、提供 AI 可以做為分析單位的理論基礎：

ANT 主張廣義對稱性原則(Latour, 2005)，顛覆目前公共治理理論以人為中心的治理典範，將 AI 納入為研究對象，擴充了分析單位，使本研究得以在不改變公共治理理論以人為本的脈絡解釋「AI 官僚」的理論基礎。

三、提供人與 AI 之間權力關係的理論基礎：

ANT 主張觀點：

技術物可以「表達」權力關係('express' power relations)、「象徵」社會等級('symbolize' social hierarchies)、「強化」社會不平等('reinforce' social



inequalities)、「輸送」社會權力(‘transport’ social power)、「客體化」不平等(‘objectify’ inequality)和「物化」性別關係(‘reify’ gender relations)；網絡中行動者的互動不是相同主題(isotopic)、同步(synchronic)、綜覽式的(synoptic)、同質的(homogeneous)、等壓的(isobaric)(Latour, 2005)。

提供 AI 對人產生影響，為公共治理理論無法解釋人與 AI 共存的現象，以及 AI 能動性(agency)影響人的行政裁量權，產生偏向 AI 決策的理論基礎。



第二節 研究架構

本研究採用公共治理理論、ANT、AI 為理論基礎。首先，從公共治理理論現有機制出發，以人為本、有限理性的觀點。其次，使用 ANT：廣義對稱性、能動性(agency)、交引纏繞(entangle)的觀點。第三，使用 AI 系統的品質與偏誤(bias)等觀點。

壹、研究假設

本研究假設及所使用名詞定義如后，後續將透過實際案例研究進一步驗證。

1. AI 技術物是分析單位：AI 是行動者，有自適應性和能動性。
2. 公共治理領域人機協作的理論基礎：人和 AI 能動性(agency)的交互作用。

若以簡單化觀點分別討論人和 AI，會忽略人和 AI 之間的重要互動關係，無法全面了解所發生的事情(Tatnall, 2003)。因此，當我們想要使用 AI 擴增人類認知、生理能力，並描述公部門使用 AI 時的官僚行為，不能單獨將 AI 用一種工具概念來描述整體官僚現象，必須考慮 AI 是行動者，具有自適應性(adaptive)與能動性(agency)，以及人的能動性、人與 AI 之間的互動關係(Human-AI Relations)。本研究提出公共治理領域人機協作的理論：人和 AI 能動性(agency)的交互作用；人和 AI 的能動性(agency)交互作用的結果：交引纏繞的新形式，有多種可能的關聯與模式，多樣化人機協作的模式。

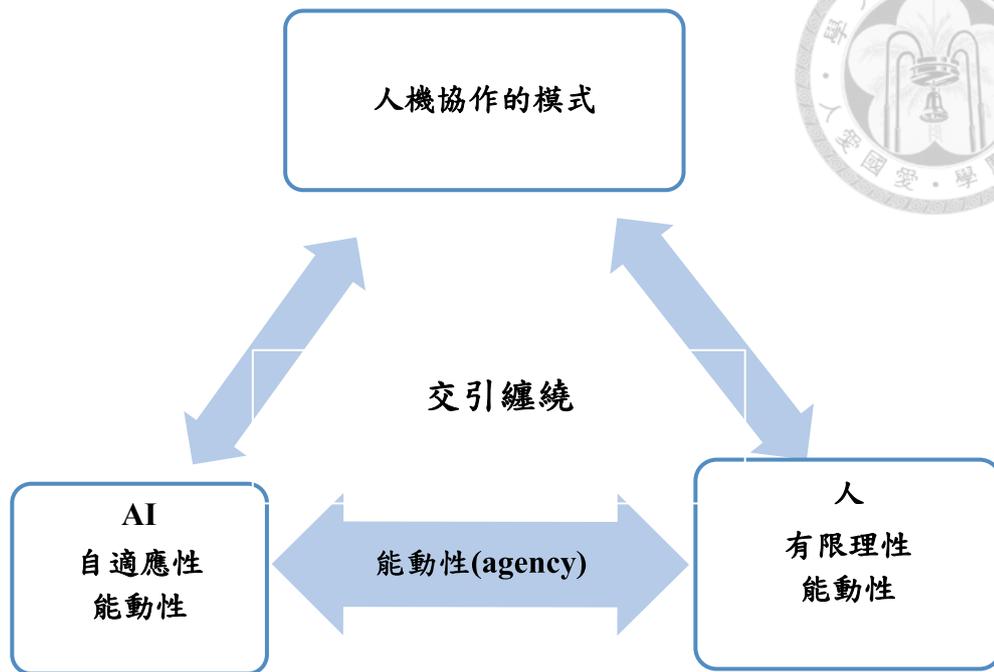


圖 3-1 人機協作的理論

資料來源：本研究整理

3. AI 的公共治理(public governance of AI)：公部門使用 AI 的理論，或公共治理導入 AI 的理論。

公部門使用 AI 時，不能僅將 AI 視為一種工具，必須考慮 AI 自適應性與能動性的作用：

- (1).單位(the units)：AI 是關鍵行動者。
- (2).單位被賦予的品質(the qualities these units are endowed with)：AI 系統品質。
- (3).關係的形式(the form the relations)：人機協作。
- (4).可能的關聯 (association)：安排最合適(most suitable)的人機協作，使具有不同的理性(rationalities)的人和 AI 可以在不同的時間軸上協力合作(synergies)，在工作不斷地交互轉移(Crosset & Dupont, 2022)、循環交錯的過程中交引纏

繞，人不能片面影響 AI，AI 也不能片面影響人的行為，但在屬於人的關係和屬於 AI 的關係之間卻有著直接的聯繫。



AI 的公共治理研究架構圖及內容，如圖 3-2 及表 3-2：

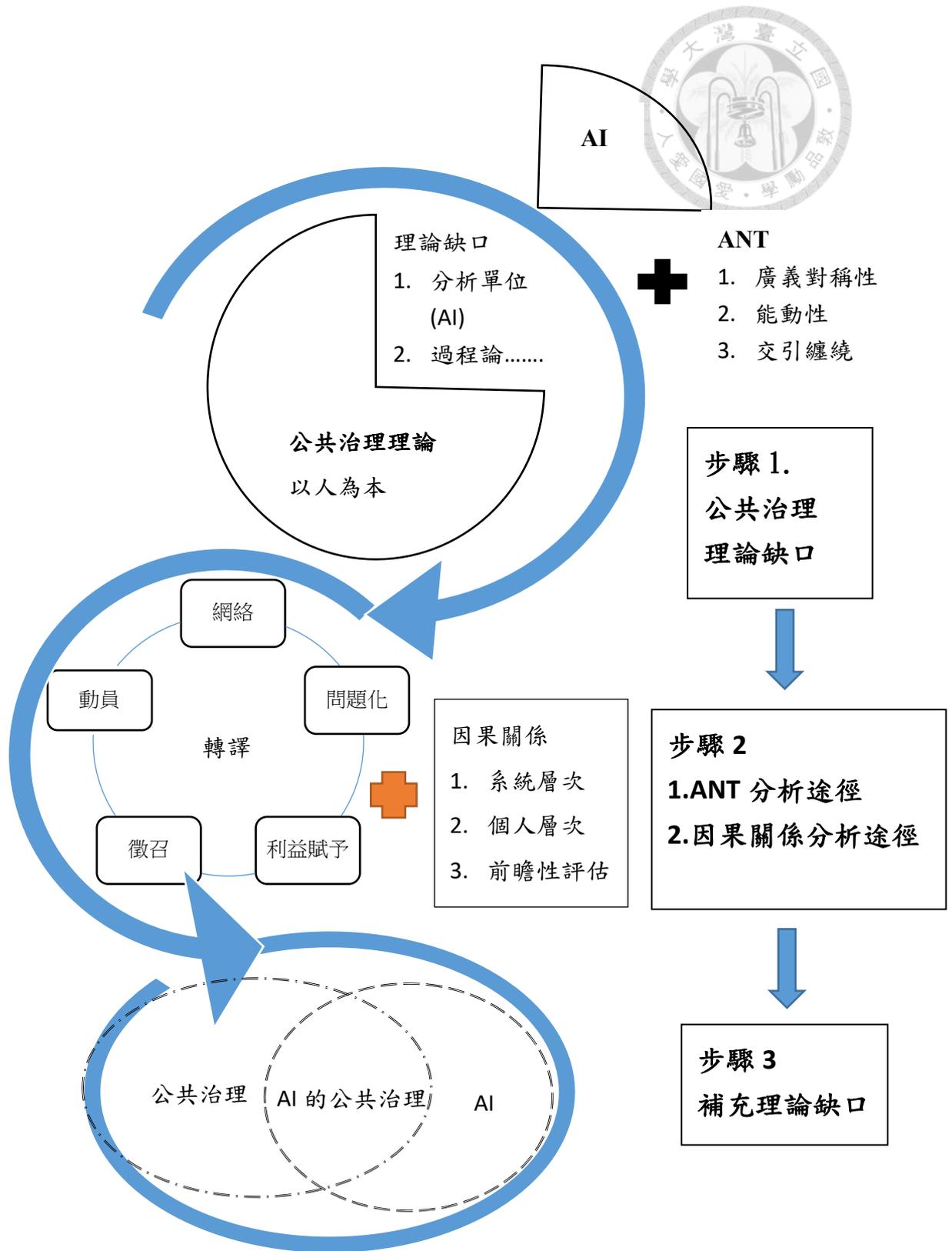


圖 3-2 AI 的公共治理研究架構圖

資料來源：本研究整理

表 3-2 AI 的公共治理研究內容

關鍵要素	公共治理理論	AI	公共治理理論缺口	ANT 分析途徑	AI 的公共治理理論
分析單位或模型	人 (有限理性)	智慧代理人 (有限理性)	1. AI 官僚 2. AI 系統的品質與偏誤	1. 廣義的對稱性 2. 網絡中的實體(人類、非人類等)	人+AI
過程論	可控制性、 次序性	連續感知並 採取行動	動態、循環 和交互	1. 瞬時關聯 2. 重新關聯 和重新組合的運動	持續規劃 與評估
執行	人類官僚的 行政裁量權	AI 自動決策的作用	1. 裁量權移轉 2. 官僚威權與 AI 的關連	AI 能動性	1. AI 官僚 2. 人機協作
分析方法	因果關係	相關性	1. 人類認知、生理能力的侷限性 2. 經驗主義客觀現實的困難	1. 對描述而非解釋的關注 2. 成功與失敗的原因	ANT 分析途徑

資料來源：本研究參考文獻(Callon, 1984; Choi et al., 2018; Latour, 2005)後整理



貳、案例選擇架構

AI 情境很重要，為理解不同情境，權衡 AI 機會與風險，例如人類不同監督程度(人類決策或 AI 自主決策)，國家權力以不同方式影響公民(不同政策工具)等多種可能性，再參考過往研究理論文獻理論、框架不足之處：

- 1.缺少 AI 自適應性(adaptive)和能動性(agency)的觀點。
- 2.缺少人類監督、AI 官僚、人機協作的觀點。
- 3.缺少國家權力涉入程度、官僚權力的觀點。

本研究提出「**entanglement × power**」的案例選擇架構說明如下：

第一維度(X軸)：人類監督程度

1.公共治理理論觀點：

在官僚政治模型中，權力為「對政府決策和行動的有效影響」；官僚權力(bureaucratic power)是官僚使用可用手段將決策過程引導到期望方向以實現官僚首選結果的能力(Hegele, 2018)。

2.AI 觀點：

AI 系統自主決策程度，向右偏向人類決策，向左偏向 AI 決策。探索文獻對人類監督程度的爭議，例如人類監督程度太高，降低公部門的科技接受度；人類監督程度太低，提高 AI 風險程度。

3.ANT 觀點：

AI 與人關聯的交引纏繞(entanglement)的新形式，其中 AI 自適應性與 AI 能動性是否對其他能動者的行動(agent's action)產生影響？使原本人類官僚的行政裁量權發生權力移轉現象。



第二維度(Y軸)：國家機關權力涉入程度

1. 公共治理理論觀點：

本文所稱國家權力係指國家統治權或政府權力。國家機關權力涉入程度使用「政策工具」的強制程度和更精細的等級分類¹²³，包括自願性工具、混合型工具、強制性，權威型、誘因型、能力建構型、學習型、象徵和勸告型等。向上偏管制性、向下偏象徵和勸告型。探索文獻對 AI 是否可使用於對人民權利和義務產生影響領域的爭議。

2. ANT 與 AI 觀點：

當 AI 與國家權力關聯¹²⁴，可能擴大權力差距，也可能縮小權力差距。探索公部門應如何妥善管理人與 AI 互動關係，避免數位裁量權和數位官僚主義擴張。

¹²³ Vedung, E. (1998)認為政策工具是政府運用權力試圖確保社會變革的一種技術 (Bali, Howlett, Lewis, & Ramesh, 2021)。Hood(1983)指出政策工具(policy instrument)又稱為治理工具(governing instrument)或政府工具(tools of government)，是政府為達成政策目標、產生政策效果的手段(Hood, 1983)，並依據「國家機關權力涉入的程度」劃分為三大類：自願性工具、混合型工具、強制性工具。Schneider 和 Ingram (1990)認為政策工具本質上是政治現象和政治行為，「試圖讓人們做他們本來不會做的事情，或者使他們能夠做他們可能不會做的事情」，因此政策工具用於克服政策相關行動的障礙，包括五大類政策工具：權威型 (authority)、誘因型(incentives)、能力建構型(capacity-building)、學習型 (learning)、象徵和勸告型(symbolic and hortatory)等(Schneider & Ingram, 1990)。

¹²⁴ Latour 在關聯的權力(The powers of association)中使用「計算關係的貨幣(the currency with which the relations are calculated)」來呈現(Latour, 1984)。

上述案例選擇架構的 4 個象限，可以涵蓋不同國家權力或政府職能，人類決策和機器決策的各自範圍，使不同研究個案具有可比性(comparability)，有利分析文獻所述爭議權衡，例如：

1. 適合 AI 的政府職能：例如 Mehr(2017)提出避免將 AI 決策用在對人民權利和義務產生重大影響領域，Bullock(2019)提出複雜性和不確定性較低的任務是 AI 處理，而複雜性和不確定性較高的任務則應由人類自行決定，Agarwal(2018)主張核心公共機構不應再使用 AI 等。
2. 監督強度：例如 Rahwan(2018) 提出社會監督等。
3. 人機協作模式：例如人類監督的程度，或是演算法如何在決策的複雜性和自動化速度之間做妥協(Johnson N, 2013; Karppi & Crawford, 2016)。

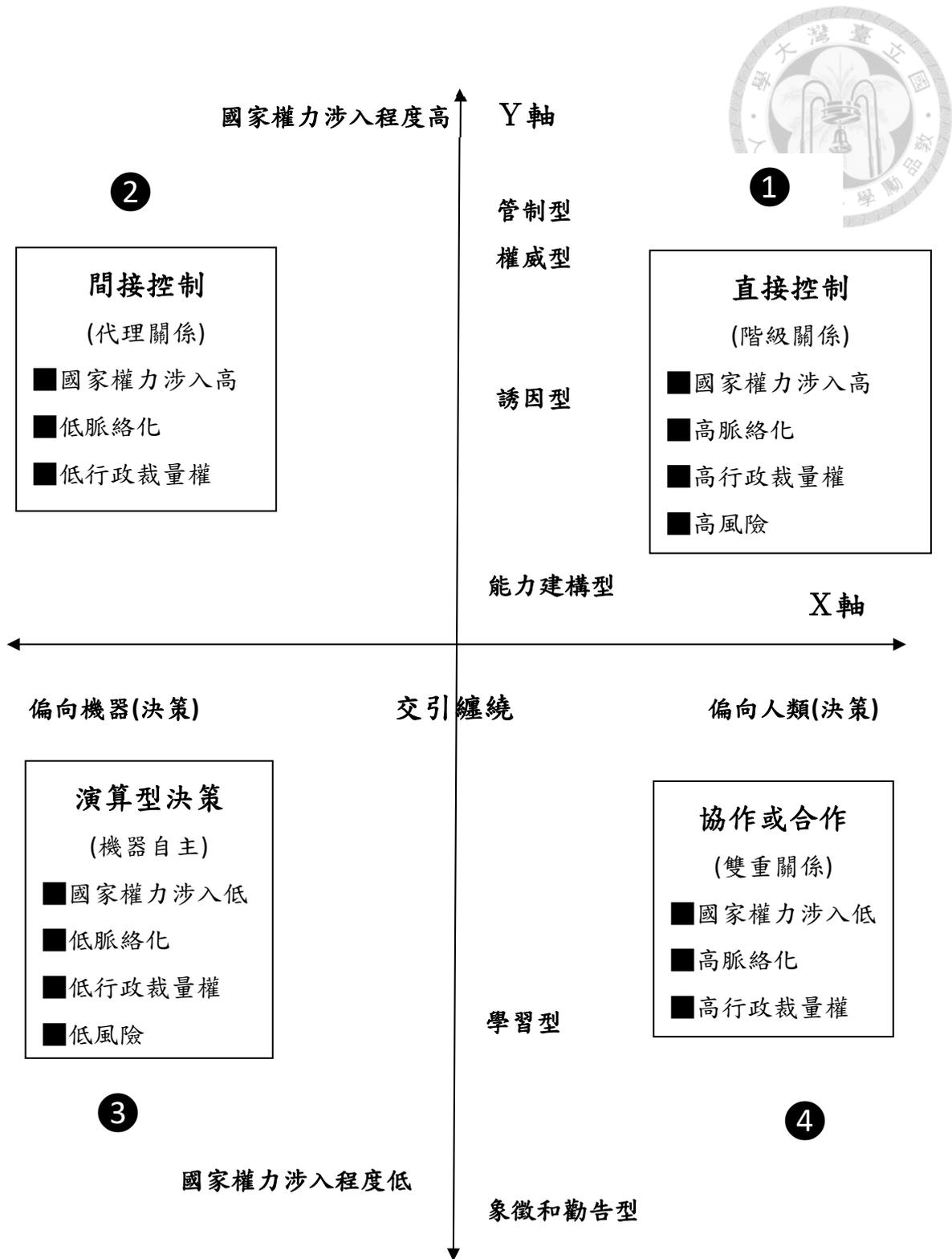


圖 3-3 案例選擇架構

資料來源：本研究設計



參、案例選擇方式

基於案例研究資料的可近性(accessibility)，案例選擇聚焦在臺灣公部門導入 AI 的實際案例，國外的案例則做為論述的佐證或補充，而不做實證上的研究。案例選擇方式採取立意選樣，選擇不同特質的案例，除避免選擇偏差，更能獲得綜觀性理解，另使研究案例都能為提供不同深度、獨特且豐富有價值資訊。篩選條件如下：

1. 政府任務領域或職能：依據美國聯邦行政會議針對演算法治理範圍中的政府運作(Engstrom et al., 2020)，以及聯合國發布政府職能的分類(UN, 1980)，選取不同政府職能。
2. 組織層級及組織型態：中央政府、地方政府、非政府組織 (Government, Local Government, Non-governmental)。
3. 計畫成熟度：歐盟案例研究使用的計畫進度，包括規劃(planning)、試驗或部分部署(piloting or partially deployed)、完全部署(fully deployed)三階段。
4. 可比性(comparability)：國家權力涉入程度高，聚焦於分配政策與管制性政策；國家權力涉入程度低，聚焦於政府提供資訊。
5. 應用普遍性或後續擴散潛力：公部門普遍的任務(task)需求(例如 AI Chatbot)或臺灣情境相關性需求(例如公文系統)。

肆、主要案例與數量

本研究提出「2x2」(entanglement × power) 的案例選擇架構共計 4 個象限，每象限各取 1-2 個案例，合計 5 個案例進行訪談。國外案例資料每象限 2-4 個案例，合計主要案例：5 案；輔助案例：7 案，合計 12 案。

主要案例與案例選擇數量如下：



1. 第一象限：主要案例「財政部智能稅務服務計畫」，輔助 2 案，共計 3 案。
2. 第二象限：主要案例「環境部智慧判煙系統」，輔助 3 案，共計 4 案。
3. 第三象限：主要案例「臺北市政府 AI 智慧交控管理」，輔助 1 案，共計 2 案。
4. 第四象限：主要案例「勞動部 1955 多元智能電服中心」、「勞動部 AI 人機協作分文」2 案，輔助 1 案，共計 3 案。。

再者，二維案例選擇架構會有「座標」、「數字」，本研究依據人與 AI 之間 entangle 的程度(向左偏機器、向右偏人類)、公部門的國家機關權力涉入程度(向上偏管制性、向下偏象徵和勸告型)，依據研究案例屬性標記座落點如圖 3-4。

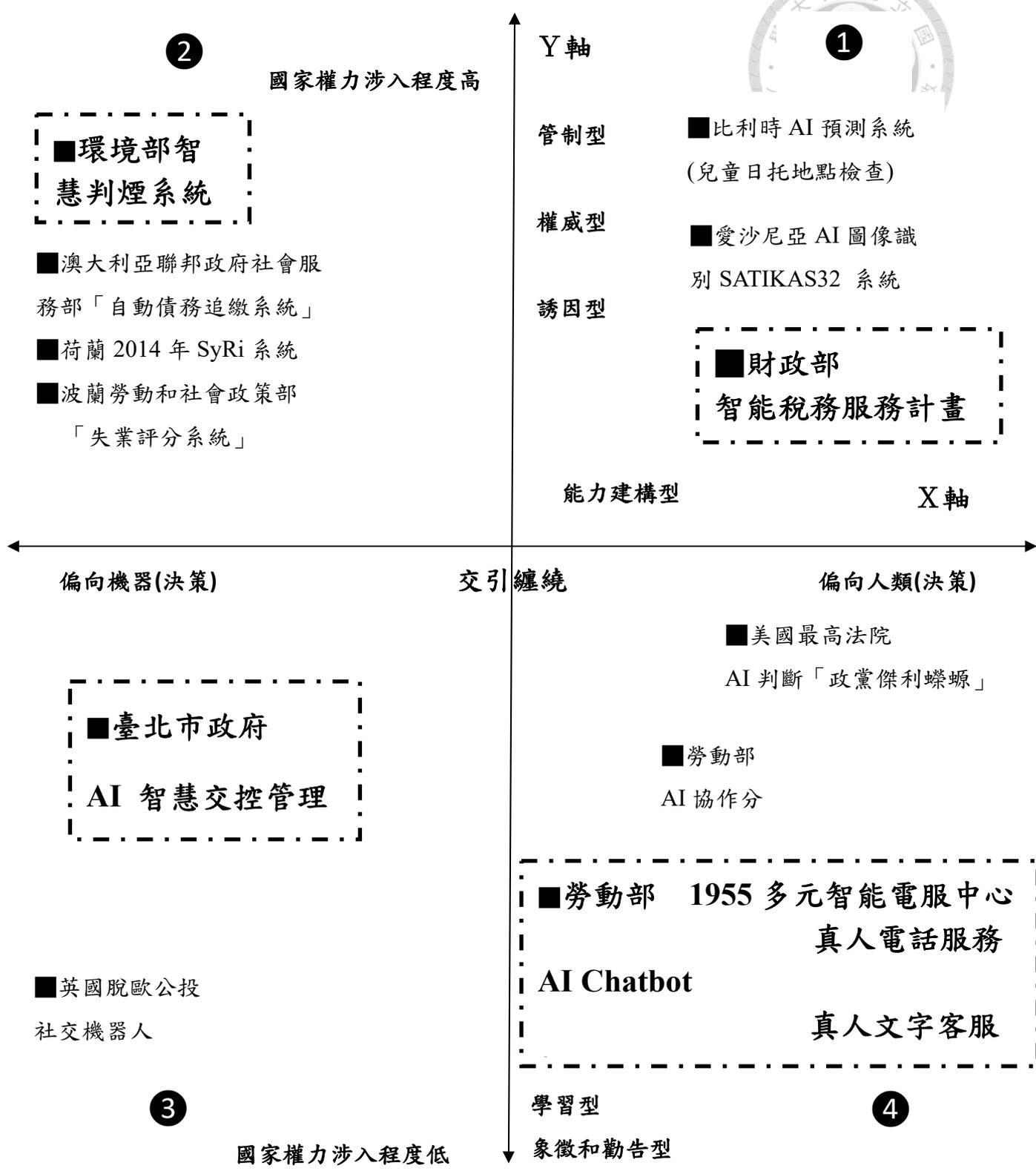


圖 3-4 研究案例

資料來源：本研究設計



第三節 研究方法

本研究採取多重研究方法，包括文獻分析方法、案例研究方法，個案分析並選擇二個分析途徑：ANT 的分析途徑及因果關係分析途徑，以 ANT 分析途徑為主，探索 AI 對公共治理理論的影響、人與 AI 的關係，目的是藉由實證研究案例推論公部門應用 AI 的公共治理理論(AI 的公共治理)、公共治理典範轉移路徑、公共治理領域人機協作的理論基礎。

壹、文獻分析方法

本研究跨學科蒐集文獻資料，包括社會科學、資訊科學、科學科技與社會學等。除蒐集學術研究成果以確保可信度，由於公部門(一般行政部門)應用 AI 較私部門緩慢，極少數文獻討論公部門 AI 發展和應用 (Desouza et al., 2020)，本研究另外著重於蒐集美國、英國、澳大利亞、愛沙尼亞、比利時、波蘭、荷蘭、歐盟及 OECD 等政府、學術單位、工商業界等組織研究報告、行政命令、建議書、營運建議書、網路資料等。另為側重於最新研究與發展，也蒐集大量技術報告、會議論文(Reis et al., 2019)。藉由文獻分析成果，根據本研究問題與目的，設計研究案例深度訪談的大綱與重點。

貳、案例研究方法

案例研究目的是對單一單元(single unit)(相對有限的現象)的意義建構與情境脈絡的深入研究，以了解更大類別類似現象的特徵。再者，案例研究方法應正確理解為定義案例的特定方式，而不是分析案例的方式或建模因果關係的方式 (Gerring, 2004)，因此，本案例研究採取混合研究(mixed-methods research)：ANT



分析途徑及因果關係分析途徑。

再者，本研究在方法論既已選擇 ANT 的分析途徑，為何選擇結合公共治理理論的因果關係分析途徑為輔助？原因在於 ANT 發展的實作本體論不同於一般本體論預設世界的本質存在與規律，實作本體論關注經驗過程中多元、異質的實作如何在具體脈絡中黑箱化(black-boxed)特定客體，因而促成(enact)特定現實實現或消失(林文源, 2017)；而社會只不過是網絡中的行動者和網絡，最終可以簡化為行動者及其關係，這種觀點不需要任何先驗假設、不區分因果關係、微觀和宏觀，而是高度探索性、描述性、非化約性(nonreductionalist)和定性方法(Latour, 2005; Waldherr et al., 2019)，著重於描述而非解釋(Latour, 2005)，適合研究發展中科技與社會之間如何互動，特別是 AI，但也限制了其規範性和批判性潛力，且由於反對經典的社會理論，會冒著淡化資源、意義構建過程和其他結構特徵的風險，例如忽略長期的權力結構和制度(Waldherr et al., 2019)。因此，本研究將 ANT 的分析途徑搭配因果關係分析途徑，除有利展現 ANT 的真正優勢，也可以更貼近真實並解決單一研究的不足與缺憾(黃麗鈴, 2019)，使本研究所找出能解釋公部門使用 AI 現象的理論具實用性與說服力。

參、研究執行方式

以深入訪談為主，輔以直接觀察或間接觀察、蒐集次級資料、官方資料等。由於 ANT 的研究方法「跟隨行動者」(Latour, 2005)詳細追蹤案例。然而，為兼顧「尊重被研究對象」及「揭露所做研究之完整過程及內容」，深入訪談研究成果依被研究對象意願選擇是否揭露單位及姓名、是否可以引述其論點等，並於後續儘量請被研究對象協助檢視，克服跨領域理解產生混淆或誤導，使本研究成果具有可信度、客觀性、嚴謹性。

再者，研究設計與以往最大的差異是分析單位增加前所未見的「AI」，雖然無



法訪談 AI，但由於 AI 也有能動性，本研究採用觀察 AI 是否對其他能動者的行動 (agent's action) 產生影響，以及人與 AI 互動的結果進行推論。有關細部研究執行方式說明如下：

一、分析單位

行動者包括：「人」、「AI」，以及其他能表達組織意志的法人、團體，例如國家 (the state)、各種利益團體、民間團體、政黨等。

二、訪談對象與方式

以人、AI 系統為中心的利害關係人，並依據視個案性質選取訪談對象，可能包括政策規劃者 (中央政府、地方政府)、AI 系統廠商、AI 系統程式設計者、資料管理者、終端使用者 (end-users) (公務員、公民等)、政策利害關係人 (例如工廠、公民) 等，訪談人數合計 30 人。分布如下：

1. 專家學者：4 人。
2. 第一象限：財政部智能稅務服務計畫 (5 人)
3. 第二象限：環境部智慧判煙系統 (4 人)
4. 第三象限：臺北市政府 AI 智慧交控管理 (3 人)
5. 第四象限：勞動部 1955 多元智能電服中心 (7 人)、勞動部 AI 人機協作分文 (7 人)

訪談期間為 2021 年起至 2023 年 6 月，由於研究期間長達約 2 年半，原有資訊系統與 AI 系統都會進步，不同環境都會影響結果，有關 AI 系統品質的數據係依據當時訪談結果而記載，可能無法精確。再者，本研究亦視個案性質採不同訪談方式，包括深入訪談 (in-depth interview)、團體訪談 (group interviews)，口頭訪



談(oral interviews)(林金定、嚴嘉楓、陳美花, 2005)、追蹤訪談(follow up interview)、文字模式線上訪談(張維安, 2001)(利用 line 通訊軟體)、直接觀察或間接觀察等方法交錯使用。

三、敘事重點和訪談大綱

ANT 分析途徑的敘事重點，以及因果關係分析方法訪談大綱，將參考歐盟及中國的研究策略：1. 詳細案例歷史(detailed case histories)：目前許多歐盟國家的研究策略，是透過詳細案例歷史的深入研究(deep dive)來獲得不同脈絡的 AI 治理問題(Molinari F., 2021)；2. AI 挑戰與風險緩解措施：中國管制機構正在探索 AI 的挑戰並規劃管制干預措施來緩解這些挑戰，並積累有關技術和潛在干預措施(potential interventions)的專業知識¹²⁵。訪談重點如下：

1. 人和 AI 的關係：人與 AI 互動，以及「人機協作」模式，原本屬於人類官僚的行政裁量權是否受到影響？是否會使官僚組織無法有效管理？
2. 公共治理理論和 AI 的關聯(association)：AI 對公共治理理論或公共政策分析的影響？AI 的機會？公部門 AI 治理的挑戰。
3. 詳細案例歷史：「公部門如何、何時、何地、為何以及為誰採用 AI¹²⁶」(Luca, Colin, Marco, Dietmar, & Francesco, 2022)，以及 AI 技術、成熟度(maturity)、數據資料、公共行政流程、公共價值或公共治理效能等(Luca et al., 2022)，以利跨國比較。
4. 成功或失敗原因：各種原始爭議如何出現、如何排除、如何穩固的歷程。

¹²⁵ 檢自 <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>

¹²⁶ 原文：It aims at answering the basic questions: how, when, where why and for whom AI is adopted in the public sector.



5. 因果關係分析：

- (1).系統層次：組織整體效能、效率、公平性、可管理性、合法性和政治可行性¹²⁷、組織管理、組織流程、對組織是否會產生負面影響等。
- (2).個人層次：個人效率、個人科技接受度、組織成員對變革是否抗拒等。
- (3).前瞻性評估((Horizon Scanning, HS)：識別及評估潛在風險(Hines, Yu, Guy, Brand, & Papaluca-Amati, 2019)，如何避免 AI 風險的因應對策。

四、研究流程

研究流程包括：1. AI 的公共治理理論分析流程：探討公共治理理論缺口與典範轉移路徑、2. AI 與人互動分析流程：探討人機協作的理論基礎。

¹²⁷ Young 等人(2019)使用 Salamon (2002) 評估治理工具的框架，開發了一個框架，根據效能(effectiveness)、效率(efficiency)、公平性(equity)、可管理性(manageability)、合法性和政治可行性(legitimacy and political feasibility)等標準來比較人工(artificial discretion)和人類的自由裁量權(human discretion)。其中「有效性」是指預期目標的成功或達成程度。「效率」是投入及其相關成本與產出或結果的平衡性。公平性從兩個維度評估：是否為所有公民提供平等待遇，以及是否有利於弱勢族群的再分配。「可管理性」，也稱為可實施性，是指操作程序的難易程度；工具越複雜，涉及的參與者越多，就越難以管理。合法性和政治可行性是為公民和政治人物是否認為工具合法（適合公共行動）和政治上可行（是否存在足夠的政治意願來實施該工具，以及反對的政治意願不足）(Lester, 2002; Young et al., 2019)。



(一)、AI 的公共治理理論分析流程

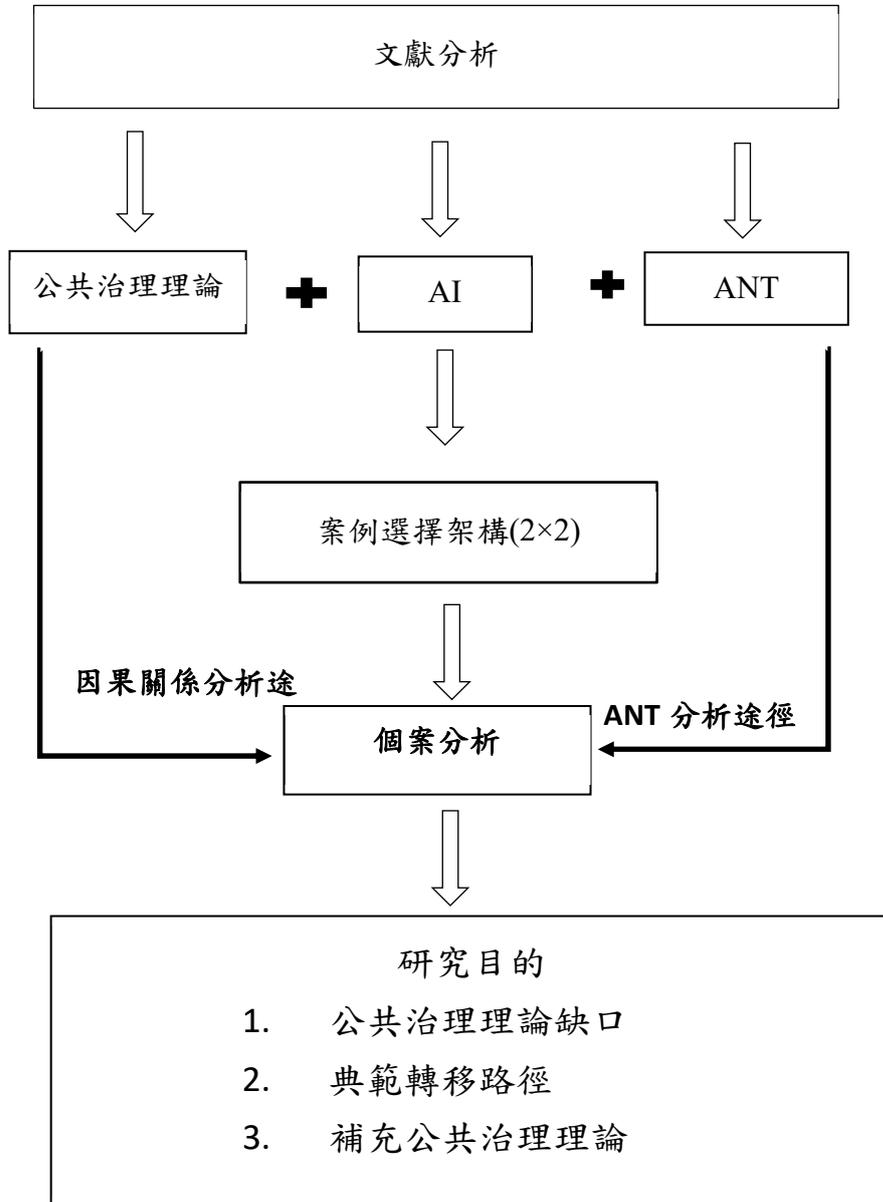


圖 3- 5 AI 和公共治理理論分析流程

資料來源：本研究設計



(二)、AI 與人互動分析流程

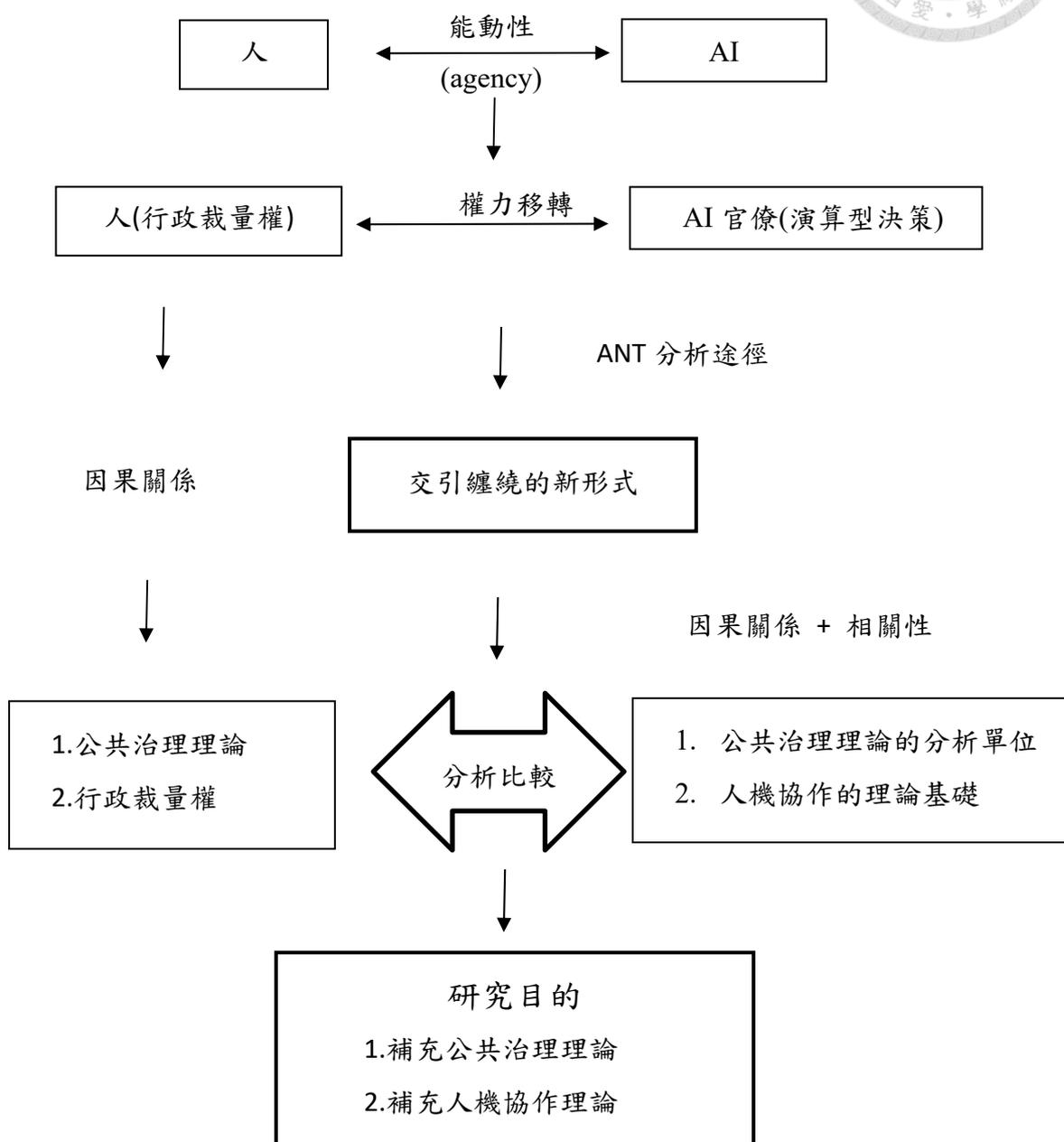


圖 3-6 AI 與公部門行動者互動分析流程

資料來源：本研究設計



第四章、臺灣公部門應用 AI 的個案分析

文獻雖然已經提出公部門使用 AI 的機會與挑戰，但在社會科學領域，很大程度上是探索性研究；再者，既有文獻對於如何因應挑戰的實務性作法較少討論。目前許多歐盟國家的研究策略，是透過詳細案例歷史的深入研究來獲得不同脈絡的 AI 治理問題(Molinari F., 2021)。臺灣同樣缺少公部門如何因應 AI 機會與挑戰的實證研究成果，為利日後跨國比較，本研究案例描述策略參考聯合國政府職能分類(UN, 1980)、歐盟針對 AI 專案使用的研究進度等。續依四個象限一一論述研究成果。

第一節 財政部「智能稅務服務計畫」

第一象限：財政部智能稅務服務計畫(Smart Tax Services Plan)，屬「國家機關權力涉入程度高」，偏管制性政策、偏人類決策，試驗或部分部署(piloting or partially deployed)，職能分類為 F4：經濟事務。本計畫目的是利用 AI 模型偵測可能逃漏稅的案件，協助增強稅務員的選案能力，解決政府無法逐案審查納稅義務人的申報資料是否有逃漏稅情況的治理困境。

諸多文獻已提出 AI 選案查核的技術可行性，例如使用類神經網路發掘營業稅逃漏稅之嫌疑案件(江美虹、吳朝欽, 2015)，或加值型營業稅逃漏稅預測(張春蓮, 2013)。



壹、案例分析—ANT 分析途徑與因果關係分析途徑分析結果

財政部財政資訊中心掌管全國民眾的財政資訊，擁有全臺灣最大的資料庫、最完善的數位治理系統。人員編制正職約 386 人、駐點約 250 人，正職中資訊人員大約占 80%，擁有數據、人才兩大優勢。2010 年「賦稅資訊系統整合再造更新整體實施計畫」提出資料探勘的概念，時任財政部財政資訊中心主任蘇俊榮 2015 年著手規畫「財政巨量資料研究中心研發計畫」¹²⁸，透過資料治理及大資料平臺做數百億筆財政資料分析，為日後智慧稅務奠定了良好的基礎。

接續於 2018 年起著手規劃未來智慧稅務服務藍圖，現有稅務資訊系統導入 AI 技術(張文熙、王培得、陳秀宇、林哲豪、鄧宜芳、曾柏勳、平曉鷗, 2021)。由於財政資訊中心被蘇建榮前部長視為是數位經濟時代提升稅捐稽徵行政效率的重要單位，在前主任退休後，蘇前部長四處尋覓具有 AI 和資訊專才的人選，從勞動部挖角當時資訊處張文熙處長至財政資訊中心擔任首長，希望借重張文熙具有專攻機器學習(Machine Learning)博士學位的 AI 專才，盼能加強財政部數位查稅力道。張文熙主任到任之後即開始推動「自動化、行動化、智慧化」：整合機器人應用的自動化、手機應用的行動化、AI 應用的智慧化，進一步加速財政部的稅務智能服務。

一、問題化

財政部在稅務系統分為國稅(所得稅、營業稅等)、地方稅(房屋稅、地價稅等)、電子發票等三項。由於查核人力短缺，實務上，稅務人員查稅時是選案查核而非普查。選案人員憑財稅行政專業及多年實務經驗，分析比對多個所得條件或模式，雖具業務實務價值，惟實務上卻有相當多的問題亟需克服，包括：

¹²⁸ 財政部財政資訊中心主任蘇俊榮：分三期推動財政雲，打造臺灣未來 10 年最重要的財政系統。檢自 <https://www.ithome.com.tw/news/94944>



1. 資料問題：

- (1). 稅務相關資料來源廣泛，收集、整理不易：稅務資料來源除納稅義務人申報之資料外，尚包括外部資料，如海關資料、戶政資料、健保署等外單位，以及因應新興交易模式的網路資料。
- (2). 第一時間發現申報資料錯誤困難：納稅義務人申報資料需與外部資料勾稽比對，因外部資料數據龐大，往往無法於第一時間發現申報資料錯誤。
- (3). 申報者全部資料不易全盤掌握：部分無法掌握申報數據來源。
- (4). 仰賴檢舉：消費形態改變，例如網紅、直播主之網路交易興起，形成的地下經濟。

2. 選案問題：

選案方式高度仰賴資深選案人員選案經驗，利用查審輔助系統輸入相關查核條件（如門檻值、風險事件權重等），系統產出涉嫌逃漏稅捐納稅義務人的「列選清單」，再進行人工試查作業。選案過程充滿重複的人工作業和人為主觀偏見，致每案均耗費相當時間才能決定是否派查，原因包括：

- (1). 選案結果範圍較大：需要動員大量稅務員進行實務鑑別，找尋更佳案件供後續派案與查核。
- (2). 經驗難以傳承：交易型態、涉嫌逃漏稅捐態樣日新月異，篩選條件調整不易。
- (3). 過程繁瑣且多為一循環迴路：人工試查作業階段，選案人員需逐案檢視、比對營業人相關申報、核定及課稅資料。

2019 年推動兩大智慧化改革，包括：第一項，提供民衆稅務便利新措施，包括 2021 年上路的手機報稅與稅務智能客服¹²⁹；第二項，建置稅務系統智慧演算

¹²⁹ 財政資訊中心主任 張文熙用科技掌握地下經濟。檢自



法，包括「提高選案命中率」，以及與賦稅署合作「稅收預測」：

1. 提高選案命中率：運用的技術包括：(1).運用監督式學習的分類技術，提高目前現有選案規則選案命中率；(2).運用非監督式學習的分群技術，找出未察覺的案件，增加應選未選案件。
2. 稅收預測：主要目的在於「數值問題」，預測稅收這數值。

為利聚焦，接續將主要探討提高選案命中率的個案，財政資訊中心找出強制性通過點(OPP)：「利用 AI 提升選案命中率，可以使每一個行動者受益。」，並定義行動者身份和他們「想要(want)」的東西：

1. 財政部前部長蘇建榮：用 AI、大數據找稅源¹³⁰。
2. 財政部財政資訊中心：利用 AI 提升選案命中率。
3. 國稅局審核員：解決選案過程過多及重複的人工選查作業、減少主觀式偏差並可滾動式修正選案方式。
4. 國稅局分局稅務員：稽徵人力短缺，能有效率辦理查緝工作。
5. AI：預測涉嫌逃漏稅捐納稅義務人的機率。

二、利益賦予

智能稅務服務計畫是「運用 AI 及數據分析」、「結合現行選案查核經驗」達成擴大稅基的目標，重點包括：

<https://www.bg3.co/a/cai-zheng-zi-xun-zhong-xin-zhu-ren-zhang-wen-xi-yong-ke-jizhang-wo-di-xia-jing-ji.html>

¹³⁰ 財長專訪／蘇建榮：用 AI、大數據找稅源。from：檢自

<https://tidatw.org/2020/07/02/%E8%B2%A1%E9%95%B7%E5%B0%88%E8%A8%AA%EF%BC%8F%E8%98%87%E5%BB%BA%E6%A6%AE%EF%BC%9A%E7%94%A8%E3%80%81%E5%A4%A7%E6%95%B8%E6%93%9A%E6%89%BE%E7%A8%85%E6%BA%90/>



- 1.健全財政資料治理：確認申報及徵納資料完整性及正確性。
- 2.整合跨部會資料：除納稅義務人主動申報資料外，並與各部會相關資料勾稽查核。
- 3.導入 AI 及大數據分析：蒐集外部資料，查核網路交易活動，找出人工不易發覺之異常交易行為。
- 4.打造智慧營運流程：提供稅務各階段之資料治理及決策資訊。

主要策略進一步說明如下：

在資料治理的面向，包含共通規範、資料安全、資料品質及資料管理等。首先以健全財稅大數據資料庫為第一步，對內整合財政部國稅(含電子發票資料)、地方稅(財產資料)及關務(進出口資料)相關資料；對外積極與經濟部、內政部、衛生福利部等部會協同合作。接續，藉由收集申報資料、跨機關跨領域的外部資料、網頁及社群軟體息等異質資料，建構完善之稅務資料治理機制。最後，為兼顧個人資料保護，發展個資假名化機制使資料運用更具安全性(張文熙、王培得、陳秀宇、林哲豪、鄧宜芳、曾柏勳、平曉鷗, 2021)。經資料可行性分析過程後，再進行 AI 建模程序。

「比較正確的資料是交易資料，例如信用卡資料、發票資料、金融資料等，都在交易當下真實記錄下來，但納稅義務人的申報資料(納稅人自己申報的)，就需要多方人工查證/電腦交查，所以國稅局會跟各政府機關或商家等蒐集資料，所以我們比較鼓勵國稅局多加入外部資料來做 AI。」

—財政資訊中心國稅組陳秀宇科長



「AI 計畫有資料蒐集組，任務是擴大蒐集資料、運用跨部會及外部資料(如網路交易及金流資料等)、洽談及研商法規限制、規劃資料介接事宜，達跨域資料加值整合運用效益，提升稅務行政服務效能。」

—財政部財政資訊中心王培得副組長

在演算法治理的面向，財政資訊中心先以國稅局稅務員過往實戰經驗所設立之風險事件及模組選案為基礎，使用決策樹(LightGBM)模型，利用所查獲「黑名單」、查核相符「白名單」進行 AI 建模。建模過程中需要投入最多時間及人力的步驟是國稅局稅務人員參與風險指標之建立與修改，估計占整體建模過程 70% 以上時間，對釐清資料、模型正確性、有效性與關聯性有直接關連。

模型建立第一步驟為蒐集資料及處理，探索性分析須針對各項特徵進行剖析，確認指標獨立性。第二步驟為製作有效特徵，並注入納稅義務人申報資料進行演算法訓練。第三步驟為模型測試，採用選案查核系統之查核回報結果，再行標註(Label)，並使用不同年度的資料交叉驗證，建模人員再依據選案人員意見、使用 AI 查案績效等作為調校、精進現有選案功能等後續模型精進之參考。最後，產生多個模型，找出最佳的參數狀況，模型分數高於建模人員標準方可進行後續上線準備，模型完成訓練並經業務主管機關驗證認可後，再投入實務應用。

三、徵召

雖然財政資訊中心幾乎全是資訊同仁，但同仁普遍不知怎麼開始做，2019 年起安排了 AI 課程，鼓勵同仁利用每週一個下午學習 AI 技能，經過 4 個月，始終參與聽課的同仁維持約 20 人，後來便將他們納為 AI 團隊成員。原則上，AI 團隊成員都能以原來工作經驗，延伸學習相關跨域所需能力，再重新組合任務角色

與分工：



「AI 目前不是必備技能條件，未來是必備條件，因此財政資訊中心的研究議題大部分是都是同仁自己成立 AI 團隊，自己研究、自己寫程式。基本上了解後才有能力提合理需求，建立驗收的能力。自己寫程式才真的懂，外包使技術流失，能寫一點最好，小事可以自己處理。」

—財政部財政資訊中心張文熙主任

一開始 AI 團隊成員都源自財政資訊中心，他們雖有好的程式背景，但不見得能完全掌握業務、法規上的細節。由於跨域整合是解決各種難題的重要趨勢，擴大稅基的工作更是如此，必須仰賴兼具法規、實務與技術的各方專家共同合作。接續由次長擔任召集人，整合財政部國際財政司、統計處、賦稅署、各地區國稅局一起加入 AI 團隊，共同組成「資料科學工作坊」，共同參與發展應用主題。賦稅署具有稅務法規方面的專業知識，國稅局則有實務經驗豐富的稅務人員，財政資訊中心則具有資訊技術專業人員，整個直接或間接參與的人員約 100-500 人。財政資訊中心透過教育訓練、分工合作模式，引導團隊成員將所學技術回饋至實作議題，具體分工模式如下：

「去年(2021 年)剛開始做 AI，大家對 AI 不熟悉不認識，所以引入中華電信(具有對業務+AI 熟悉的同仁)來帶領我們 AI 團隊從做中學，去年一開始，很多程式也是中華跳下來寫(雖然我們要求 AI 團隊自己動手)，但因為都是兼職做，且都是從 0 開始學習，很難一步到位。今年(2022 年)開始，我們跟中華電信討論逐步調整學習工具，讓學習門檻降低，所以今年的主題很多都是 AI 團隊自己動手做出來的。業務單位有 domain，

較適合由財政資訊中心提供資料+平台環境，讓業務單位自己實作 AI，這應該是理想目標，但考量 AI 導入初始期，新技術的引進得先由中心資訊人力投入，引導業務單位來做，所以我們自己要先有能力。」

—財政資訊中心國稅組陳秀宇科長

研究發現：成立跨領域實作 AI 團隊、訓練公務員自己寫程式，才有能力選取適合主題、提合理需求、檢驗 AI 模型、建立驗收的能力、避免技術流失。

AI 研究團隊的初步成果：

「從前舊做法的(選案)命中率僅 1~5 成，意謂財政部耗費人力工時做覆驗、當中高達 5~9 成屬於不具產值的作業；如今結合 AI 後，命中率拉升到 9 成以上，幾乎是一查就中，不再需要平白耗費時間和力氣。」¹³¹

—財政部財政資訊中心張文熙主任

四、動員

在本案例 AI 所構成的利益賦予策略原型，是否可以建立了一個有約束力的關係網絡，進而成功轉譯？

財政部於 2023 年開始試行 AI 選案查核，「正向方案」的動員的因素分析如下：

¹³¹ 財政部財政資訊中心主任張文熙談財政智能治理。from：檢自 <https://www.cio.com.tw/policy-lecture-hall-data-based-development-of-digital-governance/>



1. 財政部前部長蘇建榮

蘇前部長在財政部 2020 年的新春記者會表示，各國財稅監管單位也逐漸進入 AI 時代，財資訊中心也首度納入 AI 計畫，成立財政部的 AI 團隊、AI 稅務部隊，引進 AI 數據人員，提升稅捐稽徵行政效率¹³²。接續在 2020 年針對新任期的施政重點與當務之急接受專訪時亦再次表示¹³³：

「稅基的重要性大過於稅率，如果能夠抓住流失的稅基，一方面可達到租稅公平，一方面也可增加稅收。……財政部也透過 AI 人工智慧、大數據找稅源，精進選案查核能力。國稅局有許多選案查核的高手，透過他們的經驗累積，不斷精進電腦選案方式，未來準確率就會提高。」

—財政部前部長蘇建榮

2. 財政部財政資訊中心

2020 年開始推動技術、文化、組織等轉型。在組織轉型，打破公務機關不易調整組織的限制，通過任務編組成立約 20 人的 AI 團隊，以及數位鑑識的資安團隊¹³⁴。在技術轉型，主要是資訊系統再造。在文化轉型，AI 剛開

¹³² 財政部今年打造 AI 稅務部隊。檢自

<https://www.rti.org.tw/news/view/id/2048736>

¹³³ 用 AI、大數據找稅源。檢自

<https://tidatw.org/2020/07/02/%E8%B2%A1%E9%95%B7%E5%B0%88%E8%A8%AA%EF%BC%8F%E8%98%87%E5%BB%BA%E6%A6%AE%EF%BC%9A%E7%94%A8ai%E3%80%81%E5%A4%A7%E6%95%B8%E6%93%9A%E6%89%BE%E7%A8%85%E6%BA%90/>

¹³⁴ 力推手機報稅因居家防疫爆紅，財政部還要靠 AI 打造智慧稅務。檢自

<https://www.ithome.com.tw/people/145016>

始被當成獨立的新技術，但最終 AI 的應用還是要回歸到業務面：



「文化轉型目的是讓員工能對追求新科技有熱情，並將學到的技術回饋到工作，改善財政部的業務。」

—財政部財政資訊中心張文熙主任

3.國稅局審核員(選案人員)、分局稅務員(查核人員)：

查審輔助系統導入 AI，初步判別原始資料，優先排除掉無用的部份，提升選案效率：

(1).擴大選案範圍：

「智慧查覈系統會自動學習稅務人員經驗，同時結合跨部會資料，順藤摸瓜找到可能的地下收入，列為查覈範圍。」

—財政部財政資訊中心張文熙主任

(2).提高選案能力：

使用 AI 和大數據分析，就可以結合各單位的資料，例如海關的資料一起分析，可以在短時間識別人類可能錯過的模式和隱藏的關係(Höchtel et al., 2016; van Bruxvoort & van Keulen, 2021)，藉由克服人類經驗主義客觀現實的困難，擴增選案人員的選案能力，更進一步提升使用者的科技接受度：

「以前用營業人的申報資料選案，憑經驗下選案條件可能只會用 5 特徵變數、異常點去選案，使用 AI 會挖掘出憑經驗看不出來的重要異常點，可能會有第 6、第 7 個異常點出現，把更多異常點放在一起，自然

會提高命中率。」

—財政部臺北國稅局股長楊秀真



「在稅務領域使用 AI 選案查核的門檻雖然很高，但使用 AI 之後，發現對業務很有幫助，嘗到了甜頭，就回不去了。」

—財政部財政資訊中心張文熙主任

(3).精進查核效率：

利用 AI 選案，可以明確化查核方向及重點，分局稅務員比較能抓住查核方向與重點，越能節省時間。

「以前查核方式比較沒有重點或方向，要針對漏稅嫌疑人全方面查核，或依賴查核人員過去經驗，現在查核人員可以針對 AI 選案模型和大數據分析結果的異常點或特徵變數查核。」

—財政部臺北國稅局股長楊秀真

(4).經驗傳承：

逃漏稅或短漏報技巧千變變化，AI 可以從國稅、地稅、發票等一步步學習資深稅務人員的專業，新進稅務員可以利用 AI 學習資深稅務人員才懂的『眉角』：

「每一個稅都有專精的地方，但我們的工作量大、流動率高、一直輪調查不同的稅，經驗很難累積，要查到一定的深度、廣度很難。」

—財政部國稅局



「查稅領域有經驗問題，但並非人人都是老手，運用系統設定記錄資深稅務人員的篩選條件，甚至還能調整多種變數，新進同仁無需花腦筋，直接從系統獲得處置建議，未來即使是稅務菜鳥也能輕鬆找到可疑案件。」

—財政部財政資訊中心張文熙主任

在本案例中可能進行「反向方案」的動員的因素進一步分析如下：

1. 公部門缺乏 AI 技術：

(1). 在職訓練

財政資訊中心以嘉獎鼓勵同仁工作之餘參與 AI 技術數位學習課程，重點包含問題分析能力、數值分析能力(如統計分析或其他數學的基本模型)，以長期培養同仁在 AI、資安數位鑑識方面的專業能力：

「learning，機器要學、人也要學，這是一個 10 年大計，才能看到具體成效。」

—財政部財政資訊中心王培得副組長

「張主任常常親自授課，我們也要求新進同仁都要上課。資訊人員要學習 AI 知識不是問題，困難是他們要有稅務知識，跟稅務人員要有相同語言才能溝通，這個比較難。業務人員也要學習 AI 知識。」

—財政資訊中心國稅組陳秀宇科長

(2). 跨領域實作團隊



資訊單位沒有財稅領域知識，不知道怎麼訂定查稅篩選條件；業務單位有豐富領域知識，但不擅長數據分析；此時，跨領域的團隊就非常重要。在 AI 模型建構過程中，至關重要環節是資訊人員訓練 AI 模型時必須與國稅局人員協同偵測模型：

「我們透過 AI 課程知道有什麼 AI 工具可以使用，然後組成一個 team，我們提需求，資訊單位寫程式，我們合作一起建立 AI 模型。」

—財政部臺北國稅局股長楊秀真

2. 資料品質：

AI 的第一門政治課就是資料的取得，大部份資料的擁有者是不會提供的，而稅務資料還有稅稽徵法 33 條的規範，因此資料的取得是第一個問題。更進一步，該資料必須是已分類的資料(classified data)，亦即資料已經專家分類(標記)完成或有已知的結果。如果已分類的資料是正確的，學到的才有可能是對的。如果黑白不分，學到的可能就會有偏誤，模型的結果信賴度當然就低，因為學到錯的了。現有資料還包括：

「1. 現有資料特徵無法反應異常行為造成分類誤差，直接影響模型準確度。2. 擴大資料範圍、設計間接特徵指標不易。3. 人為偵測異常案例過少，形成機器學習成效不如預期。」

—財政資訊中心主任張文熙

因此，利用 AI 選案查核會遭遇無法獲得「正確真實 (accurate ground truth)」的訓練數據(Engstrom et al., 2020)。



「主要原因是課稅資料蒐集不完整，國稅局還要人工去找很多資料(例如金流)，而查核案件是黑或白，很多取決於查審人員，或是 Black Art—協商決定課稅方向及金額，把店家課死了，國稅局也沒得再課，所以要有點空間，有點藝術，在雙方可以接受的情形下課稅，所以有人說選案是一門社會科學，導致訓練資料是黑或白不是很正確。」

—財政部財政資訊中心王培得副組長

「之前沒被查到的案件，一定都沒有問題嗎？之前願意補稅的人，可能是基於其他原因而願意補稅，未必真的有問題。」

—財政部國稅局

因此，當 AI 選案查核模型使用決策樹，利用黑、白名單訓練機器，選案人員與查核人員對 AI 模型產生了信任度的問題：

「用以前人的判斷資料當成訓練資料，人也可能誤判，資深的查稅人員可能比較不容易誤判、資淺的容易誤判。所以黑名單不夠黑、白名單不夠白，把這個資料教機器，機器怎麼學得出來？。」

—財政部國稅局

「黑名單(曾經逃漏稅的營業人)不是真的黑名單，或不是很有效的黑名單，黑名單不是真的黑名單，白名單也不是真的白名單，黑、白名單數量失衡，可能使訓練結果產生偏誤。」

—財政部臺北國稅局股長楊秀真



另一個挑戰是不法行為的動態特質，必須不斷地更新逃稅態樣的訓練數據：

「我們有政策，業者就會有對策，只要出現一個漏稅點，國稅局就會發函，異常業者、或漏稅嫌疑人就會知道我們在查那一點，以後就會把這漏稅點的『資料補到完整』，永遠跟政府捉迷藏，原來建的模型就沒有用了。」

—財政部國稅局

「建立一個模型需要快一年，營業人逃漏稅樣態隨著時代背景不斷變遷，我們就要滾動式修正模型或再建立一個新的模型。」

—財政部臺北國稅局股長楊秀真

因此，很難用過去數據識別所有真陽性(true positives)，也永遠無法「知道」或全面識別假陰性(false negatives)，對於無法獲得「正確真實 (accurate ground truth)」的訓練數據，產生對 AI 模型信任度的問題應如何克服？(Engstrom et al., 2020)。

「國稅局反映的這個問題我們也知道，所以我們有提醒，選題方向及訓練的資料來源很重要。」

—財政資訊中心國稅組陳秀宇科長

首先是主題選取，運用 AI 系統必須權衡經濟邏輯與公共價值：

(1).經濟邏輯：必須權衡 AI 技術成熟度和訓練成本。主題通常具有脈絡性、獨



特性，某特定主題要建立一個高預測率的 AI 模型，必需要投入相當大的資源，例如足夠大的訓練資料、必須經過 try and error 的階段。若市場上已有成熟且類似的 AI 模型，可以依公務場域資料訓練後直接使用等（例如智能客服、智慧交通控制管理等），是比較經濟與快速的作法。若市場上沒有可用的類似 AI 模型，就必須具有經濟規模或成本效益，例如可重複使用（例如車牌辨識、人臉辨識）的特質。然而，公部門使用的 AI 通常有獨特性，常常必須從新研發；市場規模較小，導致建立 AI 模型成本太高。因此，若公共事務的屬性是高頻率、重複性高，例如每天要做，或 24 小時要做（例如交通控制、空氣品質監測），加上機器沒有情緒、不會累，就會有成本效益和開發價值。

- (2). 公共價值：公共價值不能單純以經濟邏輯或成本考量。舉例來說，美國食品和藥物管理利用 AI 預測嚴重的藥物不良反應真陽性率(true positive rate) 91%、假陽性率(false positive rate) 4.9%，屬於高預測率的 AI 模型。同樣的，大部分或所有數據主要是包含不良反應的結果(黑白名單失衡的問題)，「正確真實情況(accurate ground truth)」訓練數據難以取得，以至於難以根據可用數據預測分析來揭示藥物與不良反應之間的「因果關係」。由於不正確(inaccurate)的預測可能會導致嚴重的生死攸關的後果，即便是具有精確度(precision)和專業知識的 AI 模型，FDA 至今沒有完全成功部署 (Engstrom et al., 2020)。因此，**是否可以運用 AI 模型涉及價值選擇的問題，必須有社會共識，並不是每一個主題都適合發展 AI。**再者，公共價值有時候不能考慮成本，

「公家機關成本不是最重要，效果比較重要。」

—財政部財政資訊中心張文熙主任



應如何訓練主題選取的能力？

「持續做才會有成果，因為做了才會知道哪些題目(或資料)適合做，哪些題目(或資料)不適合做 AI，例如去年有的主題因為資料選錯了，所以**即使準確率很高也不適合上線。但選案不是非黑即白，有些還是可以靠技術克服，但避開這些題目或資料是最直接有效的方法**，去年提題目時大家沒有經驗，所以今年就決定雞婆一下有提醒國稅局不要挑這樣的題目，先提出題目讓中心審題，我們先討論過濾掉不適合的題目。」

—財政資訊中心國稅組陳秀宇科長

其次是演算法模型的選取與持續調校。

「可以有三種作法來精進模型：在學不完全對的基礎上，由資深查核人員參與模型調校；另外則是先分群，再用機器學習演算法建模；最後採專家標記方式，由資深審查人員對要訓練的資料逐一標記，再訓練。」

—財政部財政資訊中心王培得副組長

因此，財政資訊中心透過主題選取、演算法模型的選取與持續微調(fine-tuning)，舒緩沒有「**正確真實情況(accurate ground truth)**」的訓練數據的問題。

研究發現：主題選取必須權衡資料品質、經濟邏輯與公共價值。而演算法模型的選取與持續微調舒緩沒有「正確真實情況(accurate ground



truth) 」的訓練數據的問題。

3. 組織和管理挑戰：

(1). 人員管理

公部門導入 AI 選案查核，行政流程改變幅度不大，主要是數位轉型期間人力缺乏、工作負荷加重、科技接受度等，不願意改變現有的工作方式。

「就我所知，用 AI 查稅，支持與反對大約各占一半。喜歡的人覺得有趣，可以減輕工作量，但入門有點難，一開始要理解和寫程式很痛苦。不喜歡或抗拒的人，原因包括：1. 原來的的方法就可以查稅，現有工作還是要做，現在多了一件事，要增加很多額外的時間；2. 學習 AI 很困難，我們是業務單位又不是資訊單位或資訊背景的人；3. 對 AI 訓練資料的疑慮；4. 模型不一定有用，甚至可能無效，就會做白工。」

—財政部臺北國稅局股長楊秀真

「導入 AI 模型方向不錯，財政資訊中心希望我們都能寫程式，但我們平常業務都做不完了，更沒有時間去上課，很難行有餘力再去學 Python、寫程式。」

—財政部國稅局

而財政資訊中心是透過配套措施：機器人流程自動化(robotic process automation, RPA)、「人員訓練」以及 AI 團隊，組織和管理挑戰：

「靠 RPA 讓行政流程變得更自動化，RPA 已經上線的是 55 支，2024 年

會到 100 支，讓人員有較多時間投入服務改造。」

—財政部財政資訊中心張文熙主任



研究發現：機器人流程自動化(robotic process automation, RPA) 減輕現有工作負荷，降低路徑依賴及員工抗拒改變。

「一開始我們提出需求，請廠商協助我們理解那些工具可以運用，例如廠商會告訴我們可以使用決策樹分析數據，並建議門檻值，我們再用黑名單去訓練、分析、驗證 AI。以前的門檻值比較主觀，且必須以人力反覆測試及驗證，現在門檻值是機器用黑名單算出來的。雖然我不會操作但我知道可以這樣做，所以資訊人員協助很重要。」

—財政部臺北國稅局股長楊秀真

(2).組織管理：

諸多文獻建議成立專責部門或專門團隊，負責研究或促進公部門對 AI 的採用、研究 AI 的影響或 AI 新法規(Misuraca & van Noordt, 2020)。財政資訊中心則以成立 AI 團隊的方式進行：

「在政府機關欲實質地轉變組織，的確比較困難，但任務編組就相對沒有限制，財政資訊中心藉由跨組織方式來組成 AI 團隊就是彈性作法。此外，以目前 AI 的技術，公部門採用 AI，尚看不出來會對組織產生實質的影響，主要原因是政府職能、業務本質不變，AI 是輔助工具性質。」

—財政部財政資訊中心張文熙主任



但這在公部門仍可能會衍生導入 AI 的權責單位問題，

「不論是業務單位還是資訊單位主政，都逃不開 specific domain，domain 和技術一定要結合在一起，不能讓技術主導、也不能讓業務領先，我們是由次長來直接督導。」

—財政部財政資訊中心王培得副組長

至於是否應成立專責單位？民間單位有前例可循，

「我們還沒有成立 AI 專責團隊之前，先少量專家做出成績後，公司認為可行，才成立團隊。的確，成立團隊後，成員多了，成績也就更好。中華內不同的機構都有自己所屬的 AI 團隊。以中華研究院來說，至少有五組以上的 AI 團隊，每組 10-30 人不等。企業客戶分公司也有 3 組以上。」

— 中華電信企業客戶分公司廖志明股長

4. 可解釋的 AI：

在公務員科技接受度的面向，另一項非常重要的因素是「可解釋的 AI」¹³⁵。支持者何以會願意信任 AI 模型呢？對有參與 AI 設計的公務員，因為理解 AI 的優勢及限制，於多次驗證測試及試行下，傾向相信 AI 的判斷，「參與 AI 模型的開發」是解決 AI 可解釋性的一個重要因素。

¹³⁵ 檢自 <https://www.bnext.com.tw/article/53335/fujitsu-explainable-ai-with-knowledge-graph>

「我信任啊，因為 AI 模型是我跟資訊單位一起做的，選案條件、需求是我下的，我當然信任。」

—財政部臺北國稅局股長楊秀真



這樣的觀點可以呼應諸多文獻所提及資料治理與演算法治理的重要性，佛蘭芒兒童和家庭事務局 2014 年對兒童日托服務地點進行檢查所開發的 AI 預測系統 (Bongers, De Bruyn, & Verlet, 2018; Misuraca & van Noordt, 2020)，愛沙尼亞農業登記和資訊委員會 2018 年開發 AI 圖像識別 SATIKAS32 系統，檢測農業草原是否已修剪 (Boyle, Mamais, & Khabarov, 2021; de Vries et al., 2018; Misuraca & van Noordt, 2020; van Noordt & Misuraca, 2022) 等案例顯示，AI 系統使用者若能提供自己的專業領域知識，並參與 AI 模型的開發，AI 模型又能減輕原來工作負荷，是除了技術方法之外舒緩「可解釋的 AI」的一種行政式操作方法，也是公部門是否能順利導入 AI 的重要關鍵。

研究發現：領域知識者、AI 系統使用者「參與」AI 模型的開發，是除了技術方法之外舒緩「可解釋的 AI」的行政性作法。

5. 經費來源：除了調整現有預算，也申請計畫補助，並配合預算經費的提撥狀況來調整戰術；如果經費不足額、無法一次全面改，便採取分階段作業，縮小實作範圍：

「這是我們的痛，沒有編在公務預算，這是寫計畫爭取科技部科發 4 年計畫的經費，每年都要申請。」

—財政資訊中心國稅組陳秀宇科長



最後，本案例尚無公平性、倫理、隱私和數據匿名、工作權等問題的挑戰。究竟 AI 所形成的行動者網絡，「正向方案」與「反向方案」各自網絡結盟角力之後的結果究竟是如何呢？

2021-2023 年的 AI 專案計 21 案、大數據 Big DATA 計 62 案，不同稅務主題所建立 AI 模型的精確率不同，由國稅局整體驗證/評估決定 AI 模型是否有效、是否上線。2021 年有實作 8 個主題，評估上線有 5 個，未上線的主題，未來持續精進後仍可以有機會上線。

「我們都還在起步階段，財政部第一年做 AI，大部分國稅局不知道要怎麼提題目，且最後能全部產出模型，精確率(Precision)大約都有在 70%以上，且有評估上線或不上線，算是走完全過程。」

—財政資訊中心國稅組陳秀宇科長

而驗證/評估的因素是什麼？

1. 基本上比人工好就採用

「基本上、能減輕工作量、對業務有幫助、有實用價值」。

—財政部財政資訊中心張文熙主任

研究發現：AI 系統的品質評估是「比人工好就採用」，比人工好就必須有一定程度的 AI 系統品質，此時 AI 資訊的品質評估方式是和人類的能力比較。

雖然評估的方法不見得是精確率(Precision)，但是要比人工好、達到效益還是



要有一定的精確率：

「試查還沒有結束，初步評估各 AI 模型的正確率不同，因為使用方法問題，當然越高越好。人工選案正確率低於 10%，所以 AI 模型 50% 就可以用。」

—財政部財政資訊中心張文熙主任

「能不能上線，評估指標包括訓練資料完整性、選案精確率(Precision)等，我覺得具有一定的精確率(Precision)，例如五個國稅局平均大約 60-70%，只要評估『還可以接受』，大約 80-90%。」

—財政部國稅局

2.減輕工作量

「可以從題目本身、資料可否支援及實用性、有效益(例如可以減輕國稅局的負擔、降低工作量等等)，希望調到 80% 以上才算達一般可接受的水準，但仍有例外狀況，有些案例是若能降低查案量，即使未到 80% 也是可以接受的。」

—財政資訊中心國稅組陳秀宇科長

「選案可以針對一個特定行業，例如汽車維修業；也可以針對不同風險條件進行選案，例如營業稅加值率偏低或留抵稅額偏高。使用原來的『營業稅查審輔助系統』選案，必須不斷修正挑檔條件及門檻，大約需要 1 個月的時間；但使用 AI 選案，例如建立一個汽車維修業漏銷模型，1 個小時內可以完成。」

—財政部臺北國稅局股長楊秀真



「可以大幅減輕工作量，例如以前選案要 1 個星期，現在 1 天不到；或限縮查核的範圍，就可以評估上線。」

—財政部國稅局

3. 科技接受度

有關跟技術評估較無相關的因素，包括：

(1). 首長或高層長官的科技接受度：高層長官支持是非常關鍵的因素，

「AI 對業務有幫助，要讓首長看到，才能落地。」

—財政部財政資訊中心張文熙主任

(2). 業務單位的科技接受度；

「這個 AI 模型是誰的需求？如果是業務單位提出的需求，則 AI 模型的接受度會比較高，如果是資訊單位提出的需求，則 AI 模型的接受度就不高。」

—財政部財政資訊中心王培得副組長

因此，AI 主題的選定應選擇業務同仁最感興趣或最想解決的，可提高投入程度。此外，具有領域知識者的參與協作，也會傾向增加科技接受度：

「國稅局同仁以人工方式確認模型的有效程度，如果是資深的稅務員，會傾向模型是有效的，資淺的看不出來，就會覺得模型無效。」

—財政部財政資訊中心王培得副組長

4. AI 的可解釋性

對於 AI 模型可解釋性的疑慮，即使 AI 模型有極佳的表現，仍有可能考慮不上線，表示非技術性 AI 因素比技術性 AI 因素更重要(Pitt et al., 2023)：

「初步已經有很成功的案例，例如在稅收預測的實作議題，以往要花幾天才算得出來的『預測值/實際值』大約 95%。資料科學工作坊團隊利用 AI 幾分鐘推估的營業稅稅收預測『預測值/實際值』99%，這涉及 political 的問題，能不能被賦稅署及財政部接受，又回到 AI 信任度、接受度的問題，老闆敢不敢用？」

—財政部財政資訊中心王培得副組長

這意味著 AI 技術人員可能會看到改進和進步的機會，而政治家和政策制定者可能經常只看到風險，即使真正的風險可能不存在(Pitt et al., 2023)。

有些 AI 選案查核工具規劃上線實際應用，而目前暫不考慮上線實際應用，大部分都是訓練資料來源的問題，例如有些是黑白名單比例懸殊的問題，

「訓練資料 100 筆裡面有 90 幾筆是黑名單，導致學習效果不佳，意思是說機器隨便猜(只要猜是黑名單)對的機率都可以高達 90%，不該拿這資料來訓練。另一個是沒有找到重要特徵、或是找出來的特徵都不明顯。」



另一個案例，

「1.黑名單太少--->無法找出重要特徵或特徵不明顯--->學習效果不好。
2.AI 找出來有部分案件，稅法上原本就屬免稅範圍.--->原本我以為在跑 AI 前，就該先排除這些免稅案件，但國稅局說實務上要到現場查看，才知道產物是否屬於免稅，無法僅從申報資料看出是否免稅。」

—財政資訊中心國稅組陳秀宇科長

有些是稅務資料原始分類的問題，

「中文品名必須篩選汽車零件類產物才列入異常，否則會產出假異常的案件，例如負責人進口很多個人使用的衣服，但系統也判定是異常案件，所以要學自然語言分類器來處理品名問題，我們目前自行評估不上線，我們會內部討論產出清冊的試查有效性，但我個人覺得如果問題可以解決，應該可以上線」

—財政部臺北國稅局股長楊秀真

有些主題雖然評估不上線，但仍持續克服目前遇到的難題。諸多文獻已提及「沙盒(Sandbox)」測試(Engstrom et al., 2020)的概念：

「做 AI 跟開發資訊系統不一樣，開發資訊系統經過需求訪談/系統分析/系統設計等，到最後一定會有結果，但做 AI 不一定會有結果，很可能

是白忙一場(不上線)，因為還在起步階段，對那些不上線(白忙一場)的模型，我覺得學到是經驗，以後會知道如何提適合 AI 的題目。」

—財政資訊中心國稅組陳秀宇科長

貳、案例特色

從本案例觀察，利用 AI 資料探索和洞悉，針對業務問題找答案或發現規則，節省選案人員選案試查時間，協助查核人員設定查緝方向，有助於「提高公部門營運管理」、「精進管制性政策」，但由於最終採人類決策，不會作為直接裁罰或補稅的工具，因此，只能算是輔助(內部用語是「查審輔助」)。

接續針對本案例特色整理如下：

1. 高層重視：最高首長財政部長的高度重視，訂定「財政部運用人工智慧技術提升稅務行政效能作業計畫」，組成 AI 推動委員會，由次長擔任召集人，賦稅署署長、五地區國稅局局長、財政資訊中心主任擔任副召集人，另包括統計處及國際財政司等。
2. 專責任務編組：先期由資訊單位主政，再組成跨領域 AI 研究團隊、資料蒐集組，逐漸引導業務單位導入 AI。
3. 廣泛主題的應用：AI 的技術程度差異很大，系統功能改進的幅度也差異很大；在稅務領域使用 AI，門檻雖然很高，但仍能對業務很有幫助。
4. 創新模式：計畫規定五區國稅局每年都要提案實作，其他單位雖沒有強制規定，但也都有提案實作，逐漸在財政部產生多點創新，與美國聯邦政府、歐盟類似，採取一種讓百花綻放、多點試驗(pilot)的策略(Luca et al., 2022)，希望最終找出 AI 最佳創新模式。這種做法與中國不同，中國是將大量資金投入到一些預期成功率較高的 AI 計畫或主題上¹³⁶。

¹³⁶ 檢自 <https://www.brookings.edu/blog/techtank/2022/09/22/understanding-artificial->



5.將 AI 視為未來必備技能：財政資訊中心 AI 技能訓練的目標是自己寫程式，職業訓練的層次上是達到 Level 4—「效果層次」¹³⁷ (Smidt, Balandin, Sigafoos, & Reed, 2009)，有別於目前普遍對 AI 的訓練是 Level 1 -Level 3。美國聯邦政府導入 AI 的經驗顯示，53%的 AI 應用程式是由內部機構技術人員開發的；33%是透過政府採購程序從私部門購買的；14%是非商業合作（包括與研究型大學的合作夥伴關係和機構主辦的競賽) (Bignami, 2022)。由於公部門的 AI 工具必須具有可用性(usable)和法遵(compliant)；可用性是指對演算法工具所試圖解決問題的深度理解、說服持懷疑態度的公部門工作人員使用 AI 工具等；法遵是指演算法工具所產出的結果必須符合法律規範(Engstrom et al., 2020)。建立內部能力的優勢，包括：

- (1).內部人員(例如選案人員或是查核人員)能判斷哪些主題值得發展 AI。
- (2).具有領域知識較有可能以合法、符合政策和負責任的設計和實施 AI。
- (3).熟悉業務又理解 AI 工具的跨領域人才或團隊，是創新的最重要關鍵。
- (4).有些 AI 模型(例如 naive Bayes)強加很多獨立性假設，沒有領域知識和 AI 技術，很難去校正模型(calibrate model)(Engstrom et al., 2020)。
- (5).在管制性政策，違法態樣不斷推陳出新，能隨時調整 AI 模型。

相反的，如果不考慮政府職能、政府權力涉入程度，採取外包可能會遭遇的困境：

intelligence-spending-by-the-u-s-federal-government/

¹³⁷ Kirkpatrick 的「評估訓練方案的技術」中的四層次模式（依次序為「反應層次」、「學習層次」、「行為層次」和「效果層次」），「效果」層次的評估，是評估學員經過訓練後對組織所能提供的具體貢獻。



- (1).公部門內部專業知識空洞化，擴大公私技術差距。
- (2).無法了解 AI 技術限制，可能無法檢測對抗性攻擊，弱化政府對管制性政策的治理能力。
- (3).無法因應新型態逃稅態樣，即時調校 AI 模型、故障排除和調整。
- (4).外包廠商可能利用演算法，藉由公權力的行使而任意濫用權力。
- 6.完善的資料治理：擁有部會層級的財稅大數據資料庫。
- 7.機器人流程自動化：降低現有工作負荷。
- 8.推動文化轉型：推動技術、文化、組織等面向的轉型。
9. 實證可解釋 AI 的行政性作法：使用基於 AI 的決策，可解釋性與可信任性是協作必須跨越的重要門檻(蘇經天, 2023)，對公部門而言更是一個巨大的挑戰(Henman, 2020)。Dun & Bradstreet 2018 年調查已經使用 AI 的全球 2000 強公司，46%的受訪者表示對 AI 的最大挑戰是可解釋性，即理解 AI 系統如何做出結論，只有 1/3 的受訪者表示他們完全了解 AI 系統如何運作，因此需增加雇用了解 AI 的新員工¹³⁸。目前針對可解釋性 AI(XAI)是技術處理，開發能夠解釋與詮釋學習模型的 AI 演算法。本案例研究結果顯示，如果使用者高度參與 AI 模型的設計或是調校，包括自己做的，或是透過團隊做的，可以大幅提高 AI 的信任度或接受度，是除了技術方法之外舒緩「可解釋的 AI」的一種行政式操作方法，也是公部門是否能順利導入 AI 的重要關鍵。
10. AI 系統品質的評估方式：(1).基本上比人工好就採用；(2).減輕工作量；(3).科技接受度；(4).AI 的可解釋性等。
11. 如何緩解「正確真實 (accurate ground truth) 」訓練數據不易取得的問題：
 - (1).主題選取；(2).演算法模型的選取與持續調校。

¹³⁸ 檢自 <https://www.prnewswire.com/news-releases/artificial-intelligence-is-creating-jobs-dun--bradstreet-survey-finds-300774141.html>



參、其他案例

茲針對國家權力涉入程度高〔例如分配性政策和管制性，涉及標的人口利益的獲取與剝奪(羅清俊, 2020)〕、高情境相關性，挑選國外案例補充說明：

一、2018 年愛沙尼亞農業登記和資訊委員會(Agricultural Registers and Information Board, ARIB)開發 AI 圖像識別(image recognition) SATIKAS32 系統：

草地割草或放牧(mowing or grazing of the grasslands)是農業補貼政策主要條件，也是最容易違規的條件。原本是由人工進行檢查，由於人力不足，檢查比例僅 5-6%。面對不斷上升的勞動力成本和管制要求，愛沙尼亞使用 AI 檢測農業草原是否已修剪，以避免農民無法領到補貼。

在資料治理和演算法治理，SATIKAS32 系統使用包括：歐洲 COPERNICUS 計劃的衛星數據哨兵 1 號(Sentinel-1)和哨兵 2 號(Sentinel-2)光學衛星圖像、農田參考數據、歷史檢查日誌和愛沙尼亞氣象局氣象數據等資料，利用深度學習和卷積神經網絡(convolutional neural network)方法分析農業草原是否割過。

在公共行政改革，本案例將 AI 系統融入公共行政流程，起初許多人公務員擔心失去工作而對 SATIKA 系統持懷疑態度，在計畫初期的小規模試驗 (project pilots)和培訓之後，公務員知道該系統不是 100%可靠，而是必須將現有專業知識與 AI 建議相結合；現場檢查員也意識到，他們的工作不會因為 AI 系統而消失，而是會發生變化。利用 AI 有效率、正確找到符合規定補助對象，降低檢查員工作負荷同時提高檢查比例，使公務員對 AI 系統信任度增加。

在人機協作，公務員參考 AI 系統更有目標性(targeted)和數據驅動(data-driven)推薦結果，再根據自己的判斷力行使行政裁量權，使整個流程具備更高的可信度(林守德, 2021)。

本案例利用 AI 精準地找到補助對象，提升公民服務的品質，也同時提高政府的績效評估與營運管理。SATIKAS 自 2018 年以來已在農業登記和資訊委員會

(ARIB)實施，且系統仍在開發擴展其特性和功能，未來將用於識別不同類型農作物和樹木(Boyle et al., 2021; de Vries et al., 2018; Misuraca & van Noordt, 2020; van Noordt & Misuraca, 2022)。



二、2014 年比利時佛蘭芒兒童和家庭事務局(Flemish Agency for Child and Family)的 AI 預測系統(Predictive System)：

佛蘭芒兒童和家庭事務局與「福利、公共健康和家庭部的區域健康檢查組」(the Regional Health Care Inspection Unit of the Department of Welfare, Public Health and Family)合作，對兒童日托服務地點進行檢查，以提高日托服務品質及改善兒童的健康。然而，受限於檢查人力不足，佛蘭芒兒童和家庭事務局 2014 年開發的 AI 預測系統(Predictive System)，預測可能違規的日托服務地點。

在資料治理和演算法治理，兒童和家庭事務局與福利、公共衛生和家庭部(Department of Welfare, Public Health and Family)的數據科學團隊密切合作，以獲得該數據團隊文字探勘(Text Mining)的專業領域知識；以及與衛生保健檢查單位(Health Care Inspection Unit)合作，以獲得檢查相關數據。使用監督式機器學習(logistic regression 和 XGBoost)，分析醫療部門的各種內部和外部數據。

在公共行政改革，兒童和家庭事務局透過溝通，使公務員相信該 AI 系統的價值，特別是說服他們使用該模型是為了賦能(empower)公務員，增強他們的監督能力，而不是取代他們的專業知識或控制他們的工作。

在人機協作，AI 預測系統的推薦結果與現有員工經驗和專業知識相結合，有效率、準確的預測可能違規的日托服務地點。最終，AI 預測系統經統計證明是有效的(Validity)，該系統受到公務員的重視，進一步提高了終端用戶(end-users)公務員的科技接受度和支持度，並提高檢查量能和提升兒童福祉等(Bongers et al., 2018; Misuraca & van Noordt, 2020)。

從案例分析，公部門使用 AI 精進分配性政策(distributive policy)和管制性政策，「人機協作」的管理方式採人類「直接控制」，同時可以提高公部門績效評估與營運管理。



第二節 環境部「智慧判煙系統」



第二象限：智慧判煙系統(Intelligent Smoke Plume Identification System)，屬「國家機關權力涉入程度高」，偏管制性政策、偏機器決策。2024年1月1日正式實施，試驗或部分部署(piloting or partially deployed)，職能分類為F5：環保。本計畫目的是改善目前對空氣汙染違法事實，偵測方式公信力不足，同時解決政府人力不足，無法對工廠空氣汙染的行為進行長時間監控的治理困境。

空氣品質標準係指室外空氣中空氣污染物濃度限值，排放標準係指排放廢氣所容許混存各種空氣污染物之最高濃度、總量或單位原(物)料、燃料、產物之排放量。國際判煙方法主要是「目測判煙(Method 9)」、「影像判煙(ASTM7520)」：

1. 目測判煙(Method 9)

美國 1993 年使用，適用污染源：固定及移動污染源。主要是以人員肉眼，判定污染源的排放管道所排放黑煙或非黑煙、白煙、彩色煙的不透光率。

2. 影像判煙(ASTM7520)

美國 2016 年使用，適用之污染源：固定污染源。主要是使用數位影像及軟硬體並利用數位相機不透光率技術 (Digital Camera Opacity Technology, DCOT) 辨識煙霧不透光率。透過數位相機拍攝一組與背景對比之煙流之數位影像，取得對比值以判讀煙流不透光率。分析軟體會計算在固定時間內拍攝之一組煙流數位影像之不透光率並取平均值。



壹、案例分析—ANT 分析途徑與因果關係分析途徑分析結果

為防制空氣污染，依據空氣污染防制法第 20 條第 2 項規定訂定之「固定污染源空氣污染物排放標準」，其中粒狀污染物的排放標準（不透光率），若以目測判煙判斷，其標準為「不得超過不透光率 20%，停止、開始運轉時可到不透光率 40%，但一小時內超過不透光率 20%之累積時間不得超過 3 分鐘。」。

一、問題化

直轄市、縣(市)政府環境保護局稽查員的測量方式為目測判煙(Method 9)，並已經執行約 30 年，執行方式為：

1. 執法人員須經目測判煙訓練合格。
2. 仰賴稽查人員肉眼執行，並配合林格曼表判定。
3. 判煙結果須符合固定污染源空氣污染物排放標準。

然而，目測判煙的問題包括：

1. 目測判煙人員主觀判定易生爭議，導致訴願案件居高不下；

雖然目測判煙的稽查員都必須經過訓練，而且每年都要回訓，才能繼續做目測判煙的工作。

「目測判煙不是不準的問題，民眾比較不相信目測的結果，希望由儀器或機器來判定。因此目測判煙的行政救濟、訴願案子一直很多。」

—環境部環境保護司蔡孟裕司長(原大氣環境司長)

然而，實務上，目測判煙的誤差值可能達 $\pm 15\%$ ，

「目測判煙雖然需要經過目測判煙班訓練及考試，但是誤差值在 $\pm 15\%$

都會通過考試，一般目測判煙，老手幾乎沒有誤差，但生手、沒有常常在做目測判煙的人，誤差值比較大。」

—東海大學環境科學與工程學系陳鶴文教授



因而產生公信力的問題。

「我們在地方法執法的真的覺得奇怪，都什麼時代了還在目測判煙，但是中央的排放標準和檢測方式的法規沒有改，地方也只能照著做。」

—環境部大氣環境司蔡孟裕前司長(時任高雄市政府環境保護局局長)

2. 無法長時間持續監控：

實務上，工廠排放空氣污染物常常不是持續性的行為，所以工廠會利用環保稽查人員無法長時間監控的困境選擇偷排，

「常常民眾打電話到環保局檢舉，環保局稽查員就算 5 分鐘之內出門也合理，到達現場 20 分鐘，工廠排放空氣污染物的行為已經停止，現場空氣已經乾淨了，民眾就會抱怨：『你們怎來得這麼慢?』，稽查員解釋從環保局來到民眾陳情的地點 25 分鐘很合理，環保局稽查員只好開具『查無排放空氣污染物的事實』或『查無排放黑煙的事實』離開，等到環保局稽查員離開，還沒回到環保局，工廠又排放黑煙了，民眾又打電話給環保局，環保局稽查員到現場又排放完了。」

—環境部大氣環境司蔡孟裕前司長(時任高雄市政府環境保護局局長)

因此，民眾常常要求環保局稽查員到現場持續監控。然而，目前執行公權力

的環保局稽查員是常任的公務人員，除必須依據空氣污染防治法等環保法規執法，也須視案件狀況配合檢、調、警、環聯合稽查及 24 小時 3 班制方式輪值，工作辛苦離職率高。



「這樣的要求實務上有困難，環保稽查員人力有限。」

—環境部大氣環境司蔡孟裕前司長(時任高雄市政府環境保護局局長)

除了人力有限之外，人眼也無法重複一直看著影像：

「人眼平均每 30 秒看一次，5 分鐘看 10 次，目視判煙眼睛會疲勞。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

由於目測判煙已經 30-40 年了，張子敬前署長曾表示「從我當基層人員就是目測判煙，等到我當環境部長還是目測判煙，應該改一改了。」。為了讓環境稽查作業變得智慧化，增加執法效益，更重要的是提高公民對政府的信任。環境部大氣環境司找出強制性通過點(OPP)：「利用 AI 協助檢驗排放是否符合排放標準。」，並定義行動者身份和他們「想要(want)」的東西：

1. 環境部：解決目測判煙公信力的問題。
2. 廠商：處罰廠商要有具公信力的證據。
3. 民眾：持續監控有違法之虞的工廠。
4. 環保局稽查員：改善人力不足的問題。
5. AI：能判斷固定污染源空氣污染物排放標準。



二、利益賦予

環境部大氣環境司 2021 年利用「空氣污染防治基金科技研究計畫」對外徵求研究計畫¹³⁹，讓學術界研究成果與實務需求相互配合，其中具有人工智慧與資料探勘、環境工程、環境系統工業工程等領域背景的東海大學環境科學與工程學系陳鶴文教授提出人工智慧影像判煙科研計畫之可行性研究，剛好與大氣環境司的政策需求相符。賡續於 2022 年辦理「以人工智慧影像辨識技術建立黑/白煙不透光率即時量測系統」，進行硬體儀器設備、AI 判讀軟體技術開發、智慧判煙品質驗證機制，期望運用人工智慧影像辨識技術，即時並精準分析煙霧的不透光率，達到即時判斷出固定污染源的煙霧排放監測、管制狀況的目的。

智慧判煙系統程式主要分成兩大部分，1.從照片找到煙、2.判斷煙的濃度。目前試驗(pilot)成果正確率約 93%，介於目測判煙新手和老手之間：

「目前正確率約 93%，隨著測量濃度不同而有不同的偏差值，在某些排放濃度範圍內偏差值約 $\pm 0\%$ 。不成功的原因，很多是把背景白雲誤認為煙；或是同時有好幾根煙囪煙重疊。」

—東海大學環境科學與工程學系陳鶴文教授

再者，為緩解 AI 不可避免的偏誤(bias)，判定結果是採用統計檢定的概念，每 5 秒拍一張照片，6 分鐘拍了約 60-72 張照片，每一張都會有判斷結果，就會出現機率分布圖，換句話說，並非拍一張照片就開罰，而是以「1 群」照片結果降低誤差，使 AI 的判讀比目測判煙結果正確及穩定。

¹³⁹ 環署發表空污科學研究成果 人工智慧助目測判煙、各縣市減量模擬揭曉檢自 <https://e-info.org.tw/node/232712>

「目前測試結果，黑煙判識的偏差值約±7% - 7.5%之間，我們也在跟環境部討論多少的 AI 的正確率和誤差值範圍可以做為直接裁罰的依據。」

—東海大學環境科學與工程學系陳鶴文教授

智慧判煙系統取代目測判煙，具有避免主觀判定、可短時間稽查，長期監控的優勢之外，尚包括許多「技術上」的優勢：

- 1.判煙環境限制較少：(1).環境因子由設備監控；(2).可由系統選擇監控判煙標的。
- 2.程序簡單：(1).由系統自動化辨識，無視覺疲勞問題；(2).判煙步驟與程序簡化與電子化。
- 3.判煙結果穩定具有公信力：(1).AI 運算平均正確率較高；(2).有數位影像與紀錄可供佐證及查詢。
- 4.執法資訊透明化：環境部開發拍照判煙的方法、概念，廠商都可以自己開發軟體，只要能通過環境部的規格認證及系統驗證，現場考試通過就都可以使用，所以沒有開放 AI 系統原始碼的問題，符合透明化原則。

因此，藉由工業相機與鏡頭(煙霧影像拍攝)、網路攝影機(背景光譜資訊擷取)，可隨時將所欲監測工廠所排放固定污染源空氣污染物的照片傳回環保局，經過 AI 軟體(圈煙及辨識)計算之後就可以立即得知是否超過排放標準。因此，由相機取代人眼進行判煙，試驗結果已顯示以 AI 運算平均正確率較目測判煙高，也比目測判煙穩定，驗證文獻所述，由「主觀判定」修正為「科學判定」，可以減少爭議；更重要的是，AI「快」或「準」(Höchtel et al., 2016)的特質可以在短時間判斷是否超過排放標準。



三、徵召

由於環境部規劃智慧判煙目標是「直接開罰」，因此，AI 系統的判讀結果必須具有公信力：

1. AI 系統品質評估：

AI 系統可判定黑煙及白煙不透光率辨識，實務上，工廠所排放的固定污染源空氣污染物 90% 是黑煙，因此，AI 系統可以監測超過 90% 以上違規行為的固定污染源空氣污染物及範圍，技術上評估可行。

「運用智慧判煙系統準確度非常重要，我們目前測試的結果，黑煙不透光率實測結果誤差值均在+5%之內，且非常穩定，在 20%、40%、60% 的範圍甚至誤差值均近於+0%。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

白煙判定評估結果技術可行，惟仍須精進，所以目前仍在不斷的測試與校正階段。

「白煙不透光率實測結果，在大部分違規行為的範圍 20%—30%，誤差值也均在+5%之內，但超過 30% 誤差值則較大。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

2. 訂定標準量測方法

影響 AI 判煙結果的因素很多，環境部參照美國 ASTM 7520 及 method 9 標準方法(含儀器架設及軟硬體認定規範與程序)，訂定標準量測方法，內容包括：適用範圍、干擾因子、拍照系統內容、判煙程序、準確度及誤差等。再

由工業技術研究院量測技術發展中心建置驗證實驗室驗證，並採雙重把關機制，包括：

- (1).設備認證：「照度計—經濟部標準檢驗局」、「風速風向—交通部中央氣象局」、「GPS—財團法人全國認證基金會(TAF)」、「相機或攝影機—環境部指定驗證實驗室」；主要硬體設備，照度計應經國家度量衡標準實驗室(NML) 或 TAF 實驗室認證。
- (2).系統驗證：工業技術研究院量測技術發展中心針對 AI 系統調校與校正訂定 SOP，包括調校與校正的偏差範圍、次數和頻率。

「智慧判煙系統辨識模組校驗系統建置的方式是利用林格曼表¹⁴⁰的灰階標準來調校 AI 判讀的結果，經過統計分析找出容忍值(tolerance)，並考慮所有可能影響 AI 判讀的影響因素。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

驗證方式包括：

- (1).影像辨識軟體：於實驗室，以標準件林格曼表或煙霧表進行驗證室內驗證；
- (2).判煙系統：於判煙場所，以發煙機產生標準煙霧進行驗證戶外驗證。

「法規規範智慧判煙系統每 2 年必須校正及驗證，確保 AI 系統穩定」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

¹⁴⁰ 灰階標準：林格曼表→ 進化為煙霧表→ 影像判煙經由亮度對比計算不透光率，反射亮度則再追溯至 NML 國家反射率標準。



3. AI 判煙合法化：

利用 AI 取代公務員的判斷，為使地方環保局稽查員有 AI 執法的法遵依據，環境部修正「固定污染源空氣污染物排放標準」，修正重點：

- (1). 排放管道「目測判煙」修正為「目測及拍照判煙」
- (2). 拍照判煙方式適用所有行業別
- (3). 訂定 AI 調校與儀器校正的 SOP

環境部國家環境研究院主責量測方法(含儀器架設及軟硬體認定規範與程序)的訂定。採雙軌並行的執行策略：先從目測判煙和拍照判煙雙軌並行，長期是朝向拍照判煙的方式，以後就不再使用目測判煙。

4. 防弊措施與防干擾措施：

利用智慧判煙系統，拍照結果可以降低「人」的因素，從而避免以往稽查時可能發生的人為勾結。然而，人也可能利用修圖技巧影響拍照及 AI 判讀結果，防弊措施為不得修改拍攝影像資訊，且拍攝的影像具可追溯性的影像資訊檔(EXIF 檔)，原始影像資料(raw 及 jpg 檔)不得修改。防干擾措施為同時拍攝背景及煙流對比度，對比度不足例如能見度不佳時，不得進行智慧判煙。

5. 人員訓練：

環境部國家環境研究院增設智慧判煙系統訓練班，進行判煙人員 AI 技能訓練，環保稽查的工作內容和工作方法將和以前不一樣，變成如何利用 AI 工具更有效率的完成工作等。

「以後所需職能是 AI 系統產出結果的判讀(分析 AI 系統產出的數據)、AI 系統的儀器校正(調校與校正)、檢測(包括建置檢測方法)、驗證(包括建置驗證實驗室)等。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)



綜上分析，環境部規劃的配套措施是：「法規修正、軟硬體驗證、防弊措施、人員訓練」。

在資料治理以及演算法治理，環境部大氣環境司整合相當多的權責單位，包括東海大學(AI 技術廠商)、環檢所、工研院、標檢局、氣象局、TAF 財團法人全國認證基金會等，確保 AI 系統的判讀結果具有一定的正確率(accuracy)、AI 的偏誤(bias)範圍穩定，且具公信力的研發成果：

「我們已獲得臺灣的專利權：煙流不透光率數位影像辨識系統(Digital Optical Identification System using in Smoke Plume Opacity)」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

四、動員

智慧判煙系統「正向方案」的動員的因素分析如下：

1. 環境部：環保稽查單位利用「AI 官僚」，預期可以減少目測判煙人員主觀判定爭議，降低出錯率與環保稽查訴願案件數量，再者，透過 AI 科技執法長期監控，更可有效嚇阻有違法之虞的工廠，降低「負的外部性」¹⁴¹。

「針對民眾有疑慮的工廠，我就放一臺 AI 機器直接監控廠商，時間不用久，就持續監控 6 小時，就會有嚇阻效果。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

¹⁴¹ 行為人的行為會對不相關第三者產生負面影響，使第三者間接承擔成本。

- 
2. 環保局稽查員：使用目測判煙，一個稽查員同一時間只能看一個煙道；利用拍照及智慧判煙系統，一個稽查員可以同時看多支煙道，可以大幅提高稽查效率。再者，稽查人員可以領稽查獎金，預期可以大幅提高稽查人員使用 AI 工具的誘因，增加使用者(公務員)的科技接受度。

「稽查人員可以領稽查獎金，各地比例不同，額度不同。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

最重要的是，AI 可以 24 小時連續監測、節省環保稽查人力負荷，

「AI 不會累，人會累，環保稽查員的肉眼沒辦法持續看 6 小時。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

「我們很期待環境部的智慧判煙系統上路。」

—臺中市政府環境保護局空氣品質及噪音管制科

研究發現：AI 官僚也可以合法行使行政裁量權，且平均正確率、效率較人類官僚高。

本案例中可能進行「反向方案」的動員的因素分析如下：

1. 關鍵行動者，例如有違法之虞、可能被裁罰的工廠的演算法趨避：

環境部首先以行政規則規範 AI 判讀的結果在管制點附近不會開罰，例如當 AI 檢測結果達「(排放標準值+ AI 偏誤)的上限值」才會開罰，AI 的偏誤 (bias) 由國家或是法規吸收；或是 AI 判讀結果有疑慮也不會開罰，在於環保稽查

單位判斷工廠是否違反排放標準，並依照法律規定行政裁罰時，要有具公信力的證據：



「處罰廠商要有證據，這是必要條件，加上檢測的誤差，確保檢測一定超標，才會心服口服。排放標準值 20%，假設 AI 的正確率(accuracy)是 95%、且 AI 偏誤(bias) 的範圍「非常穩定」，例如在『+5%』之內，則 AI 測出來的結果在『25%』 (20% +5% = 25%)以上才會開罰。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

其次是「透明度」，裁罰證據的「再現性」，被裁罰之工廠可以申請調閱照片及 AI 系統自動決策結果，使整個裁決過程高度透明化以杜絕爭議：

「工廠若對 AI 裁罰有疑慮，環保稽查單位可以提供 AI 判讀結果的數據；或是提供照片，廠商也可以自行判斷或進一步驗證。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

對於工廠而言，是否能接受環境部的論述及作法？關鍵在於「公信力」。

「重點是『公信力』，而不是用什麼方法測，若 AI 的正確率夠高，例如 95%以上，會願意接受；如果環境部將願意可能的偏誤(bias)，例如 5%和標準值(20%)相加，則接受度會更高。若 AI 的正確率 93%，如果環境部也願意將可能的偏誤(bias)，例如 7%和標準值相加到 27%，也可以接受。」

—臺灣上市公司化工廠廠長



研究發現：設立風險值是避免演算法趨避和自動化偏誤的行政性作法。

2. 公部門缺乏 AI 技術等跨領域的問題：智慧判煙系統的開發，涉及環保、AI 技術知識，本案例是利用科技研究計畫導入學界的 AI 技術，再透過政府採購法導入具 AI 技術的廠商，再納入驗證實驗室等，組成產、官、學界的跨領域團隊。
3. AI 官僚的合法化：環境部依照空氣污染防制法第 20 條第 2 項授權，2023 年 3 月 16 日修正「固定污染源空氣污染物排放標準」第 2、3 條，新增「影像判煙：指檢查人員依中央主管機關規定之方法，以數位影像拍攝設備及辨識軟體進行固定污染源排放管道排放口廢氣不透光率之判定。」，使智慧判煙系統具合法性。
4. AI 決策的課責性：若工廠不服的「AI 官僚」的裁罰，可依據現行法制作業程序提起訴願，環保稽查單位必須提出 AI 系統通過驗證證明、AI 系統判讀結果等資料說明，因此，公民仍有救濟管道，權利不會受到侵害。
5. 經費來源：每一個 AI 裁罰工具成本約 50 至 76 萬元(硬體設備約 25 至 41 萬元、圈煙及辨識軟體約 25 至 35 萬元)，成本在可接受範圍，具執行效益，且未來環境部將補助或由地方環保機關自行購置。此外，空氣品質採用科技執法已受到諸多地方縣市首長支持，例如桃園市、苗栗縣、南投縣等，有高層支持，加上環境部經費補助，經費不是問題。



6. 政府採購法：智慧財產的採購都很難適用政府採購法。

「例如之前有些單位要求實報實銷，但我們是用頭腦、寫程式做決策管理，就幾張紙很難計價，加上 AI 技術差異很大，如果要再提高正確率，價格更難計算。我們做這個主要是使命的問題，希望環境變好、希望 AI 判煙也可以提供給工業區、工廠自主管理使用。」

—東海大學環境科學與工程學系陳鶴文教授

7. 政治可行性：地方政府是否願意接受 AI 系統判讀的結果，並直接開罰。

實務上，諸多直轄市、縣市政府已有使用 AI 監控空氣品質，目前礙於法規無法直接開罰，因此，具有政治可行性。舉例來說，臺中市政府運用智慧科技治理空氣品質，使用 AI 影像判讀，一旦偵測到異常排放情事後，透過 Line bot 推播(每 5 分鐘偵測)給指定人員，稽查人員可第一時間直接調閱影像資料、氣象資料、IOT 數據及告警事件相關資訊，即可立即派員現場稽查驗證，若查獲違規屬實再予以裁罰(臺中市政府, 2022)。

環境部與臺中市政府規劃內容最大的不同點，環境部的目標是將 AI 當成「直接裁罰的工具」，亦即「AI 決策」，而臺中市政府環保局是把 AI 當成「篩選工具」，所以是「人類決策」，主要原因是受限於「固定污染源空氣污染物排放標準」第 12 條目測判煙規定，依據中央的法規，AI 只能當作篩選工具。

「目前依法沒有辦法直接開罰，環境部有通盤檢視依據 AI 判讀結果可直接開罰的可能性，但會涉及法令的訂定或修正，環境部 2022/9/21 的業務檢討會中有提到他們正在研議，以後只要 AI 判斷出來工廠排黑煙就可據以處分，可取代目前稽查人員需受訓取得證書後執行目測判煙

的部分」

臺中市政府環境保護局空氣品質及噪音管制科



對於法規上目前無法利用 AI 演算型決策直接開罰，環境部表示：

「以往目測判煙產生的公信力問題，以 AI 直接開罰一定要有足夠的正確率和穩定偏誤(bias)範圍，我們花了 1-2 年的時間做 AI 調校與儀器校正，就是希望 AI 的判讀結果正確率高、穩定，具有公信力。」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

至於如何處理機器學習的偏誤(bias)，臺中市政府環保局的做法是「人類決策」：

「目前 AI 判斷出來的東西，原則上要經人員再次確認，依法不能單憑 AI 的資訊，直接據以裁罰。」

—臺中市政府環境保護局空氣品質及噪音管制科

即使 AI 只能當作篩選工具，仍然成效斐然，自 109 年 10 月上線至 111 年 7 月止，共偵測到 49 件異常排放，查獲違規屬實裁罰 10 件，共計裁罰 613 萬 5,200 元(臺中市政府, 2022)。苗栗縣政府環保局布建 500 臺空品微型感測器進行空氣品質監測，2022 年查獲 13 件次工廠排放汙染，共裁處 277 萬元罰鍰¹⁴²。

「有些縣市政府環保局對 AI 判煙很有興趣，也都在觀望，已經在期待

¹⁴² 檢自 <https://money.udn.com/money/story/5612/6652382>

環境部智慧判煙系統的法規和技術上路。」

—東海大學環境科學與工程學系陳鶴文教授



羅維(Theodore J. Lowi)和薩里斯堡(Robert.H. Salisbury)認為政策可依據「對社會的影響」，分為分配性(distributive)、管制性(regulative)、重分配性(redistributive)、自我管制性(self-regulative)政策(賴維堯, 1994)，其中自我管制政策的定義是權威當局並未制訂鉅細靡遺的規則，而是設定一些原則性規定，委由各政府機關與標的團體自行採取行動(羅清俊, 2020)。環境部政策朝向「好學生條款」，2022年6月6日修正發布「固定污染源自行或委託檢測及申報管理辦法」，以「量少、質精及有效管理」為重點，鼓勵業者做好自我管理，在降低污染排放量同時，給予適度調整定期檢測頻率之誘因，減少非必要檢測數量¹⁴³。因此，智慧判煙系統的終極目的不在裁罰，而是自主管理，彰顯釐清AI科技雙重用途的重要性：

「我們希望未來這套系統不僅為稽查工具，也是廠商自我監測，調整操作及製程參數的利器」

—環境部大氣環境司蔡孟裕司長(現環境保護司長)

研究發現：釐清AI科技雙重用途的重要性，例如AI系統的目的不在裁罰，而是自主管理。

因此，智慧判煙系統可以提供給業者做好自主管理、也可以節省稽查人力，

¹⁴³ 「量少、質精、有效管理」 環保署發布固定污染源定檢精進新措施。from：檢自 <https://air.epa.gov.tw/News/news.aspx?ID=1535>

同時即早發現或掌握污染行為，預期將可以大幅提升管制性政策的管制效率。



研究發現：AI 系統具可信任性，包括

1. 個人資料保護的「合理性」。
2. AI 系統產出資訊的「非歧視性」。
3. AI 系統產出資訊平均正確率、穩定度比人高的「公信力」。
4. AI 系統成本、效率比人高的「效能性」。
5. AI 演算法可解釋性(explainable/explainability)、可理解性(interpretable/interpretability)的「透明度」。
6. AI 行使行政裁量權的「合法性」。
7. AI 官僚直接決策的「可課責性」。
8. AI 官僚直接行政裁罰證據的「再現性」。
9. AI 雙重用途的「明確性」。
10. 利害關係人科技接受度的「政治可行性」。

貳、案例特色

環境部導入 AI，作為固定污染源排放空氣污染物超過排放標準時的裁罰工具，但 AI 系統仍必須依照法規進行 AI 調校與校正，以及環保單位的遠端監控。以公共行政的觀點，公務員何以願意釋出行政裁量權？智慧判煙可減少目測判煙人員主觀判定爭議、降低目測判煙之視覺疲勞，遠端監控可以節省人力，甚至針對有違法之虞的工廠進行長時間持續監控，對公務人員而言應有正向誘因。

研究發現：公務員何以願意釋出行政裁量權？增加公務員的公共治理能力。



因此，當 AI 的演算型決策取代公務員的行政裁量權，表示 AI 不再是簡單的 AI 程式，而是實際上是一類新的組織代理人或組織能動者(a new class of organisational agents)(Luca et al., 2022)。原本官僚制(bureaucracy)或「科層制」的層級節制原理，權責自上而下傳遞，科層化的管理方式：「人員的管理—減少人類主觀式偏差管理」，轉變成管理「AI 官僚的偏誤(bias)」。

接續針對本案例特色整理如下：

1. 「AI 官僚」的正當性：程序的正當性(due process)仍然是行政裁決的核心挑戰一，當 AI 的判讀結果直接成為裁罰工具時，必須符合行政法(administrative law)和程序公正原則(procedural fairness principles)(Henman, 2020)，以及問責制(Engstrom et al., 2020)。為避免 AI 的演算型決策沒有典型的審計、申訴和問責機制(Whittaker et al., 2018)，本研究個案透過修正「固定污染源空氣污染物排放標準」，使 AI 官僚具有合法性；其次，民眾不服機器決策結果，仍然可依據現行法制作業程序提起訴願，不會因為行政單位應用 AI 裁罰工具而受到公民權利的侵害。
2. AI 演算型決策直接作為行政裁罰依據：依目前 AI 技術的進程，諸多文獻指出 AI 不應單獨直接用於決策，而是篩選工具(screening tool) (Obermeyer et al., 2019)。實務上，若 AI 使用在和公民權利義務相關的領域時，AI 大多數是當作篩選的工具(例如 AI 選案查核等)。本案例係屬管制性政策工具，且目標是「直接開罰」，後續執行若能成功，對提升政府治理能力將有巨大潛力。
3. AI 系統產出資訊的「透明度」：環境部只訂一個最低規格，符合的設備，例如環境部規定 1 萬畫素，廠商設備如果是 2 萬，3 萬畫素，都可以用。AI 系統的目標是透明執法，成為稽察員稽查、廠商自我監測的工具，並非裁罰。
4. 申請專利：由政府機關主導研究計劃，與學術界共同研發 AI 模型，並取得專利：「煙流不透光率數位影像辨識系統(Digital Optical Identification System

using in Smoke Plume Opacity)」，後續將擴大授權加速加大政策擴散效果。

5. 重視 AI 系統的正確率：欲將 AI 系統判讀結果當成直接裁罰的依據，AI 系統判讀結果必須具有公信力，所以花了很多時間進行調校、驗證、申請專利，與大部分案例「基本上比人工好就採用」（例如智能稅務服務計畫）不同。
6. 降低人類主觀判定的爭議：AI 執法工具可以解決長久以來目測判煙的公信力爭議問題，甚至是更透明的執法機構(Engstrom et al., 2020)。
7. 實證避免演算法趨避的行政性作法：設立風險值，行政規則規範 AI 系統判讀結果在管制點附近，即 AI 可能偏誤(bias)範圍，不視為超過排放標準。
8. 工作效率與薪資制度關聯：稽查人員可以領稽查獎金，利用 AI 工具提高稽查效率可以增加稽查獎金，預期可提高終端使用者的科技接受度。

參、其他案例

茲針對國家權力涉入程度高(對人民的權利具有較強的干預效果)、高情境相關性，挑選國外案例補充說明：

一、澳大利亞聯邦政府社會服務部(Services Australia)「自動債務追繳系統(Automated Debt Raising and Recovery System)」

2016 年澳大利亞聯邦政府社會服務部(Services Australia)[前為：民政服務部(Department of Human Services, DHS)]社會福利聯絡中心(Australia's social security delivery agency, Centrelink)與澳大利亞稅務局(Australian Tax Office, ATO)合作推動「檢查和更新過去資訊 (the Check and Update Past Information, CUPI)」計畫，包括「自動債務追繳系統(Automated Debt Raising and Recovery System)」，該系統在社交媒體和主流媒體上被標記為「Robodebt」，所以又稱為「債務機器人(Robodebt)」。

2015 年以前，DHS 為確認是否有「超額付款(overpayment)」給接受社會福



利金(Social Security payments)的人，以人工逐案詳細審查及核對，DHS 和 ATO 系統個人收入的數據差異。Robodebt 具有判斷民眾是否溢領社會福利金、與民眾確認收入、發出債務通知，以及隨後債務追繳過程等自動化功能，大幅為節省人工成本、增加案件審查數量以及提高超額付款的追繳率，使必須追繳人數大幅增加。統計 2016 年至 2019 年受影響的人，52%無法再收到社會福利金補助，主要特徵是女性、35 歲以下的人。

然而，Robodebt 演算法未能依照法律規定執行(例如 ATO 計算年平均收入、Centrelink 計算二週平均收入，兩者概念不同不能直接比較)、違反行政程序相關規定(例如舉證責任由 DHS 移轉到 Centrelink 和受補助者)，使得 Robodebt 原本被視為技術管理工具，看似不屬於政治領域，卻因加劇社會不平等並懲罰最弱勢族群引起政治關注。

在媒體、社交媒體廣泛討論後，參議院介入調查。Gordon Legal 律師事務所 2019 年 9 月為受 Robodebt 影響的人提起集體訴訟，挑戰 Robodebt 系統的法律基礎¹⁴⁴。聯邦法院裁定要求政府使用人工重新計算超過 500,000 筆個人債務，2019 年聯邦法院宣布債務「無效」，政府承認 Robodebt 是「非法」¹⁴⁵。2020 年 11 月集體訴訟宣布和解，這是澳大利亞歷史上最大的集體訴訟和解案，政府支出費用總額為 12 億美元，其中包括向 373,000 人退還 7.21 億美元、1.12 億美元賠償金和取消 3.98 億美元債務，以上費用並沒有包括重新計算超額付款所衍生的任何行政成本，CUPI 不僅未能達成節省，而且破壞社會保障體系的完整性(Whiteford, 2021)。

¹⁴⁴ 檢自 <https://www.theguardian.com/australia-news/2019/sep/17/robodebt-class-action-shorten-unveils-david-and-goliath-legal-battle-into-centrelink-scheme>

¹⁴⁵ 檢自 <https://www.theguardian.com/australia-news/2019/nov/27/government-admits-robodebt-was-unlawful-as-it-settles-legal-challenge>



二、荷蘭政府 2014 年開發 SyRi (Systeem Risico Indicatie)系統

荷蘭政府 2014 年開發 SyRi (Systeem Risico Indicatie)系統，利用現有政府系統中的各種風險指標：稅收、健康保險、居住、教育等，檢測詐欺或濫用社會福利資源的高風險地址，同時以法律明定政府可以蒐集、存儲和共享哪種數據。

在「問題化」階段，SyRi 核心目標是解決福利欺詐、減少公共資金濫用、減少管理成本、改善社區安全等，荷蘭政府找出強制性通過點(OPP)：「SyRi 系統是否可以解決福利欺詐的問題？」。在「利益賦予(interessement)」階段，由於政府對 SyRi 的解釋很少，例如，政府未公開 SyRi 風險模型及其指標、使用什麼 AI 技術(例如機器學習)、公務員使用自動決策結果的程度，數據主體(公民)不知道自己的資料是否被蒐集。再者，原本政策設計該系統不會做出最終決定，而是推薦公務員進一步查核的資料，但由於數據資料過時未能及時更新、帶有歧視風險的數據資料來源，使得 SyRi 系統推薦的結果通常是特定的、不發達和較貧窮的地區、推薦成功率很低，不符合成本效益(SyRi 的成本估計每年超過 325,000 歐元)。在「徵召」階段，由於 SyRi 系統缺乏透明度，受影響的人無法了解數據及推薦依據，引發各種組織反對使用該系統，理由包括：侵犯太多隱私權，對貧困和脆弱公民具有歧視性等，也引發了荷蘭社會的廣泛辯論。在「動員」階段，SyRi 的代表性受到質疑，包括受到貧困和脆弱公民的拒絕等；甚至遭到「異議」：「聯合國人權事務報告員對使用 SyRi 表達對人權構成重大威脅的關切」。在四個時刻結束時無法建立了一個有約束力的關係網絡，SyRi 轉譯的結果失敗。最後，基於應用 AI 技術的好處與對數據主體隱私權的潛在干擾之間應取得適當的平衡，以及可能產生的歧視風險，荷蘭法院於 2020 年初裁定 SyRi 使用不符合歐洲人權公約(European Convention on Human Rights, ECHR)第 8 條規定，防止欺詐的集體經濟利益不足以與隱私的社會利益權衡，SyRi 透明度和可驗證性(verifiable)不足，荷蘭政府表示不會對該判決提出上訴後已取消 SyRi 的使用



(Misuraca & van Noordt, 2020; van Bekkum & Borgesius, 2021)。

三、2012 年波蘭勞動和社會政策部(Polish Ministry of Labour and Social Policy, MLSP)的「AI 失業評分系統(unemployment scoring system)」

MLSP 面對 340 個勞動辦公室(PUP-Powiatowe Urzędy Pracy)低效率、人員不足、無法適應現代勞動力市場、無法大幅增加預算等挑戰。PUPs 找出強制性通過點(OPP)：「AI 失業評分系統提供個人化就業服務的建議」。AI 失業評分系統依據初次面試收集的數據(例如年齡，性別，身心障礙者和失業期間)以及電腦測驗結果分類失業者，並提供對應的安置計畫(例如就業安置、職業訓練、學徒制、積極性津貼)。

然而，AI 失業評分系統受到內部和外部組織的批評。在內部組織，原本失業評分系統是標準化決策的建議工具，在 AI 計算出失業者個人資料後，就業諮詢人員對失業分類結果有最終決定權，可以接受或拒絕 AI 系統決定。然而，即使 44% 的就業中心認為 AI 系統對他們的工作毫無用處，80% 的人認為應該改變 AI 系統，**但非常弔詭的是，諮詢人員超過 99% 依據失業評分系統的推薦分類，AI 系統成為實質上的決策者。**究其原因，包括諮詢人員相信 AI 系統的演算法、員工沒有時間詳細考慮決策、擔心如果沒有依照 AI 系統決策執行會受到質疑、害怕上司影響等。

在外部組織，由於 AI 系統演算法不透明，失業者不知道自己的分數和分數如何算出來，也無權了解失業評分系統的決策依據(例如 AI 設計邏輯、數據來源以及這些數據如何影響決策)，向行政法院投訴分類不公正。波蘭數據保護機構、公務員、法院、人權專員、民間社會組織和公民等也提出質疑。

最高審計署 (Najwyższa Izba Kontroli)對失業評分系統分析機制進行全面審查，提

出 AI 系統的無效性及可能導致歧視¹⁴⁶。最後，人權事務專員就程序問題向憲法法庭提出正式申訴，2018 年底，憲法法庭裁定 AI 失業評分系統違反波蘭憲法，2019 年 6 月 14 日該系統已被政府正式拆除(Misuraca & van Noordt, 2020; Niklas, 2019)。

綜上分析，當開發出一種演算法來進行敏感性分類時，顯然不應該直接依據 AI 的決策，而該演算法應該嵌入到社會技術系統中，具有人工評估、推翻、定期檢查、定期更新的程序(van Bruxvoort & van Keulen, 2021)；或是篩選工具(screening tool) (Obermeyer et al., 2019)，並應由人類「直接控制」。然而，此 3 個案例卻使用「機器決策」，亦即 AI 演算型決策直接對對標的團體(target group)管制或干預，侵害社會最弱勢群體的公民權利，最後導致 AI 系統均取消使用。

因此，若要善用 AI 比人「快(faster)」與「準(accuracy)」的技術優勢進行管制性政策，政府職能必須是「任務所需脈絡分析程度」低，「人機協作」採用「間接控制」，才能提升政府管制效率。

¹⁴⁶ 根據評分規則，女性的評估方式與男性不同，屬於社會最弱勢群體（單身母親、身心障礙者和農村居民）獲得援助的機會減少。From 檢自 <https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/>

第三節 臺北市政府「AI 智慧交控管理」



本計畫屬第三象限「國家機關權力涉入程度低」，偏象徵和勸告型、偏機器決策、完全部署(fully deployed)、職能分類為F2：公共秩序與安全。「AI 智慧交控管理」。本計畫的目是解決當交通量變化大時，人力無法即時調整定時號誌(Time of Day, TOD)的治理困境，改善塞車情形及確保行人安全。

壹、案例分析—ANT 分析途徑與因果關係分析途徑分析結果

臺灣 95%以上路口採用定時號誌，依照過去的尖、離峰交通量將一天分成若干時段，每天週期性地執行固定時制計畫。然而，臺北市因路口幹支道流量差異懸殊，以及臺北市路幅較寬之主要幹道，因顧及行人安全穿越路口所需時間規劃較長之綠燈秒數，或支道行車通過路口綠燈秒數等需求，以致於在離峰且支道人車較少時段(如深夜)，常發生沒有行人或行車穿越時幹道車流影響路口疏解效率之情形。實務上，都市交通業務單位不定期檢討調整既有時制計畫，且在重要瓶頸路口或是交通特性改變時，進行號誌時制重整(劉瑞麟、王耀鐸、郭豐璋、鄭柏舜、董尚義、劉德添、陳學台、彭振聲, 2022)。

雖然交通控制中心可自動偵測並觀測現場路況監視器(CCTV)，若無裝設 CCTV 路口則由民眾主動反應通報，交通勤務單位將會引導現場車流動線，或手動調整增加相應方向的綠燈時間，待壅塞情況稍微舒緩，將綠燈時間恢復至定時號誌狀態，這樣的運作方式既浪費人力亦無效率。再者，當交通量變化大時，以人力處理，無法調整單一或多個路口的定時號誌的需求，因此，定時號誌(Time of Day, TOD)的時制策略無法因應短時間交通服務需要。

一、問題化：

交通壅塞及衍伸之空污排放、駕駛人旅行時間浪費，是都市交通相關業務單

位必須面臨的課題，另一個極端是，於白天時段雖然支道行人與行車較多，夜間時段由於人車較少，因此以往較容易發生紅燈空等之行為：



「柯市長說，現在推動智慧城市，在離峰或深夜時，明明沒有人、沒有車，卻要空等交通號誌變綠燈才能通行，能不能改善紅燈空等的情況？」

—臺北市交通管制工程處交通控制中心

臺北市政府前市長柯文哲在主持市府交通會報時表示：「交通 AI 化是必然過程。」¹⁴⁷，行人的安全更是重中之重，並指示「對症下藥，不再讓行人挨撞」¹⁴⁸。臺北市交通管制工程處開始規劃 AI 智慧交控管理，包含智慧控制中心、智慧控制器、事件自動偵測以及智慧號誌等：

「利用 AI 技術來解決既有時制的問題，是交通控制中心的同仁提出來的構想，一開始是為解決市長提出紅燈空等的問題。」

—臺北市交通管制工程處交通控制中心

臺北市政府交通管制工程處找出強制性通過點(OPP)：「利用 AI 改善時制策

¹⁴⁷ 主持交通會報 柯文哲:交通 AI 化是必然過程 要加速推動智慧號誌的建置。

from：檢自

https://sec.gov.taipei/News_Content.aspx?n=49B4C3242CB7658C&sms=72544237BBE4C5F6&s=C6FDF437F852F5E9

¹⁴⁸ [對症下藥，不再讓行人挨撞]。from：檢自

<https://www.facebook.com/DoctorKoWJ/photos/a.136856586416330/1717506818351291/?type=3>



略」，並定義行動者身份和他們「想要(want)」的東西：

1. 柯文哲前市長：改善紅燈空等的問題。
2. 臺北市交通管制工程處交通控制中心：利用 AI 改善時制策略。
3. 行人：安全。例如部分支道人、車於綠燈秒數快結束前進入路口，此時剩餘秒數不足以提供人、車安全通過路口，可能發生進入紅燈時，人、車仍未安全通過路口。

「臺灣 2025 年進入超高齡社會，老年人過馬路的時間可能需要比一般人更長，利用 AI 技術偵測斑馬線上仍有行人通行，支道可以自動延長 10 秒鐘，確保行人行車通過路口之安全。」

—臺北市交通管制工程處交通控制中心

4. 駕駛：節省旅行時間。
5. 環保團體：減少整體車流延滯，可減少油耗與 CO₂ 排放。
6. AI：改善時制策略。

二、利益賦予

臺北市長前市長柯文哲主持 2021 年市府交通會報，交通管制工程處報告「2020 年臺北市智慧號誌執行情形」時表示：「臺北市交通要再進步，應該要進入 AI 時代。」。如經評估具建置效益的路口就要加速辦理，至於專業問題，建議與其直接委託廠商規劃，不如邀集此領域之專家學者、各大學教授研議，直接「買服務」；此外，美國等先進國家一定有現成的交通 AI 系統可參考。

交通管制工程處表示智慧號誌建設重點不在硬體設備，而在於交通問題的解析，此有賴有經驗之工程師來做分析。臺北市車流中含有很多機車，若要引進國

外系統，需要有段外國系統轉化在地化分析的過程，且若以國外系統執行，後續在修正維護及臺灣本地系統擴展性上都會出現疑慮，不過若要以臺灣廠商用臺灣系統執行，當前碰到的問題是遇國內廠商量能不足，完成期程也會拖得太長。

此外，交通控制中心擁有很多監控設備和大數據，在資料治理部分沒有問題，至於演算法治理的部分則需要進一步導入有 AI 技術專長的外部單位或資源，因此，交通控制中心參考國外建製智慧號誌的案例，開始著手規劃設計，並找廠商一起規劃設計，北市交控中心主任王耀鐸表示：「除參考國外經驗外，還須納入台灣本土特色。」¹⁴⁹，

「現在 AI 技術太新、太快，我又是交通背景，所以要找廠商直接提供技術上的意見，跟有 AI 技術背景的廠商一起合作規劃，組成跨領域的合作團隊，一起來規劃臺北市 AI 智慧交控管理。」

—臺北市交通管制工程處交通控制中心

研究發現：領域知識者參與 AI 模型開發提高 AI 專案成功率。

AI 演算法的設計如下：

1. PaSO Plus¹⁵⁰動態號誌控制系統：

(1). PaSO Plus 為採用 PaSO 模擬軟體為開發基礎工具，結合旅次起迄樹之動態

¹⁴⁹ 檢自 <https://www.chinatimes.com/realtimenews/20210730004080-260421?chdtv>

¹⁵⁰ 臺灣號誌控制軟體 (PaSO) 示範驗證與推廣計畫，from：檢自 https://books.google.com.tw/books/about/%E5%8F%B0%E7%81%A3%E8%99%9F%E8%AA%8C%E6%8E%A7%E5%88%B6%E8%BB%9F%E9%AB%94_PaSO_%E7%A4%BA%E7%AF%84%E9%A9%97.html?hl=zh-TW&id=Pl0BEAAAQBAJ&redir_esc=y



控制策略模組，演算目標為對路徑所需綠燈時間之最佳化，確保群組路口主要車流路徑皆順暢通行，並搭配 AI 影像辨識偵測器，蒐集旅行時間及偵測關鍵路口之停等延滯與停等車輛數。PaSO Plus 系統依據即時偵測結果及不同交通情境下風險值，計算(約 1 分鐘以內)最佳時制計畫。

- (2).為驗證 PaSO Plus 動態號誌控制邏輯策略之有效性，透過實際現場調查與蒐集系統上線(2021/10/13)前後 1 個月之資料分析顯示，交通壅塞越嚴重，改善績效會更加顯著，有利於駕駛的節省旅行時間。

2.感應性號誌控制系統：

- (1).現場端辨識、運算與控制，於支道近端號誌桿件安裝 AI 影像辨識偵測器，橫交幹道行穿線端點佈設行人觸動延長按鈕，透過 AI 影像偵測及觸動按鈕蒐集支道欲通過路口之人車狀態，取得路口時制即時運作資訊(以 RS232 方式傳輸)，依照資料蒐集與運算結果，採用步階控制方式即時調整現場號誌時制計畫，增加路口運作彈性，減少幹道車輛紅燈空等狀況。另為配合行人闖紅燈違規警示功能，於對向桿件安裝戶外 LED 可變標示牌及與同向號誌桿件安裝戶外告警喇叭，提供給違規行人聲、光之警示提醒。
- (2).臺北市於 2019、2020 年分別於 8 個行政區共 23 處路口導入感應性號誌控制系統，透過半觸動控制邏輯，AI 影像辨識偵測器與行人觸動延長按鈕偵測支道行人、行車之狀態，即時調整路口號誌，於白天離峰或夜間支道人車較少之時段效果尤其顯著。此外，伴隨高齡化趨勢，行人通過綠燈秒數不足亦或是行人闖紅燈所衍生交通肇事有日益增加，藉由已建置之 AI 偵測設備偵測行人闖紅燈情形，並連動告警牌面與聲響進行即時告警，達到遏止違規行為、降低潛在交通肇事之風險。



3. 事件告警系統

- (1).以攝影機影像結合 AI 影像辨識技術與交通事件判斷邏輯，即時進行道路事件自動化偵測並發報通知；以 AI 偵測器計算平均畫面車速，每 5 分鐘產出壅塞分析結果。
- (2).常見之道路交通事件包含交通事故及交通壅塞，而交通事件資訊內容時常影響用路人對交通路徑之選擇。目前事件偵測資料已串接至交控中心資料庫，並透過臺北市交通控制平台，呈現相關資訊及推播告警資訊，減少值班人員查看道路 CCTV 的繁瑣流程，由系統主動告警，可立即通知相關單位及早反應，縮短現場處理時間，且可於上下游道路之電子看板(CMS)即時發布訊息，提醒用路人提早判斷行進路線並修正。因此，事件告警系統有利於駕駛的節省旅行時間、縮短事件現場處理時間、減少值班人員查看道路 CCTV 的繁瑣流程，降低交通控制中心「既有時制調整」的人力負荷。

三、徵召

臺北市政府 2019 年交控管理導入 AI，一開始是先選擇重點區域路廊與部分路口試辦導入智慧號誌：

「利用 AI 技術解決既有時制問題是很新的做法，**我們無法確保會成功，於是運用『試辦』，由示範地點與壅塞區域逐步拓展。**」

—臺北市交通管制工程處交通控制中心

試辦的結果如何呢？由於 AI 依賴主要路口所設置「動態號誌」收集車輛偵測器的即時流量、智慧影像攝影機的即時流量及停等車隊長度資料、AI 影像偵測



支道有無行人及行車等¹⁵¹數據，以內部反覆迭代運算構建最佳化號誌時制。因此，外在環境的因素，例如戶外天氣、陽光，造成物件辨識的損失等情況的資料偏誤(bias)，則 AI 生成預測的模型可能產生偏誤(bias)。

「目前 AI 系統相當穩定，正確率約 9 成，會影響 AI 正確率的另外 1 成主要是外在環境因素，例如戶外天氣、陽光，造成物件辨識的 loss，這個真的要遇到的時候，或有重大異常通報的時候才會把有偏誤(bias)的錄影時段擷取下來分析，進一步微調(fine-tune) AI，否則就會維持現在的版本。」

—資拓宏宇國際股份有限公司有限公司

在交控管理，重要的是趨勢，且每提高 1% 的正確率，要多花更多的時間、更多的成本訓練 AI，因此通常不需要百分之百精準。

「AI 影像辨識技術要辨識『物件』，例如辨識車子來計算車流數量、速度，加上政府的標案都會有專案的時程限制，因此正確率約 9 成是**考慮『效率、時間、成本』**的結果。」

—資拓宏宇國際股份有限公司

「臺北市交通控制中心一開始訓練 AI 時，準確度(accuracy)約 90%，我們逐步訓練機器人讓準確度(accuracy) 提高，目前白天約 96%、晚上約

¹⁵¹ 智慧化號誌實施成效良好，行車更有效率！檢自

https://www.dot.gov.taipei/News_Content.aspx?n=D739A9F6B5C0AB95&sms=72544237BBE4C5F6&s=9C8FFE863D70D15B

92%，因為 AI 現在很穩定，也不太需要再訓練。」

—臺北市交通管制工程處交通控制中心



在 AI 具有「快」與「準」的技術特性中，本研究個案顯然「快」比「準」更有價值：

「用哪種架構主要是考慮業主(機關)需求，所需的即時處理性有多高，以及機關頻寬的穩定度及品質，交通控制即時性非常重要，目前是採用邊緣運算(Edge Computing)。」

—資拓宏宇國際股份有限公司

AI 官僚蒐集人類資訊，透過 AI 符號(例如 AI 邊緣運算結果的資訊)代表人類表達具有法律約束力的行政行為(例如交通號誌)，使人類改變行為，AI 官僚再蒐集人類資訊(例如人類移動資訊)。因此，公共政策過程的並非一個完全線性的過程(官美吟、彭立沛, 2018)，而是循環交錯、動態互動的過程，形成一個共同網絡的集體(collective)。

研究發現：人不能片面影響 AI，AI 也不能片面影響人的行為，人與 AI 是彼此互相影響；人與 AI 工作不斷地交互轉移、循環交錯，交引纏繞的過程。

智慧交控管理幾乎直接使用 AI 演算型決策，而 AI 偏誤(bias) 的管理，臺北市交通管制工程處是使用『配套措施』。

「因為車子物體比較大，比較不會漏偵測，偏誤(bias) 可能原因是行人衣服、角度、雨、天候、光線等影響，可能會漏偵測，『**配套**』**很重要，因為交通安全非常重要。**」



—臺北市交通管制工程處交通控制中心

相關配套包括：

1. 風險值：為避免 AI 偵測誤判，設計最大幹道綠燈風險值(預設 7 分鐘)，幹道綠燈執行時間大於等於最大幹道時間時，不進行幹道綠燈延長且執行支道行人最短綠燈。
2. 降階運作¹⁵²：若 AI 自動決策系統 3-5 分鐘沒有回傳訊息，系統立即直接切換到定時號誌。
3. 行人觸動號誌。
4. 每日審查：交通控制中心每日審查 AI 自動決策系統是否正常運作。
5. 人類監督：交通控制中心透過交通路況監控錄影監視系統隨時監控 AI。

研究發現：AI 系統的品質是考慮『效率、時間、成本』，再搭配 AI 風險緩解機制(配套措施)。

四、動員

在本案例中可能進行「正向方案」的動員的因素分析如下：

1. 首長支持，有充足經費來源：臺北市是以建置完成的路口數為 KPI，並逐步擴大建置，一個交通路口所需的感應性號誌控制系統約 1 百多萬，除了臺北市

¹⁵² 亦有使用逾時機制(Timeout)(資訊領域)、降階控制(Reduced-Order Control)(電機領域)、保護性停止(protective stop)(職業安全衛生領域)等。

政府的自籌款，亦獲得交通部「智慧運輸系統發展建設計畫」¹⁵³的經費挹注。



2. 公民支持：民眾感受交通通行效率提升。

「有 AI 交控管理後，號誌有連線，會隨著交通的流量變化而改變。」

—每天上班經過臺北市的新北市民

若再從另一個面向觀察，由於陳情服務係重要民意表達途徑，亦為人民與政府間的主要溝通工具之一(汪林玲, 2014)，在交通管理面向，主要是以負面管理，降低民眾陳情為主：

「民眾不會知道有沒有用 AI 做交通控制，只要塞車就會來陳情。」

—臺北市交通管制工程處交通控制中心

「剛開始導入 AI 時因為會漏偵測，會有少量民眾等綠燈等了很久陳情，隨著訓練資料量越來越大，AI 正確率逐步提升，到現在幾乎沒有接到漏偵測而陳情。」

—臺北市交通管制工程處交通控制中心

另參照 2022 年第 1 次(4 月)臺北市交通民意調查報告顯示整體而言，民眾對臺北市交通狀況之評價以傾向滿意居多，占 74.2%，相較去年調查增加 5.3 個百

¹⁵³ 智慧運輸系統發展建設計畫—改變未來交通移動力。from：檢自
<https://www.ey.gov.tw/Page/448DE008087A1971/b61ca63b-b210-4f6c-9718-c40316c6a65a>

分點；傾向不滿意比率占 20.8%，較去年同期調查減少 3.4 個百分點¹⁵⁴；以及交通部「智慧運輸系統發展建設計畫(110 至 113 年)」的報告：臺北市重點路口肇事率減少約 43%¹⁵⁵等；雖然交通事項眾多，對交通狀況的評價直接推論是 AI 智慧交通控制的因素太過勉強，但仍能看出一些端倪。

在本案例中可能進行「反向方案」的動員的因素分析如下：

1. 公部門缺乏 AI 技術等跨領域的問題：公部門要導入 AI 是跨領域的專長，對主要承辦人員和直屬主管來說真的很辛苦。

「利用 AI 技術來解決既有時制的問題涉及跨領域專業，規畫過程中遇到很多問題，但是交通管制工程處有交通管理、資訊工程師、電機等不同工程背景的人，都可以請教與詢問，也都給予很多專業上的協助。」

—臺北市交通管制工程處交通控制中心

2. 工作負荷：由於本案例屬定時號誌智慧化，情境無關且樣本多、情境相關性程度低，因此無較無涉公共行政改革、公平性、合法性和政治可行性、倫理、隱私和數據匿名、工作權等問題的挑戰。接著就必須得到使用者(公務員)的支持，有關公務員部分，主要的困境是工作負荷。雖然從長期來看，AI 系統穩定後可以節省人力，但是規劃期間產生額外工作負荷的組織管理問題，

¹⁵⁴ 交通民意調查報告。From：檢自

<https://www.dot.gov.taipei/News.aspx?n=ACFB485A398A1BEB&sms=BF42F24130B56DF1>

¹⁵⁵ 檢自 <https://www.ey.gov.tw/Page/448DE008087A1971/b61ca63b-b210-4f6c-9718-c40316c6a65a>

是反對導入 AI 的重要因素。



「表面上雖然減少值班人員查看道路 CCTV 的繁瑣流程、既有時制調整的人力負荷，但為了規劃和拓展 AI 所增加的業務更多，原來承辦人的業務必須分給其他人來做，所以除了原來主辦 AI 業務的承辦人每天都在加班，所有的人工作負荷都加重了，繁重的工作負荷難免讓同仁有怨言。」

—臺北市交通管制工程處交通控制中心

3. 採購制度：現有的採購制度也有很多窒礙難行之處。

「AI 技術持續演進，不是一個成熟的技術，政府的招標方式或採購制度必須更加彈性、包容性，應該以創新技術的角度去衡量。公部門在徵選廠商時，第一步必須篩選廠商 AI 的技術能力、實績證明；第二步廠商必須敘明技術架構，提出「概念驗證(Proof Of Concept, POC)」，驗證廠商的『概念』或『理論』是否可以正確執行新產物設計，藉以評估廠商的 AI 技術能力；而廠商實績可以知道廠商有這方面的產物開發經驗，已經有建置好的模型以及類似場景的相關資料，例如廠商原來的模型正確率(accuracy)為 60%，再利用公部門特殊場域的資料再提高至 90%，如果廠商沒有這方面的經驗就會變成新產物開發，很容易失敗，而目前的評選方式可能無法很快判斷廠商的 AI 技術能力與實績。」

—資拓宏宇國際股份有限公司

在採購策略，由小做起，先概念驗證(POC)再逐步擴展軟硬體的投资，可以提



高公部門使用 AI 的成功率。

因此，雖然有「反向方案」的動員，例如少數民眾陳情、規劃、內部工作負荷、採購制度等推動障礙，由於試辦成效斐然，持續擴大辦理，2019 年至 2022 年利用 AI 影像辨識技術建置 40 處智慧「感應性號誌」、103 處「動態號誌」¹⁵⁶，未來將針對關鍵路口、區域範圍進行 2 階段擴大建置，完成建置 1480 處的目標^{157 158}。

研究發現：由小做起，先試辦再逐步擴展，或先概念驗證(POC)再逐步擴展軟硬體的投资，提高公部門應用 AI 的成功率。

臺北市自 2019 年於重點區域路廊與部分路口導入智慧號誌，2020 年於市民高架建置 AI 交通事件偵測系統，結合 AI 辨識技術偵測行人與行車等物件，其成效如下：行車時間改善 7-15%、路口延滯減少 10-20%、部分路口夜間空等紅燈時段降低 32%及縮短 9 分鐘事件現場處理時間，每年節能減碳約近 1 億元之效益。

¹⁵⁶ 檢自

<https://tw.news.yahoo.com/news/%E8%87%BA%E5%8C%97%E5%BB%BA%E7%BD%AE%E6%99%BA%E6%85%A7%E8%99%9F%E8%AA%8C%E8%88%87%E5%84%AA%E5%85%88%E8%99%9F%E8%AA%8C-200009696.html>

¹⁵⁷ 檢自 <https://www.chinatimes.com/realtimenews/20210730004080-260421?chdtv>

¹⁵⁸ 檢自 <https://www->

[ws.gov.taipei/Download.ashx?u=LzAwMS9VcGxvYWQvMzkwL3JlbGZpbGUvMTk2OTQvODQyNzEzMi9iOWFhNzZhNi03MGZjLTRmZGYtOGMxOS1jOTVjMWZhMTMwMjgucGRm&n=MTEwMDczMOS6pOmAmuacg%2BWgseacg%2BitsOizh%2BawmV8xMTDlubTnrKw25qyh5Lqk6YCa5pyD5aCx5pyD6K2w6LOH5paZLnBkZg%3D&icon=.pdf](https://www.gov.taipei/Download.ashx?u=LzAwMS9VcGxvYWQvMzkwL3JlbGZpbGUvMTk2OTQvODQyNzEzMi9iOWFhNzZhNi03MGZjLTRmZGYtOGMxOS1jOTVjMWZhMTMwMjgucGRm&n=MTEwMDczMOS6pOmAmuacg%2BWgseacg%2BitsOizh%2BawmV8xMTDlubTnrKw25qyh5Lqk6YCa5pyD5aCx5pyD6K2w6LOH5paZLnBkZg%3D&icon=.pdf)



2021 年於 29 處路口實施動態號誌及調整時制，改善區域路廊因交通變異大而常過飽和的交通壅塞問題。另於 12 處路口實施感應性號誌控制技術，結合 AI 影像辨識即時偵測支道行人行車到達與停等狀態，即時調節綠燈秒數，不僅能降低幹道紅燈空等情形與提升路口常綠續進效果，更可彈性地確保行人行車通過路口之安全，達到行人行車效率兼顧的智慧控制目標。

貳、案例特色

接續針對本案例特色整理如下：

1. 加速公部門導入 AI 的效率及成功率：採購市場上 AI 智慧號誌控制系統，使用公部門特殊場域的資料進行演算法調校(the fine-tuning of the algorithm)以提高正確率(accuracy)，將市場上模型正確率(例如 60%)提高至 92%-96%。

研究發現：採購市場上成熟的 AI 模型，再使用公部門特殊場域的資料進行演算法調校以提高正確率(accuracy)。

2. 完善的配套措施：依序採用(1).風險值、(2).降階運作、(3).行人觸動號誌、(4).每日審查、(5).人類監督等配套措施，避免 AI 偏誤(bias)，確保交通安全。
3. AI 系統品質評估：在交控管理重要的是趨勢，權衡『效率、時間、成本』，可以在 AI 的正確率與成本效益取得一個平衡，大幅提升行政效率與效能。

參、其他案例

茲針對國家權力幾乎不涉入(例如「政策工具光譜」屬象徵和勸告型)、低情境相關性，挑選國外案例補充說明：

英國脫歐公投(Brexit referendum)社交機器人(Social bots)

2016年6月23日英國通過公投決定離開歐盟，脫歐派(the Leave camp)(希望英國脫離歐盟陣營)以52%打敗留歐派(the Remain camp)(希望英國繼續留在歐盟陣營)48%，出人意料是大部分民意調查¹⁵⁹均未能預測英國脫歐投票結果，例如歐盟2016年6月16日至22日期間進行公投民意調查結果¹⁶⁰留歐52%、脫歐48%，公投結果對於脫歐派和留歐派都感到意外(Bouko & Garcia, 2020)。

因此，2016年英國脫歐公投被描述為「頭腦 heads」與「心靈(hearts)」、「理性(reason)」與「情感(emotion)」之間的戰鬥，數位情感資本主義(digital affective capitalism)比經濟本身能產生更迅速、更可靠經濟效果(Bouko & Garcia, 2020; Karppi, Kähkönen, Mannevu, Pajala, & Sihvonen, 2016; Massumi, 2021)，AI所驅動的社交媒體影響公民的情緒(affect)、價值(appreciation)和判斷(judgment)，從而擴大留歐者(Remainer)與脫歐者(Brexiters)的政治對立(political polarization) (Bouko & Garcia, 2020)，選民傾向相信自己感受而不是專業知識，引發關於民主政治未來辯論(Moss et al., 2020)。

英國脫歐派(the Leave)於脫歐公投期間¹⁶¹利用殭屍網絡(Botnet)和使用者生成超黨派新聞(User-Generated Hyperpartisan New)，無論是否為嚴格意義上假新聞(fake news)，或是媒體系統性侵犯認知權利，經過社交媒體上模仿人類活動的政治運動社交機器人(Social bots in Political Campaigns)瀑布狀的轉發 (retweet cascades) [使用者到機器人與機器人到機器人(user-to-bot vs. bot-to-bot)]、自動生

¹⁵⁹ 檢自 <https://zh.wikipedia.org/zh-hant/%E8%8B%B1%E5%9C%8B%E5%8E%BB%E7%95%99%E6%AD%90%E7%9B%9F%E5%85%AC%E6%8A%95%E6%B0%91%E6%84%8F%E8%AA%BF%E6%9F%A5>

¹⁶⁰ 檢自 <https://whatukthinks.org/eu/opinion-polls/poll-of-polls/>

¹⁶¹ 不同研究選取研究期間不同，例如 Bastos & Mercea(2019)選擇公投前後2週研究比較分析，Rosa & Ruiz(2020)選擇前3週。

成、轉推的功能，機器人成為有效資訊傳播者(Bastos & Mercea, 2019)。

再者，計算機媒介通信(computer-mediated communication, CMC)是鼓勵而不是抑制情緒表達(Derks, Fischer, & Bos, 2008)，訴諸情感和貶低對立觀點更能引起更高程度參與、更高影響力，使激情勝過理性(passion trumps reason)(Rosa & Ruiz, 2020)。更甚者，機器人以近似人類行為創造「對特定想法集體共識的錯誤感覺」(Bouko & Garcia, 2020)，對溝通與交流產生不利影響(Bastos & Mercea, 2019; Watson, 2018)，使得原本哈伯馬斯所設想理性(rational)、審議式對話(deliberate dialogue)、排除情感反應(affective responses)所組成的公共領域(public sphere)變成一個不允許真正對話的公共空間(public space)，無法反映英國公民真正期望(Wogu et al., 2020)。

雖然 AI 所驅動的社交媒體究竟是否會影響公投結果？目前尚無大規模定量研究(Bouko & Garcia, 2020)，以至對於社交媒體所能產生效果見解不一，有研究認為還沒有證據支持可以實質性地改變公投結果(Bastos & Mercea, 2019)，有研究指出 AI 已被證明對於競選、預測和操縱選舉結果是可行且非常有效，如果不是最後一刻 AggregateIQ (AIQ)等「AI 政治代理人(agents of AI politicking)」的 AI 負面廣告政治方法，結果可能會有所不同(Moss et al., 2020; Wogu et al., 2020)。2024 年世界經濟論壇(WEF)的全球風險報告(The Global Risks Report 2024)指出，未來兩年，尖端 AI 科技產出的錯誤和假訊息(Misinformation and disinformation)可能會干擾選舉過程，使公民對媒體和政府資訊的不信任，有可能侵蝕民主並使社會兩極化。^{162 163}

綜上分析，雖然「國家機關權力涉入程度低」，對公民個人權利義務的權力影響低，「人機協作」採用「演算型決策」，善用 AI 比人「快(faster)」與「準

¹⁶² 檢自 <https://www.weforum.org/publications/global-risks-report-2024/>

¹⁶³ 檢自 <https://time.com/6553908/ai-misinformation-threat-wef/>



(accuracy)」的技術優勢，強化與大眾溝通能力(例如 AI Chatbot)。然而，誠如 AI 雙重用途的特質，AI 也可能就選民既存或潛在的特殊政治偏好／偏見進行分類，鎖定特定群體，進行精準投放，削弱選民對候選人與政治議題判斷能力(陳柏良, 2021)。因此，為避免使用 AI 技術(AI-based technologie)的「垃圾郵件機器人(spambots)」、「支付機器人(payment bots)」、「自動按讚機器人(autolikers)」和其他「假帳號(inauthentic accounts)」傳播虛假信息、炒作社會參與，除了需要更多的大眾參與監督濫用或操縱真正公共對話的空間(public space)、公共關係(Public Relations)的廣告政治、專業的操縱者、宣傳者，政府也應協助公民識別操縱技術(recognizing manipulative techniques) 以抵制宣傳和廣告的力量(the powers of propaganda and advertising)(Rubin, 2022)。

研究發現：

1. *AI 是行動者，應該是分析單位。*
2. *AI 雙重用途的特性，不同 AI 情境應有不同 AI 風險緩解機制：在國家權力涉入程度低，看似無害的 AI，當成為政治代理人(例如社交機器人)，雙重用途的特性仍有可能造成極高的 AI 風險(例如可以強化政府與大眾溝通能力，也可以操縱公共對話的空間)，因此仍須依照政府職能領域、情境相關性，設計相對應的 AI 風險緩解機制。*

第四節 勞動部「1955 多元智能電服中心」、「AI 人機協作 分文」



1955 多元智能電服中心(1955 Omnichannel Intelligent Contact Center)屬第四象限「國家機關權力涉入程度低」，偏象徵和勸告型、偏人類決策、完全部署(fully deployed)，職能分類為 F1：一般公共服務。

壹、案例分析—ANT 分析途徑與因果關係分析途徑分析結果

2020 年的 Covid-19 疫情更是加速此系統應用的外生性因素，疫情導致電話瞬間大量湧進，促使該系統於 2021 年 1 月始試營運、5 月正式上線，本計畫目的是解決政府人力不足無法接聽大量的勞動諮詢電話，而身心障礙的電話服務人員也反映工作負荷太重的治理困境。

一、問題化：

勞動部 1955 電話服務中心 2001 年成立，為僱用身心障礙者從事勞動法令諮詢服務的政策示範。2019 年中心約 48 人，其中 93% 為身心障礙者，多數為視覺障礙者(35%)，約有 22% 的電話服務人員年資超過 19 年，部分資深電話服務員已逐漸邁入中高齡。身心障礙加上中高齡雙重身分，體力逐漸無法負荷全職工作，產生退休年齡普遍早於一般勞工，而新進人員獨立上線所需訓練時間約為 1 年。

傳統電話服務中心由真人提供服務，配合上下班時間無法提供 24 小時服務，一旦產生新興議題，電話諮詢量大幅增加時，無法迅速擴充，民眾電話無法接通，或等待通話時間增加，立即產生民怨。

勞動議題雖然複雜，但諮詢內容有大量重複性，適合導入 AI。『人工智慧來了』提到 15 年內，40-50% 崗位的工作將被取代，其中第五名是電話接線員(李開

復、王詠剛, 2017), 傳統諮詢電話服務員擔心工作被 AI 取代：



「一直有新聞報導說 AI 的發展『可能』讓電話服務員這個職類萎縮，因此一直放在心上，在教育訓練課程導讀李開復的書，看完真的嚇了一跳，AI 的發展『確實』會讓電話服務員這個職類萎縮。」

—伊甸社會福利基金會袁天英經理

勞動部找出強制性通過點(OPP)：「利用 AI 擴增身心障礙者能力。」，並定義行動者身份和他們「想要(want)」的東西：

1. 勞動部：提高電話諮詢服務量能、提升電話服務人員工作效率。
2. 電話服務人員：避免被 AI 取代。
3. 公民：不同管道、7/24 的服務。
4. AI：提供勞動議題的問答服務。

二、利益賦予

雖然智能客服系統使用 AI 技術(自然語言處理 NLP— Bert 模型)已開放原始碼，然而，不同專業領域知識將影響 AI chatbot 的正確率(accuracy)與覆蓋率(coverage)，欲將電話服務資訊系統升級為智能客服系統，必須先克服缺乏勞動領域問答語料庫的問題。其次，使用者介面將影響「有用性(Perceived Usefulness, PU)」(相信使用特定系統可以增進工作績效的程度)以及「易用性(Perceived Ease of Use, PEOU)」(相信使用特定系統不需體力或智力付出的程度)的特性(F. D. Davis, 1989; F. D. Davis & Venkatesh, 1996; 洪新原、梁定澎、張嘉銘, 2005)，目前也無適合身心障礙者，特別是視覺障礙者的使用者介面的智能客服系統。因此，有關 AI 系統的建置與架構如下：



勞動部同仁(領域知識專家)建立標準問答，再利用資訊擷取、資訊分類與資訊檢索技術中的自然語言處理，定時、自動、跨裝置的檢索勞動法規及最新發布或修正之法規命令等資料並自動文件分類，有效節省原來人工資料彙整及搜尋的時間。

AI 模型的建置是利用 Google 於 2018 年釋出自然語言處理的預訓練模型 BERT-(Bidirectional Encoder Representations from Transformers)作為意圖辨識的基礎，並以勞動領域標準問答進行微調(fine-tune)訓練問答系統，使 AI 自動決策系統產出勞動領域的問答資料。

AI 模型建置的目標是賦能(empower)身心障礙者(Fjeld, Achten, Hilligoss, Nagy, & Sri Kumar, 2020)，成就「不會裁員的智能客服中心」，採取「以人為本」的設計，關鍵的程序是電話服務人員全程參與 AI 模型建置，包括提供無障礙者使用介面需求、原有電話服務中心操作流程等提供給 AI 程式設計人員，在協作過程中即時回饋使用經驗讓 AI 工程團隊調整模型，使 AI 系統達成符合無障礙規範 AA 等級，對身心障礙者具有用性以及易用性。簡言之，AI 模型的設計過程，是利害關係人共同設計(co-design)、共同創造(co-created)、共同生產(co-production)、共同遞交(co-delivering)或共同評估(co-evaluating) (Molinari F., 2021)。回應文獻所述，利害關係人的「參與」是解決數位轉型過程中所導致權力不平衡等棘手問題(wicked problems)的具體方法(J. Fountain, 2019)。

AI 演算法的訓練資料來源，除了勞動部的知識管理系統(Knowledge Management Systems, KMS)之外，當 AI chatbot 回答不正確，客服人員可立即檢誤與蒐集反饋為訓練資料，透過不斷效能調校(Performance Tuning) 或微調(Fine-tune)，大幅提高 AI 模型的正確率(accuracy)與覆蓋率(Coverage)。

三、徵召

勞動部需開拓身心障礙者就業模式、提升電話服務量能，而電話服務員擔心



被 AI 取代，因此，數位轉型是共識。勞動部、AI 技術廠商與電話服務中心三方合作，將電話服務中心數位轉型為「1955 多元智能電服中心」，期待三方與公民都能共同獲益同時避免風險。

勞動部 2019 年開始導入 AI 技術，在不降低電話接聽量，辦理電話服務中心全員「職務再設計(Job Accommodation)」、AI 技術「在職訓練(On-Job-Training)」，即使全盲電話服務員也接受職業訓練。AI 系統建置過程，重新思考和重新設計公共行政流程(Elmi, Broekaert, & Larsen, 2018)、動態調整工作流程、工作方式、邊做邊學(Learning by doing)，藉由人機協作(human-machine collaboration)提高電話服務人員工作價值，擔任機器人訓練、標音、知識庫資料蒐集等工作。同時，勞動部參考市場價格，逐步於制度上調整薪資；尚未轉型前 10 年電服員僅調薪 2 次，且平均每年調薪不到 1%，至 2022 年整體薪資調幅約為 25%，更進一步推升電話服務人員 AI 科技接受度。

「一開始覺得他們薪水很低，促進身心障礙者穩定就業是低標，更重要的是身心障礙者應該有合理的薪資報酬，但在公部門沒有多做事很難多拿到預算、很難加薪。我要想辦法加薪，最重要的是他們願意多做事，也不怕多做事，我們就有加薪的理由。」

—勞動部

研究發現：工作效率與薪資制度關聯，提高終端使用者的科技接受度。

由於 AI chatbot 的覆蓋率為 96%，為克服 AI 偏誤(bias)，人機協作服務模式包括二層：第一層由 AI chatbot(文字客服系統)，即所謂「AI 官僚」負責第一線回應，可提供 24 小時全年無休的服務。在上班時間輔以人工監督及修正回應內

容，當 AI chatbot 無法解決民眾問題，則接續第二層服務。第二層是 AI chatbot 推薦答案給電話服務人員，不論資深或新進電服員都能即時取得標準問答，



「樂博(勞動部 AI Chatbot)現在也可以作為電服員的查詢工具之一，舉例來說，有一次一個民眾問職場霸凌問題，其實明明樂博有答案，在民眾還是按轉真人，可能是民眾的問法樂博無法辨識，等到真人文字客服服務時，就問樂博有關職場霸凌的答案，再把樂博比較口語版的答案貼給民眾，以這個例子來看，樂博現在可以算是口語版的 km。」

—伊甸社會福利基金會袁天英經理

提供包括(1).真人文字客服(線上文字交談)；或(2).真人電話諮詢服務等。

四、動員

本案例中可能進行「正向方案」的動員的因素分析如下：

1. 電話服務人員：

AI 協助電話服務人員提升服務品質，可以分為二部分：

(1).穩定電話服務人員的情緒：電話服務人員被要求回答問題必須同時注視勞動法規條文，避免僅憑記憶力而提供錯誤資訊。原本以人工方式進行法規條文資料的蒐集及整理，必需熟記資料夾路徑，憑記憶搜尋需要 20-30 秒，有了 AI chatbot 和 KMS，只需要 1 秒，對提升電話服務品質非常有幫助：

「電話服務人員找資料的速度很快、心情會比較穩定，心情穩定很重要，如果一時找不到資料，服務人員會處於焦慮狀態，回答內容就會忘

東忘西；1 秒就找到資料，回答問題時比較從容、比較完整。」

—伊甸社會福利基金會袁天英經理



(2).提升電話服務人員回答精準度

「客服最重要的是答案要精準，熱忱雖然很重要，但如果只有熱誠卻不精準是沒有用的。因此我們會很認真看待『真人文字客服』，我們的回答反而會更謹慎，因為所有的文字都會留下紀錄。」

—伊甸社會福利基金會袁天英經理

AI 系統也針對電話服務人員提供客製化的協助，例如每個電話服務人員都有較專精和比較不熟悉的勞動領域知識，『我的最愛』可以更快速找到比較不熟的資料，大幅降低搜尋記憶資料夾路徑的時間，實質延長身心障礙者的工作年限，以及新進人員的訓練時間。

2.勞動部：提升電話諮詢服務的品質與量能。

「目前我們還在訓練 AI，所以還沒有看到 AI 對我們大量幫助的回饋，但是隨著時間的累積，可以看出樂博可以在疫情時期幫助中心消化服務量，現在即使沒有疫情時的話量樂博服務數量仍四萬多件，也就是在協助中心消化服務量了。」

—伊甸社會福利基金會袁天英經理

由於本案例屬無較無涉公共行政改革、公平性、合法性和政治可行性、可管



理性、組織管理、組織流程、倫理、隱私和數據匿名等問題的挑戰。接續針對可能進行「反向方案」的動員的因素分析：

1. 終端使用者(身心障礙者)的科技接受度

電話服務人員全程參與 AI 模型建置，因此信任 AI charbot 和 KMS 所推薦的答案，AI 如果錯誤也能立刻判斷，並列入後續進行 AI 系統調校的訓練資料，形成人類與 AI 互相協作的人機協作模式。

「我們相信 AI 的推薦啊，AI 模型是我們自己建的，訓練資料是我們提供的，我們自己維護資料庫，如果有錯誤，我們自己會修正。如果是別人在維護的，我們可能就會先思考維護人員是否具有領域知識，如果不是領域知識的專業人員，我們就會有某種程度的懷疑。」

—伊甸社會福利基金會經理袁天英、組長蒲坤城

2. 勞動力替代

一般 AI chatbot 的導入是採購服務，容易發生勞動力替代的風險。勞動部藉由政策溝通，強調導入 AI 是使身心障礙者生理增強降低工作障礙、認知增強來強化職能，避免被 AI 取代：

「雖然一開始心裡很擔心、害怕、很抗拒，後來經過很多教育訓練和學習，知道 AI 可以是協助角色，幫助我們工作更好，人不是那麼容易被取代，就像按摩椅沒有辦法取代按摩師一樣，人類不一定會被取代。」

—1955 多元智能電服中心組長蒲坤城、黃雅絹

本研究案例甚至培訓出臺灣第一個全盲 AI 數據標註師：



「我覺得很榮幸，成為全臺灣第一個學會『AI 數據標註服務』的視覺障礙者，對我之後的履歷是很棒的幫助。」

—1955 多元智能電服中心黃雅絹 (全盲)

聯合國電子政府調查報告(UN 2018 E-Government Survey)明確指出，數位科技應用之程度與社會排除(social exclusion)呈現負相關(UN, 2018)，若能運用 AI 賦能獲得尊嚴勞動(decent work)的工作，將有助於加速達成 SDGs(Aquaro, 2019)。

「學了 AI 這個全新的知識，這是蠻新奇的經驗，之後跟別人聊天聊 AI 這個話題，因為這個領域是我懂的，相對比其他的人可以有切入的點，別人會覺得我懂得不少，覺得自己很先進。」

—1955 多元智能電服中心組長蒲坤城

研究發現：擁有 AI 技術應該成為身心障礙者基本人權，加速達成 SDGs。

本案例提升公共治理效能：

1. 提升電話諮詢服務效能：

- (1). 提高電話諮詢服務量能：2021 年電話服務量為 58 萬 6,728 通(2020 年為 62 萬 4,838 通)，AI chatbot 回覆為 99 萬 1,423 通次(覆蓋率為 95%)，AI chatbot 服務量已經超過電話服務人員 (接近 2 倍)，且真人電話服務量並未明顯減少。2021 年總計服務量為 158 萬 8,151 通次，換句話說，在電話服務人力總數不變，2021 年服務量較 2020 年擴增約 2.53 倍。



(2).增加電話諮詢服務管道：AI chatbot 提供公民全年 365 天，24 小時諮詢服務，強化政府與公民的互動品質。在智能客服品質，2020 年使用「標準問答」調校，每月需花 352 小時、滿意度 68%；2022 年使用「通話數據資料」調校，每月花 38 小時(節省 314 小時)、智能客服滿意度達 86%。

(3).提高電話諮詢服務品質：簡單問題 AI chatbot 回答，複雜問題真人回答，提高電話諮詢服務品質，AI 的目的不僅僅是節省政府資源，相反的，是將資源用於需要更深入處理的案件，以便做出公正，公正的決定，是通過 AI 來變得更加人性化(Djeffal, 2020)。

「一開始，我們對於 AI 在我們的業務中扮演的角色感到懷疑，但隨著時間的推移，逐漸體驗到它帶來的價值，儘管目前仍在訓練中，但已經開始看到它在我們業務中的幫助。在行為方面，它展現了出色的表現，回答迅速且準確性也慢慢提升，並且能夠處理大量的服務請求，消化不少的服務量，這可以讓我們的電服員能夠專注於更複雜的問題及人性化的感動服務，所以我們對於 AI 的潛力感到興奮及滿意，我們相信隨著時間的積累和 AI 的不斷進步，它將成為我們業務的重要一環，為我們提供更多價值和競爭力。」

—1955 多元智能電服中心組長蒲坤城

2. 提升身心障礙電話服務人員工作效率：AI chatbot 和 KMS 搜尋資料的速度從 20-30 秒降低至 1 秒，克服人類生理能力的侷限性，減輕工作負荷。

簡言之，勞動部電話服務中心導入 AI 技術後，達成提高公部門績效評估與

營運管理、提升公民服務的品質，進而提升公共治理效能。





貳、方案特色

1955 勞動諮詢服務專線提供服務的管道包括 AI chatbot、「真人文字客服」、「真人電話客服」，其中 AI chatbot 是目前公部門最常看到、最典型的「AI 官僚」或「隱藏官僚 (hidden bureaucrat)」。¹⁶⁴ AI chatbot 的覆蓋率(coverage)為 96%，AI 偏誤(bias)的配套措施為轉「真人文字客服」、「真人電話客服」。因此，有關人與機器的互動關係，AI chatbot 在人類的監督之下直接提供服務；AI chatbot 也是輔助電話服務人員的致能技術(an enabling technology)。

因此，原本「人員的管理—減少人類主觀式偏差管理」，當使用 AI 的演算型決策時，必須增加「AI 的管理—AI 的偏誤(bias)管理」。

接續針對本案例特色整理如下：

1. 創新模式：勞動部推動人工智慧專案評獎計畫，鼓勵多點創新。
2. 跨域治理：利用公私協力強化政府治理能力，參與的單位包括：勞動部、非政府組織(NGO)、AI 技術外包廠商；在勞動部機關內部，則為多部門(Multi-departments) (秘書處、資訊處、其他業務司處單位、所屬機關構)合作開發 AI 模型。
3. 建立勞動領域語料庫：自然語言處理(Natural Language Processing, NLP)模型正確率(accuracy)要提高，要使用領域資料庫而不是通用型資料庫。中央研究院雖然已經建置有漢語平衡語料庫(Sinica Corpus)¹⁶⁴，但在涉及更專業 (specialized)、技術(technical)和行業術語 (jargon)的勞動領域，並沒有大量標準的標記數據集(labelled datasets)及勞動語料庫(labour-specific corpora)。本研

¹⁶⁴ 檢自 https://ckip.iis.sinica.edu.tw/project/sinica_corpus

究案例，熟悉勞動法令的電話服務員，經過訓練之後即有能力從事勞動客服語料庫的標記工作，未來將可更進一步建置勞動語料庫。



4. 工作效率與薪資制度關聯：制度上，若效率和薪資無法直接關聯，當原來的工作方式就能辦理業務，會降低終端使用者使用 AI 工具的誘因。本案例同步處理「增加工作價值」與「調整薪資制度」，從而提高終端使用者的科技接受度。
5. 避免未接受訓練的終端使用者使用 AI：AI 不可避免的產生偏誤(bias)，若終端使用者未具備 AI 技術及領域知識，

「新進人員是無知狀態，只能選擇相信。」

—伊甸社會福利基金會經理袁天英

特別政府部門具有行使統治權作用之行為，過於依賴這些 AI 機器的危險，遠比機器本身自主學習產生的危險更甚(Barth & Arnold, 1999)。

6. AI 是賦能(enable)技術：AI 可能改變工作的性質，產生新的工作類型(OECD, 2019)，若是拿技術來做賦能的工具，**在 AI 技術設計階段就應該充分考量「以人為本的 AI 設計」**、「職務再設計(Job Accommodation)」、「在職訓練(On-Job-Training)」、「提升工作價值後加薪」等策略，讓員工以不同的方式做事，產生更加培力(empowerment)的結果，不僅能避免勞動力替代，更開拓全盲身心障礙者就業新模式：AI 訓練師



「雖然現在 AI 還不夠聰明，需要大家的訓練才會越來越茁壯，然後訓練 AI 的過程很燒腦，會多出好幾根白頭髮，但是當 AI 能回答的問題越來越多，需要優化的問題越來越少，會非常的感動，不枉費多出來的白頭髮跟十條皺紋。」

—1955 多元智能電服中心黃雅絹 (全盲)

AI 的最終效果，不是取決於技術能力，而是取決於如何被使用。在勞動力容易被取代的智能客服領域，如果，我們只專注於發展完全自動化，AI 純粹替代人力和某些類型的工作，則 AI 只會帶來負面影響(Zeng, 2020)；但如果，我們將擁有 AI 技術視為人類的基本人權¹⁶⁵，甚至是身心障礙者的基本人權，並量身訂做賦能的 AI 技術，AI 不會取代人類，反而有機會達成「聯合國永續發展目標」(United Nations Sustainable Development Goals, SDGs)¹⁶⁶，消除貧窮、減少不平等的目標。

研究發現：若是拿技術來做賦能的工具，在 AI 技術設計階段就應該充分考量「以人為本的 AI 設計」，使人機協作的設計重點是人與 AI 能合作互利共生地工作來擴增和強化彼此的能力。

¹⁶⁵ AI 360: Expert evaluation of future implications of AI and how to steer for societal benefit；檢自 <https://www.humanbrainproject.eu/en/follow-hbp/news/ai-360-expert-evaluation-of-future-implications-of-ai-and-how-to-steer-for-societal-benefit/>

¹⁶⁶ 聯合國於 2015 年正式通過「聯合國永續發展目標」(United Nations Sustainable Development Goals, SDGs)，希望從 2015 年到 2030 年以綜合方式解決社會、經濟和環境的發展問題。



參、其他案例

茲針對國家機關權力幾乎不涉入〔例如「政策工具光譜」屬自願性政策工具〕、「任務所需脈絡分析程度」、情境相關性，視個案情況不一，以技術的觀點是 AI 可能擅長或不擅長的任務，挑選其他國內、外案例補充說明：

一、勞動部「AI 人機協作分文 (AI e-Document Distribution System, AI-eDDS)」

在臺灣，每一個公務機關都會有分文的工作。據本研究初步探詢是第一個部會層級利用 AI 智能協作分文提升內部管理行政效率，完全部署(fully deployed)。公文系統包括收文、發文及檔案作業，每天重複性最高的工作之一是收文及接續的分文作業，工作內容是看似簡單卻大量重複、單調及繁瑣。勞動部每年公文量即約 9-10 萬件，指派錯誤影響行政效率，原本分文仰賴分文人員以關鍵字查詢以往分文歷程以及經驗判斷，屬於工人智慧(artificial artificial intelligence)。

由於勞動部已經累積了超過 20 年的人工分類公文的資料庫，成為建置 AI-eDDS 訓練資料庫的重要來源。勞動部透過公私協力，與 AI 技術外包廠商合作開發 AI 模型，AI 模型是以過去人工分文紀錄，以人工標註(labeled)後作為訓練資料，使用的 AI 技術有：(1).規則式系統(Rules Based System)，處理具有可歸納規則的公文；(2).機器學習(Machine Learning)：採用 Google 建構之 BERT 模型進行機器學習，利用自然語言(NLP)功能加上 AI 監督式學習(supervised learning)分類，處理規則式系統無法判斷之公文。該 AI 模型自 2021 年 10 月起上線截至 2022 年底止，正確率(Accuracy)平均約 85%- 90%，人機協作方式採用人類監督(human in the loop)提高機關分文效率及正確率。根據機關統計已節省分文人力約 2/3，2022 年估計減少 1,250 小時，縮短分文作業時間約 60-80%。內容單純的例行性公文僅需收文人員快速目視掃描確認後立即「一鍵送出」，即使新進人員經過簡易訓練亦可辦理分文，分文作業可專注於處理爭議複雜案件，減少公文錯誤

率。此外，經檢視產生 AI-eDDS 偏誤的公文，多屬人類亦難以直接判斷如何分類的公文。因此，針對公部門大量重複性的工作，可利用 AI 減少傳統官僚機構運作所必須的繁文縟節，提高政府內部管理的治理能力。



研究發現：政府使用 AI 消除繁文縟節。

二、美國最高法院 Rucho v. Common Cause 案，利用 AI 判斷黨派「政黨傑利蠮螋 (PARTISAN GERRYMANDERING)」的現象

美國的選區每 10 年重新劃定一次，在超過三分之二的州，黨派立法者控制國會選區的重新劃分。「政黨傑利蠮螋」指以不公平的選區劃分方法操縱選舉，使投票結果有利於某方；例如，將特定選民集中在某一選區，使其支持之候選人以過高之選票當選；或將特定選民分散在不同選區，使其支持之候選人均落選，有效率地剝奪百萬美國人行使選舉權(許炳華, 2022)。

在 Rucho v. Common Cause 案，黨派立法者利用美國按選區計票和贏者通吃原則，以 AI 繪製不公平的選區來操縱選舉結果。政治和法律專家利用 AI 和數據庫，依據影響選區劃分的人口比例、城市和縣級保護、緊湊度、黨派效應等因素，以人機協作在數小時而不是數月內，繪製成對特定某方有利的傑利蠮螋選區地圖。當地區沒有競爭性時，基本上選舉結果是預先確定的，儘管選民仍然能夠投票，但他們卻無法影響選舉，選民被剝奪了選舉權，這也是 2016 年希拉蕊敗給川普的原因之一。

利用 AI 繪製選舉地圖的案例顯示 AI 技術已對人類行動者的行為產生影響，技術已侵蝕人的領域，人格概念(concept of personhood)的完整性受到威脅(Garvey, 2007)。人類能動性的行使包括利用自我意識反思想什麼和做什麼所決定的判斷和



行為¹⁶⁷，從這個角度來看，AI 繪製選舉地圖繞過決策者的反思或審議過程來威脅到個人的自主權(Reijula & Hertwig, 2022)；或是經由分析人類的選擇和行為，利用 AI 干擾或壓抑潛認知(cognition of unknown knowns)，從注意力的被剝奪到選擇的被剝奪，顛覆人類的民主制度(蘇經天, 2023)。最後，最高法院認為黨派關係可以用於重新劃分範圍，但不能「過度使用(excessive use)」，過分注入政治是非法的(unlawful)，當黨派競爭已經達到了不公平的極端、選區劃分不公時，法院必須進行干預。因此，最高法院也利用 AI 發現黨派或是政客在技術中隱藏的看似合法但是卻是不合理分配的潛在意圖，提供司法上可管理的標準來判斷選區地圖是否過於偏黨派因素，開始啟動去除黨派或政客控制的因素(Tam Cho, 2019)，這案例不僅展現 AI 技術物政治性的特徵，更充分展現 AI 技術物雙重用途(dual use)。

¹⁶⁷ 原文：the exercise of human agency consists in judgment and behavior ordered by self-conscious reflection about what to think and do.

第五章、AI 與公共治理的交引纏繞



接續從公共治理理論現有機制出發，以臺灣公部門使用 AI 案例的研究結果，並以國外案例輔助，回應本文研究問題與目的。

第一節 AI 與公部門行動者的新關係

本節為回應在公共治理理論層次的研究目的：

1. 公共治理理論的單位，除了人，是否應該包括 AI？
2. 在公共治理領域中人機協作的理論基礎？

當原有公共治理理論無法解釋 AI 官僚、人與 AI 共存的現象，例如 AI 和人藉由且通過(with and through)技術與彼此關聯(王志弘、高郁婷譯, 2023)在一起：「AI+」與「H+」的新組合。當人和機器相結合擴增(augment)和強化(enhance)彼此的能力(Miller, 2018)，達成人機共同協作才能具備的某些能力(周信輝、方世杰、李慶芳、蔡馥陞、劉亭蘭, 2022)；未來的組織必須整合人和 AI 協作來創造價值(Dwivedi et al., 2019; Stead, 2018)，而管理者需評估人與 AI 合作的最佳工作狀態(Kaplan & Haenlein, 2019)，迫使我们應轉向重新思考，公共治理理論以人為中心的典範應轉移，AI 與公部門行動者必須共同工作(needs to be done together “with” them.)、共存治理(governing with AI) (Molinari F., 2021)。

壹、重新思考 AI 與公部門行動者共存的關係—交引纏繞(entanglement)

Breit 等人(2021)曾透過對第一線就業服務人員的觀察指出，這不是從人類服



務到數位服務的轉變，而是人類和數位服務新的交引纏繞(a new entanglement of human and digital services) (Breit, Egeland, Løberg, & Røhnebæk, 2021)。研究發現，AI 與人的分工看似明確，例如：AI 推薦&人類決策(AI recommendation & human decision)(智能稅務服務計畫的 AI 選案查核)、AI 自動決策&人類檢閱(AI's autonomous decision & human review)(AI 智慧交控管理、智慧判煙系統)、AI 第一線&人類第二線 (AI as the first-line and human as the second-line)(1955 多元智能電服中心)、AI 預處理&人類決策(AI preprocesses and human decision)(AI 智能協作分文)。然而，當人類難以理解演算法，不同的人可能會採取不同的態度和行動：

一、AI 系統是實際決策者

1. 公務員願意釋出行政裁量權。公務員為何願意主動釋出行政裁量權？

(1). 增加公務員公共治理能力的理性選擇。「智慧判煙系統」AI 官僚平均正確率、效率較人類官僚高。

(2). 不願質疑 AI 系統的決定，放棄行政裁量權的非理性選擇。2012 年波蘭勞動和社會政策部的「失業評分系統」，工作人員不願質疑計算機系統作的決定，超過 99% 依據失業評分系統的推薦分類，幾乎放棄行政裁量權。

2. 公務員與 AI 協作，AI 官僚第一線&人類第二線。1955 多元智能電服中心的 AI chatbot 無法回答時，可立即轉真人服務。

3. 新進人員只能選擇相信。當領域知識不足又難以理解演算法決策方式。1955 多元智能電服中心受訪者表示：「新進人員是無知狀態，只能選擇相信」。

研究發現：公務員具有領域知識和 AI 技術知識，才能透過協作來創造價值，達到人和 AI 相結合擴增和強化彼此的能力，避免人對 AI 信任度過高或過低的問題。



二、刻意忽視 AI 的建議。智能稅務服務計畫受訪者表示：「黑名單不夠黑、白名單不夠白，把這個資料教機器，機器怎麼學得出來？」。

三、AI 能動性(agency)對終端使用者的感知(perception)或認知(cognition)產生影響。AI 人機協作文受訪者表示：「失去 AI 分文之感受是不習慣，且對於較無把握的分文，無法比對 AI 分文建議，略有不安感。」。

本研究從 ANT 的視角彙整人與 AI 的關係與人機協作模式如表 5-1，結果顯示，公務員對 AI 的認知、行為或感受比我們所能預期的要複雜和微妙，讓人類以新的、意想不到的方式理解和概念化世界，如同文獻所說，：AI 於人機互動之中並非僅是扮演被動的角色，而是在互動中扮演平等和對稱的角色(Semeraro et al., 2023)，已經超越公共治理理論以人為主的治理典範，不得不從本質上重新考慮我們如何理解 AI 與人的關係，以及「AI+」與「H+」的新組合。

因此，公共治理理論的單位，除了人，應該包括 AI。而 AI 與公部門行動者是交引纏繞的關係，人與 AI 互動的結果係依據 AI 和人協商和劃定的結果，因此，產生的結果可能有非常多種態樣，案例所產生的結果包括階級關係(hierarchy)、代理關係(agency)、機器自主(autonomy)、雙重關係(dual)。

研究發現：公共治理理論的單位，除了人，應該包括 AI 技術物。AI 與公部門行動者的關係是交引纏繞(entanglement)。因此，公共治理領域人機協作的理論：人和 AI 能動性(agency)的交互作用，能完整詮釋 AI 與人的關係，AI 對公共治理的影響。



表 5-1 AI 與人的互動與感知彙整表

案例	AI 與人的互動與感知現象	人和 AI 能動性(agency)的交互作用	
		人與 AI 關係	人機協作模式
智能稅務服務計畫	<ol style="list-style-type: none"> 1. AI 增強人類認知能力(AI 挖掘重要異常點)。 2. 領域知識影響人對 AI 的感知(資深傾向模型有效, 資淺傾向無效)。 	階級關係: AI 模型輔助, 最終人類決策。	<ol style="list-style-type: none"> 1. AI 推薦&人類決策 2. 直接控制: AI 不會作為直接補稅的工具。
智慧判煙系統	<ol style="list-style-type: none"> 1. AI 增強人類生理能力(AI 可長時間監控)。 2. AI 增強人類認知能力(AI 判斷平均比目測判煙正確及穩定)。 	代理關係: AI 直接裁決取代目測判煙。	<ol style="list-style-type: none"> 1. AI 自動決策&人類檢閱 2. 間接控制: 訂定儀器校正的 SOP。
AI 智慧交控管理	<ol style="list-style-type: none"> 1. AI 增強人類生理能力(AI 可長時間監控)。 2. AI 增強人類認知能力(交通流量計算 AI 比人快與準)。 	機器自主。	<ol style="list-style-type: none"> 1. AI 自動決策&人類檢閱 2. 演算型決策
1955 多元智能電服中心	<ol style="list-style-type: none"> 1. AI 增強人類認知能力(協助新進人員分文)。 2. AI 增強人類生理能力(處理大量簡單問題)。 	雙重關係: AI chatbot、真人文字客服、真人電話服務同步進行。	<ol style="list-style-type: none"> 1. AI 第一線&人類第二線 2. 協作或合作
AI 人機協作分文	<ol style="list-style-type: none"> 1. AI 增強人類認知能力(人可比對 AI 分文建議)。 2. AI 增強人類生理能力(簡單案件 AI 分文速度較人快)。 	雙重關係: 簡單案件人快速掃描「一鍵送出」, 人處理複雜案件。	<ol style="list-style-type: none"> 1. AI 預處理&人類決策 2. 協作或合作

資料來源: 本研究整理



貳、人機協作如何互利共生—影響 AI 與公部門行動者互動的變數

AI 技術讓我們能以不同的方式做事(王志弘、高郁婷譯, 2023), 從而衍生公部門行動者與 AI 的分工議題, 第二章第四節已梳理三個觀點:

- 一、從公共行政觀點, 是將機器預測(Machine Predictions)變成人類決策(Human Decisions), 是分工觀點。
- 二、從資訊科學、工程、控制等領域的觀點, Sheridan 從計算機如何接管人類的執行任務, 表示自動化程度, 是計算機系統的自治權(the autonomy of computer systems) (Floridi et al., 2018) 觀點。
- 三、從 ANT 觀點, 人與 AI 互動, 彼此的能動性(agency)或作用(agency)交互影響, 可能產生授權主體改變, 是權力觀點。

雖然公共行政、資訊科學等領域提出「分工」、「自動化程度」的觀點, 研究發現, 人與 AI 互動是在互相提示(prompt)及協作(collaboration)、循環交錯過程中進行, 可能顯示人類與 AI 協同作用的強大形式(Malone et al., 2023), 例如可能產生互利共生(mutualism), 使得人和 AI 合作的表現比單獨的人或單獨的 AI 做得更好(1955 多元智能電服中心、智能稅務服務計畫), 也可能是彼此相互減弱的過程(波蘭失業評分系統)。因此, **ANT 觀點更能完整詮釋 AI 與公部門行動者互動的關係。**

人機協作如何在循環交錯過程中互利共生? 從 ANT 的視角研究案例, 在「智能稅務服務計畫」中, 人與 AI 透過協作增強彼此的能力, 首先是資深稅務人員提出風險指標, 提高 AI 系統的品質,

「國稅局稅務人員參與風險指標之建立與修改, 對釐清資料、模型正確性、有效性與關聯性有直接關連。」

—財政部財政資訊中心張文熙主任



接續是 AI 提升稅務人員的選案能力，

「AI 會挖掘出憑經驗看不出來的重要異常點，自然會提高命中率。」

—財政部臺北國稅局股長楊秀真

人類認知能力的侷限性，要找出原因與結果之間的聯結，呈現邏輯上所需的必要條件、充分條件、充分且必要條件，確實存在著客觀現實的困難(李聖傑, 2002)。若能以「機率因果關係定義(probabilistic definition of causation)」，以對於結果影響力的機率高低認定因果關係；或是至少可認為因果關係「無法排除」(何建志, 2010)，克服人類認知能力的侷限性。

而影響人與 AI 互動關係或結果的變數是什麼？經梳理文獻及本研究案例觀察結果，分別從公共治理理論、ANT、AI 各擇一重要因素分述如下：

一、AI 資訊品質(AI information quality)：AI/技術情境(context)相關性

「AI 系統品質」，包括正確性(accuracy)、及時性(timely)、時效性(currency)、呈現方式(presentation)和完整性(completeness)(Prentice et al., 2020)。由於本文研究案例均是應用系統功能加 AI 應用模組，因此本文之前名詞使用 AI 系統品質(AI system quality)。此處主要在微觀層次上討論影響人與 AI 互動關係或結果的變數，為能更精準描述 AI 演算型決策對人的認知所產生的影響，在此使用「AI 資訊品質」。

AI 系統功能的增強與人類依賴 AI 的程度有直接關係(Bullock, 2019; Prentice et al., 2020)。而影響 AI 資訊品質(AI information quality)成因包括資料來源和 AI 技術限制(例如語意分析、自然語言技術限制等)：

1. 「**情境相關性**」和「資料可得性」(陳昇瑋、溫怡玲, 2019)。
2. 公共政策問題的「**複雜性(complexity)**」和「**不確定性(uncertainty)**」



(Bullock, 2019)。

3. 「任務特異性(task specificity)」和「環境複雜性(environmental complexity)」

二個軸度，並提出 AI 裁量(artificial discretion, AD)的關鍵任務維度：所需行政裁量權程度和分析的程度(Busch & Henriksen, 2018; Young et al., 2019)等。

AI 資訊品質直接影響 AI 與人的互動關係和結果。舉例來說，智慧判煙系統，情境相關性低，黑煙不透光率實測結果誤差值均在+5%之內，在某些排放濃度範圍內偏差值約±0%，採用 AI 自動決策「直接裁決」，AI 與人是代理關係。另一個案例，AI 人機協作文，情境相關性高，正確率約 85-90%，AI 與人是「AI 預處理&人類決策」的雙重關係。

二、AI 能動性：ANT/ 交引纏繞後「偏向人類決策」或「偏向機器決策」(人類監督程度)

AI 能動性對公務員行政裁量權產生的作用，使 AI 的自動決策結果產生法律實質效果的可能性。研究發現，在智慧判煙系統，AI 與人之間是代理關係(agency)；在 AI 智慧交控管理，AI 與人之間是機器自主(automony)；在智能稅務服務計畫，AI 與人之間是階級關係(hierarchy)；在 1955 多元智能電服中心，AI 與人之間是雙重關係(dual)。

三、國家權力：公共治理理論 / 國家威權(national authority)或官僚權力(bureaucratic power)

AI 強化國家或官僚威權(authority)的統治能力或治理能力。在智慧判煙系統，強化公務員對違法事項的長期監控(Long-term monitoring)的治理能力；在 AI 智慧交控管理，AI 強化公務員瞬時、動態監控的治理能力。

影響公部門行動者與 AI 互動關係的 3 個變數：AI 資訊品質、AI 能動性、國家權力。



在分析公共治理領域中影響 AI 與公部門行動者互動關係和結果的變數後，在具體實踐層次，人機協作的最佳化模式是什麼？

參、在公共治理領域中人機協作的模式—人機協作魔力象限

從 ANT 的視角，顯示幾乎所有的 AI 風險緩解機制均包括人類監督，隱含 AI 挑戰的最終解法莫屬於人類監督，只是監督程度的差別。然而，人類監督程度太高，勢必降低 AI 比人「快」與「準」的技術優勢、以及公部門的科技接受度；人類監督程度太低，可能提高 AI 風險程度、社會的科技接受度。如何權衡人類監督的程度，成為是否能發揮 AI 潛力的重要關鍵。研究發現，較務實的做法是將技術限制與行政實務相結合，**策略性的運用「人機協作」和「AI 風險緩解機制」**，綜合評估**最佳化人類監督程度，權衡：**

1.AI 比人「快」與「準」 VS AI 風險。

2.公部門的科技接受度 VS 社會的科技接受度。

接續以本研究發現，提出公部門人機協作的可能模式：

一、第一象限：人類「直接控制」

當政府職能屬於國家權力涉入高、情境相關性高，屬高風險，可能提高公部門採用 AI 的風險。為避免新型態歧視(Dignam, 2020)，AI 不應單獨直接用於決策，而是篩選工具(screening tool) (Obermeyer et al., 2019)，採用「AI 推薦&人類決策(AI recommendation & human decision)」，可以提高公部門績效。



財政部「智能稅務服務計畫」、比利時「AI 預測系統(兒童日托地點檢查)」、愛沙尼亞「AI 圖像識別 SATIKAS32 系統(草地割草或放牧)」等案例顯示，公務員知道該系統不是 100%可靠，而是必須將現有專業知識與 AI 建議相結合。

二、第二象限：人類「間接控制」

當政府職能屬於國家權力涉入高、情境相關性低，採用「AI 自動決策&人類檢閱(AI' s autonomous decision & human review)」，可精進管制性政策的有效性。

環境部「智慧判煙系統」案例，國家權力涉入高、情境相關性低，採用 AI 自主決策「直接裁決」，AI 風險緩解機制包括 AI 設備認證、系統驗證，確保 AI 系統的判讀結果具有公信力。

另一方面，澳大利亞聯邦政府社會服務部「自動債務追繳系統」、波蘭勞動和社會政策部「失業評分系統」，政府職能屬於國家權力涉入高，情境相關性高，依照本研究的案例選擇架構屬第一象限，AI 不應單獨直接用於決策，而是篩選工具(screening tool) (Obermeyer et al., 2019)，採用人類「直接控制」較佳。然而，「自動債務追繳系統」採用債務追繳過程自動化；公務員幾乎直接採用「失業評分系統」的 AI 推薦，因而侵害社會最弱勢群體的公民權利，目前均已被政府正式拆除。

三、第三象限：「演算型決策」

當政府職能屬於國家權力涉入低、情境相關性低，屬低風險，利用 AI 比人「快」與「準」的技術優勢，採用「AI 自動決策&人類檢閱(AI' s autonomous decision & human review)」，可提升決策速度、監測和分析效率(AI 智慧交控管理)、公民服務品質(1955 多元智能電服中心 AI Chatbot 的 24 小時服務)。



然而，雖然在國家權力很少或幾乎沒有政府干預的情況，或基於自願的基礎上，以公共行政的角度偏向低風險。然而，從英國脫歐公投「社交機器人」，以及越來越多政府機構或政黨利用 AI、自動化的社群媒體帳號(social media account automation)、線上評論團隊(online commentary teams)操縱國內輿論(Bradshaw & Howard, 2018)等案例，顯示 AI 有政治性與雙重用途，因此，仍不可忽視 AI 可能傷害民主的風險，此時，政府應強化公民識別 AI 技術(AI-based technologie)傳播虛假信息及成為操縱技術的能力(Rubin, 2022)。

四、第四象限：「協作(collaborate)或合作(cooperate)」

當政府職能屬於國家權力涉入低、情境相關性高，採用「AI 第一線&人類第二線」(1955 多元智能電服中心的真人文字客服、真人客服)，可強化政府與大眾溝通品質；或「AI 預處理&人類決策(AI preprocesses and human decision)」(AI 人機協作分文)提高官僚組織效率。

從 1955 多元智能電服中心、AI 人機協作分文、美國最高法院 AI 判斷「政黨傑利蠟蝟」等案例顯示，協作或合作可有效利用人力資源、提高工作價值等。

然而，雖然在國家權力很少或幾乎沒有政府干預的情況，或基於自願的基礎上，以公共行政的角度偏向低風險。但從美國最高法院 AI 判斷「政黨傑利蠟蝟」案例，仍應關注 AI 技術物政治性、雙重用途(dual use)的特徵。舉例來說，諸多文獻呼籲政府應利用 AI 提高公民參與的效率(Eggers, Schatsky, & Viechnicki, 2017)，以及讓公民積極參與 AI 治理制度設計(designing AI governance regimes)(Wilson, 2022)，透過人與 AI 的協作或合作，使 AI 成為強化公民參與的利器(Helbing, 2019)。然而，人也可能利用 AI 優化搜尋引擎、個人化廣告和助推(nudging)，繞過決策者的反思或審議過程來威脅到個人的自主權(Reijula & Hertwig, 2022)，此時，AI 變成專業操縱者操縱認知的操縱技術(manipulative



technologies)，提高 AI 可能的風險。

公部門與私部門應用 AI 最大的不同點，是國家具有統治(rule)和控制(control)的權力，官僚制度(bureaucracy)使 AI 與人的互動結果更加複雜與難以預料，可能擴大權力差距，也可能縮小權力差距。因此，公部門可以利用影響公部門行動者與 AI 互動關係的 3 個變數：AI 資訊品質、AI 能動性、國家權力，評估 AI 與人共存和協同工作的最佳條件，確保 AI 技術的機會和風險合理分配。再者，本研究在公共治理領域提出影響 AI 與人互動關係的 3 個變數，可以補充既有 AI 治理理論或框架，偏向普遍應用性概念的理論缺口。

另參考資訊科學常用 Gartner 魔力象限(gartner magic quadrant)(Richardson, Schlegel, Hostmann, & McMurchy, 2008)，提出「人機協作魔力象限(Human-Robot Collaboration Magic Quadrant)」(圖 5-1)，並利用實際案例找出人機協作的可能模式，包括直接控制、間接控制、演算型決策、協作或合作，協助公部門的管理者應用 AI 的初步評估，選擇人與 AI 共存和協同工作的最佳條件。

最後，必須強調的是，人機協作的設計重點是人與 AI 合作互利共生地工作來擴增和強化彼此的能力，惟不同案例各有其領域知識和情境相關性，公部門實際應用 AI 時，仍須依政府職能採用適合的人機協作模式及 AI 風險緩解機制。

人機協作魔力象限提供公部門應用 AI 的初步評估，選擇人與 AI 共存和協同工作的最佳條件。

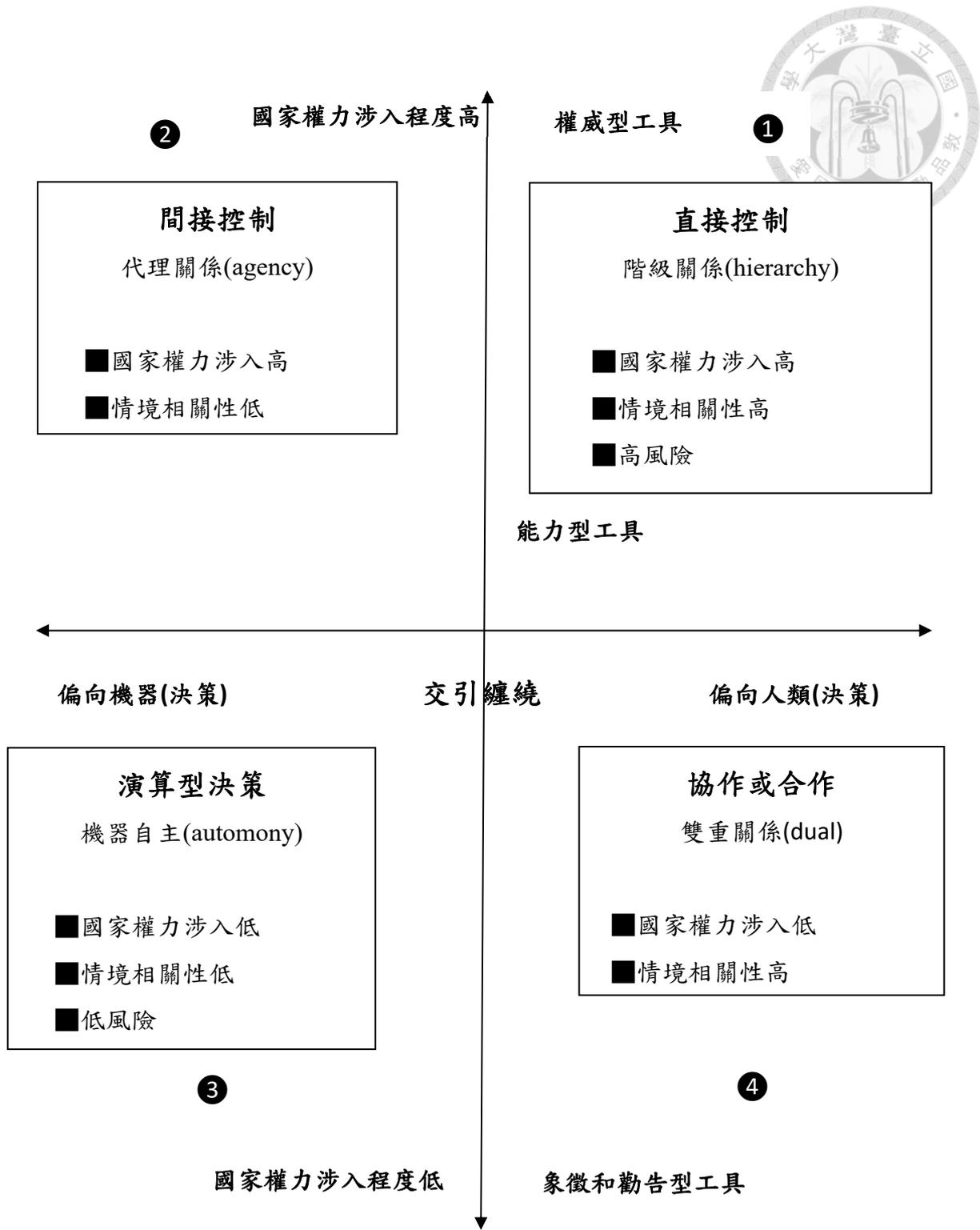


圖 5-1 人機協作魔力象限

資料來源：本研究設計



第二節 公共治理理論的典範轉移路徑

本節為回應本文在理論層次及政府使用 AI 的政策分析方法的研究目的：

1. AI 對公共治理理論產生的影響，以及應補充的理論基礎？
2. 驗證 ANT 是否是公共政策分析的利器？

接續從關注非人行動者(AI)如何影響公共政策模型、過程、執行、評估，再回饋到公共治理理論的不足之處，提出公共治理理論的典範轉移路徑，以及有助於延伸公共治理理論的描述和解釋力的推論。

壹、公共政策模型

公共治理理論探討公共政策的制定由誰決定？從而衍生不同公共政策模型，包括描述性模型、規範性模型等，亦即公共政策的規劃與分析時，必須正視「誰的意見該被考量？」，如何讓公共政策決策建構在被統治者之廣泛的同意之上，是否能代表人民聲音的「利害關係人指認」是政策能否實踐的關鍵(陳敦源、劉宜君、蕭乃沂、林昭吟, 2011)。接續以 AI 使公共政策模型產生理論不足之處，並利用 ANT 分析途徑提出如何補充的論述。

環境部 2023 年 6 月 14 日修正「固定污染源空氣污染物排放標準」，新增拍照判煙，由 AI 判定固定污染源是否符合標準，賦予「智慧判煙系統」—「AI 官僚」行使行政裁量權合法性。接續提供國外案例：「荷蘭政府 2014 年開發 SyRi (Systeem Risico Indicatie)系統」，說明將 AI 視為分析單位的重要性。

荷蘭政府 2014 年利用政府系統中的各種風險指標：稅收、健康保險、居住、教育等開發 SyRi 系統，同時以法律明定政府可以蒐集、存儲和共享哪種數據。



SyRi 系統目標是檢測詐欺或濫用社會福利資源的高風險地址，以解決福利欺詐、減少公共資金濫用、減少管理成本、改善社區安全等。

基於系統安全的問題，荷蘭政府對演算法保密，未公開 SyRi 風險模型及其指標、使用的 AI 技術、公務員使用自動決策結果的程度，公民無法得知 SyRi 是否有逾越或濫用權力之情形。原本政策設計 AI 系統不會做出最終決定，而是推薦公務員進一步查核的資料。由於訓練數據資料未能及時更新、帶有歧視風險，使得 SyRi 系統推薦的結果通常是特定的、不發達和較貧窮的地區、推薦成功率很低，造成 SyRi 的「AI 黑箱」「強化」社會不平等。再者，SyRi 的成本估計每年超過 325,000 歐元，不符合成本效益。

荷蘭政府對個人資料蒐集、處理或利用未尊重當事人之權益，數據主體(公民)不知道自己的資料是否被蒐集，引發社會的廣泛辯論及反對使用該系統，理由包括：侵犯太多隱私權，對貧困和脆弱公民具有歧視性等。最後，基於應用 AI 技術的好處與對數據主體隱私權的潛在干擾無法取得適當的平衡，以及可能產生的歧視風險，荷蘭法院於 2020 年初裁定 SyRi 使用不符合歐洲人權公約(European Convention on Human Rights, ECHR)第 8 條規定，防止欺詐的集體經濟利益不足以與隱私的社會利益權衡，SyRi 透明度和可驗證性(verifiable)不足，荷蘭政府表示不會對該判決提出上訴後取消 SyRi 的使用(Misuraca & van Noordt, 2020; van Bekkum & Borgesium, 2021)。

SyRi 的自動決策產生法律效果，成為新的 AI 官僚。然而「現行的一般法律原則是規範人類而非機器的行為」(黃心怡、曾冠球、廖洲棚、陳敦源, 2021)，顯示演算法不能簡單地視為行政過程自動化，必須將 AI 技術物視為分析單位：

1. 在法規層次，參考「智慧判煙系統」、德國¹⁶⁸的做法，將 AI 系統演算型決策

¹⁶⁸ 檢自 <https://ai.iias.sinica.edu.tw/ai-and-public-department-in-germany/> 「人工智慧科技於公部門之應用及其法制框架：德國國情報告」：「德國行政程序法第 35a



法制化。

2. 在管理層次，公部門對於「人員的管理—減少人類主觀式偏差的管理」，必須因應 AI 官僚增加「AI 的管理—減少 AI 偏誤的管理」。

綜上分析，公共治理理論應該超越社會學解釋人類能動性的理論應，給予 AI 廣義對稱性原則。**政策利害關係人的定義，亦即分析單位除了「人」，應包括「AI」，能完整詮釋公部門使用 AI 時的現象，補充公共治理理論之不足。**

貳、公共政策執行

公共治理理論探討科層體制的控制過程，立法機關訂定法規，並給予行政機關及基層官僚行政裁量空間，這些基層官僚是具有實質行政裁量權(substantial discretion)的街頭官僚(street-level bureaucrats)。接續以 AI 使公共政策執行產生理論不足之處，並利用 ANT 提出如何補充的論述。

AI 如何影響公共政策執行？

一、AI 取代公務員，補足公務員做不到的部分

人有主動性、AI 有自適應性，AI 可以自主識別行動、自主決策的行為，與人類行動者執行的社會行動(social actions)相當，也是馬克斯·韋伯(Max Weber)所理解的社會行動(Ågerfalk, 2020)。韋伯說：「該行為將被稱為『社會性(social)』，其含義是一個或多個行動者考慮到他人的行為所意圖(intended)，從而在其過程中具有『導向性(oriented)』¹⁶⁹」(Weber, 1978)，AI 於人機互動之中並非

條：行政處分得依法完全由自動化設備作成之，但以行政機關無裁量權或判斷餘地者為限。類似的規定出現在：社會法典第十部第 31-1 條；租稅法典第 155 條第 4 項：羈束決定、專法保留（必須有專法授權導入自動化）。」

¹⁶⁹ 原文：That action will be called 'social' which in its meaning as intended by the actor or actors, takes account of the behaviour of others and is thereby oriented in its course.



僅是扮演被動的角色，而是具有能夠產生行動的能力(capacity for actions)，從互動過程中產生集體能動性(collective agency)(周信輝、方世杰、李慶芳、蔡馥陞、劉亭蘭, 2022)。

交通標誌與道路標誌一樣具有法律約束力，也是行政行為(Djeffal, 2020)；在 AI 智慧交控案例中，AI 作為符號系統(semiotic system)，透過符號表達行動者意圖，能夠代表人類和組織執行社會行動的數據行動者(digital actors)。在「1955 多元智能電服中心」案例，AI 第一線&人類第二線，AI chatbot 是取代公務員的「隱藏官僚」。

二、AI 能動性(agency)對公務員的決策產生影響

人有能動性、AI 也有能動性，「AI 人機協作分文」的受訪者無法比對 AI 分文建議時略有不安全感。「1955 多元智能電服中心」的新進人員只能選擇相信 AI 的推薦。「智能稅務服務計畫」的資深的稅務員會傾向模型是有效的，資淺的會覺得模型無效等案例，顯示 AI 能動性(agency)對公務員的感知(perception)或認知(cognition)產生影響。AI 智慧交控案例中，AI 和人在高度時間依賴性敏感性(sensitive to time dependencies)的相互作用(interplay) (Semeraro et al., 2023)過程中產生集體能動性(collective agency)。

綜上分析，**公共政策執行時，在概念上區分人有主動性、AI 有自適應性，人有能動性(agency)、AI 也有能動性(agency)，能完整詮釋公部門使用 AI 時的現象，補充公共治理理論之不足。**例如 AI 有主動性(或自適應性)，詮釋 AI 的演算型自動決策(Automated decisions)取代公務員的行政裁量權現象；AI 也有能動性(agency)，詮釋 AI 的認知技術影響公務員的行政裁量權，產生行政裁量權移轉的現象。



參、公共政策過程

公共政策過程的研究途徑，階段論是主流，重視政策過程的可控制性、步驟發展的必然性與各階段的次序性(丘昌泰, 2008)。政策評估是調查一項進行中的計畫，就其實際成就(actual accomplishment)與預期成就(anticipated accomplishment)的差異加以衡量，對未來作相關的調整(曹俊漢, 1983)。按照 Max Weber 的定義，官僚制(bureaucracy)或「科層制」的層級節制原理，權責自上而下傳遞，科層化的管理方式。

AI 改變了公共政策過程階段論中次序性的概念(Elmi et al., 2018)，評估成為整個政策週期中不可或缺的一部分，而不是在決策過程結束時的單獨步驟(Höchtel et al., 2016)，持續評估的概念縮短傳統的政策週期。接續以 AI 使公共政策過程產生理論不足之處，並利用 ANT 分析途徑提出如何補充的論述。

在「AI 智慧交控管理」，AI 依賴主要路口現場所偵測的數據，在沒有直接人工干預的情況下，即時運算最佳化號誌時制，自動調整交通號誌，使得**公共政策過程呈現「政策規劃—政策執行—政策評估—政策規劃」的循環交錯過程。**

1955 多元智能電服中心案例，電話服務人員利用 AI chatbot 和 KMS 所推薦的答案提高電話服務品質，AI 如果錯誤也能立刻判斷，並列入後續進行 AI 系統調校的訓練資料，形成人與 AI 互相協作的人機協作模式。國外的案例：「愛沙尼亞(Estonian) SATIKAS32 系統」、「比利時 AI 預測系統(Predictive System)」。公務員知道 SATIKAS32、AI 預測系統不是 100%可靠，AI 模型提供更有目標性(targeted)和數據驅動(data-driven)的推薦結果，公務員利用專業知識和自己的判斷力做出最終決定。因此，**在人機協作過程中，呈現出人與 AI 的動態互動、人和 AI 能動性(agency)彼此交互作用，並形成共享數據、共同設計、共同執行等共同生產的過程。**

綜上分析，公共政策過程呈現循環交錯，人機協作過程呈現人和 AI 能動性



(agency)的交互作用，均與公共治理理論中典型公共行政過程次序性的流程、層級節制的官僚組織背道而馳(Mergel et al., 2021)。因此，**公共治理理論的過程論應包括循環交錯的政策過程、人機協作過程中人和 AI 能動性(agency)的交互作用，能完整詮釋公部門使用 AI 時的現象，補充公共治理理論之不足。**

肆、公共政策分析

AI 能動性改變人的認知權利(Watson, 2018)，AI 政治性改變權力結構(Winner, 2017)，應如何分析非人行動者(AI 系統)控制公務員、或行政裁量權在公務員、非人行動者(AI 系統)之間轉移的現象？以及 AI 對公共治理理論的影響？接續以 AI 使公共政策分析方法不足之處，並提出如何補充的論述。

在智慧判煙系統，AI 智慧交控管理，AI 是實際決策者。在 1955 多元智能電腦中心、AI 人機協作分文，AI 能動性(agency)對終端使用者的感知(perception)或認知(cognition)產生影響。國外的案例，2012 年波蘭勞動和社會政策部的「AI 失業評分系統」、澳大利亞聯邦政府社會服務部「自動債務追繳系統」，AI 能動性使 AI 成為實際上的決策者等案例，均顯示**技術的意義要在使用中發現**，「公務員『應如何』與 AI 互動」與「公務員『實際』與 AI 互動」，理論和實作產生落差。再者，AI 智慧交控管理、智慧判煙系統，強化公部門監控的治理能力，顯示一個表徵的原始權力並不比次級的對立者重要。

因此，本研究的方法論使用 ANT 的分析途徑來做公共政策分析，通過 ANT 的視角，重建這些關聯和關係的網絡，觀察 ANT 提供權力是如何融入這些網絡(Bueger & Stockbruegger, 2017)、權力關係的解構與重構(雷祥麟, 2004)的敘述、AI 與人之間權力移轉的複雜情況。

研究發現，「AI 是行動者(actors)」、「AI 能動性(agency)」的觀點，可以分析與解釋 AI 官僚、AI 賦能公務員治理能力、公務員和 AI 之間權力轉移等現象。再

者，ANT 重描述而非解釋的分析方法，可以揭示行動者網絡中隱藏的權力結構 (Waldherr et al., 2019)、技術隱匿的面向，解釋理論和實作之間為何產生落差、權力差距為何擴大、如何消除，技術導入成功與失敗的原因等，進而分析 AI 對公共治理理論的影響。

舉例來說，2012 年波蘭的「失業評分系統」，ANT 的分析途徑提供 AI 能動性影響公務員的專業判斷，導致 AI 系統控制公務員、AI 系統成為實質上的決策者。澳大利亞聯邦政府社會服務部「自動債務追繳系統」，ANT 的分析途徑提供 AI 是行動者，AI 系統成為實質上的決策者等案例，侵害社會最弱勢群體的公民權利，擴大了權力差距。

伍、小結：公部門使用 AI 的理論

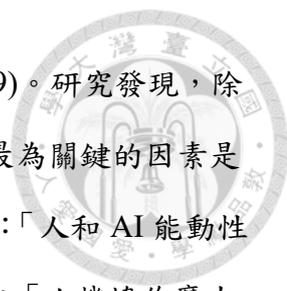
公共治理理論長久以來的的方法論、「政策利害關係」以「人」為主的治理典範，公共政策模型所欲達成權衡不同人類價值觀，公共政策過程的次序性、可控制性，公共政策執行時官僚行使裁量權的獨立決策自由度的詮釋邏輯等，面對 AI 所產生的影響，無法有效提出合理的解釋。

本研究結合公共治理理論、ANT、AI 提出公共治理理論的典範轉移路徑，已藉由臺灣實證研究案例驗證，亦可解釋本研究所補充或佐證的國外案例，因而具有強固性，茲依序論述如下：

一、重新定義公共治理理論的分析單位：「AI」應為分析單位

研究發現，公部門採用 AI，並不在於以往對技術的思考：「最適的技術引進機制」，AI 是一種新型態參與者，公共治理理論的分析單位，除了「人」，應包括「AI」。

二、重新審視 AI 與人互動的關係：AI 資訊品質、AI 能動性、國家權力的影響



AI 與人互動，文獻已提出 AI 資訊品質的影響(Bullock, 2019)。研究發現，除了 AI 資訊品質，應包括 AI 能動性、國家權力的影響，而其中最為關鍵的因素是 AI 能動性。因此，本研究提出公共治理領域中人機協作的理論：「人和 AI 能動性(agency)的交互作用」。再者，針對人機協作的模式，本研究提出「人機協作魔力象限(Human-Robot Collaboration Magic Quadrant)」(圖 5-1)，幫助公部門應用 AI 的初步評估與理解 AI 和人互動各種可能性，除了「分工」、「自動化程度」，還有其他的可能性，例如階級關係(hierarchy)、代理關係(agency)、機器自主(autonomy)、雙重關係(dual)等複雜的關係，以及所衍生人機協作的模式可能包括直接控制、間接控制、演算型決策、協作或合作等。

三、重新審視公共政策的過程：公共政策的過程呈現循環交錯的現象，人與 AI 的能動性(agency)的交互作用

研究發現，公部門使用 AI，人與 AI 之間是彼此能動性(agency)的交互作用，人機協作更頻繁決策功能與速度的過程與現象，並不是「政策規劃—執行—評估」次序性、可控制性的線性思維，而是 AI 即時處理大量數據、自主決策、即時評估，循環交錯的現象的，已改變公共政策過程階段論中次序性的觀點。

四、重新審視公共治理理論原本的方法論：ANT 分析途徑

研究發現，公部門使用 AI 時，ANT 的分析途徑途徑顯現出優異的公共政策分析能力。



第三節 與文獻對話：公部門的 AI 挑戰

有關 AI 在政府治理的挑戰，第二章第三節已梳理公共治理相關理論，並聚焦政府是使用者，將公部門的 AI 挑戰整合分成 5 類：

1. 公共行政改革
2. AI 官僚的合法性
3. AI 系統的品質(資料治理與演算法治理)
4. 在公共治理領域中人機協作的管理
5. 勞動力影響(勞動力替代與轉型)

文獻雖然提出公部門使用 AI 的挑戰，但很大程度上是探索性研究，對於如何因應挑戰的實務性作法較少討論。本研究第四章依據 ANT 的分析途徑探索臺灣公部門使用 AI 的經驗故事，研究發現，公部門使用 AI 實非易事。本節著重於與文獻對話，探究在臺灣尚未受到公共治理學術界關注領域的理論缺口：臺灣公部門實證案例的 AI 治理挑戰。

茲整理研究個案在公部門職能應用領域、優勢或前提、挑戰，以利後續探討臺灣公部門如何因應使用 AI 的治理挑戰。

壹、財政部「智能稅務服務計畫」

- 一、公部門職能應用領域：縮小稅務員查核重點和範圍，強化賦稅公平的治理能力。
- 二、推動 AI 的優勢或前提：建立內部能力、資料治理(財稅大數據資料庫)。
- 三、挑戰



本案例 AI 挑戰如表 5-2。主要是資料來源和可解釋性(interpretation)的挑戰。資料來源的挑戰，包括：人為偵測異常案例過少；不法行為的動態特質，必須不斷地更新逃稅態樣的訓練數據等問題。

財政資訊中心的策略，包括：訓練公務員尋找適合 AI 的稅務主題能力、AI 系統設計者與資深稅務審查人員合作選取適合的演算法模型、持續進行演算法調校以提高正確率。此外，研究發現，有參與 AI 設計的公務員，理解 AI 預測的優勢及限制，於多次驗證測試及試行下，傾向相信 AI 的判斷。因此，**參與是解決 AI 可解釋性的一個重要因素。**

表 5-2 智能稅務服務計畫的 AI 挑戰

AI 治理挑戰	內容
1.公共行政改革	1.技能：缺乏 AI 專業知識 2.人員管理和組織管理 3.財務可行性：沒有編公務預算
2.AI 系統的品質	1. 隱私/安全性：個人資料保護 2. 資料治理：數據資料不一致(data inconsistency)、黑名單和白名單 3. 演算法治理：精確率(60-70%)待提升，希望 80%以上 4. 可解釋性(interpretability/explainability)：資料洞悉分析(data insight analysis)、終端使用者的科技接受度
3.人機協作的管理	演算法趨避和自動化偏誤：終端使用者(公務員)的科技接受度

資料來源：本研究整理



貳、環境部「智慧判煙系統」

一、公部門職能應用領域：環境的持續性監測，強化政府信任、社會安全的治理能力。

二、推動 AI 的優勢或前提：

- 1.研發 AI 工具並取得臺灣專利權。
- 2.AI 智慧判煙較目測判煙更具成本效益、穩定和高正確率、節省人力，且提高違法行為的監管能力。

三、挑戰

本案例的 AI 挑戰如表 5-3。主要是 AI 官僚的合法性、AI 系統的品質。

實務上，諸多直轄市、縣市政府已將 AI 當成「篩選工具」，再由公務員後續裁決(人類決策)；環境部則將 AI 當成「直接裁決(adjudication)的工具」(AI 決策)，並修正法規，新增「影像判煙」，將 AI 官僚決策合法化，顯示 **AI 應為分析單位**。

其次，如何確保 AI 系統的品質高正確率、低偏誤(bias)、判讀結果穩定？

環境部的策略，將「人員的管理—減少人類主觀式偏差管理」，原本目測判煙執法人員須訓練合格；新增「AI 的管理—AI 的偏誤(bias)管理」，AI 自動決策判定結果的管理。管理方式包括：採用統計檢定、AI 判讀結果在管制點附近不會開罰、防干擾措施、AI 決策結果資訊透明化、裁罰證據具再現性等。最後，為避免 AI 能動性的影響，加強稽查人員操作 AI 系統的教育訓練、訂定儀器校正的 SOP，避免公務員對 AI 信任度過高或過低的問題。



表 5-3 智慧判煙系統的 AI 挑戰

AI 治理挑戰	內容
1.公共行政改革	1. 技能：缺乏 AI 專業知識、環保稽查工作內容變化 2. 財務可行性：一個 AI 工具成本約 50 至 76 萬元
2.AI 官僚的合法性	1. AI 法律與法規的控制和治理、責任和課責：法規測量方式為目測判煙，AI「直接裁決」需具有合法性、以及不服「AI 官僚」裁罰的救濟管道 2. 社會信任/公信力或透明度：終端使用者(廠商)的科技接受度
3.AI 系統的品質	1. 資料治理：背景白雲誤認為煙，或同時有好幾根煙 鹵煙重疊，利用修圖技巧影響拍照及 AI 判讀結果 2. 演算法治理：正確率約 93%
4.人機協作的管理	如何使用 AI 系統、儀器校正的 SOP

資料來源：本研究整理

參、臺北市政府「AI 智慧交控管理」

一、公部門職能應用領域：交通監控、紓解高交通流量，強化社會安全、低碳排
治理能力。

二、推動 AI 的優勢或前提：採購市場上 AI 智慧號誌控制系統，使用公部門場域
資料進行演算法調校，將模型正確率從 60%提高至 92%-96%。

三、挑戰

本案例的 AI 挑戰如表 5-4。主要是資料品質、AI 技能、資料品質（漏偵



測)。臺北市政府的策略：

1. 資料品質：交控管理重要的是趨勢，不需要百分之百精準，AI 正確率考慮『效率、時間、成本』，針對 AI 偏誤(bias)，當 AI 偵測錯誤，除資料清洗(Data cleansing)外，設置專業閾值參數、降階運作、行人觸動號誌、每日審查 AI 系統、人類隨時監控 AI 等方式，避免影響交通安全。
2. AI 技能挑戰：公私協力、政府採購、跨單位等跨域治理，導入 AI 技術過程中，公務員邊做邊學的經驗學習(Experiential Learning)克服 AI 技能落差。

表 5- 4 AI 智慧交控管理的 AI 挑戰

AI 治理挑戰	內容
1.公共行政改革	1.技能：交控中心人員專長多數是交通管理 2.組織和管理：規劃期間額外工作負荷 3.政府採購法：AI 開發不易適用政府採購法 4.創新風險：AI 交控是很新的做法，無法確保會成功
3.AI 系統的品質	1. 資料治理：AI 模型擷取戶外行人和車子的即時動態影像，可能會受到雨、天候、光線、行人的衣服、拍攝角度等影響而漏偵測。 2. 演算法治理：正確率白天約 96%、晚上約 92%

資料來源：本研究整理

肆、勞動部「1955 多元智能電服中心」

一、公部門職能應用：公民服務，強化社會公平、SDGS 的治理能力。

二、推動 AI 的優勢或前提：跨域治理、AI 系統「以人為本」的設計、人機介面



整合(Human-machine interface integration)。

三、挑戰

本案例的 AI 挑戰如表 5-5。主要是勞動力替代與轉型、人機協作的管理、訓練資料來源。勞動部的策略：

1. 勞動力替代與轉型：AI 模型的設計過程，利害關係人(包括身心障礙的終端使用者) 全程參與 AI 模型建置，達成設計 AI 系統初始目標→「賦能身心障礙者，讓電話服務人員以不同的方式做事，而非取代」，此時，「技術是器官、是完全的夥伴」(王志弘、高郁婷譯, 2023)，AI 技術成為身心障礙者基本人權。再者，電話服務中心全員「職務再設計(Job Accommodation)」、AI 技術「在職訓練(On-Job-Training)」，動態調整工作方式、工作流程、邊做邊學(Learning by doing)，為電話服務人員開拓新職業類別：AI 資料分析師、AI 訓練師、AI 標註員、知識庫建構等工作。最後，工作效率與薪資制度關聯，提高身心障礙者的科技接受度。
2. 人機協作的管理：使用者介面無障礙設計，AI 系統符合無障礙規範 AA 等級。評估人與 AI 合作的最佳工作狀態是人與 AI 的協作或合作，AI 第一線&人類第二線的服務架構；第一線 AI chatbot 取代公務員成為「隱藏官僚」，當 AI chatbot 無法回答時，AI chatbot 成為電話服務人員的賦能工具，藉由整合人和 AI 協作來創造價值，使人和 AI 相結合擴增和強化彼此的能力，達成人機共同協作才能具備的某些能力。
3. 訓練資料來源：利用「AI 訓練 AI」¹⁷⁰提高 AI chatbot 正確率，此時，AI 成為參與社會型塑的行動者。「AI 訓練 AI」比「人訓練 AI」更有效，AI 成為 AI Chatbot 的「訓練師」，顛覆公共治理理論以人為中心的觀點，回應 ANT 廣義的對稱性原則。

¹⁷⁰ 利用 AI 語音辨識(Speech-To-Text, STT)、生成式 AI，訓練 AI Chatbot。



表 5-5 1955 多元智能電服中心的 AI 挑戰

AI 治理挑戰	內容
1.公共行政改革	1. 組織和管理：組織路徑依賴，抗拒數位轉型
3.AI 系統的品質	1. 資料治理：建構勞動領域問答語料庫及知識庫 2. 演算法治理：2021 年 AI chatbot 正確率 99%、覆蓋率 95%、滿意度 83%
4.人機協作的管理	1. 終端使用者的接受度：如何提高身心障礙者對 AI chatbot 和 KMS 的科技接受度 2. 使用者介面：無障礙設計
5.勞動力影響	電話服務員擔心被 AI Chatbot 取代

資料來源：本研究整理

伍、勞動部「AI 人機協作分文」

一、公部門職能應用：提高官僚效率、公部門營運管理(提高分文效率)，強化人力資源運用(節省分文人力及時間)的治理能力。

二、推動 AI 的優勢或前提：建構事先分類的資料庫。

三、挑戰

本案例的 AI 挑戰如表 5-5。主要是政策制定者的創新困境與可解釋性(interpretation)。勞動部的策略：

1. 政策制定者的創新困境：電子公文系統(EDMS)是公部門的繁文縟節，AI 人機協作分文在臺灣其他政府部門沒有類似案例，AI 系統開發成效無法預期。AI 系統的建置是利用跨部門(cross - sectors) (AI 廠商、政府)、跨單位

(cross-departments) (業務單位、資訊部門)的合作，克服政策制定者的創新困境。



跨部門、跨單位合作克服政策制定者的創新困境。

2. 可解釋性：研究發現，有參與 AI 系統設計的公務員，較理解 AI 系統決策的優勢和侷限性，有較佳的人機協作關係和工作績效；沒有參與的，即使 AI 系統分文正確率(Accuracy)約 85%- 90%，傾向不信任 AI 系統或覺得 AI 系統沒有用。

表 5- 6 AI 人機協作分文的 AI 挑戰

AI 治理挑戰	內容
1.公共行政改革	1.組織管理：政策制定者的創新困境，增加參與 AI 模型建置公務員的工作負荷 2.技能：公務人員多數是勞工行政背景
3.AI 系統的品質	1.演算法治理：正確率(Accuracy)約 85%- 90% 2.可解釋性(interpretability/explainability)：終端使用者的科技接受度
4.人機協作的管理	演算法趨避和自動化偏誤：公務員對 AI 的信任度過高或過低

資料來源：本研究整理

陸、小結

從 ANT 的視角觀察研究案例，研究發現，*AI 情境很重要(AI context matters)*，AI 可以用於不同的政府職能或公共服務應用領域，產生意想不到的 AI 挑戰，而各公部門設計不同的配套措施因應 AI 的治理挑戰，回應 Jastroch 所提 AI 風險緩解很重要(risk mitigation matters) (Jastroch, 2021)。然而，Jastroch 雖然針對 AI 風險進行廣泛性的分析，但缺乏針對公共治理領域的 AI 風險緩解措施，且仍屬探索性研究，具體操作性不足。



第四節 公部門 AI 風險緩解機制

技術在複雜的脈絡中運行，使得 AI 技術自身及其性質並非充分條件，設計最好的 AI 技術不見得會贏，甚至出現技術的意圖無法實現的非意圖結果，有文獻即建議研究公部門使用 AI 時的風險管理實際做法(Zuiderwijk et al., 2021)。有鑑於公部門應用 AI 必須更加謹慎，除了技術性作法，本節根據 ANT 的分析途徑歸納臺灣公部門 AI 風險緩解機制、爭議權衡的做法，提出共通性原則，併同技術性作法，回應本文在實踐層的研究目的：

1. 提升公共治理效能及控制風險，提高 AI 專案的成功率？
2. 爭議權衡的實踐做法？如何權衡 (1). 科技創新原則、(2). 人類監督強度、(3). AI 系統的品質、(4). AI 在政府職能的應用領域、(5). 價值選擇等。

壹、公部門 AI 風險緩解機制的定義與意涵

AI 的自適應性與能動性會不會操控公務員？或是影響公務員的認知能力及專業判斷？成為公共治理的棘手問題(Wicked problem)。當 AI 從工具或輔助者的概念，轉變具有主動性、自主決策(autonomous decision)時，應如何管理官僚組織中人與 AI 共存的世界？應如何提出前瞻式的評估機制，在行政上建構前導式系統以及防火牆？應如何驗證 AI 系統的可信任性，防備 AI 可能的風險？開發 AI 系統最大挑戰之一，是必須理解真實世界情況相關的脈絡意義(Jarrahi et al., 2022; Oleinik, 2019)和糾正平衡(Pitt et al., 2023)。

Gartner 針對企業機構提出 10 大戰略技術趨勢，2023 年、2024 年連續 2 年均包括對 AI 的信任、風險和安全管理(AI trust, risk and security management, AI

TRiSM)¹⁷¹。AI TRiSM 主要是避免 AI 模型可能產生脫離控制的*技術性作法*，內容包括模型運維(ModelOps)、資料保護、AI 特定安全、模型預測控制。Gartner 預測到 2026 年，採用 AI TRiSM 可以篩除多達 80%的錯誤來提高決策的準確性。另一方面，從新興 AI 治理研究社群提出的理論、框架、策略和指南，或歐盟的 AI 法案，甚至臺灣的指引，追根究底，某種程度上都是識別和緩解 AI 風險，或是有類似的因子在裡面。然而，這些倫理原則、指南、框架和政策策略等，很大程度上是探索性研究，且缺乏系統性(unsystematic)、太高層次(too high-level)、高度抽象(a highly abstract level)，無法具體實踐(Henman, 2020; Whittlestone et al., 2019)，茲說明原因如下：

1. 缺乏系統性：原則太多，而且這些 AI 倫理原則中有許多相互矛盾(B. Zhang & Dafoe, 2020)，並未整合成明確的概念或定義。
2. 太高層次：過度追求倫理準則，可能陳義甚高，阻礙創新。
3. 高度抽象：多屬探索性研究，無具體操作細節，不易執行。

因此，若欲在公部門使用 AI，究竟應如何緩解 AI 風險？本文歸納臺灣公部門緩解 AI 風險的具體作法，研究過程中，受訪者多以「配套措施」表達，本文統一以「公部門 AI 風險緩解機制(AI Risk Mitigation Mechanism of public Sector)」表達，歸納後提出公部門的 AI 風險緩解機制意涵。

本研究提出公部門 AI 風險緩解機制意涵：「*基於 AI 是行動者、有能動性，設計風險管理措施，使 AI 系統產出的結果在可以控制的範圍內*」；概念圖 5-2。

¹⁷¹ 檢自 https://www.profetai.com/en/post/gartner_2023_aia_recap



AI 風險緩解機制意涵詳細的概念說明如下：

1. 理論層次：

- (1). AI 是行動者、有能動性：使 AI 系統產出的結果在可以控制的範圍內，降低 AI 風險。
- (2). AI 偏誤的補救：權衡 AI 系統的正確率、成本、效能、公共價值。

2. 原則層次¹⁷²：

- (1). 識別不正確或誤導的結果：公務員監督或使用自動審查機制。
- (2). 公部門使用 AI 的管理：建立治理和符合法規的框架。
- (3). 保障資料安全：禁止公務員洩漏組織和個人敏感資訊。
- (4). 避免違規使用 AI：使用安全控制措施偵測可疑的違規行為。

3. 具體實踐層次

包括公共行政方法及資訊科技風險管理方法：

- (1). 公共行政方法：依據公部門 AI 治理的 5 種挑戰：公共行政改革、AI 官僚的合法性、AI 系統的品質、人機協作的管理、勞動力影響等，設計降低 AI 風險的措施。
- (2). 資訊科技風險管理方法：依據風險管理生命週期，從威脅識別到定量和定性風險分析，利用接受(Accept)、避免(Avoid)、轉嫁(Transfer)、防範(Deter)、拒絕(Reject)等，設計降低 AI 風險的措施。

¹⁷² 1. The implementation of manual or automatic result reviews is crucial in identifying incorrect or misinformed outcomes. 2. It is imperative to establish a governance and compliance framework that governs the utilization of AI-embedded applications in the public sector. 3. To ensure data security, it is essential to prohibit civil servants from disclosing sensitive organizational and personal information. 4. Monitoring unsanctioned use of AI-embedded applications using existing security controls is vital in detecting any suspicious policy violations.

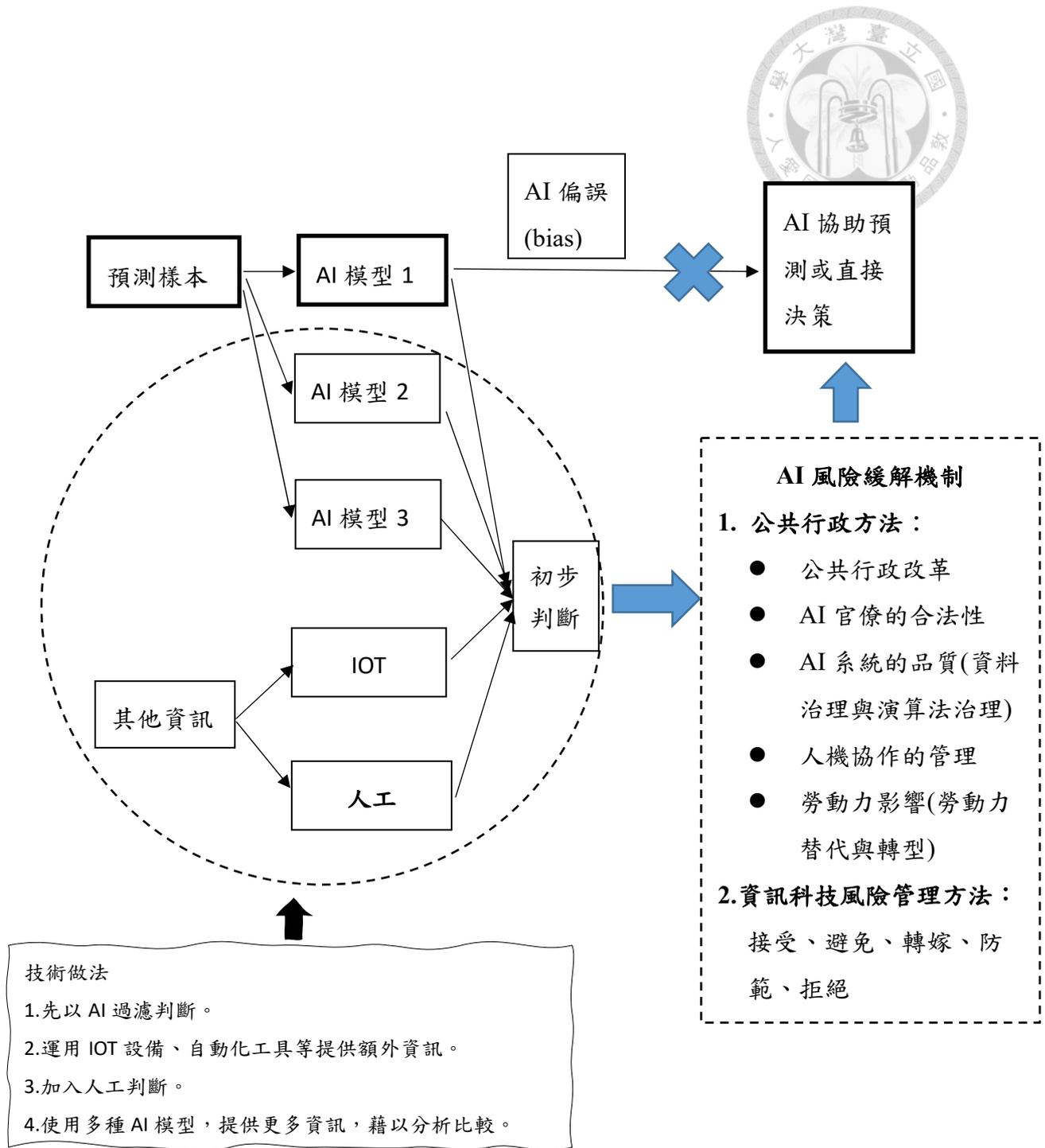


圖 5-2 公部門 AI 風險緩解機制概念圖

資料來源：本研究整理[AI 技術廣泛且進程一日千里，本概念圖係以任務導向的 AI 模型(task oriented AI moduls)示範，對通用性生成式 AI 不適用。]



經梳理公部門 AI 風險緩解機制意涵後，接續歸納研究案例的 AI 風險緩解方法：公共行政分析方法、資訊科技風險管理方法。

貳、公部門 AI 風險緩解機制—公共行政分析方法

依據公部門 AI 治理的 5 種挑戰：公共行政改革、AI 官僚的合法性、AI 系統的品質(資料治理與演算法治理)、人機協作的管理、勞動力影響(勞動力替代與轉型)，歸納研究案例的 AI 風險緩解機制(彙整詳如表 5-7)，並就詳細作法論述如後。

表 5-7 公共行政方法之 AI 風險緩解機制彙整表

AI 挑戰	主要項目	AI 風險緩解機制
1.公共行政改革	1.組織和管理	首長或高層支持、推動文化轉型、專責任務編組(AI 團隊)、機器人流程自動化(RPA)、編列常規預算、鼓勵創新機制、調整薪資制度、常態化的 AI 在職訓練、政策溝通、公務員公共服務動機與熱誠。
	2.財務可行性	調整現有預算、申請計畫補助、配合經費狀況調整戰術。
	3.科技接受度	1.若 AI 使用在和公民權利義務相關的領域時，AI 當作篩選的工具。 2.策略性的運用「人機協作」和「AI 風險緩解機制」，適度降低人類監督的程度。
	4.政策制定者的創新困境	跨域創新、多點創新、從資料出發、從現有系統發展、從小做起、從非核心職能開始
	5.政府職能或公共服務領域	大多數政府職能都適合導入 AI，仍需審慎研究可行性。
2.AI 官僚的合法性	1.個人資料保護	1.個人資料保護符合相關法律和倫理規範。 2.隱私/安全性、社會信任、個資假名化機制。
	2.AI 官僚合法性	1.修正法規使 AI 官僚合法化。 2.課責和責任。
	3.可解釋性	1.參與：領域知識者參與 AI 模型開發。 2.透明化：AI 決策結果資訊透明化。
3.AI 系統的	1.AI 系統品質	1.主要評估方式「比人工好就採用」；比人工好

品質		<p>就必須有一定程度的 AI 系統品質。</p> <p>2.AI 系統品質評估：「有一定程度的 AI 系統品質，以及 AI 風險緩解機制。」</p>
	2.資料治理	<p>1.選擇適合的 AI 主題。</p> <p>2.建立以 AI 為目標導向的大數據資料庫或資料資源。</p> <p>3.演算法模型的選取與持續調校。</p>
	3.演算法治理	<p>1.選擇適合的演算法模型。</p> <p>2.領域知識者參與資料治理與 AI 模型調校。</p> <p>3. AI 系統資訊品質的治理重點：不是在源頭，重點是在結果。</p>
4.人機協作的管理	1. 終端使用者的科技接受度	<p>以人為本的 AI 設計、人機介面整合、選擇人機協作模式、評估人與 AI 合作的最佳工作狀態、工作效率與薪資制度關聯。</p>
	2.演算法趨避和自動化偏誤	<p>1.行政性作法：設立風險值</p> <p>2.技術性作法：驗證能動性。</p>
5.勞動力影響	1.缺乏 AI 專業知識和專家	<p>1.多元化導入策略：建立內部 AI 能力、政府主導研究計劃、政府採購、公私協力、跨域治理。</p> <p>2.越複雜政府職能，或與公民權利高度相關領域，越需要政府自己培養人才。</p>
	2.文官體系的 AI 轉型	<p>建構 AI 生態系統使 AI 成為賦能工具。</p>

資料來源：本研究整理



一、公共行政改革

(一)、組織和管理：高層支持

主要挑戰，數位轉型期間人力缺乏、工作負荷加重、不願意改變現有的工作方式、終端使用者的科技接受度等。AI 風險緩解機制包括：

1. 首長或高層支持
2. 推動文化轉型
3. 專責任務編組(AI 團隊)
4. 常態化的 AI 在職訓練
5. 編列常規預算
6. 調整薪資制度
7. 鼓勵創新機制
8. 機器人流程自動化
9. 政策溝通
10. 公務員公共服務動機與熱誠

(二)、財務可行性：調整現有預算

主要挑戰，沒有預算。AI 風險緩解機制包括：

1. 調整現有預算
2. 申請計畫補助
3. 配合經費狀況調整戰術(例如分階段作業，縮小實作範圍等)

此外，智慧判煙系統案例，AI 工具成本比人力低，在可負擔範圍；智能稅務服務計畫首長或高層支持調整現有預算等策略，都是克服缺乏預算的方法。

(三)、科技接受度：

主要挑戰，人類監督程度太高，勢必降低 AI 比人「快」與「準」的技術優



勢、以及公部門的科技接受度；人類監督程度太低，可能提高 AI 風險程度、社會的科技接受度。AI 風險緩解機制包括：若 AI 使用在和公民權利義務相關的領域時，AI 當作篩選的工具(智能稅務服務計畫)；或策略性的運用「人機協作」和「AI 風險緩解機制」，適度降低人類監督的程度。

(四)、政策制定者的創新困境：跨域創新等原則

主要挑戰，是 AI 系統開發成效無法預期。AI 風險緩解機制包括：公部門 AI 創新包括：跨域創新、多點創新、從資料出發、從現有系統發展、從小做起、從非核心職能開始。

1. 跨域創新：跨部門(cross - sectors) (AI 廠商、學術界、政府)、跨層級(cross - levels)、跨單位(cross-departments) (業務單位、資訊部門)合作創新模式。
2. 多點創新：財政部內每單位每年都要提案實作，勞動部有 AI 評獎實施計畫，類似美國聯邦政府、歐盟的多點試驗(pilot)的策略(Luca et al., 2022)。
3. 從資料出發：盤點現有資料庫，評估那些資料適合發展 AI 主題(topics)? 是否有高品質和足夠的數據資料和資料庫？再者，AI 系統開發有 60-80%(視個案情況)的時間是花在資料工程(data engineering)(資料準備)、特徵工程(feature engineering) (資料清理、分類數據、特徵提取或特徵發現、特徵儲存等)。
4. 從現有系統發展：先從現有系統(AI 協作分文、1955 多元智能電服中心、智能稅務服務計畫、AI 智慧交控管理等)，或現正推動之業務為標的(智慧判煙系統)作為創新實驗方案，系統設計必須涵蓋可擴充性及彈性的空間考量，可提高成功率(丁玉珍、林子倫, 2020)。
5. 從小做起：透過試辦(AI 智慧交控管理)，累積後續政策擴散的能量。
6. 從非核心職能開始：基於 AI 應用於公部門相對新穎，幾乎都是在經驗學習

中培養 AI 技能；再者，政府非核心職能工作多屬重複性高、繁文縟節，藉由導入 AI 系統提供公務員經驗學習的機會，也避免失敗時影響政府核心職能的運作。



(五)、政府職能或公共服務領域：大多數政府職能都適合導入 AI，惟仍需審慎研究可行性

主要挑戰，Mehr、Bullock、Agarwal (Agarwal, 2018; Bullock, 2019; Mehr et al., 2017)等人對於 AI 在某些政府職能(function)的應用提出諸多關切(concern)。本研究發現，行政機關不論是低階重複性工作(例如 AI 智慧交控管理，AI 智能協作分文)、高階政策性決策(例如智能稅務服務計畫)、與人權利義務相關的行政裁量權工作(智慧判煙系統)、行政諮詢服務工作(例如 1955 多元智能電服中心)等，都可以應用 AI。有鑑於 AI 的採用將造成的權力不平衡(Pitt et al., 2023)，仍需審慎研究可行性。

大多數政府職能都適合應用 AI，仍需審慎研究可行性。

另依據 Engstrom et al.(2020)替美國聯邦行政會議撰寫的調查報告(Engstrom et al., 2020)，將 AI 應用於政府機關的業務分五類¹⁷³：

¹⁷³ 原文：1.enforcement：Tasks that identify or prioritize targets of agency enforcement action.

2.regulatory research, analysis, and monitoring：Tasks that collect or analyze information that shapes agency policymaking.

3. adjudication：Tasks that support formal or informal agency adjudication of benefits or rights.



1. 強制執行：幫助公部門辨識或優先考慮強制執行的目標。
2. 監理分析與監控：蒐集或分析影響機構決策的資訊。
3. 裁決：支持正式或非正式機構對福利或權利進行裁決的任務。
4. 公共服務：直接向公民提供服務、或為監管、或其他目的促進與公民溝通的任務。
5. 內部管理：維持機構資源管理，包括員工管理、採購和技術系統維護。

本文依據上述分類搜尋臺灣公部門導入 AI 的案例，發現不同政府職能、層級、組織型態(中央、地方政府、非政府組織)均有使用 AI 的案例(詳如表 5-8 臺灣公部門導入 AI 案例)。另有關內部管理，雖源自於內部管理的需求，但外溢效果可以提高公民服務品質、政府透明度、公平性等(例如提供身心障礙者不同障別知的權利)。

因此，本研究發現：「大多數政府職能都適合導入 AI」，與「維持美國 AI 領導地位的行政命令」指出 AI 將影響幾乎所有的行政機關(構)的任務，和 Gartner 2021 年調查美國、德國和英國已部署 AI 或打算在 3 年內部署 AI 的組織，80% 的高層認為 AI 可以應用於任何業務決策¹⁷⁴的觀點相同。此外，文獻提及現階段主要問題是政府職能應用領域太集中、進展分散且不平衡，並在公共治理中產生

4. public services and engagement : Tasks that support the direct provision of services to the public or facilitate communication with the public for regulatory or other purposes.

5. internal management : Tasks that support agency management of resources, including employee management, procurement, and maintenance of technology systems.

¹⁷⁴ 檢自 <https://www.gartner.com/en/newsroom/press-releases/2022-08-22-gartner-survey-reveals-80-percent-of-executives-think-automation-can-be-applied-to-any-business-decision>

相對集中的影響(Dunleavy & Margetts, 2023)，臺灣公部門亦有類似情況，例如提供政府資訊的聊天機器人。





表 5-8 臺灣公部門導入 AI 案例

職能	案例
1.強制執行	1. 財政部(智能稅務服務計畫) 2. 內政部(AI 智慧無人機橋梁檢測、AI 巡檢道路路況) 3. 衛生福利部(醫療申報資料 AI 協助審查作業)
2.監理分析 與監控	1. 臺北市政府(AI 智慧交控管理) 2. 桃園市政府(分析淹水區域及早佈署調派) 3. 臺中市政府(AI 智慧水線灑水抑制河川空污揚塵) 4. 衛生福利部(建構老人保護個案之 AI 預警分析) 5. 交通部(天氣風險預測) 6. 勞動部(中小企業職場危害監測、實境場域提升勞工防災知能)
3.裁決	1. 司法院(AI 量刑資訊系統) 2. 環境部(智慧判煙系統) 3. 環境部及各縣市政府「聲音照相-科技執法」噪音車取締 4. 交通部及各縣市政府 AI 偵測違規及交通科技執法
4.公共服務	1. 勞動部(1955 多元智能電服中心、職業安全衛生資訊整合串連及 一站式應用分析服務 職業安全衛生資訊整合分析平台) 2. 經濟部(智慧電網、AI 即時淹水預報) 3. 經濟部水利署(智慧防災應答機器人) 4. 交通部(智慧型運輸系統、自動駕駛車輛) 5. 農業部(神農 GPT—農業知識檢索系統、智慧農業) 6. 財政部及各縣市政府稅務諮詢服務機器人

5.內部管理	<ol style="list-style-type: none"> 1. 勞動部(AI 人機協作文) 2. 勞動部(利用生成式 AI 檔案管理，例如立案、編目；AI 智慧徵才，例如 AI 生成履歷、AI 就服員、AI 分流面試叫號) 3. 衛生福利部「中央流行疫情指揮中心嚴重特殊傳染性肺炎記者會」提供即時字幕服務聽覺障礙者 4. 法務部各檢察機關導入中文語音辨識系統產生逐字稿 5. 立法院(2021 年院會議事直播、總質詢與各委員會會議公報產製等)(廖元甫, 趙偉成, 楊智合, & 楊仲捷, 2020)。 6. 公設財團法人機構公共電視「2020 總統大選辯論會」 7. 各政府內部應用生成式 AI(Generative AI, GAI)：生成文字、影像、摘要、會議記錄、致詞稿、翻譯、繪圖、寫程式等
--------	---

資料來源：本研究依據 Engstrom et al.(2020)分類整理

二、AI 官僚的合法性

(一)、個人資料保護需符合相關法律和倫理規範

主要挑戰，公部門使用大數據發展 AI 仍受到許多限制，其中一項重要因素是公民對個人資料隱私的喪失感到不安，成為是否能夠順利採用 AI 的關鍵因素(O'Leary, 2014)。舉例來說，我國 2022 年憲判字第 13 號 [健保資料庫案] 作成憲法判決，就個人健康保險資料在原始蒐集目的外之利用違憲¹⁷⁵。

AI 風險緩解機制包括：

1. 個人資料保護符合相關法律和倫理規範
2. 隱私/安全性、社會信任、個資假名化機制等。

除此之外，本研究認為，現階段與其陷入「把人當人，而非把人當成資料的

¹⁷⁵ 檢自 <https://cons.judicial.gov.tw/docdata.aspx?fid=38&id=309956>

來源。」¹⁷⁶的價值爭論，公部門可先從非個資依賴型(來自於非人類的資料)資料先著手，例如環境部智慧判煙系統、臺北市政府 AI 智慧交控管理、1955 多元智能電服中心均屬非個資依賴型資料。



現階段公部門使用大數據發展 AI 仍受到許多限制，從非個資依賴型資料先著手。

(二)、修正法規使 AI 官僚合法化

主要挑戰：現行的一般法律原則是規範人類而非機器的行為。AI 風險緩解機制包括：

- 1.修正法規：使 AI 官僚合法化，例如環境部智慧判煙系統修正「固定污染源空氣污染物排放標準」。
- 2.課責和責任：民眾不服「AI 官僚」決策，可依法制作業程序提起訴願。

(三)、AI 的可解釋性(interpretation)行政性作法

主要挑戰：資料洞悉分析(data insight analysis)、終端使用者的科技接受度等。AI 風險緩解機制包括：

1. 領域知識者「參與」AI 模型的開發是可解釋 AI 的行政性作法，可使「可解釋性和 AI 系統的效能/正確率呈正相關」

文獻指出，AI 系統的效能/正確率(performance/accuracy)和可解釋性(explainability)存在反比關係(Dwivedi et al., 2019)。研究發現，公務員(領域知識者)參與資料治理，與 AI 模型的協作(collaboration)，可同時提高 AI 系

¹⁷⁶ 法學專家怎麼看健保資料庫憲法訴訟？檢自 <https://ai.iias.sinica.edu.tw/national-health-insurance-db-constitutional-litigation-minutes/>

統的效能/正確率和可解釋性/科技接受度(智能稅務服務計畫、1955 多元智能電服中心)。



2. AI 決策結果資訊「透明化(transparency)」

文獻指出，管制性的 AI 工具，若開放技術細節的原始碼，當受管制廠商已知演算法使用特定變量或截止值(cutoffs)時，可以操縱這些變量、博弈(gaming)、對抗性學習(adversarial learning)，或調整自己行為等方式，破壞執法工具的有效性(Engstrom et al., 2020)。因此，管制性 AI 工具的原始碼，常常以商業秘密(Yeung, 2018)，或受商業秘密法等相關法律保護為理由，採取不透明的作法(Whittaker et al., 2018)。

AI 風險緩解機制(以智慧判煙系統為例)包括：AI 系統產出資訊的「透明度」、裁罰證據的「再現性」，所以沒有開放 AI 系統原始碼的問題。

最後，雖然目前針對可解釋性人工智慧(XAI) 是技術處理，開發能夠解釋與詮釋學習模型的 AI 演算法，但技術開發成效無法預期，現階段可先使用可解釋 AI 的行政性作法：參與與透明化。

可解釋 AI 的行政性作法：領域知識者「參與」AI 模型的開發，AI 決策結果資訊「透明化」。

三、AI 系統的品質(資料治理與演算法治理)

(一)、AI 系統品質的評估

主要挑戰：AI 模型的品質(例如正確率等)，受到下列因素限制：

- 1.模型建立階段：主題選取、訓練樣本的取得、訓練樣本的多寡、使用的機器學習演算法、是否有足夠的運算資源及時間(權衡經濟邏輯與公共價值)。
- 2.模型預測階段：預測樣本資料的完整性、預測樣本所處環境干擾、是否有足



夠的運算資源及時間。

AI 系統功能與效率的品質評估(Quality Assessment)非常困難，在技術領域已有相當多評估 AI 系統品質的研究(J. Davis & Goadrich, 2006; Ekelund, 2017; Flach & Kull, 2015; Kuwajima & Ishikawa, 2019)。

研究結果顯示，主要評估方式是：「比人工好就採用」，亦即 AI 系統品質(正確率、精確率等)的評估方式是**跟人類比較**，從 80-99%均有。即使以環境部智慧判煙系統案例，認為 AI 系統品質的正確率很重要為例，AI 系統品質的評估方式之一亦是「比人工好」。而 AI 風險緩解機制，諸多文獻指出使用人類監督(Human monitoring)，亦即人工判斷電腦決策是否合理；然而，過度追求 AI 系統軟體品質及功能(例如正確率等)來降低機器學習的偏誤(bias)，將大幅提高 AI 模型的建置成本，必須在 AI 系統的品質與成本之間權衡。因此，不一定要追逐高標準的 AI 系統品質，關鍵在於識別 AI 的偏誤(bias)與可能的風險，在公共行政流程中搭配經常性的『配套措施』，確保能實踐公共價值與倫理準則(合法性、公信力、透明度、非歧視、個人資料保護等)。綜上分析，AI 系統品質的評估在於將技術限制與行政實務相結合：

除特定政府職能與用途需要關注 AI 系統的品質，AI 系統品質評估：

「比人工好就採用」，比人工好就必須有一定程度的 AI 系統品質，以及

「AI 風險緩解機制。」

上述對 AI 系統品質的評估方式觀點，可以回應諸多文獻對於政府職能或公共服務應用領域適合使用 AI 等的諸多論辯 (Busch & Henriksen, 2018; Kuziemski & Misuraca, 2020)，例如有文獻指出公部門應了解哪種 AI 技術適合哪種政府職能的

任務(Duan et al., 2019)；或當「人工智慧裁量(artificial discretion)」取代「人類裁量(human discretion)」時，應仔細考慮政府職能品質的適當閾值(appropriate threshold) (Young et al., 2019)；或是 AI 使用在和公民權利義務相關的領域時，技術正確率(Technical accuracy)非常重要(Henman, 2020)等觀點。

(二)、資料治理

主要數據挑戰包括：資料不一致(data inconsistency)、黑名單(blacklist)和白名單(whitelist)、測量偏誤(measurement bias)、資料清洗(Data cleansing)、知識庫建構(knowledge base construction)等。

AI 風險緩解機制包括：採用適合 AI 的主題(topics)、演算法模型的選取與持續調校、訂定標準量測方法、室內和戶外驗證、防弊措施、防干擾措施、不得使用 AI 的情況等。不同案例處理方式各有其領域知識和脈絡性，但可以歸納共通性或關鍵要素：

1. 選擇適合的 AI 主題：找到適合訓練 AI 的資料，或高品質的訓練資料。
2. 建立以 AI 為目標導向的大數據資料庫或資料資源：AI 是 60-80%時間花在數據工程(data engineering)，公部門如果未來想做什麼主題的 AI 應用，現在就要開始準備資料以及資料治理。
3. 演算法模型的選取與持續調校：選擇更適合的演算法模型，並由領域知識人員對訓練資料標記與參與模型調校。

(三)、演算法治理

主要挑戰是「正確真實 (accurate ground truth)」訓練數據不易取得的資料品質問題、AI 模型的可解釋性。

AI 風險緩解機制，包括：



1. 選擇適合的演算法模型。
2. 領域知識者參與資料治理與 AI 模型調校，避免 AI 演算法隱含偏誤(implicit biases)。
3. AI 決策結果資訊透明化。

從技術的觀點，雖然已致力於「可解釋 AI (Explainable Artificial Intelligence, XAI)」的研究，然而，演算法的模型設定時是不是有正式的偏好？會產生正式的偏好？也許設計者可能有，目前仍無法明確釐清。所以應如何避免演算法可能產生的偏誤或歧視？**實務上較具可行性的作法，關切 AI 系統決策產出的結果，及其真正影響性，讓結果可以被控制，設計避免 AI 結果含主觀偏見或歧視的控制措施。**一般對於管理的概念是著重於源頭管理，但對於 AI，因為大部分使用者沒有辦法從源頭判斷，即便是學資訊科學，若不是學演算法也沒有辦法判斷是不是有選擇上的偏好，建議著重於管理結果。

AI 系統資訊品質的治理重點：不是在源頭，重點是在結果。

四、人機協作的管理

主要挑戰：終端使用者的科技接受度、演算法趨避和自動化偏誤。AI 風險緩解機制，包括：

(一)、終端使用者的科技接受度

1. 以人為本的 AI 設計、人機介面整合：AI 是賦能的工具，在設計階段就應該充分考量「以人為本的 AI 設計」、使用者介面具有用性及易用性、符合終端使用者的需求，例如身心障礙者的需求，符合無障礙規範 AA 等級。
2. 選擇人機協作模式、評估人與 AI 合作的最佳工作狀態：搭配工作流程設計(workflow design)、職務再設計(Job Accommodation)、在職訓練(On-Job-



Training)，使終端使用者理解 AI 的優勢和侷限性。

3.工作效率與薪資制度關聯

當使用原來的工作方式就能辦理業務，且學習 AI 新技術很困難時，會使終端使用者抗拒改變。AI 風險緩解機制包括：藉由 AI 提高工作能力來增加工作價值，再利用「調整薪資制度」(1955 多元智能電服中心)、績效獎金(智慧判煙系統的稽查獎金)，當工作效率與薪資制度關聯，將提高終端使用者的科技接受度。

(二)、避免演算法趨避和自動化偏誤

1.行政性作法：設立風險值，超過風險值的因應措施。

(1).使用行政規則排除 AI 偏誤(bias)：例如 AI 系統判讀結果在管制點附近，不視為違反規定(智慧判煙系統)。

(2).降階運作：例如超過風險值，立即切換到原始設定(AI 智慧交控管理)。

(3).人類監督(human-in-the-loop)：例如 AI chatbot 無法回答(風險值)，轉接真人文字客服或改由人類客服(1955 多元智能電服中心)。

2.技術性作法：驗證能動性

在人機協作的過程中，人類對 AI 的信任度過高或過低都會產生負面影響(Okamura & Yamada, 2020; Schmidt & Biessmann, 2020)。有研究提出人機協作技術性的干預方式：驗證能動性以改善人類對 AI 的信任度過高或過低所產生負面影響(Okamura & Yamada, 2020; Schmidt & Biessmann, 2020)，透過監測使用者依賴行為，當檢測到過度信任或信任不足時，向使用者提出認知提示：「信任校準提示 TCC(trust calibration cues, TCC)」，提醒他們重新校正其信任度，可以顯著改善過度信任或不信任的問題(Okamura & Yamada, 2020)。

雖然在本文研究案例中並未發現使用驗證能動性的作法，然而，2023年10月5日立法院司法委員會立委林淑芬質詢 AI 草擬裁判草稿時，表示「如何了解法官參採 AI 比例高低？」¹⁷⁷，已提出驗證能動性來增加公民、公務員對 AI 的信任度的需求性與必要性。

五、AI 對社會影響(勞動力替代和轉型)

主要挑戰是缺乏 AI 專業知識和專家、勞動力替代和轉型。AI 風險緩解機制包括：

(一)、多元化 AI 採用策略：分為導入階段、執行階段、混合階段

1. 導入階段：採用多元化導入策略

- (1). 建立內部 AI 能力：訓練公務員自己設計 AI 模型(智能稅務服務計畫)。
- (2). 政府主導研究計劃：學術研究成果與實務需求結合，共同研發 AI 模型(智慧判煙系統)。
- (3). 政府採購：利用政府採購法採購市場上 AI 技術成熟且類似的 AI 模型，使用公部門場域資料進行演算法調校(AI 智慧交控、AI 人機協作分文)。
- (4). 公私協力、跨域治理：政府、AI 技術外包廠商、NGO 團體合作，公私協力合作開發 AI 模型(1955 多元智能電服中心)。

然而，不論哪一種方式，公務員均積極參與提供公部門場域訓練資料，或參與演算法調。越複雜的政府職能，或是與公民權利高度相關的領域，越傾

¹⁷⁷ AI 草擬裁判草稿計劃說不清。from：檢自

<https://tw.news.yahoo.com/ai%E8%8D%89%E6%93%AC%E8%A3%81%E5%88%A4%E8%8D%89%E7%A8%BF%E8%A8%88%E5%8A%83%E8%AA%AA%E4%B8%8D%E6%B8%85-%E7%B6%A0%E5%A7%94%E8%BD%9F-%E4%B8%8D%E8%B2%A0%E8%B2%AC%E4%BB%BB-%E8%83%A1%E8%AA%AA%E5%85%AB%E9%81%93-061940472.html>

向政府自己培養人才，特別強調的是，沒有一個案例完全依賴外包廠商或外部專業知識。



越複雜政府職能與公民權利高度相關領域，政府越傾向自己培養人才。

2.執行階段：增設 AI 技能訓練班。

3.混合階段包括：工作流程設計、職務再設計、AI 在職訓練等。

(二)、文官體系的 AI 轉型：建構 AI 生態系統使 AI 成為賦能(empowerment)工具

過往探討如何提升政府治理能力，主要認為需透過政府文官人力資源體系的變革，著重於官僚、組織競爭力、策略性人力資源管理和能力管理等理論(施能傑, 2006)，並未探討利用 AI 賦能公務員及提升政府治理能力的觀點。若要發展 AI 準備的勞動力(AI-ready workforce)，歐盟委員會(European Commission)訂定「AI 系統應遵循以人為本的設計原則並為人類選擇留下有意義的機會」¹⁷⁸的原則(Giest & Klievink, 2022)。因此，AI 是「削減論點(curtailment thesis)」，還是「賦能論點(enablement thesis)」(Buffat, 2015; Busch & Henriksen, 2018)？取決於建構一個由硬體和軟體組成的 AI 生態系統(AI entire ecosystem)：

1.協作設計(collaboration design)：人與機器互動的設計。

2.機器設計(machine artifact design)：以人為本 AI 設計、人機介面整合設計。

3.機構設計(institution design)：公共行政流程設計、工作流程設計。

4.職務再設計(job accommodation)：政府職務能力設計、工作方法設計。

5.職業訓練設計(vocational training design)：跟隨 AI 系統建置過程的經驗學習

¹⁷⁸ 原文：AI systems should follow human-centric design principles and leave meaningful opportunity for human choice.



設計。

參、公部門 AI 風險緩解機制—資訊科技風險管理方法

資訊科技風險管理探討風險管理生命週期，從威脅識別到定量和定性風險分析，再到風險緩解策略(Kaur & Lashkari, 2021)，包括：¹⁷⁹

- 1.接受(Accept)：IT 系統意識到風險並接受風險以繼續運作。可接受的風險所造成的潛在損失是可以承受的。
- 2.避免(Avoid)：透過消除風險的原因和後果來避免風險。
- 3.轉嫁(Transfer)：將風險轉嫁給保險公司等第三方，以補償由此可能造成的損失。
- 4.防範(Deter)：透過實施最大程度降低不利影響的糾正或預防控制措施來降低風險。
- 5.拒絕(Reject)：最後的選擇是拒絕風險並繼續運作，就像什麼都沒發生一樣。

若依據資訊科技風險管理方法之風險緩解措施分析研究個案，彙整如下：

¹⁷⁹ 原文：1.Accept: The IT system is aware of the risk and accepts it to continue functioning. The potential loss from an accepted risk is bearable. 2.Avoid: Risk is avoided by eliminating its cause and consequences. Shutting down a system and isolating it in case of a targeted attack is an example of avoiding risk. 3.Transfer: Transferring the risk to a third party such as insurance company to compensate for the potential losses incurred due to it. 4.Deter: Risk is reduced by implementing the corrective or preventive controls designed to minimize its adverse impact. 5.Reject: The final option is to reject the risk and continue functioning as if nothing has happened.



表 5-9 資訊科技風險管理方法之 AI 風險緩解機制彙整表

風險緩解措施	研究案例
接受	<p>AI 智慧交控管理：交通流量重要的是趨勢</p> <p>■每提高 1% 的正確率，要多花更多的時間、更多的成本訓練 AI，因此通常不需要百分之百精準。</p>
避免	<p>智慧判煙系統：確保檢測結果一定超標</p> <p>■AI 的正確率(accuracy)是 95%、且 AI 偏誤(bias) 的範圍「非常穩定」，例如在『+5%』之內，則 AI 測出來的結果在『25%』 (20% +5% = 25%) 以上才會開罰。</p>
轉嫁	<p>1955 多元智能電服中心：改由人類接手</p> <p>第一層由 AI chatbot 回應，當 AI chatbot 無法解決民眾問題，接續第二層由真人回答</p>
防範	<p>AI 智慧交控管理：行人交通安全非常重要</p> <p>■諸多風險防範機制，包括：1.風險值；2.降階運作；3.行人觸動號誌；4.每日審查；5.人類監督。</p>
拒絕	<p>AI 人機協作分文：人類監督直接修正錯誤</p> <p>■正確率平均約 85%- 90%，收文人員快速目視掃描確認後立即「一鍵送出」，錯誤分類的公文直接修正轉送正確單位</p>

資料來源：本研究整理



肆、公部門 AI 風險緩解機制—共通性原則

由於本研究設計以公共治理、ANT 及 AI 理論為基礎的案例選擇架構，涵蓋不同政府職能、人和 AI 決策、AI 資訊品質等，能獲得整合性觀點。因此，雖然針對個別案例探究 AI 風險緩解機制，仍可以歸納共通性原則：

一、AI 風險緩解機制應融入公共行政流程，成為經常性的工作。

為了實現 AI 的可信任和可持續性，誠如文獻所說：「汽車需要定期檢查、維護和持續狀態監控，為什麼 AI 技術應該有所不同？」，即強調從穩健性角度，就 AI 生命週期形式化和實踐 AI 維護的重要性，並在整個 AI 生命週期中主動檢測和緩解 AI 風險(P.-Y. Chen & Das, 2023)。在技術層面，智慧判煙系統建立固定頻率的 AI 系統量測標準。在行政層面，AI 智慧交控管理的配套措施亦是經常性的工作。因此，AI 風險緩解機制包括公共行政方法及資訊科技風險管理方法：

- 1.公共行政方法：依據公部門 AI 治理的 5 種挑戰：公共行政改革、AI 官僚的合法性、AI 系統的品質(資料治理與演算法治理)、人機協作的管理、勞動力影響(勞動力替代與轉型)，設計 AI 風險緩解措施。
- 2.資訊科技風險管理方法：依據風險管理生命週期，從威脅識別到定量和定性風險分析，利用接受、避免、轉嫁、防範、拒絕等策略，設計 AI 風險緩解措施。

二、AI 情境很重要(AI context matters)，不同案例應設計不同 AI 風險緩解措施。

AI 部署的大部分挑戰都在實施階段(Campion, Hernandez, Jankin, & Esteve, 2020; Jimenez-Gomez et al., 2020)，不同案例應設計不同 AI 風險緩解措施。舉例來說，目前許多公共管理上懸而未決、爭議權衡的難題，例如如何權衡 (1).科技創新原則、(2).人類監督強度、(3). AI 系統的品質、(4). AI 在政府職能的應用領域、(5).價值選擇等問題。研究發現，AI 技術可以用於截然不同的目的，並產生



截然不同的結果，較務實的做法是將技術限制與行政實務相結合，依據不同政府職能、人類決策和 AI 決策，綜合評估設計適合的 AI 風險緩解機制，控制 AI 可能產生的風險，而無需陷入如何權衡的爭議。

三、當 AI 自主決策用於直接裁決(adjudication)，或對人民權利、義務和安全影響越大，採用 AI 風險緩解機制越多，或監督程度越強。

當 AI 自主決策用於直接裁決，或對人民權利、義務和安全影響越大，採用 AI 風險緩解機制越多(智慧判煙系統、AI 智慧交控管理有 5 種機制設計)，或監督程度越強(AI 選案查稅採人類監督)。例如當 AI 做為與「公民權利高度相關」的裁決時，AI 風險緩解機制(人類監督或人類做最後判斷)提供因果關係的判斷(AI 選案查核)，彌補「是基於可能性(相關性)而不是事實(因果關係)」的風險。當「AI 官僚」行使行政裁量權時，AI 風險緩解機制(AI 官僚的合法化)，彌補目前法規是規範公務員行為而非機器行為的風險(智慧判煙)。當「AI 官僚」可能造成勞動力影響時，AI 風險緩解機制(以人為本的 AI 設計、AI 技術訓練)，使 AI 技術成為賦能工具(1955 多元智能電服中心)，彌補 AI 可能造成技術性失業的風險。

四、公務員具有領域知識及 AI 技術能力，有能力設計公部門的 AI 風險緩解機制，避免受到 AI 自適應性和能動性的影響。

AI 不依照簡單的「若 P 則 Q 邏輯(if-then logic)」，設計公部門的 AI 風險緩解機制需具有脈絡化的領域知識，不能以單純技術外包的概念看待。公務員經過考試、任用、政府體系職務歷練，具有如何避免政策或政府職能產生風險的領域知識。因此，公務員具有 AI 技術的基本知識，理解其運作原理，才有能力設計 AI 風險緩解機制，並將 AI 技術及風險緩解機制融入公共行政流程，包括：

- 1.提出適合的 AI 主題需求。



- 2.提出適合 AI 訓練的數據品質需求。
- 3.參與 AI 系統的設計、開發可信任和以人為本的 AI 系統(a trustworthy and human-centric use of AI)、檢驗 AI 模型、執行驗收作業。
- 4.進行 AI 合法化的法制作業。
- 5.設計 AI 風險緩解機制。
- 6.設計新公共行政流程。
- 7.設計 AI 生態系統。
- 8.設計終端使用者的 AI 技術訓練課程。

再者，以生成式 AI 為例，生成式 AI 的能力很廣，人沒有辦法在任何領域都能判斷真偽，此時，公務員具有領域知識才能與具有自適應性和能動性的 AI 共存，例如擔任 AI 模型檢驗員(P.-Y. Chen & Das, 2023)。當未具有領域知識，或未接受 AI 技能訓練的公務員使用 AI 行使行政裁量權時，將會產生更糟(Barth & Arnold, 1999)的結果，甚至因為 AI 的雙重用途(dual use) (Floridi, 2018)產生權力差距擴大的風險。雖然在 AI 導入初期，公私協力或跨領域 AI 團隊是避免風險的一種方式，長期而言，制度性的讓具有領域知識的公務員具有 AI 技術能力，是降低風險的最重要關鍵。

公務員具有領域知識與 AI 技術能力，有能力設計 AI 風險緩解機制，避免受到 AI 能動性的影響。



第六章、結論

本研究使用行動者網絡理論(ANT)的分析途徑，實證研究臺灣公部門使用 AI 的 5 個案例。以往文獻有關公共治理領域的 AI 理論研究，集中於探索性的概念框架，理論化不足，也缺少證據的支持。本研究則以循證為基礎(evidence-based)，推演公部門使用 AI 的公共治理理論。研究個案包括「財政部智能稅務服務計畫」、「環境部智慧判煙系統」、「臺北市政府 AI 智慧交控管理」、「勞動部 1955 多元智能電服中心」、「勞動部 AI 人機協作分文」。另為使本研究所提出的理論具有強固性，再利用文獻分析法分析國外公部門使用 AI 的 7 個案例佐證本研究的推論成果。

本研究屬跨領域研究，包括公共政策(公共治理理論)與資訊科學(AI)。有關研究方法的選擇，依循以往跨學科(interdisciplinary)和應用研究(applied research)中常見的做法：從其他研究領域借用理論(Berkovich, 2020; 陳金貴, 2013)，本研究再跨領域從科學科技與社會(Science, technology and society, STS)研究領域借用 ANT，使本研究涵蓋 3 個領域的研究範疇，包括公共治理理論、AI、ANT。

本研究從 ANT 的視角，並以過往有關公共治理與 AI 的理論、研究基礎和研究建議為基礎，包括公共治理研究應整合技術與公共治理理論(Bailey & Barley, 2020)、從現有機制設計創新(Misuraca, 2020)、關注 AI 在政府的應用(Misuraca & van Noordt, 2020; Misuraca & Viscusi, 2015)、研究公部門使用 AI 時的風險管理實際做法(Zuiderwijk et al., 2021)、人機協作的理論(Androutsopoulou et al., 2019; Molinari F., 2021; Schoemaker & Tetlock, 2017)、以循證為基礎的實施路線圖(Duan et al., 2019; Misuraca & van Noordt, 2020; Molinari F., 2021)等，藉由臺灣公部門 5



個實證案例、國外公部門 7 個案例，合計 12 個案例，重新審視公共治理理論的方法論，包括公共政策模型、過程論、政策分析方法，以及 AI 與人的合作關係(人機協作)等，逐項探討公共治理理論的缺口及應補充觀點。接續依研究成果提出研究發現、研究貢獻、政策建議，未來研究方向。

第一節 研究發現

依據研究問題與目的，提出主要研究發現。

壹、在公共治理理論層次

一、AI 是行動者並具有能動性，應成為分析單位

現有公共治理理論忽略 AI 是行動者並具有能動性，因而無法有效解釋公部門使用 AI 的現象；從 ANT 的視角，當公部門使用 AI 時，**公共治理理論的分析單位，除了人，應包括 AI。**

本研究從 ANT 的視角發現，AI 能以「AI 官僚」行使裁量權(例如智慧判煙系統)；或以演算型決策影響公部門行動者的感知(perception)、認知(cognition)(例如 AI 協作分文)、影響人類官僚行政裁量權的獨立決策自由度(例如波蘭 2012 年開發 AI 失業評分系統)；或利用「AI 訓練 AI」(例如 1955 多元智能電服中心)；更甚者，利用 AI 操縱選舉結果(美國最高法院 *Rucho v. Common Cause* 案，AI 繪製不公平的選區)，AI 成為政治代理人，可能侵蝕民主並使社會兩極化(英國脫歐公投社交機器人)等案例均顯示，當 AI 呈現直觀且顯而易見、相當強度的行動者特徵(自適應性及能動性)時，促使我們不能再忽視 AI 技術物的作用性。

因此，本研究認為，公共治理理論之所以無法有效解釋公部門使用 AI 的現象，在於過去公共治理理論學術社群彼此溝通的世界觀、知識論與方法論的基本

假設：「以人為本」的治理典範，**忽略AI應該是行動者並具有能動性。**

雖然公共治理理論不曾迴避「非人」，在分析層次或分析單位也會以組織、科技為分析目標。然而，以科技為例，過去科技在社會關係中主要扮演沉默(silent)的行動者，而AI可以在行動者網絡中展演主動性(active)、自動決策(automated decision)，甚至自主決策(autonomous decision)，原因包括：

1. AI可能超過人類可以預測和控制的範圍

目前最受認可的AI學習系統(AI-enabled learning systems)是自適應性學習系統(Adaptive Learning Systems) (Kabudi et al., 2021)。自適應性系統(Adaptive systems)能在情境感知機制(context-aware mechanisms)的驅動下自動改變其行為(Fischer, 2023)。

「AI自適應性(Adaptive AI)：需學習(有資料、能回饋)、可調適(能進步、有彈性)、有限制(有偏誤、會犯錯)」。

再者，傳統的科技或技術物需要人類干預之後產生作用，以腳踏車、原子彈為例，腳踏車、原子彈不會自己動，我們要去「騎」腳踏車，去「按」原子彈的按鈕，那些技術才會有作用和功能。但AI可以在沒有人類干預、沒有被寫入程式的情況下自主學習，不需要人類下指令或操作就能自主決策(autonomous decision)，甚至能自主做出致命的決定¹⁸⁰，使AI可能超過人類可以預測和控制的範圍。

2. AI通用技術的應用範圍超越之前的科技

以原子彈為例，原子彈應用在特定領域，但AI幾乎是全面性翻轉。

¹⁸⁰ 當人工智慧應用在戰場上，未來各國要比的是：誰更會開發AI戰爭工具。檢自 <https://www.wealth.com.tw/articles/52e0d420-8a52-48b3-bde3-788072cfl1d3d>



因此，從公共治理理論的角度，目前正處在因應 AI 發展的早期階段，必須從方法論上徹底檢討，及早因應，彌補公共治理「以人為本」的理論缺口。本研究從 ANT 的視角探索 12 個案例，發現傳統公共治理理論缺少的關鍵要素，並找到可行的解決方案：使用 ANT 主張技術物是行動者(actor)、也有能動性(agency)的觀點。

科學科技與社會(STS)領域中的行動者網絡理論(ANT)，主張任何事物，當然包括非人類和人類都是能動者(agents)，具有改變現狀的能力；也都有潛在的能動性，能對其他能動者的行動產生影響，甚至能動者之間能交互影響(Latour, 2005)。雖然 ANT 主張非人類是行動者也有能動性(agency)，但並不都具有與彼此或特定人類相同類型(type)和強度(intensity)的能動性(Sayes, 2014)，使以前技術物是行動者且有能動性的觀點未受到應有的重視。

然而，當公部門使用 AI 時，ANT 的非人類概念可以解釋我們與他人和非人類之間複雜關聯性(Sayes, 2014)，且非人行動者也同樣擁有能動性時，轉譯的因果(why)與內涵(how)才能不偏頗的被呈現(周信輝、蔡志豪, 2013)。例如 AI 為何能以「AI 官僚」行使行政裁量權，AI 為何能對人的感知(perception)或認知(cognition)能力產生影響。

因此，**本研究發現 ANT 分析途徑對 AI 有極佳的解釋力**。舉例來說，原本公共治理理論無法解釋公部門使用 AI 的現象，例如不同的人對演算型決策可能會採取不同的行動，包括：(1). AI 系統是實際決策者：公務員主動釋出行政裁量權(智慧判煙系統)、公務員與 AI 協作、公務員只能選擇相信(1955 多元智能電服中心)等；(2). 刻意忽視 AI 的建議(智能稅務服務計畫)；(3). AI 能動性對公務員的感知或認知產生影響(2012 年波蘭失業評分系統)。對純粹沒有使用 AI 決策的人，當 AI 變成同事之後，在某些案例中竟然發生權力移轉的現象，迫使我们不得不正視 AI 技術物的存在。此時，**本研究使用 ANT，在理論的層次上，ANT 主張技**



術物有主動性，詮釋AI自適應性(Adaptive AI)的技術特性，可以解釋AI系統是實際決策者的現象；ANT主張技術物有能動性，詮釋AI對公務員的感知或認知產生影響，可以解釋AI影響人類官僚行政裁量權的獨立決策自由度的現象，ANT的方法論能為公共治理理論目前無法解釋的困境提出可行的解決方案。因此，依循本研究的理論發現，接續以ANT的分析途徑探索及梳理12個研究案例，獲得人機協作的理論，以及在應用的層次上人類官僚應如何與AI共存治理的方法。

本研究從ANT視角進一步論述某些政策會失敗的原因，是對AI官僚(AI bureaucracy)或者街頭官僚(street level bureaucracy)的理解不夠完整，忽略AI自適應性與AI能動性，以至於無法提出有效的AI風險緩解措施。舉例來說，2016年澳大利亞聯邦政府社會服務部債務機器人Robodebt，**忽略AI是行動者，具有主動性、自適應性**，導致演算法未能依照法律規定執行；波蘭2012年開發AI失業評分系統，**忽略AI的能動性**對公務員的感知或認知產生影響，影響人類官僚行政裁量權的獨立決策自由度，導致工作人員不願質疑AI作的決定，超過99%依據AI的推薦分類，幾乎放棄行政裁量權，即使他們的專業判斷表明並非如此。因此，為填補以往理論缺陷，本文研究提出以下觀點：

若欲對AI官僚有完整理解，不應忽略「AI是行動者與AI能動性」。

雖然現在部份文獻，似乎沒有認為AI是行動者，本研究認為原因在於AI技術廣泛，有些AI技術是關鍵的行動者，有顯而易見的自適應性及能動性，例如英國脫歐公投的社交機器人；有些是沉默的行動者，例如AI智慧交控管理，就像人一樣，有不同態樣，但不能因此而否定AI是行動者，具有自適應性及能動性的存在。因此，從ANT的視角，當公部門使用AI時，前瞻性評估(Horizon



Scanning, HS)時應包括理解 AI 是行動者，有自適應性及能動性，可能產生偏誤 (bias)，進而設計 AI 風險緩解措施：

1. AI 是行動者：設計 AI 風險緩解機制，避免 AI 自適應性產生預期之外的結果。
2. AI 有能動性：驗證 AI 能動性是否對人類官僚的感知或認知產生影響，使人類官僚產生對 AI 的信任度過高或過低的問題。

因此，本研究使用 ANT 給予技術物廣義對稱性原則、ANT 分析途徑分析案例，以實證研究指出 *AI 是行動者並具有能動性、應該成為分析單位*。接續提出公共治理理論必須重新定義分析單位，並以此做為本研究最重要的理論基礎及研究發現。

未來公共治理理論「政策利害關係」以「人」為分析單位的治理典範，當公部門使用 AI 時，應包括「人」、「AI」。

本研究並以 AI 為分析單位的觀點開始，接續提出公部門使用 AI 的理論缺口與補充觀點。

二、AI 鑲嵌的公共政策過程可能呈現循環交錯的現象

公共政策過程階段論主張政策過程具有可控制性、次序性的概念，當公部門使用 AI 時，公共政策過程並非完全線性，可能呈現循環交錯的現象。

公共政策過程的研究途徑以階段論為主流，階段論將政策過程的階段劃分為政策規劃、合法化、執行、評估四個步驟分明的階段，每一個階段都有其先後次序性，政策過程具有可控制性、步驟發展的必然性與各階段的次序性(丘昌泰, 2008)。然而，本研究從 ANT 的視角，在 AI 智慧交控管理、1955 多元智能電服



中心等案例，AI 協助人類公務員更快速決策，當 AI 有偏誤(bias)，人類公務員運用自己的判斷力修正、反饋為 AI 訓練資料，提高 AI 模型的正確率，使 AI 以更高功能的表現協助人類公務員擴增認知與生理能力。此時，AI 與人類公務員是動態互動、共同進化，擴增(augment)和強化(enhance)彼此能力的過程，而人類公務員與 AI 之間複雜的動態互動關係所產生即時決策、即時評估、即時規劃、即時修正的現象，不同於過往公共政策過程階段論次序性的概念，而是循環交錯的過程。另一方面，2012 年波蘭勞動和社會政策部(MLSP)「失業評分系統」的案例，超過 99% 的公務員依據 AI 系統的推薦分類，幾乎放棄行政裁量權，即使他們的專業判斷表明並非如此，導致 AI 系統變成實質決策者(Misuraca & van Noordt, 2020)，此時，公共政策過程呈現線性的過程。

本研究從 ANT 的視角發現，ANT 交引纏繞 (an entanglement of interactions) 的社會世界：「假設社會是一種瞬時關聯的典型，物體成為行動過程的參與者，並參與重新關聯和重新組合的運動，強調社會和技術之間是動態、循環和交互，共同進化過程(Bueger & Stockbruegger, 2017; Latour, 2005; Waldherr et al., 2019)」，可以詮釋 AI 鑲嵌的公共政策過程所呈現循環交錯的現象，補充公共治理理論過去對過程的理解偏向靜態、直線分析的理論缺口。

公共政策過程階段論可控制性、次序性的概念，當公部門使用 AI 時，「政策規劃—政策執行—政策評估」呈現循環交錯的現象。

三、AI 能動性(agency)會透過人機協作對人類官僚的認知能力產生影響

對照以往工業機器人著重於物理系統中的人機協作，本研究提出認知層面的人機協作。

當公部門使用 AI 時，如何在循環交錯的政策過程中分析人與 AI 的關係，



以及人與 AI 共存治理的現象，成為公共治理的新挑戰，要有進一步有研究和證據(Androutsopoulou et al., 2019; Schoemaker & Tetlock, 2017)。本研究梳理文獻關於 AI 與人的互動關係，包括：

- 1.從公共行政觀點是「分工」。
- 2.從資訊科學等領域觀點，表示自動化程度，是計算機系統的自治權(the autonomy of computer systems)」(Floridi et al., 2018)。

從 ANT 視角研究案例，人類公務員與 AI 在高度時間依賴性敏感性的相互作用中呈現多樣化官僚行為，**AI 的作用性**可能取代官僚行使裁量權的獨立決策自由度(智慧案判煙系統)，也可能維持人類官僚決策(智能稅務服務計畫)。人與 AI 的共存治理，除了「分工」、「自動化程度」，還有其他可能性，包括階級關係(hierarchy)(例如智能稅務服務計畫)、代理關係(agency)(例如智慧案判煙系統)、機器自主(autonomy)(例如 AI 智慧交控管理)、雙重關係(dual)(例如 1955 多元智能電服中心)；以及衍生人機協作模式，包括直接控制、間接控制、演算型決策、協作或合作。本研究認為，產生多樣化官僚行為的原因之一是 AI 能動性(agency)影響人類官僚的認知能力，衍生人類公務員的科技接受度、演算法趨避和自動化偏誤。因此，依據案例的研究結果顯示：**在公共治理領域中探討 AI 與人的關係時，必須關注 AI 能動性(agency)可能對人類官僚的認知能力產生影響。**

接續，本文根據研究發現提出**公共治理領域中人機協作的理論基礎**：「人和 AI 能動性(agency)的交互作用」，著重於以往所忽略的 AI 能動性(agency)。本研究再依據 AI 技術的特徵、人的基本生理限制進一步論述人機協作的意旨：「AI 的自適應性(adaptive)與能動性(agency)，和人的有限理性(bounded rationality)與能動性(agency)的交互作用」，而共同進化的結果，可能相互增強彼此能力(例如智能稅務服務計畫)，也可能相互減弱彼此能力(例如波蘭勞動

和社會政策部的「失業評分系統」)。對照以往工業機器人著重於物理系統中的人機協作，本研究發現呈現出認知層面的人機協作。

另一方面，若從理論的角度論述，從 ANT 闡述行動者網絡中交引纏繞的新形式(new shapes)，可以理解為權力控制幅度的變化：「互動的可能性和調動的幅度得到協商和界定」(Latour, 2005)，顯示人與 AI 互動的結果、共存治理的現象，係依據 AI 和人協商和劃定的結果。因此，本研究依據 ANT 提出人機協作的理論：「人和 AI 能動性(agency)的交互作用」，除了能詮釋研究案例的 4 種可能性，包括 AI 自動化決策是如何改變基層官僚的自主行政裁量權，進而產生權力移轉的現象(例如 2012 年波蘭「失業評分系統」、2016 年澳大利亞聯邦政府「自動債務追繳系統」)，也可以詮釋其他官僚行為的可能性。

綜上分析，不論從 ANT 分析途徑研究案例的結果，或是 ANT 理論的角度，均顯示以 ANT 的視角詮釋 AI 和人互動的結果，能全面性的理解人與 AI 共存治理的現象，而以往文獻所指「分工」、「自動化程度」，僅是部分態樣。

探討 AI 與人類公務員的關係時，必須考慮 AI 能動性(agency)；公共治理領域中人機協作的理論基礎：「人和 AI 能動性的交互作用」。

為利公部門使用 AI 時規劃人機協作的實務需求，本研究提出「人機協作魔力象限(Human-Robot Collaboration Magic Quadrant)」(圖 5-1)，提供公部門應用 AI 的初步評估，選擇人與 AI 共存和協同工作的最佳條件。

四、「AI 官僚」也能行使行政裁量權

現行法律是規範人類而非 AI 的行為，把 AI 當成人來管理，修法讓「AI 官僚」行使行政裁量權，改變典型官僚組織層級節制的控制性、基層官僚具有獨立



決策自由度行使裁量權的詮釋邏輯。

「智慧判煙系統」AI 演算型決策直接作為行政裁罰的依據、愛沙尼亞使用 AI 審理小型訴訟案件等案例，機器人可以與現實世界直接互動，自動觸發其他系統來實施某種處罰(penalty)或執法程序(enforcement procedure)，顯示「機器人國家(Robotic state)」已經來臨(Dunleavy & Margetts, 2023)。然而，現行法律原則是規範人類而 AI 的行為，AI 行使公權力超過法律對行政裁量權的授權範圍(Henman, 2017)。為使「AI 官僚」也能行使行政裁量權，環境部(智慧判煙系統)修正法規，使 AI 演算型決策的直接裁決具合法性、可課責。

再者，AI 能動性影響行使裁量權的主體，可能是人類官僚、AI 官僚，或 AI 能動性(agency)影響基層官僚的感知或認知的「隱藏官僚(hidden bureaucrat)」。原本官僚制(bureaucracy)或「科層制」的層級節制原理，權責自上而下傳遞，科層化的管理方式，無法因應 AI 官僚的管理需求。

依據 ANT 主張給予技術物廣義對稱性原則，要將「非人類」行動者(例如 AI)整合到我們對社會世界理解的觀點，官僚組織的管理就應包括 AI 官僚。應如何管理 AI 官僚？**管理 AI 官僚的概念如同管理人類官僚，就是「把 AI 當成人來管理」**，只是管理方式不同(詳如表 6-2)。舉例來說，人類官僚的管理，包括避免人類主觀式偏見的管理，設定行政裁量權範圍等；而 AI 官僚的管理，包括 AI 偏誤(bias)的管理、設計 AI 風險緩解機制等。

其實把 AI 當成人來管理並非新鮮事，最早期定義 AI 是軟體代理人(software agent)，用一個軟體來執行一個代理人的功能，主體是人，代理人就是仲介的功能，以特定功能或業務的代理人來思考最為直觀，接下來就是要管理仲介(智慧代理人)。因此，管理 AI 的方法也如同管理仲介，如同公共行政以人為主體，現有各種行政業務都有專業代理人，如報稅代理人、律師、代書等，業務均不同，接下來就是管理專業代理人。而管理專業代理人，也有多種方法，例如合法化，保

險代理人有保險法及保險代理人管理規則、記帳士暨記帳及報稅代理人有記帳士法及相關管理辦法。

再者， AI 技術涵蓋範圍很廣，通用性生成式 ChatGPT 和以任務導向的 AI 模型(task oriented AI moduls)有不同的偏誤，有些偏誤是事先就決定的，有很多 AI 是可預測，並非每個 AI 都不可預測。如同管理人類官僚，有些可預測，有些不可預測，所以有行政程序法等相關法規規範人類官僚一樣。



表 6-1 人類官僚和 AI 官僚的管理方式對照表

管理對象	管理概念	管理作法
人類官僚	<p>以人為主的治理</p> <p>人員的管理—避免人類主觀式偏見管理</p>	<p>1. 層級節制</p> <p> 權責自上而下傳遞，科層化的管理方式</p> <p>2. 訂定行政裁量權範圍</p> <p>(1). 公務員行使裁量權之結果，應在法令規範之範圍內，不得超越法律授權範圍。</p> <p>(2). 逾越或濫用裁量權，行政法院得予撤銷。</p>
人類官僚、AI 官僚共存	<p>人與 AI 共存治理</p> <p>1.AI 是行動者的管理—AI 偏誤的管理</p> <p>(1).AI 自適應性</p> <p>(2). AI 能動性</p> <p>2.人機協作的管理—一人與 AI 協作的最佳工作狀態</p>	<p>1.行政性作法</p> <p>(1).AI 官僚合法化</p> <p>(2).AI 資訊品質的管理</p> <p>(3).設立風險值(含超過風險值措施)</p> <p> ➤ 行政規則排除</p> <p> ➤ 降階運作.....</p> <p>(4).人機協作模式</p> <p> ➤ 選擇人類監督程度</p> <p>2.技術性作法</p> <p>(1).AI 系統設計</p> <p> ➤ 以人為本的 AI 設計</p> <p> ➤ 人機介面整合</p> <p>(2).驗證能動性</p>

資料來源：本研究整理



貳、在公共治理分析層次

使用 ANT 分析途徑分析公部門使用 AI 的案例，對 AI 與人的關係、人機協作的現象、AI 對公共治理理論的影響等，有極佳的分析能力。

現有公共政策分析方法，例如因果關係的推理邏輯，無法有效分析公部門使用 AI、AI 與人共存治理的現象、AI 對公共治理理論的影響。

ANT 研究「行動化(actor-ing)」的敘事(narrativity)手法，跳脫一般想法與假定(方念萱, 2016)，適合研究發展中科技(例如 AI)與社會之間如何互動。本研究使用 ANT 分析途徑研究案例，研究發現可以從 ANT 的視角描述 AI 與人互動、共同變化的過程，人與 AI 合作互利共生地工作來擴增(augment)和強化(enhance)彼此能力的過程(智能稅務服務計畫)，或是彼此相互減弱的過程(波蘭失業評分系統)；更能進一步分析公部門導入 AI 成功與失敗的原因¹⁸¹(例如忽略 AI 是行動者、有能動性)、AI 對公共治理理論的影響(例如 AI 智慧交控管理的政策過程呈現循環交錯的現象)。

使用 ANT 分析途徑為何對分析公部門使用 AI 有極佳的分析能力？本研究認為，ANT 與社會科學理論的假設不同，非常適合分析 AI，特別是 AI 不同於

¹⁸¹成功的案例：愛沙尼亞 2018 年開發 SATIKAS32 系統，目前仍在持續開發並擴展其特性和功能；比利時 2014 年佛蘭芒兒童和家庭事務局(Flemish Agency for Child and Family)開發的 AI 預測系統(Predictive System)，目前仍在使用中；公務員知道 SATIKAS32、AI 預測系統不是 100%可靠，AI 模型提供更有目標性(targeted)和數據驅動(data-driven)的推薦結果，公務員利用專業知識和自己的判斷力做出最終決定。失敗的案例：2012 年波蘭「失業評分系統」諮詢人員超過 99% 依據失業評分系統的推薦分類，AI 系統成為實質上的決策者，2019 年正式停止；2016 年澳大利亞聯邦政府「自動債務追繳系統」，債務追繳過程自動化，使原來的核心法律原則遭到破壞；均是忽略 AI 是行動者、有能動性。



以往科技，除了是行動者，有能動性，更具有自適應性，會產生人類預期之外的結果。而 ANT 將技術物視為分析單位，不需要任何先驗假設、著重於描述而非解釋(Latour, 2005) 的分析途徑，能解決公共治理理論以人為主的治理典範、關注因果推理邏輯的侷限性，因而無法有效分析公部門使用 AI 的現象。

從 Latour 的觀點，Latour 認為「沒有社會這樣的東西(There is no such a thing as a society)」，沒有特定的社會現象、「社會(society)」、「社會秩序(social order)」、「社會實踐(social practice)」、「社會維度(social dimension)」或「社會結構(social structure)」等，沒有社會的(social)與社會有關 (pertain to society)的「社會力量(social force)」可用於「解釋(explain)」其他領域無法解釋的特徵，應該以經驗來「描述」而不是「解釋」社會活動(Latour, 2005)。但由於 ANT 與社會理論區分開(Sayes, 2014)，以致於 ANT 較少在主流理論(例如公共治理理論)被使用。

然而，**本研究認為，Latour 對技術最大的貢獻是給予技術物廣義對稱性原則、物體也有能動性，特別是在 AI 時代，尤其當我們一時沒有更好的解決方案時，更凸顯 ANT 的可用性。**因此，未來公共治理領域有關 AI 的研究，或分析公部門使用 AI 的案例，使用 ANT 分析途徑有極佳的公共政策分析能力。

綜上分析，本研究提出公共治理理論應補充 4 個觀點(表 6-1)：

表 6-2 公共治理理論缺口與補充觀點

分析項目	公共治理理論	理論缺口	補充觀點
1.公共政策模型	以「人」為分析單位的治理典範	1. 缺少 AI 是行動者，有自適應性 (adaptive) 2. 缺少 AI 有能動性	1. 公共治理分析單位，除了人，應包括 AI 2. AI 是行動者、有能動性(agency)
2.公共政策過程、政策評估及人機協作	1.階段論(次序性、可控制) 2. 「政策規劃—執行—評估」線性思維	1. AI 即時決策、即時評估、即時修正、即時規劃 2.人機協作、人與 AI 的動態互動	1. 政策過程呈現循環交錯的現象 2. 人機協作：人和 AI 能動性(agency)的交互作用
3.政策執行	人類官僚行使裁量權及獨立決策自由度	1. AI 官僚也能行使行政裁量權 2. AI 能動性影響人類官僚的感知或認知能力	1. AI 的作用性 2. AI 官僚
4.政策分析	1. 以人為分析單位 2. 定量或定性方法論 3. 因果關係推理邏輯	經驗主義客觀現實，不易提出人和 AI 共存的因果推理假設	ANT 分析途徑有極佳的公共政策分析能力 1. 將技術物視為分析單位 2. 不需要任何先驗假設 3. 著重於描述而非解釋

資料來源：本研究整理



參、在公共治理實踐層次

本研究從 ANT 的視角，藉由臺灣公部門 5 個、國外公部門 7 個，合計 12 個案例的研究成果，提出具體實踐的策略。研究發現如下：

一、策略性運用 AI 風險緩解機制，提升公共治理效能及控制風險

基於 AI 風險的疑慮，公部門使用 AI 的執行速度和案例成熟度不足，策略性運用 AI 風險緩解機制，將技術限制與行政實務相結合，提升公共治理效能及控制風險。

AI 應用於公部門相對新穎，且 AI 的應用和挑戰是彼此密切相關且可能產生協同效應(synergy effects)，基於 AI 風險的疑慮，特別是 AI 的倫理問題，例如個人資料保護、勞動力替代、道德困境(moral dilemmas)、AI 歧視、透明度等。在缺乏公部門使用 AI 風險管理的實務研究，且政府使用 AI 失敗可能會對政府和社會產生重大負面影響(Zuiderwijk et al., 2021)，使得公部門使用 AI 的執行速度和案例成熟度不足。

研究發現，要提高 AI 專案的成功率及控制風險，除了公部門行動者、AI、情境相關性、制度等綜合因素正向增強之外，重要關鍵之一是 AI 風險緩解機制，包括公共行政方法及資訊科技風險管理方法。而公部門使用 AI 可以提升的公共治理效能，包括：

1. 系統層次：提升組織整體的效能(effectiveness)、效率(efficiency)和公平性(equality)。「1955 多元智能電服中心」不增加第一線服務人數，提供公民全年 365 天，24 小時諮詢服務，2021 年服務量較 2020 年擴增約 2.53 倍。
2. 個人層次：使公務員認知和生理增強。「智能稅務服務計畫」增加稅務人員的選案查核能力；「AI 智慧交控管理」增加交控人員交通控制能力；「智慧判煙系統」消除環保局稽查員無法長時間監控的生理限制等。
3. 前瞻式評估機制：運用前瞻式評估機制排除 AI 轉型的阻力。「智能稅務服務



計畫」逐步調整學習工具，讓學習門檻降低，或導入機器人流程自動化(RPA)減輕工作負荷，降低員工抗拒數位轉型。「1955 多元智能電服中心」以人為本的 AI 設計，避免勞動力替代；「增加工作價值」與「調整薪資制度」提高電話服務人員的科技接受度，避免路徑依賴。

二、強化文官人力資源的策略：AI 賦能的新官僚

AI 使公務員「認知增強」、「生理增強」，未來探討提升政府文官人力資源體系時，應包括 AI 賦能公務員的觀點。

過往探討如何提升政府治理能力，主要是政府文官人力資源體系的變革，著重於官僚、組織競爭力、策略性人力資源管理和能力管理等理論(施能傑, 2006)。研究發現，AI 提升文官體系的治理能力，包括：

1. 認知增強：獲得憑經驗看不出來的重要關聯，提升查核的精確率(precision)和效率(efficiency) (智能稅務服務計畫)；AI 決策比人快與準 (AI 智慧交控管理)。
2. 生理增強：持續性監測可能的違法，提升政府信任、社會安全 (智慧判煙系統)；消除身心障礙者的生理限制，提升社會公平、SDGS (1955 多元智能電服中心)。

未來文官體系的競爭力，除了強化公務員本身能力，也必須增加 AI 賦能公務員、公務員與 AI 協作提升治理能力的策略。再者，從研究案例觀察，呼應文獻所提出 AI 可提升的公共治理應用潛力的面向(丁玉珍、林子倫, 2020)，包括：

1. 提高公部門績效評估與營運管理：消除繁文縟節、提升組織效能(AI 協作分文)。
2. 提升公民服務品質：改善電話諮詢服務(1955 多元智能電服中心)。
3. 精進管制性政策：空氣污染物監控(智慧判煙系統)、賦稅公平(智能稅務服務)



計畫)等。

4. 強化公共政策決策的品質：交通監控、低碳排放(AI 智慧交控管理)等。

AI 有潛力解決當今公共治理困境，未來公共治理領域探討政府文官人力資源體系的變革、政府必須有更好的治理能力時，應包括 AI 賦能公務員的觀點。

三、緩解 AI 治理爭議的方法：以循證為基礎

當今很多降低 AI 風險作法的爭議，建議從實際案例獲得如何權衡的作法，避免過度臆測 AI 風險而阻礙 AI 的採用。

很多 AI 風險都是猜測的，所提出降低 AI 風險的作法，缺乏系統性[倫理原則太多]、太高層次[歐盟可信任 AI 的倫理準則 7 項要求第 1 項是人類代理與監督 (human agency and oversight)，AI 系統需實現人類重要的目標或價值等]、太抽象 [Rahwan(2018) 提出的社會監督]。另一方面，從實證研究結果，大多數挑戰都不是技術或數據(Sun & Medaglia, 2019)，而是如何在官僚環境中應用 AI (Höchtel, Parycek, & Schöllhammer, 2016)，公部門採用 AI 最大的挑戰即是公共治理的回應：未被充分利用(underused)產生的機會成本，或過度使用和濫用(overused and misused)產生的風險(Floridi et al., 2018)。因此，已有文獻建議未來的研究應該超越概念和推測層面，加強實證研究，實證檢驗政府使用 AI 對公共治理的影響 (Zuiderwijk et al., 2021)。

本研究依循文獻的建議，從實證研究出發，以 ANT 的視角分析 12 個研究案例，並梳理如何權衡 AI 治理爭議作法：

1. 公部門 AI 科技創新原則：

政策制定者與管制者常面臨兩難：「預防原則」與「不需許可的創新」

(Thierer et al., 2017)的爭議。研究發現，目前公部門 AI 科技創新的共通性原是

「跨域創新、多點創新、從資料出發、從現有系統發展、從小做起、從非核心職能開始」，並未爭議「預防原則」與「不需許可的創新」等原則。



公部門 AI 科技創新原則：「跨域創新、多點創新、從資料出發、從現有系統發展、從小做起、從非核心職能開始」。

2. 監督強度：

Rahwan(2018)建議需要人類監督(human-in-the-loop, HITL)，甚至社會監督(society-in-the-loop, SITL)：「將社會置於循環中監督，以增強社會的決策能力」(Rahwan, 2018)。然而，監督程度太高，降低公部門的科技接受度；監督程度太低，提高 AI 風險程度。研究發現，**發揮 AI 潛力的重要關鍵是「如何權衡監督程度」，策略性的運用「人機協作」和「AI 風險緩解機制」，可以適度降低人類監督、社會監督的程度。**

舉例來說，國家權力涉入程度高(對人民的權利具有較強的干預效果)，政府職能是「任務所需脈絡分析程度高」，應由人類「**直接控制**」(智能稅務服務計畫)；但若政府職能是「任務所需脈絡分析程度低」，「人機協作」採用「**間接控制**」，可以提升政府管制效率(智慧判煙系統計畫)。再者，國家機關權力涉入程度低，對公民個人權利義務的權力影響低，若政府職能是「任務所需脈絡分析程度低」，「人機協作」採用「**演算型決策**」，善用 AI 比人「快(faster)」與「準(accuracy)」的技術優勢，可以提高公民服務品質(AI 智慧交控管理)、強化與大眾溝通能力(AI Chatbot)；但若政府職能是「任務所需脈絡分析程度高」，從 1955 多元智能電服中心、AI 人機協作分文、美國最高法院 AI 判斷「政黨傑利蠟蝟」等案例顯示，「**協作或合作**」可以提高工作價值、有效利用人力資源、避免選舉政治操作等。



然而，「國家機關權力涉入程度低」，例如自願性工具、或象徵和勸告型的政策工具，雖然對公民個人權利義務的權力影響低，誠如 AI 雙重用途的特質，AI 也可能就公民既存或潛在的偏好／偏見進行分類，鎖定特定群體，傳播虛假信息(英國脫歐公投社交機器人)；此時，就應該發揮「**社會監督**」(Rahwan, 2018)的力量，由大眾參與監督濫用或操縱公共對話空間的廣告政治、專業的操縱者、宣傳者等，避免 AI 技術(AI-based technologie)的深度偽造(Deepfake)、錯誤資訊(Misinformation)、造假資訊(Disinformation)，以及「垃圾郵件機器人(spambots)」、「支付機器人(paybots)」、「自動按讚機器人(autolikers)」和其他「假帳號(inauthentic accounts)」傳播虛假信息、炒作社會參與。再者，政府應協助公民識別 AI 操縱技術，理解 AI 的作用性、能動性(agency)能對人的認知能力產生影響，使公民有能力抵制宣傳和廣告的力量，避免 2017 年牛津大學研究員維亞切斯拉夫·波隆斯基(Vyacheslav Polonski)在世界經濟論壇的文章斷言，AI「默默地接管了民主(silently took over democracy)」的場景發生。

3. AI 系統的品質

AI 系統的品質的爭議，包括現有資訊系統的原理和方法均無法對 AI 軟體有效進行質量評估(Quality Assessment) (Kuwajima & Ishikawa, 2019)；AI 系統的效能/正確率(performance/accuracy)和可解釋性(explainability)之間存在反比關係的情況下，需要進行權衡(Dwivedi et al., 2019)；演算法必須具有一定的透明度和可解釋性，但是這樣做會削弱『機器學習』測試不同模型的能力，以及建立因果關係的能力，是否需要這種折衷方案？(Dwivedi et al., 2019)；AI 系統功能的增強與人類依賴它們的趨勢之間有著直接的關係(Bullock, 2019)，以及人類對 AI 的信任度過高或過低都會產生負面影響(Okamura & Yamada, 2020; Schmidt & Biessmann, 2020)的問題。



本研究從 ANT 的視角，梳理案例如何評估、權衡、提高 AI 系統的品質，彙整如下：

- (1). AI 系統品質評估：「有一定程度的 AI 系統品質，以及 AI 風險緩解機制。」
- (2). 利害關係人「參與」AI 模型建置：可同時提高「AI 系統的效能/正確率」和「可解釋性/科技接受度」。
- (3). 避免演算法趨避和自動化偏誤：設立風險值、驗證 AI 能動性。

4. AI 在政府職能的應用領域與價值選擇

文獻對 AI 在政府職能的應用領域有諸多論辯，包括避免將 AI 決策用在對人民權利和義務產生重大影響領域(Mehr et al., 2017)；複雜性和不確定性較低的任務是 AI 處理，而複雜性和不確定性較高的任務則應由人類自行決定(Bullock, 2019)；核心公共機構不應再使用 AI (Agarwal, 2018)。究竟哪些政府職能或公共服務應用領域適合導入 AI？(Busch & Henriksen, 2018; Kuziemski & Misuraca, 2020)。再者，有關價值選擇的諸多論辯(Kleinberg et al., 2018)，例如「科技優勢 VS 法律原則」，涉及「能力可行性及規範合理性」的「倫理辯證」，「AI 法官」的價值衝突等。

本研究從 ANT 的視角研究實際案例，無論國家權力涉入程度高或低，人類決策或 AI 決策，情境相關性高或低，臺灣案例或是國外案例等，研究結果均顯示，大多數政府職能都適合導入 AI。因此，與其陷入不同領域的觀點之爭，不如將技術限制與行政實務相結合，降低 AI 偏誤的效果，控制 AI 可能的風險。

大多數政府職能都適合導入 AI，仍需審慎研究可行性，設計相對應的 AI 風險緩解機制。

第二節 研究貢獻



本研究從 ANT 的視角所提出的理論，已藉由臺灣實證研究案例驗證，亦可解釋本研究所補充或佐證的國外案例，因而具有強固性。茲簡單描述本研究所提出 3 大研究貢獻如下，再詳細分述如后：

首先，**將公部門應用 AI 的決策及行為理論化，提出 AI 的公共治理理論應增加 AI 是行動者、有能動性的觀點**。實證研究發現公共治理理論困境的主要原因，是公部門使用 AI 時忽略 AI 是行動者、有能動性。據此，本研究嘗試將公部門應用 AI 的決策及行為理論化，提出公共治理理論典範轉移的路徑，將研究進程從探索性研究往前推進至理論層次。

其次，**提出「AI 風險緩解機制」的意涵與具體實踐做法**。基於 AI 風險的疑慮、缺乏公部門使用 AI 風險管理的實務研究，且政府使用 AI 失敗可能會對政府和社會產生重大負面影響(Zuiderwijk et al., 2021)等因素，使公部門使用 AI 的執行速度和案例成熟度不足的困境。本研究提出了「AI 風險緩解機制」的意涵與具體實踐做法，避免目前新興 AI 治理所遭遇缺乏系統性(unsystematic)、太高層次(too high-level)、高度抽象(a highly abstract level)，無法具體實踐(Henman, 2020; Whittlestone et al., 2019)的批判。

最後，**提出修正人機協作的理論，包括認知層面人機協作**。過往文獻關注於物理系統中的人機協作，探索人機互補性問題(Ansari, Hold, & Khobreh, 2020)；人機協作在公部門不僅少見，更缺少關注認知層面人機協作(*cognitive human-robot collaboration*)的理論基礎(Semeraro et al., 2023)。本研究提出公共治理領域的人機協作理論，除了補充公共治理理論缺乏人機協作的理論，更可以涵蓋以往所忽略的認知層面。

本研究所提出的公共治理理論典範轉移路徑，有別於目前主要研究途徑：1.



目前公共治理理論領域：對 AI 技術關注很少；或是 2.新興 AI 治理：單純從技術角度出發的治理框架，容易忽略對現有治理結構的影響。

本研究所提出理論架構的優勢，包括：1.將 AI 技術物理論化，提高公共治理理論的解釋力；2.新增人機協作、AI 與人的互動關係等相關理論；3.著重於可近性及務實性，本研究所提出的理論架構已藉由實際案例驗證理論的可用性。再者，本研究依據實證案例研究成果所提出的理論意義和具體實踐作法，有助於臺灣公部門參考，對其他地區具備高度可適用性，例如亞洲公共行政體系，特別是政治制度和文化類似的國家(Yu-Jen, Hsini, Tze-Luen, & Wen-Hsi, 2023)，也可以提供國際比較參考。

最後，本研究也具體回應了 Casares(2018)提出認知機器時代恢復韌性民主社會的建議：「AI 與公共治理必須共存(coexistence)」¹⁸² (Casares, 2018)。

壹、將公部門使用 AI 的決策及行為理論化，提出公共治理理論增加 AI 是行動者、有能動性的觀點

將 AI 理論化後，未來公部門使用 AI 時，可以推演、解釋和預測公部門使用 AI 的現象，設計 AI 風險緩解機制，降低 AI 風險。

公部門使用 AI，目前的理論困境是公共治理理論忽略技術，新興 AI 治理忽略對現有治理結構的影響。原有公共治理理論無法有效解釋 AI 官僚、AI 和人結合的新組合(Miller, 2018)；而新興 AI 治理又缺乏系統性、太高層次、高度抽象，無法具體實踐(Henman, 2020; Whittlestone et al., 2019)。

本研究從 ANT 的視角研究 12 個案例，提出公共治理理論無法有效解釋公部門使用 AI 的現象，是忽略 1.AI 是行動者、2.AI 有能動性。接續提出 AI 應成為

¹⁸² A proposal for a resilient democratic society in the age of cognitive machines: the necessary coexistence of private AI agents and public AI systems of governance.



分析單位，必須關注 AI 的特性¹⁸³、人機協作¹⁸⁴對公共治理理論的影響，從而提出以循證為基礎的理論：從以人為本的典範轉移到人與 AI 共存的社會世界觀(圖 6-1)。

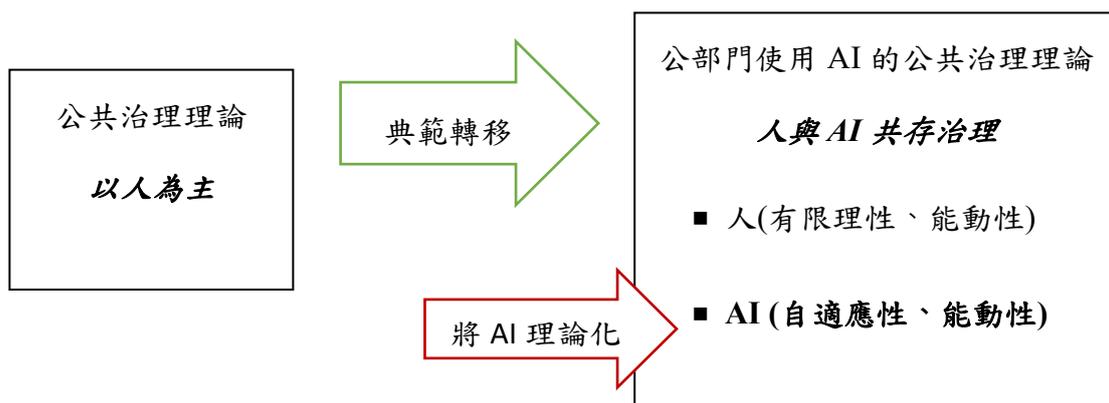


圖 6-1 公部門應用 AI 的決策及行為理論化概念圖

資料來源：本研究整理

接續從實際案例驗證可以有效解釋公部門使用 AI 的現象，包括 AI 官僚的行為等，也可以說明公共治理理論的困境，是對 AI 作用性理解不夠完整，以至於無法有效解釋為何某些政策會成功、或是某些政策失敗。

AI 的公共治理理論：當公部門使用 AI 時，公共治理理論的分析單位，除了人，應包括 AI。

¹⁸³ 自適應性(Adaptive)及能動性(agency)。

¹⁸⁴ 人和 AI 能動性(agency)的交互作用：AI 自適應性(adaptive)與能動性(agency)與人有限理性(bounded rationality)和能動性(agency)的交互作用。



再者，一旦將公部門應用 AI 的決策及行為理論化，能補充原先公共治理理論的缺口，並能推演、解釋與預測公部門使用 AI 的現象。舉例來說，理解 AI 有侷限性，人機協作有風險¹⁸⁵，不同 AI 技術可能產生不同的偏誤(bias)，就可以在公共行政流程中設計 AI 風險緩解機制，管理 AI 產出的結果，使 AI 風險在可以控制的範圍內。

貳、提出「AI 風險緩解機制」的意涵與具體實踐做法

歸納臺灣公部門 AI 風險緩解機制的意涵與具體實踐做法，提供公部門實踐參考。

不論是新興 AI 治理研究社群提出的理論、框架、策略和指南，或歐盟的 AI 法(草案)、臺灣生成式 AI 的參考指引，追根究底都是識別和緩解 AI 風險，或是有類似的因子在裡面。然而，這些倫理原則、指南、框架和政策、策略等，很大程度上是探索性研究，且缺乏系統性、太高層次、高度抽象，無法具體實踐 (Henman, 2020; Whittlestone et al., 2019)。

本研究從 ANT 的視角研究 12 個案例，提出 *AI 風險的來源之一*：「*AI 能自主決策(autonomous decision)*、*AI 有偏誤(bias)*、*AI 有自適應性(adaptive)*；*AI 能動性(agency)能影響其他行動者*」。接續梳理實證案的 AI 風險緩解機制、意涵與具體實踐做法，提出公部門 AI 風險緩解機制意涵：「*AI 是行動者、有能動性，設計風險管理措施，使 AI 系統產出的結果在可以控制的範圍內。*」，以及實證做法，可提供臺灣公部門實踐參考，避免新興 AI 治理無法具體實踐的批判，

¹⁸⁵ 例如演算法趨避和自動化偏誤。



參、提出修正人機協作的理論，包括認知層面人機協作

公共治理領域的人機協作理論，詮釋官僚組織中 AI 與人共存治理的現象。

實務上 AI 系統旨在促進人與機器之間的協作(Mills et al., 2021)，過往文獻關注於物理系統中的人機協作，探索人機互補性問題(Ansari, Hold, & Khobreh, 2020)，因此，人機協作在工業機器人、製造業、營造業、服務業等領域蓬勃發展(辛炳隆, 葉俊偉, 林金玲, & 陳秋蓉, 2023; 林葳均, 余佩儒, & 陳佳珍, 2021)。然而，人機協作在公部門不僅少見，尚未充分分析人類與 AI 共存和協同工作的最佳條件(Seeber et al., 2020)，特別是**缺少關注認知層面人機協作(cognitive human-robot collaboration)的理論基礎**(Semeraro et al., 2023)。

從 ANT 的視角研究 12 個案例，發現善用人機協作，特別是認知層面的人機協作，可以提升公共治理效能。例如在「智能稅務服務計畫」，「1955 多元智能電服中心」、「AI 協作分文」的案例，人與 AI 協作增強彼此的能力。因此，探討 AI 與人的關係時，必須考慮 AI 能動性(agency)，本研究據此提出**公共治理領域中人機協作的理論：「人和 AI 能動性(agency)的交互作用」**，包括**認知層面人機協作(cognitive human-AI collaboration)**，詮釋人與 AI 如何協作增強彼此的能力，補充以往文獻主要關注智慧工廠、物理生產系統中的人機協作的缺陷。

再者，人與 AI 的關係，從公共行政觀點，是將機器預測(Machine Predictions)變成人類決策(Human Decisions)，是分工觀點；從資訊科學、工程、控制等領域的觀點，Sheridan 從計算機如何接管人類的執行任務，表示自動化程度，是計算機系統的自治權(the autonomy of computer systems)」(Floridi et al., 2018)觀點。**本研究提出公共治理領域人機協作的理論：人和 AI 能動性(agency)的交互作用**，意旨人機協作將有眾多的可能性，幫助公部門使用 AI 時，理解 AI 和人互動各種可能性，除了文獻所指「分工」、「自動化程度」的觀點，還有其他可能性，例如階級關係(hierarchy)、代理關係(agency)、機器自主(autonomy)、雙重關係(dual)等複

雜的關係，以及衍生人機協作的模式，可能包括直接控制、間接控制、演算型決策、協作或合作等。因此，本研究從 ANT 視角提出公共治理領域人機協作的理論，能完整詮釋官僚組織中 AI 與人互動的關係，人機協作如何互利共生，進而分析人類與 AI 共存和協同工作的最佳條件，補充以前文獻對人與 AI 之間關係理解與論述的不足。



第三節 政策建議

本研究依據 12 個案例的研究成果，針對公部門使用 AI 提出政策建議，包括：

1. 因應 AI 系統建構課責機制，讓 AI 成為可信任的 AI。
2. 公部門導入 AI 時應設計 AI 風險緩解機制。
3. 建立公部門以 AI 為目標導向的大數據資料庫或資料資源。
4. 增加「公私協力關係型」的創新採購制度。
5. 掌握有效 AI 社會對話的前提。
6. 公務員的 AI 技能訓練應達 Level 4—「效果層次」。
7. 訂定公部門可信任 AI 參考指引。

茲分別說明如下：

一、因應 AI 系統建構課責機制，讓 AI 成為可信任的 AI (AI trustworthiness)

課責制度是可信任 AI 的關鍵因素(Doshi-Velez et al., 2017)。

AI 在過去幾十年的成功，很大程度是 AI 系統正確率的提升，然而，從近期 AI 的發展與應用，尤其臺灣是高度民主化的公共政策對話環境，使過去以正確率為績效標準(accuracy-based performance measurements)面臨挑戰，「可信任 AI」重要性已逐漸超越 AI 預測正確率，甚至是成敗關鍵，迫使我们應該從績效驅動的 AI (performance-driven AI)轉向信任驅動的 AI (trust-driven AI) (Li et al., 2023)。

雖然目前學術研究仍主要關注改善演算法模型強化可信任 AI，然而，僅演算法進步不足以建立可信任 AI，除了技術，應該考慮正確率以外的因素。有文獻整理與實際應用密切相關的可信任 AI 原則，分 2 大類：



- 1.技術要求：包括穩健性(robustness)、可解釋性(explainability)、透明度(transparency)、可重複性(reproducibility)和泛化性(generalization)。
- 2.倫理(ethical)要求：包括公平(fairness)、隱私(privacy)和課責制(accountability)。

同時指出上述原則應適用整個 AI 系統生命週期的整體情況、各個階段、各特定面向；又由於各項因素彼此密切相關，也必須關注彼此組合(combination)和相互作用(interaction)，以系統性的角度聯合優化(joint optimization)[例如相互增強(mutual enhancement)]或權衡(tradeoffs)，且任何一個環節或面向的信任破壞都會損害整個系統的可信任度，並依此提出可信任 AI 系統參考指引或操作指南(operational guidebook)(Li et al., 2023)。然而，本研究認為仍屬探索性研究及普遍應用性，對公部門而言，可操作性仍不足。

回顧臺灣在公共政策問題面臨重大爭議時，例如經濟發展與環境影響的權衡時，曾利用公開說明會、成立監督小組、風險評估計畫等方式提高公民對政府的信任。未來公部門將 AI 運用於公民權利高度相關的領域時，雖然仍可使用上述措施，然而，依據本研究成果「把 AI 當成人來管理」，比較簡單務實的做法是「管理 AI 產出的結果」。

先回到以人為主的世界，我們如何管理人類官僚，以及管理人類官僚產出的結果(例如行政裁量權)?傳統以來，政府是一個獨占管制權力的機構，*利用完善的法律和制度*，對人類官僚「有效的課責機制」提振民眾對政府的信任，而「*行政程序控制*」實現民主課責的主要的方式之一(陳敦源, 2009)。因此，當官僚組織的成員從人類官僚，變成人類官僚和 AI 官僚共存時，依照本研究成果，「把 AI 當成人來管理」的概念，仍然可以比照人類官僚的方式，利用 *完善的法律和制度，行政程序控制，對 AI 官僚課責*，提振民眾對政府使用 AI 的信任。舉例來說，在智慧判煙的案例，環境部修正法規把 AI 官僚變成課責制度中的一員。再

者，本研究第二章曾提及，民進黨立委林淑芬表示，發展 AI 的前提是「要能課責」，更顯示課責制度是可信任 AI 的關鍵因素。

因此，未來公部門使用 AI，視政府職能領域與情境，適時對 AI 系統建構課責機制。而提振民眾對政府使用 AI 信任的政策溝通方式，建議運用「公部門使用 AI，AI 官僚也納入課責機制，可依法提起訴願，不會讓數位裁量權和數位官僚主義傷害公民權力。」，無論是對內或對外溝通皆然。

二、公部門導入 AI 時應設計 AI 風險緩解機制

傳統以來，為避免官僚體制的擴張或弱化，立法機關會訂定行政法規來拘束行政機關作成的行政處分，保障人民權益，增進人民對行政之信賴，例如行政程序法規範行政行為應受法律及一般法律原則之拘束，非有正當理由，不得為差別待遇。同樣的，為避免 AI 官僚擴張或弱化，本研究提出「把 AI 當成人來管理」，視政府職能領域與情境，適時對 AI 系統建構課責機制。然而，課責仍屬後端的補救措施，積極的做法是利用前瞻式評估機制，設計「AI 風險緩解機制」，包括公共行政分析方法、資訊科技風險管理方法，拘束 AI 系統產出的結果：

1. 公共行政分析方法：

公部門 AI 治理 5 種挑戰的 AI 風險緩解措施：公共行政改革、AI 官僚的合法性、AI 系統的品質(資料治理與演算法治理)、人機協作的管理、勞動力影響(勞動力替代與轉型) 等。

2. 資訊科技風險管理方法：

從威脅識別到定量和定性風險分析的 AI 風險緩解措施：接受(Accept)、避免(Avoid)、轉嫁(Transfer)、防範(Deter)、拒絕(Reject) (Kaur & Lashkari, 2021)。



三、建立以 AI 為目標導向的大數據資料庫或資料資源

依據典型 AI 產品開發生命週期，包括：資料準備(data preparation)、演算法設計(algorithm design)、開發部署(development–deployment)與管理(management) (Ashmore, Calinescu, & Paterson, 2021; Li et al., 2023)，每個 AI 專案都從資料(data)開始。財政資訊中心有財稅大數據資料庫、財政巨量資料研究中心；勞動部公文系統有超過 20 年人工建構事先分類的資料庫，都是成就 AI 系統的重要前提。臺灣發展大型語言模型(large language model, LLM)的 4 大挑戰首要挑戰是繁體中文語料不足¹⁸⁶，顯示建立以 AI 為目標導向的大數據資料庫或資料資源的重要性。

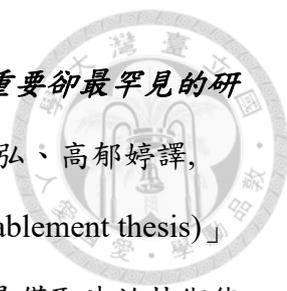
四、掌握有效 AI 社會對話的前提

人類歷史上有多次面臨新科技出現而引發各種正反意見的論戰，然而，沒有任何一次比這次 AI 所引發的還要更影響深遠(蘇經天, 2023)。目前主流觀點是政府若要負責任地使用 AI，必須利用「公民參與」和「事先與社會進行辯論」(Wilson, 2022)，如果事先不使用參與的方式促進社會的辯論，政府將無法擴大採用 AI 技術(de Sousa et al., 2019; Mikhaylov et al., 2018)。然而，欲促進有效的 AI 社會對話，建議先掌握社會對話的前提：1.釐清 AI 技術設計的用途；2.以循證為基礎；3.掌握跨領域術語或專有名詞的真正意涵；4.審議式民主。茲分述如下：

1.社會對話的開端—釐清 AI 技術設計的用途，避免雙重用途

對於公共治理諸多懸而未決的難題，例如 AI 在政府職能的應用領域，或是勞動力替代和轉型產生增加或減少員工需求等論辯。研究發現，該等爭點產生的原因在於 AI 雙重用途的特質，**如同歐威爾(Orwell)指出，我們絕對不該混淆了技術**

¹⁸⁶ 【一年從無到有，臺灣生成式 AI 發展下一步如何擴大？】臺灣發展 LLM 的 4 大挑戰。檢自 <https://www.ithome.com.tw/news/160090>



能做的事情，和它實際用來做的事情，這引導我們邁向技術最重要卻最罕見的研究面向：「人們拿技術來做什麼？」(Pinch & Trocco, 2004; 王志弘、高郁婷譯, 2023)，進而衍生「削減論點(curtailment thesis)或「賦能論點(enablement thesis)」(Buffat, 2015; Busch & Henriksen, 2018)，顯示 AI 的最終效果不是僅取決於技術能力，還取決於如何被使用。舉例來說，1955 多元智能電服中心，AI「以人為本」的設計，AI 技術成為賦能工具，消除身心障礙者的生理限制，而不是取代電話服務人員，同時提高政府公共治理效能。AI 量刑資訊系統中協助法官整理大量裁判資料，提升量刑的公平性、妥適性、透明性及可預測性¹⁸⁷，而不是「有罪產生器」¹⁸⁸。因此，若不討論清楚 AI 的用途，將無法展開有意義的社會對話。

2. 社會對話的重點—以循證為基礎，避免過度臆測 AI 風險

從技術角度，大多數的科學社群(包括資訊工程)學者、資料科學家對公部門使用 AI 抱持較樂觀的態度；但從社會科學角度，多數抱持較謹慎的觀點。使得現今理論性的爭論，阻礙了 AI 在公部門的應用潛力，但很少有實證研究來證實 AI 危險性和益處(Sun & Medaglia, 2019; Valle-Cruz et al., 2020)。從本研究成果發現，解決 AI 爭議的作法是以循證為基礎，才能避免過度臆測 AI 風險，建議累積不同脈絡的案例研究成果，展開基於客觀事實衡量的社會對話。

¹⁸⁷ 檢自 <https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html>

¹⁸⁸ 檢自 <https://opinion.cw.com.tw/blog/profile/553/article/14016>



3. 社會對話的效率—掌握跨領域術語或專有名詞的真正意涵，避免主觀感知討論

AI artifacts、agency、adaptive¹⁸⁹、dual use¹⁹⁰、decision¹⁹¹、bias 等，AI 技術領域人員和公共政策領域人員的定義和理解重點不同。以 bias 為例，在社會科學領域，「偏差(bias)」的概念涉及相關行為者間的特定利益分配方式(蕭全政, 1991)，即使是源於合理的選擇(reasonable choices)，仍然可能產生非常有傷害性的偏差(bias) (Obermeyer et al., 2019; Young et al., 2019)。資訊科學領域，機器學習有 5 種偏誤(bias) (TDS, 2021)。因此，跨領域社會對話前，先釐清術語或專有名詞，避免不同領域各自解讀，或是主觀的感知討論，讓跨領域的社會對話在同一個定義或立論基礎下進行，才能進行有效率的社會對話。

4. 社會對話的方式—審議式民主，避免權力不對稱

AI 通用技術、雙重用途、自適應性(adaptive)、能動性(agency)或作用、政治性、偏差或偏誤(bias)等特性，可能產生權力差距擴大的風險。而審議式民主的社會對話方式及政策分析途徑，讓公民可以直接參與公共政策或議題的討論，破除權力不對稱關係的現象，探詢公民在該時空系絡下皆能接受的政策內容(林子倫、陳亮宇, 2009)，可以成為 AI 時代社會對話的學理基礎。

¹⁸⁹ 在氣候風險定義，調適能力 (adaptive capacity) 指個人、組織、區域及社會上所具有調適氣候變遷衝擊所需具備的能力，包含減少氣候變遷衝擊的能力、可利用的資源等。在 AI 領域，自適應性系統(Adaptive systems)是在情境感知機制(context-aware mechanisms)的驅動下自動改變其行為¹⁸⁹(Fischer, 2023)。

¹⁹⁰ 在 AI 演算法或數學定義，「對偶使用」是用對偶、對稱的方法去解這個問題，先解列再解行，對偶先解行再解列，答案一樣，但可能其中一種方式比較快。

¹⁹¹ 在 AI 領域決策指「決策變數的選擇」，做完決策變數的選擇就結束了。跟制定公共政策的決策不一樣，公共政策的決策還需要進一步加工。



五、增加「公私協力關係型」的創新採購制度

研究發現，採購策略由小做起，透過試辦或先概念驗證(POC)再逐步擴展軟體的投資，可以提高公部門使用 AI 的成功率。然而，從制度面分析，現行政府採購制度，以行政院公共工程委員會訂定「資訊服務採購作業指引」為例，機關可以在招標文件中載明服務水準及品質需求、要求廠商投標時載明執行規劃，機關與廠商之間是「**需求端、供給端**」關係，政府採購著重於公平、公開之採購程序。

然而，公部門使用 AI，採購的不確定因素甚廣，包括：

1.發展階段

(1).廠商是否有足夠的量能。AI 的三大核心要素，算力、演算法和數據，

Open AI 的 GPT-4 的參數數量 1.8 萬億。

(2).研發成本的不確定性。廠商要為公部門開發量身定制的任務，招標金額有時難以預估，也影響廠商投標意願。

2.導入到應用階段

AI 模型的正確率受下列相關因素影響：(1).數據品質，(2).資料可得性，(3).情境相關性，(4).公務員持續「參與」，(5).終端使用者的演算法趨避和自動化偏誤等。以上均非廠商單方面能掌握與控制的元素。

從以上分析，公部門採購 AI 是新的公私合作夥伴關係，AI 系統是機關與廠商公私協力進行「**共同設計、共同創造、共同生產**」；再者，廠商的量能、AI 的開發成本更是關鍵，舉例來說，國科會即採成本高、步驟最完整的方法自行打造的臺版 ChatGPT¹⁹²等，已顛覆過往政府「需求端、供給端」的採購邏輯。因此，

¹⁹² 打造臺版 LLM 供企業和公部門免費使用。檢自

<https://www.ithome.com.tw/news/160092?fbclid=IwAR2a1B0zQCLtrgPlvBPHn1ynfdZ>



建議政府目前資訊服務採購架構、「資訊服務採購作業指引」必須因應 AI 滾動修正，增加「公私協力關係型」的採購制度。

六、公務員的 AI 技能訓練應達「效果層次」

從最新技術發展，例如 ChatGPT¹⁹³藉由人腦與 AI 互下指令培養人類「批判性思考」¹⁹⁴的能力，以及本研究發現，人才是推動 AI 的最重要關鍵之一，顯示未來擁有 AI 技能是必備條件。另從美國聯邦政府導入 AI 的經驗，53%的案例是內部人員的努力成果(Engstrom et al., 2020)，加上 Low-Code / No-Code(低代碼 / 無代碼)時代已來臨，寫程式將成為基本技能。

然而，目前對於公務員應該補充哪些 AI 工作技能的具體資訊很少(Acemoglu & Restrepo, 2019)，研究發現，公務體系對 AI 的訓練集中於 Level 1 -Level 3，建議應達到 Level 4—「效果層次」，公務員能對組織提供具體的貢獻。初期強化文官體系適應 AI 新工具的能力，例如，如何問對話式機器人問題等。

七、訂定公部門可信任 AI 參考指引

為加速公部門導入 AI 的速度，同時秉持負責任及可信任之態度，建議訂定公部門可信任 AI 參考指引。雖然行政院已訂定「行政院及所屬機關(構)使用生成式 AI 參考指引」，目前仍有不足之處，包括：

1.AI 技術廣泛，生成式 AI 是其中的一種，缺少廣泛應用的參考指引。

5jm9n33RuD40t_xwHo5QERV8iOm-

LefQ_aem_AU3wN2x7Momr2rmniMCNCokQIVhxxrQnHAbFhriOQOeg3xs58GqpCy
1UN0Vw5z_xMEg

¹⁹³ 臺大公布 ChatGPT 指引。檢自 <https://udn.com/news/story/6885/7027728>

¹⁹⁴ 數學界菲爾茲得主陶哲軒看 ChatGPT 的錯誤答案：總給我靈感找到新想法
From：檢自 <https://www.businessweekly.com.tw/careers/blog/3012198>

2.對於公部門而言，上述參考指引仍太抽象，缺少能具體實踐的原則。

再者，為強化 AI 系統的安全性(safty)，參考指引應區分技術性及行政性參考指引。有關 AI 技術性參考指引，國際標準化組織(International Organization for Standardization, ISO)訂定 ISO/IEC 42001 作為 AI 管理系統標準¹⁹⁵，數位發展部推動 AI 評測制度與可信任 AI 環境發展¹⁹⁶，亦有文獻從穩健性(robustness)的角度提出如何維護(maintenance)AI 系統(P.-Y. Chen & Das, 2023)，本研究著重於行政性參考指引。

經梳理研究案例善用 AI 潛力與因應 AI 挑戰的實際做法，並著重於文獻缺少、或爭議的議題，提出「公部門可信任 AI 參考指引(草案)」。再者，該指引從公共治理現有機制出發，也與新興 AI 治理的概念契合，可做為行政性參考指引討論的起點。內容詳如表 6-3：

¹⁹⁵ 檢自 <https://www.iso.org/standard/81230.html>

¹⁹⁶ 數位發展部「AI 產品與系統評測中心」啟動 推動我國 AI 評測制度與可信任 AI 環境發展。檢自 <https://moda.gov.tw/ADI/news/latest-news/9295>

表 6-3 公部門可信任 AI 參考指引(草案)



公部門可信任 AI 參考指引(草案)

公共治理領域越來越廣泛使用 AI 作為提高政府治理能力，達成永續治理目標的手段，預期政府將成為 AI 的最大採用者之一。然而，AI 的應用和挑戰是彼此密切相關且可能產生協同效應。再者，公部門有統治權、公民無法選擇退出、公部門使用大數據受到許多限制、公部門需要最大化公共價值，使公部門使用 AI 必須比私部門更加謹慎。

為提升公共治理效能及控制風險，提高 AI 專案成功率，參考各國政府之審慎因應作法，並依據臺灣公部門使用 AI 的實證研究成果，研訂「公部門可信任 AI 參考指引」(以下簡稱本參考指引)，供各機關參考。

本參考指引共計十點如下：

一、加強組織的數位韌性

1. 推動文化轉型：建立鼓勵 AI 發展的創新制度、編列常規預算、專責任務編組、AI 在職訓練、政策溝通等。
2. 推動機器人流程自動化(RPA)：減輕公務員現有工作負荷，降低路徑依賴及抗拒改變。

二、創新原則

1. 跨域創新：跨部門、跨層級、跨單位(業務單位、資訊單位)合作的創新模式。
2. 多點創新：多點試驗，找出 AI 可能的創新模式。
3. 從現有系統發展：從現有資訊系統導入 AI 技術作為創新實驗方案。
4. 從資料出發：

- 
- (1).建置以 AI 為目標導向的大數據資料庫，或建構事先分類的資料庫。
 - (2).盤點現有資料庫，評估那些資料適合發展 AI 主題。
 - (3).公部門使用大數據發展 AI 仍受到許多限制，可以從非個資依賴型資料開始。

5. 從小做起：透過試辦，或先概念驗證(POC)再逐步擴展軟硬體的投资，累積後續政策擴散的能量。
6. 從非核心職能開始：政府非核心職能工作多屬重複性高、繁文縟節。再者，AI 應用於公部門相對新穎，初期導入 AI 系統時，利用經驗學習訓練公務員 AI 技能，避免 AI 專案失敗時影響政府核心職能運作。

三、審慎選擇應用的政府職能

1. 大多數政府職能都適合導入 AI，仍需審慎研究可行性。
2. 若導入 AI 的政府職能具敏感性，或涉及個人資料保護法「敏感性個人資訊」，建議先形成社會共識。

四、AI 合法化

當 AI 自主決策(autonomous decision)用於直接裁決(adjudication)，需關注：

1. 程序正當性：修正法規使 AI 具合法性。
2. 可課責：公民對 AI 決策可依據法制作業程序提起訴願。
3. 可驗證：訂定 AI 系統調校、量測儀器校正標準作業流程(Standard Operating Procedures, SOP)；或公正獨立第三者之機構進行評估、驗證(Certification)或認證(Accreditation)。

五、導入策略

1. 建立機關內部人員 AI 技術能力

成立 AI 團隊，或公務員自己設計 AI 系統。公務員理解 AI 技術運作原理，較有能力提出合理 AI 專案需求、提出適合 AI 訓練的數據品質需



求、參與 AI 系統設計、檢驗 AI 模型、識別 AI 風險、建立驗收能力等，避免技術流失。

2. 政府採購

AI 專案與傳統 IT 或軟體開發專案不同，傳統 IT 或軟體採購，開發成效容易預期，公部門與廠商之間是「需求端、供給端」關係；AI 專案採購，公部門與廠商之間是「共同設計、共同創造、共同生產」關係，採購策略包括：

- (1).先試辦再逐步擴展。
- (2).先概念驗證(POC)再逐步擴展軟硬體的投资。
- (3).善用採購法的彈性機制，依採購特性因「案」制宜，訂定彈性、合宜的招標文件、選擇適當招標方式，選出具有足夠能力廠商辦理。

3. 公私協力或跨域治理

政府主導研究計劃，先評估 AI 專案可行性，再透過政府採購導入廠商的 AI 技術能力和資源等。

六、AI 系統的設計

1. 公務員參與 AI 專案的規劃與設計：公務員有 AI 技術的基本知識並參與 AI 系統設計，同時提高 AI 系統可解釋性和 AI 系統的效能/正確率。
2. 以人為本的 AI 設計：AI 是賦能工具，避免勞動力替代
3. AI 系統品質的評估：

不一定需要高標準的 AI 系統品質，關鍵在於將技術限制與行政實務相結合，包括：

 - (1).比人工好就採用。
 - (2).依不同政府職能和人機協作方式，評估 AI 系統品質
 - (3).權衡公共價值、資料品質與經濟邏輯(效率、時間、成本)。

(4).識別 AI 偏誤(bias)與可能風險，確保能實踐公共價值與倫理準則。

七、人機協作的管理

人機協作的重點是人與 AI 合作擴增和強化彼此的能力：

1.管理 AI：管理 AI 資訊品質在控制範圍內。

2.管理公務員：避免演算法趨避和自動化偏誤

(1).行政作法：避免人對 AI 的信任度過高或過低產生負面影響。

(2).技術作法：設立風險值或驗證能動性。

3.管理 AI 與公務員的協作：

(1).人機介面整合：一端是 AI 系統使用者介面設計，若使用者包括身心障礙者，須符合無障礙規範；另一端是公務員職務再設計，職務能力和行為的調整、工作重新設計等。

(2).人機協作模式：選擇合適的人機協作模式，例如直接控制、間接控制、演算型決策、協作或合作等各種可能性。

(3).權衡監督程度：過度監督降低 AI 應用潛力，且需要大量監督人力降低公部門 AI 科技接受度；若 AI 自主決策無涉直接裁決(adjudication)、無涉人民權利、義務和安全之政府職能，善用 AI 比人快與準的優勢。

4.公共行政流程再設計：原公共行政流程 + AI 系統 + AI 風險緩解機制。

八、建立 AI 專案的風險緩解機制

不同案例有不同 AI 風險緩解機制，共通性原則如下：

1. AI 風險緩解機制應融入公共行政流程，成為經常性的工作。

包括公共行政方法及資訊科技風險管理方法：

(1).公共行政方法：依據公部門 AI 治理的 5 種挑戰：公共行政改革、AI 官僚的合法性、AI 系統的品質(資料治理與演算法治理)、人機協作的管理、勞動力影響(勞動力替代與轉型)，設計 AI 風險緩解措施。

(2).資訊科技風險管理方法：依據風險管理生命週期，從威脅識別到定量和定性風險分析，利用接受、避免、轉嫁、防範、拒絕等策略，設計 AI 風險緩解措施。

2. AI 情境很重要(AI context matters)，依據 AI 專案設計 AI 風險緩解措施。
3. AI 自主決策用於直接裁決(adjudication)，或對人民權利、義務和安全影響越大，採用 AI 風險緩解機制越多，或監督程度越強。
4. 公務員具有領域知識及 AI 技術能力，有能力設計公部門的 AI 風險緩解機制，且可避免受到 AI 主動性(自適應性)和能動性的影響。

九、為使 AI 系統具可信任性，建立 AI 系統使用前的檢核制度：

1. 個人資料保護的「合理性」。
2. AI 系統產出資訊的「非歧視性」。
3. AI 系統產出資訊平均正確率、穩定度比人高的「公信力」。
4. AI 系統成本、效率比人高的「效能性」。
5. AI 演算法可解釋性(explainable/explainability)、可理解性(interpretable/interpretability)的「透明度」。
6. AI 行使行政裁量權的「合法性」。
7. AI 官僚直接決策的「可課責性」。
8. AI 官僚直接行政裁罰證據的「再現性」。
9. AI 雙重用途的「明確性」。
10. 利害關係人科技接受度的「政治可行性」。

十、基於 AI 情境很重要，各機關仍須依照政府職能、情境相關性、人機協作模式，參酌本參考指引另訂使用規範或設計相對應的 AI 風險緩解機制。

資料來源：本研究整理



第四節 未來研究方向

依據本研究目前成果不足之處，研提未來研究方向建議：

1. 擴大研究範圍，從政府是使用者，擴展至推動者、促進者和管制者。
2. 公部門應用 AI 的案例研究。
3. 雙軸轉型的研究。
4. 參與全球 AI 治理或國際 AI 政策的研究。

茲分別詳細說明如下：

一、擴大研究範圍，從政府是使用者，擴展至推動者、促進者和管制者

本研究聚焦政府是使用者，但公共治理範圍甚廣，政府可以身兼使用者、推動者、促進者和管制者的多重角色，建議未來擴大研究範圍。

二、公部門應用 AI 的案例研究

研究發現：「AI 情境很重要」，AI 技術可以用於截然不同的政府職能或公共服務領域，產生截然不同的結果，許多意想不到的挑戰將會出現。未來研究可繼續累積不同情境的 AI 治理案例，包括：

1. 因應 AI 新技術的發展：例如生成式 AI 的政府應用等。
2. 不同政府職能或任務領域：本研究立意取樣 4 個不同政府職能的代表性案例，包括 F1：一般公共服務、F2：公共秩序與安全、F4：經濟事務、F5：環保。然而，聯合國政府職能分類(COFOG)分為 10 類，建議可以針對 F3：國防；F6：住房和社區設施；F7：健康、F8：娛樂、文化和宗教；F9：教育；F10：社會保障等職能領域進行研究。

3. 相同案例持續追蹤研究：公部門應用 AI 在初期階段的案例，會因執行時間不足，導致所推行的理論架構產生缺陷的可能性。有研究者認為演算法的問題和基本價值往往在遇到有問題案例出現(Mittelstadt et al., 2016)，或長期對多個使用者進行調查，或是必須花 1.5 至 3 年才能進行衡量及科學驗證(Misuraca et al., 2019)。例如波蘭勞動和社會政策部 2012 年設計的「AI 失業評分系統」，2018 年憲法法庭裁定違反波蘭憲法，2019 年拆除。因此，建議對相同案例持續追蹤研究，獲得完整 AI 工具生命週期的觀點。

三、雙軸轉型的研究

歐盟數位歐洲計劃 2021 年底提出歐洲數位轉型概念：「雙軸轉型」(Twin Transformation)的概念，亦即永續轉型(sustainability transformation)與數位轉型(Digital Transformation)，這兩種轉型對於組織的發展都至關重要。然而，研究和實踐通常單獨關注其中一種轉變，或者優先考慮其中一種轉變，缺乏對兩種轉型相互作用的綜合觀點¹⁹⁷。未來可研究兩種轉型相互作用的綜合觀點，加速達成 SDGS 的政策目標。

四、參與全球 AI 治理或國際 AI 政策的研究

有鑑於 AI 數位生態系統的跨境性質，純粹的國家管制制度效率低且成本高昂，有研究呼籲在現有架構中容納國家和非國家行動者的治理措施，展開全球 AI 治理架構。然而，專門討論全球 AI 治理或國際 AI 政策的文獻稀少，以及公部門實際上並沒有參與全球 AI 治理討論的困境(Schmitt, 2022)。

再者，臺灣企業的 AI 模型部署率(開發的模型用於實際生產比例)在 50-75%

¹⁹⁷ 檢自 https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1254&context=ecis2023_rp

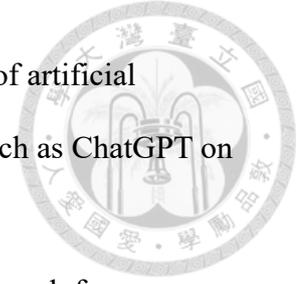
之間，高於全球平均(部署率約 20%)(葉哲良、李育杰, 2022)，臺灣有實力及能力把公部門應用 AI 的經驗與 AI 領先國家分享及建立合作關係，並積極參與全球 AI 治理或國際 AI 政策的研究，提升臺灣 AI 的國際影響力。



參考文獻



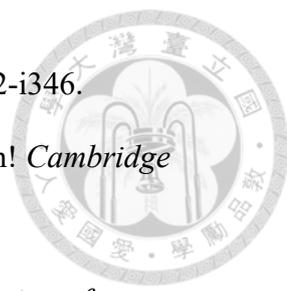
- Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424-438.
- Acemoglu, D., & Restrepo, P. (2019). Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda* (pp. 197-236): University of Chicago Press.
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6, 52138-52160.
- Agarwal, P. K. (2018). Public administration challenges in the world of AI and Bots. *Public Administration Review*, 78(6), 917-921.
- Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, 29(1), 1-8.
- Almeida, P., Santos, C., & Farias, J. S. (2020). *Artificial Intelligence Regulation: A Meta-Framework for Formulation and Governance*. Paper presented at the Proceedings of the 53rd hawaii international conference on system sciences.
- Alqooti, A. A. (2020). Public governance in the public sector: literature review. *International Journal of Business Ethics and Governance*, 3(3), 14-25.
- Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, 36(2), 358-367.

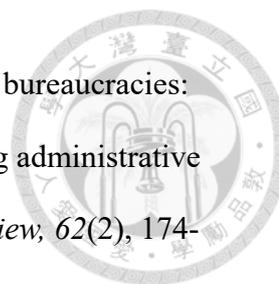


- Ang, T. L., Choolani, M., See, K. C., & Poh, K. K. (2023). The rise of artificial intelligence: addressing the impact of large language models such as ChatGPT on scientific publications. *Singapore Medical Journal*, 64(4), 219.
- Ansari, F., Hold, P., & Khobreh, M. (2020). A knowledge-based approach for representing jobholder profile toward optimal human-machine collaboration in cyber physical production systems. *CIRP Journal of manufacturing science and technology*, 28, 87-106.
- Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, 37(4), 101490.
- Aquaro, V. (2019). Can digital government accelerate the sustainable development goals? A pathway to the future. *Dubai Policy Review*. Retrieved from <https://dubaipolicyreview.ae/can-digital-government-accelerate-the-sustainable-development-goals-a-pathway-to-the-future/>
- Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35(3), 611-623.
- Ashmore, R., Calinescu, R., & Paterson, C. (2021). Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Computing Surveys (CSUR)*, 54(5), 1-39.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., . . . Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, 30(2), 100286.
- Bali, A. S., Howlett, M., Lewis, J. M., & Ramesh, M. (2021). Procedural policy tools in

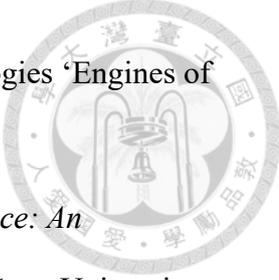


- theory and practice. *40*(3), 295-311.
- Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. *Government Information Quarterly*, *31*(1), 119-128.
- Barcevičius, E., Cibaitė, G., Codagnone, C., Gineikytė, V., Klimavičiūtė, L., Liva, G., . . . Vanini, I. (2019). *Exploring Digital Government transformation in the EU - Analysis of the state of the art and review of literature*: Publications Office of the European Union, Luxembourg.
- Barth, T. J., & Arnold, E. (1999). Artificial intelligence and administrative discretion: Implications for public administration. *The American Review of Public Administration*, *29*(4), 332-351.
- Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*, *37*(1), 38-54.
- Ben-Haim, Y., Osteen, C. D., & Moffitt, L. J. (2013). Policy dilemma of innovation: An info-gap approach. *Ecological Economics*, *85*, 130-138.
- Berditchevskaia, A., Malliaraki, E., & Peach, K. (2021). Participatory AI for humanitarian innovation. In: Nesta.
- Berkovich, I. (2020). Something borrowed, something blue: reflections on theory borrowing in educational administration research. *Journal of Educational Administration*, *58*(6), 749-760.
- Bhatnagar, S., Alexandrova, A., Avin, S., Cave, S., Cheke, L., Crosby, M., . . . Ó hÉigearthaigh, S. (2018). *Mapping intelligence: Requirements and possibilities*. Paper presented at the Philosophy and theory of artificial intelligence 2017.
- Bignami, F. (2022). Artificial Intelligence Accountability of Public Administration. *The*

- 
- American Journal of Comparative Law*, 70(Supplement_1), i312-i346.
- Bijker, W. E. (2010). How is technology made?—That is the question! *Cambridge journal of economics*, 34(1), 63-76.
- Bijker, W. E., Hughes, T. P., & Trevor, J. (1987). *The Social Construction of Technological Systems - New Directions in the Sociology and History of Technology*: The MIT Press Cambridge, Massachusetts London, England.
- Bloom, P. (2020). *Identity, Institutions and Governance in an AI World: Transhuman Relations*: Springer.
- Bobadilla-Suarez, S., Sunstein, C. R., & Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of risk and uncertainty*, 54(3), 187-202.
- Bongers, F., De Bruyn, K., & Verlet, D. (2018). Small kids, big data: toepassing analytics in de kinderopvang in Vlaanderen. *VLAAMS TIJDSCHRIFT VOOR OVERHEIDSMANAGEMENT*, 2018(1), 63-72.
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573-1576.
- Bouko, C., & Garcia, D. (2020). Patterns of emotional Tweets: The case of Brexit after the referendum results. In *Twitter, the Public Sphere, and the Chaos of Online Deliberation* (pp. 175-203): Springer.
- Bovaird, T., & Löffler, E. (2001). Emerging trends in public management and governance. *BBS Teaching and Research Review*, 5(5), 1-9.
- Bovaird, T., & Löffler, E. (2002). Moving from excellence models of local service delivery to benchmarking ‘good local governance’. *International Review of Administrative Sciences*, 68(1), 9-24.



- Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174-184.
- Boychenko, K. (2019). *Agency of Interactive Architecture in socio-technological relationship through Actor-Network Theory*. Paper presented at the “Hello, Culture” 18th International Conference, CAAD Futures 2019.
- Boychenko, K., Teixeira, F. F., Kuzovkova, T., & Boychenko, I. (2020). *Role of interactive space in social relationships through actor-network theory*. Paper presented at the 2020 International Conference on Engineering Management of Communication and Technology (EMCTECH).
- Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: how might the New Zealand government respond? *Policy Quarterly*, 13(4).
- Boyle, D., Mamais, L., & Khabarov, N. (2021). A Case Study Grassland Monitoring in Estonia. (EARSC). Retrieved from https://earsc.org/sebs/wp-content/uploads/2021/05/Grassland-Monitoring-in-Estonia_vfinal.pdf
- Bradshaw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The computational propaganda project*, 1, 1-26.
- Breit, E., Egeland, C., & Løberg, I. B. (2019). Cyborg bureaucracy: Frontline work in digitalized labor and welfare services. In *Big Data*: Edward Elgar Publishing.
- Breit, E., Egeland, C., Løberg, I. B., & Røhnebæk, M. T. (2021). Digital coping: How frontline workers cope with digital service encounters. *Social Policy & Administration*, 55(5), 833-847.

- 
- Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies ‘Engines of growth’? *Journal of econometrics*, 65(1), 83-108.
- Brundage, M. (2019). *Responsible Governance of Artificial Intelligence: An Assessment, Theoretical Framework, and Exploration*. Arizona State University,
- Bueger, C., & Stockbruegger, J. (2017). Actor-network theory: objects and actants, networks and narratives. In *Technology and World Politics* (pp. 42-59): Routledge.
- Buffat, A. (2015). Street-level bureaucracy and e-government. *Public Management Review*, 17(1), 149-161.
- Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7), 751-761.
- Bullock, J. B., Hansen, J. R., & Houston, D. J. (2018). Sector Differences in Employee’s Perceived Importance of Income and Job Security: Can These be Found Across the Contexts of Countries, Cultures, and Occupations? *International public management journal*, 21(2), 243-271.
- Burke, B., Cearley, D., Jones, N., Smith, D., Chandrasekaran, A., Lu, C., & Panetta, K. (2019). Gartner top 10 strategic technology trends for 2020-Smarter with Gartner. In: Gartner.
- Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, 23(1), 3-28.
- Cabaneros, S. M. S., Calautit, J. K. S., & Hughes, B. R. (2017). Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO2 pollution. *Energy Procedia*, 142, 3524-3530.
- Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The sociological review*, 32(1_suppl),



- 196-233.
- Callon, M. (1986). Éléments pour une sociologie de la traduction: la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L'Année sociologique (1940/1948-)*, 36, 169-208.
- Calo, R. (2017). Artificial intelligence policy: a primer and roadmap. *UCDL Rev.*, 51, 399-435.
- Campion, A., Gasco-Hernandez, M., Jankin Mikhaylov, S., & Esteve, M. (2022). Overcoming the challenges of collaboratively adopting artificial intelligence in the public sector. *Social Science Computer Review*, 40(2), 462-477.
- Campion, A., Hernandez, M.-G., Jankin, S. M., & Esteve, M. (2020). Managing artificial intelligence deployment in the public sector. *Computer*, 53(10), 28-37.
- Carrascosa, A., & Banjac, B. (2021). AI & CIVIC TECH BOOST YOUTH PARTICIPATION. Retrieved from <https://ai4da.com/ai-civic-tech-boost-youth-participation/>
- Casares, A. P. (2018). The brain of the future and the viability of democratic governance: The role of artificial intelligence, cognitive machines, and viable systems. *Futures*, 103, 5-16.
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180080.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, 24(2), 505-528.
- Cepiku, D. (2013). Unraveling the concept of public governance: A literature review of

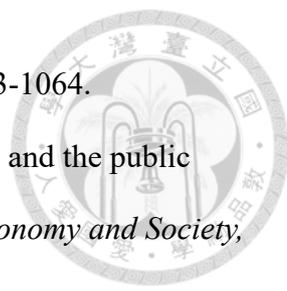


- different traditions. *Conceptualizing and researching governance in public and non-profit organizations*, 1, 3-32.
- Chatfield, A. T., & Reddick, C. G. (2019). A framework for Internet of Things-enabled smart government: A case of IoT cybersecurity policies and use cases in US federal government. *Government Information Quarterly*, 36(2), 346-357.
- Chen, H. (2019). *Success factors impacting artificial intelligence adoption: Perspective from the Telecom Industry in China*. Old Dominion University, Retrieved from <https://www.proquest.com/docview/2302014868?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20These>
- Chen, P.-Y., & Das, P. (2023). AI Maintenance: A Robustness Perspective. *Computer*, 56(2), 48-56.
- Choi, B. K., Yeo, W.-D., & Won, D. (2018). The implication of ANT (Actor-Network-Theory) methodology for R&D policy in open innovation paradigm. *Knowledge Management Research & Practice*, 16(3), 315-326.
- Cohen, J. E. (2019). Between truth and power. In J. E. Cohen (Ed.), *Information, Freedom and Property* (pp. 24): Oxford University Press.
- Couldry, N. (2008). Actor network theory and media: Do they connect and on what terms? In A. Hepp, Krotz, Friedrich, Moores, Shaun and Winter, Carsten, (eds.) (Ed.), *Connectivity, Networks and Flows: Conceptualizing Contemporary Communications* (pp. 93-110): Hampton Press Inc.
- Crawford, T. H. (2020). Actor-network theory. In *Oxford Research Encyclopedia of Literature*.
- Crosset, V., & Dupont, B. (2022). Cognitive assemblages: The entangled nature of

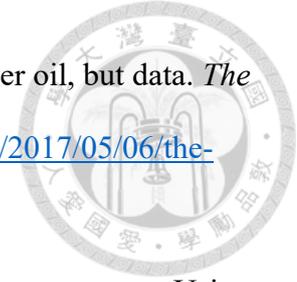
- 
- algorithmic content moderation. *Big Data & Society*, 9(2), 20539517221143361.
- Cuau, C. (2019). Artificial Intelligence and Citizen Participation, a Case-Study. (CitizenLab's Blog). Retrieved from <https://www.citizenlab.co/blog/civic-engagement/youth-for-climate-case-study/>
- Dafoe, A. (2018). AI governance: a research agenda. *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK.*
- Dahl, R. A. (1957). The concept of power. *Behavioral science*, 2(3), 201-215.
- Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., . . . Witteborn, S. (2019). Artificial intelligence, governance and ethics: Global perspectives. *The Chinese University of Hong Kong Faculty of Law Research Paper(2019-15).*
- Davenport, T. H., & Kirby, J. (2016). Just how smart are smart machines?
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International journal of human-computer studies*, 45(1), 19-45.
- Davis, J., & Goadrich, M. (2006). *The relationship between Precision-Recall and ROC curves*. Paper presented at the Proceedings of the 23rd international conference on Machine learning.
- de Boer, N., & Raaphorst, N. (2021). Automation and discretion: explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, 1-21.
- de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government*



- Information Quarterly*, 39(2), 101666.
- de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, 36(4), 101392.
- De Spiegeleire, S., Maas, M., & Sweijjs, T. (2017). *Artificial intelligence and the future of defense: strategic implications for small-and medium-sized force providers*: The Hague Centre for Strategic Studies.
- de Vries, M., Moeller, I., Peralta, G., Morris, E., Stanica, A., Scricciu, A., . . . van Wesenbeeck, B. (2018). Earth observation and the coastal zone: from global images to local information. In *The ever growing use of Copernicus across Europe's regions: A selection of 99 user stories by local and regional authorities* (pp. 148-149): European Commission, NEREUS, European Space Agency.
- Derks, D., Fischer, A. H., & Bos, A. E. (2008). The role of emotion in computer-mediated communication: A review. *Computers in human behavior*, 24(3), 766-785.
- DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization science*, 5(2), 121-147.
- Desouza, K. C. (2018). Delivering artificial intelligence in government: challenges and opportunities. *IBM Center for the Business of Government*.
- Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, 63(2), 205-213.
- Desouza, K. C., & Jacob, B. (2017). Big data in the public sector: Lessons for

- 
- practitioners and scholars. *Administration & society*, 49(7), 1043-1064.
- Dignam, A. (2020). Artificial intelligence, tech corporate governance and the public interest regulatory response. *Cambridge Journal of Regions, Economy and Society*, 13(1), 37-54.
- Djeffal, C. (2020). Artificial Intelligence and Public Governance: Normative Guidelines for Artificial Intelligence in Government and Public Administration. In *Regulating Artificial Intelligence* (pp. 277-293): Springer.
- Dong, Y., Hou, J., Zhang, N., & Zhang, M. (2020). Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. *Complexity*, 2020, 1-10.
- Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., . . . Weinberger, D. (2017). Accountability of AI under the law: The role of explanation. *arXiv preprint arXiv:1711.01134*.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71.
- Dunleavy, P., & Margetts, H. (2023). Data science, artificial intelligence and the third wave of digital era governance. *Public Policy and Administration*, 09520767231198737.
- Dunn, W. N. (2015). *Public policy analysis*: Routledge.
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., . . . Eirug, A. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.

Economist, T. (2017). The world's most valuable resource is no longer oil, but data. *The Economist*. Retrieved from <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>



Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). AI-augmented government. Using cognitive technologies to redesign public sector work. *Deloitte Center for Government Insights*, 1, 24.

Ekelund, S. (2017). Precision-Recall Curves—What Are They and How Are They Used. *Radiometer Medical ApS: Denmark, March*.

El-Haddadeh, R., Weerakkody, V., Osmani, M., Thakker, D., & Kapoor, K. K. (2019). Examining citizens' perceived value of internet of things technologies in facilitating public sector services engagement. *Government Information Quarterly*, 36(2), 310-320.

Elmi, N., Broekaert, K., & Larsen, A.-M. E. (2018). *Agile Governance: Reimagining Policy-Making in the Fourth Industrial Revolution*. Paper presented at the White Paper. World Economic Forum, January.

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. *NYU School of Law, Public Law Research Paper*(20-54).

Escalante, H. J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., . . . van Lier, R. (2018). *Explainable and interpretable models in computer vision and machine learning*: Springer.

Fatima, S., Desouza, K. C., & Dawson, G. S. (2020). National strategic artificial intelligence plans: A multi-dimensional analysis. *Economic Analysis and Policy*, 67, 178-194.

Feenberg, A. (1992). Subversive rationalization: Technology, power, and democracy.

Inquiry, 35(3-4), 301-322.

Fischer, G. (2023). *Adaptive and Adaptable Systems: Differentiating and Integrating AI and EUD*. Paper presented at the International Symposium on End User Development.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*(2020-1).

Flach, P., & Kull, M. (2015). Precision-recall-gain curves: PR analysis done right. *Advances in neural information processing systems*, 28.

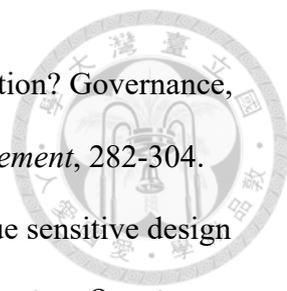
Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180081.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . Rossi, F. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707.

Fountain, J. (2019). The wicked nature of digital transformation: A policy perspective. *Dubai Policy Review*, 1, 40.

Fountain, J. E. (2004). *Building the virtual state: Information technology and institutional change*: Brookings Institution Press.

FRA. (2020). *Artificial Intelligence, Big Data and Fundamental Rights Country Research Estonia 2020*. Retrieved from The EU Agency for Fundamental Rights https://fra.europa.eu/sites/default/files/fra_uploads/fra-ai-project-estonia-country-research_en.pdf

- 
- Frederickson, H. G. (2005). Whatever happened to public administration? Governance, governance everywhere. *The Oxford handbook of public management*, 282-304.
- Friedman, B., Kahn, P. H., Borning, A., & Huldgtren, A. (2013). Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95): Springer.
- Fung, A. (2015). Putting the public back into governance: The challenges of citizen participation and its future. *Public Administration Review*, 75(4), 513-522.
- Garson, G. D. (2003). Technological teleology and the theory of technology enactment: The case of the international trade data system. *Social Science Computer Review*, 21(4), 425-431.
- Garvey, J. (2007). The moral use of technology. *Royal Institute of Philosophy Supplements*, 61, 241-260.
- Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58-62.
- Gellers, J. C. (2020). *Rights for robots: artificial intelligence, animal and environmental law (edition 1)*: Routledge.
- Gerring, J. (2004). What is a case study and what is it good for? *American political science review*, 98(2), 341-354.
- Ghazinoory, S., & Hajishirzi, R. (2012). Using actor-network theory to identify the role of IT in cognitive science in Iran. *Procedia-Social and Behavioral Sciences*, 32, 153-162.
- Giest, S. N., & Klievink, B. (2022). More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts. *Public Management Review*, 1-20.

Gillespie, T. (2010). The politics of 'platforms'. *new media & society*, 12(3), 347-364.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*:

Harvard University Press.

Grant, R. (1995). Values, means and ends. *Royal Institute of Philosophy Supplements*, 38, 177-188.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D.

(2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5), 1-42.

Guilbeault, D. (2016). Automation, algorithms, and Politics| growing bot security: an ecological view of bot agency. *International journal of communication*, 10, 19.

Habermas, J. (1968). Technik und Wissenschaft als» Ideologie «? *Merkur*, 22(243), 591-610.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in human behavior*, 119, 106730.

Hall, P. A. (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative politics*, 275-296.

Hegele, Y. (2018). Explaining bureaucratic power in intergovernmental relations: A network approach. *Public Administration*, 96(4), 753-768.

Helbing, D. (2019). *Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies*: Springer.

Henman, P. (2017). The computer says 'DEBT': Towards a critical sociology of algorithms and algorithmic governance.

Henman, P. (2020). Improving public services using artificial intelligence: possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, 42(4), 209-221.

Hernández-Orallo, J. (2014). AI evaluation: past, present and future (2014). *arXiv preprint arXiv:1408.6908*.

Hernández-Orallo, J. (2017). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review*, 48(3), 397-447.

Hill, C. J., & Lynn, L. E. (2004). Is hierarchical governance in decline? Evidence from empirical research. *Journal of public administration research and theory*, 15(2), 173-195.

Hill, R. K. (2016). What an algorithm is. *Philosophy & Technology*, 29(1), 35-59.

Hines, P., Yu, L. H., Guy, R. H., Brand, A., & Papaluca-Amati, M. (2019). Scanning the horizon: a systematic literature review of methodologies. *BMJ open*, 9(5), 1-9.

HLEG. (2019). *High-level expert group on artificial intelligence: A definition of AI: Main capabilities and scientific disciplines*. Retrieved from <https://www.aepd.es/sites/default/files/2019-09/ai-definition.pdf>

Höchtel, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 147-169.

Holdren, J., & Smith, M. (2016). *Artificial intelligence, automation and the economy*. Retrieved from <https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy>

Hood, C. (1983). *Tools of government*: Macmillan International Higher Education.

House, W. (2019). *2016–2019 Progress Report: Advancing Artificial Intelligence R&D*.



Retrieved from <https://www.nitrd.gov/pubs/AI-Research-and-Development-Progress-Report-2016-2019.pdf>

Hughes, T. P. (1993). *Networks of power: electrification in Western society, 1880-1930*. JHU press.

Insights, O. (2020). Government AI Readiness Index 2020. Retrieved from <https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f6dc236342b2a77d225b0f3/1601028752121/AI+Readiness+Index+2020+-+full+report.pdf>

Insights, O. (2021). Government AI Readiness Index 2021. Retrieved from <https://www.oxfordinsights.com/government-ai-readiness-index2021>

Jackson, M. H., Poole, M. S., & Kuhn, T. (2002). The social construction of technology in studies of the workplace. *Handbook of new media: Social shaping and consequences of ICTs*, 236-253.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 37(3), 101493.

Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society*, 9(2), 1-6.

Jastroch, N. (2021). *Applied artificial intelligence: risk mitigation matters*. Paper presented at the IFIP International Conference on Product Lifecycle Management.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, 3(4), 305-360.

Jimenez-Gomez, C. E., Cano-Carrillo, J., & Lanás, F. F. (2020). Artificial intelligence in

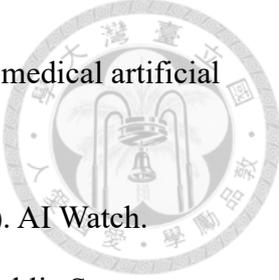


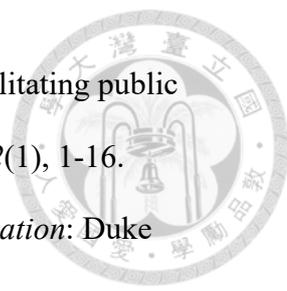
- government. *Computer*, 53(10), 23-27.
- Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics*, 127(4), 707-715.
- Johnson N, Z. G., Hunsader E, Qi H, Johnson N, Meng J and Tivnan B. (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports* 3(1), 2627. Retrieved from <https://www.nature.com/articles/srep02627>
- Johnston, R. B. (2001). Situated action, structuration and actor-network theory: an integrative theoretical perspective. *ECIS 2001 Proceedings*, 73.
- Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2023). Who Answers It Better? An In-Depth Analysis of ChatGPT and Stack Overflow Answers to Software Engineering Questions. *arXiv preprint arXiv:2308.02312*.
- Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, 2, 100017.
- Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.
- Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. In (Vol. 36, pp. 304-309): Elsevier.
- Karppi, T., & Crawford, K. (2016). Social media, financial algorithms and the hack crash. *Theory, culture & society*, 33(1), 73-92.
- Karppi, T., Kähkönen, L., Mannevu, M., Pajala, M., & Sihvonen, T. (2016). Affective capitalism. *Ephemera: Theory and Politics in Organization*, 16(4), 1-13.
- Kaur, G., & Lashkari, A. H. (2021). Information Technology Risk Management. In *Advances in Cybersecurity Management* (pp. 269-287): Springer.
- Kettl, D. F. (2000). The transformation of governance: Globalization, devolution, and

- 
- the role of government. *Public Administration Review*, 60(6), 488-497.
- Kettl, D. F. (2006). *The global public management revolution: A report on the transformation of governance*: Brookings Institution Press.
- Kitchner, K. S. (1983). Cognition, metacognition, and epistemic cognition. *Human development*, 26(4), 222-232.
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The quarterly journal of economics*, 133(1), 237-293.
- Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *new media & society*, 20(12), 4653-4670.
- Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, 37(3), 101488.
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388-409.
- Kranzberg, M. (1991). Science-technology-society: It's as simple as XYZ! *Theory into Practice*, 30(4), 234-241.
- Kulshreshtha, P. (2008). Public sector governance reform: The World Bank's framework. *International Journal of Public Sector Management*, 21(5), 556-567.
- Kuwajima, H., & Ishikawa, F. (2019). *Adapting SQuaRE for quality assessment of artificial intelligence systems*. Paper presented at the 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW).
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings.



- Telecommunications policy*, 44(6), 101976.
- Latour, B. (1984). The powers of association. *The sociological review*, 32(1_suppl), 264-280.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*: Harvard university press.
- Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale welt*, 369-381.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*: Oxford university press.
- Latour, B. (2011). Networks, societies, spheres: Reflections of an actor-network theorist. *International journal of communication*, 5, 15.
- Latour, B., & Woolgar, S. (2013). *Laboratory life: The construction of scientific facts*: Princeton University Press.
- Lee, M. K., & Rich, K. (2021). *Who Is Included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust*. Paper presented at the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, 31, 611-627.
- Lester, S. M. (2002). *The tools of government. A guide to the New Governance*. New York: Oxford University Press.
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., . . . Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1-46.

- 
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629-650.
- Luca, T., Colin, V. N., Marco, C., Dietmar, G., & Francesco, P. (2022). AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector.
- Lund, S., Manyika, J., Segel, L. H., Dua, A., Hancock, B., Rutherford, S., & Macon, B. (2019). The future of work in America: People and places, today and tomorrow. Retrieved from <https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-in-america-people-and-places-today-and-tomorrow>
- Lynn Jr, L. E. (2006). *Public management: Old and new*: Routledge.
- Lynn Jr, L. E., Heinrich, C. J., & Hill, C. J. (2001). *Improving governance: A new logic for empirical research*: Georgetown University Press.
- Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, 83(1_suppl), 120-135.
- Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60.
- Malone, T., Vaccaro, M., Campero, A., Song, J., Wen, H., & Almaatouq, A. (2023). A Test for Evaluating Performance in Human-AI Systems.
- Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature*, 568(7751), 163-165.
- Marmolejo-Ramos, F., Workman, T., Walker, C., Lenihan, D., Moulds, S., Correa, J.

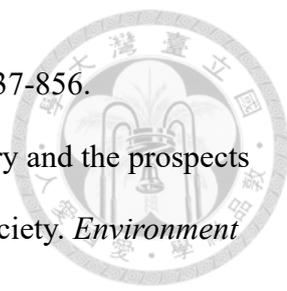
- 
- C., . . . Sonna, B. (2022). AI-powered narrative building for facilitating public participation and engagement. *Discover Artificial Intelligence*, 2(1), 1-16.
- Massumi, B. (2021). *Parables for the virtual: Movement, affect, sensation*: Duke University Press.
- McLean, S., Read, G. J., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5), 649-663.
- McQuillan, D. (2020). Deep bureaucracy and autonomist AI.
- Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. *Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch.*, no. August, 1-12.
- Meijer, A., & Löfgren, K. (2015). The neglect of technology in theories of policy change. *International Journal of Public Administration in the Digital Age (IJPADA)*, 2(1), 75-88.
- Mele, C., Spina, T. R., Kaartemo, V., & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, 129, 949-960.
- Mergel, I., Kattel, R., Lember, V., & McBride, K. (2018). *Citizen-oriented digital transformation in the public sector*. Paper presented at the Proceedings of the 19th annual international conference on digital government research: Governance in the data age.
- Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928-937.



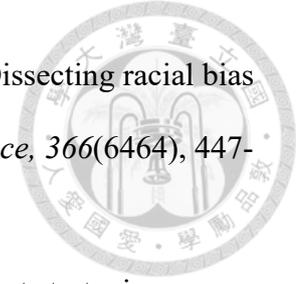
- Merhi, M. I. (2023). An evaluation of the critical success factors impacting artificial intelligence implementation. *International Journal of Information Management*, 69, 102545.
- Meszaros, J., Ho, C.-h., & Corrales Compagnucci, M. (2020). Nudging Consent and the New Opt-Out System to the Processing of Health Data in England. *Legal Tech and the New Sharing Economy*, 93-113.
- Mikhaylov, S. J., Esteve, M., & Champion, A. (2018). Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170357.
- Miller, S. M. (2018). AI: Augmentation, more so than automation. *Asian Management Insights (Singapore Management University)*, Vol. 5 Iss. 1 (Singapore Management University, Centre for Management Practice), 1 - 20. Retrieved from <http://works.bepress.com/steve-miller/2/>
- Mills, S., Duranton, S., Santinelli, M., Hua, G., Baltassis, E., Thiel, S., & Muehlstein, O. (2021). *Are You Overestimating Your Responsible AI Maturity*. Retrieved from <https://www.bcg.com/publications/2021/the-four-stages-of-responsible-ai-maturity>
- Minn, S. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence*, 3, 100050.
- Misuraca, G. (2020). *Rethinking Democracy in the “Pandemic Society” A journey in search of the governance with, of and by AI*. Paper presented at the CEUR Workshop Proceedings.
- Misuraca, G., Barcevicus, E., & Codagnone, C. (2019). *Exploring digital government transformation in the EU - Understanding public sector innovation in a data-*



- driven society*: Publications Office of the European Union, Luxembourg,.
- Misuraca, G., & van Noordt, C. (2020). *AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU*. Retrieved from <https://ideas.repec.org/p/ipt/iptwpa/jrc120399.html>
- Misuraca, G., & Viscusi, G. (2014). *Digital governance in the public sector: challenging the policy-maker's innovation dilemma*. Paper presented at the Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance.
- Misuraca, G., & Viscusi, G. (2015). Shaping public sector innovation theory: an interpretative framework for ICT-enabled governance innovation. *Electronic Commerce Research*, 15(3), 303-322.
- Mitchell, M. (2019). Artificial intelligence hits the barrier of meaning. *Information*, 10(2), 51.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Mok, L., & Hyysalo, S. (2018). Designing for energy transition through Value Sensitive Design. *Design Studies*, 54, 162-183.
- Molinari F., v. N. C., Vaccari L, Pignatelli F. and Tangi L. (2021). AI Watch. Beyond pilots: sustainable implementation of AI in public services. (Publications Office of the European Union, Luxembourg).
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics*, 26(4), 2141-2168.
- Moss, J., Robinson, E., & Watts, J. (2020). Brexit and the everyday politics of emotion:

- 
- Methodological lessons from history. *Political Studies*, 68(4), 837-856.
- Murdoch, J. (1997). Inhuman/nonhuman/human: actor-network theory and the prospects for a nondualistic and symmetrical perspective on nature and society. *Environment and planning D: Society and Space*, 15(6), 731-756.
- Murdoch, J. (2000). Networks—a new paradigm of rural development? *Journal of rural studies*, 16(4), 407-419.
- NAPA. (2020). Make Government AI Ready. (National Academy of Public Administration). Retrieved from <https://napawash.org/grand-challenges/make-government-ai-ready>
- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication theory*, 24(3), 340-360.
- Niklas, J. (2019). Poland: Government to scrap controversial unemployment scoring system. (AlgorithmWatch). Retrieved from <https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/>
- Nikolinakos, N. T. (2023). A European Approach to Excellence and Trust: The 2020 White Paper on Artificial Intelligence. In *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies-The AI Act* (pp. 211-280): Springer.
- Nisar, M. A., & Masood, A. (2018). From Street-level to Cyborg Bureaucrats: Theory and Evidence on Socio-materiality in Public Administration. Available at SSRN 3886338.
- O'Leary, D. E. (2014). Embedding AI and crowdsourcing in the big data lake. *IEEE Intelligent Systems*, 29(5), 70-73.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.



OECD. (2014). Recommendation of the council on digital government strategies. *Organisation for Economic Co-operation and Development*.

OECD. (2019). *Artificial Intelligence in Society*. Retrieved from <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

OECD. (2020). Recommendation of the Council on Artificial Intelligence. Retrieved from <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Ojo, A. (2019). *Next generation government-hyperconnected, smart and augmented*. Paper presented at the Collaborative Networks and Digital Transformation: 20th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2019, Turin, Italy, September 23–25, 2019, Proceedings 20.

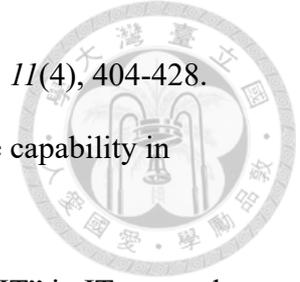
Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). *A realist perspective on AI-era public management*. Paper presented at the Proceedings of the 20th Annual International Conference on Digital Government Research.

Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *Plos one*, 15(2), e0229132.

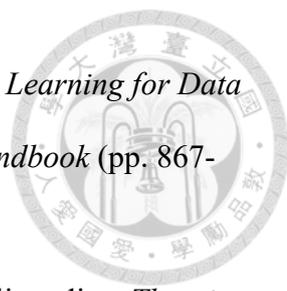
Oleinik, A. (2019). What are neural networks not good at? On artificial creativity. *Big Data & Society*, 6(1), 2053951719839433.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization science*, 3(3), 398-427.

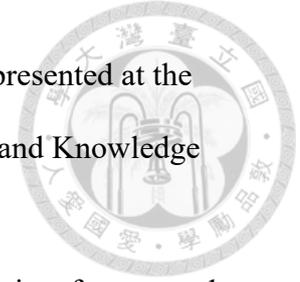
Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens



- for studying technology in organizations. *Organization science*, 11(4), 404-428.
- Orlikowski, W. J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science*, 13(3), 249-273.
- Orlikowski, W. J., & Iacono, C. S. (2001). Desperately seeking the “IT” in IT research— a call to theorizing the IT artifact. *Information systems research*, 12(2), 121-134.
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170359.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3), 286-297.
- Patel, J., Manetti, M., Mendelsohn, M., Mills, S., Felden, F., Littig, L., & Rocha, M. (2021). AI Brings Science to the Art of Policymaking. (BCG). Retrieved from <https://www.bcg.com/publications/2021/how-artificial-intelligence-can-shape-policy-making>
- Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI—A transformational shift for government: So, what next for research? *Public Policy and Administration*, 35(1), 24-44.
- Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of public administration research and theory*, 6(1), 5-22.
- Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review*, 367-373.

- 
- Pessach, D., & Shmueli, E. (2023). Algorithmic fairness. In *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook* (pp. 867-886): Springer.
- Peters, B. G., & Zittoun, P. (2016). Contemporary approaches to public policy. *Theories, controversies and perspectives*. UK: Palgrave Macmillan.
- Pinch, T., & Trocco, F. (2004). *Analog days: The invention and impact of the Moog synthesizer*: Harvard University Press.
- Pitt, C., Paschen, J., Kietzmann, J., Pitt, L. F., & Pala, E. (2023). Artificial intelligence, marketing, and the history of technology: Kranzberg's laws as a conceptual lens. *Australasian Marketing Journal*, 31(1), 81-89.
- Pollitt, C. (2011). Mainstreaming technological change in the study of public management. *Public Policy and Administration*, 26(4), 377-397.
- Prentice, C., Weaven, S., & Wong, I. A. (2020). Linking AI quality performance and customer engagement: The moderating effect of AI preference. *International Journal of Hospitality Management*, 90, 102629.
- Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5-14.
- Ramati, I., & Pinchevski, A. (2018). Uniform multilingualism: A media genealogy of Google Translate. *new media & society*, 20(7), 2550-2565.
- Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, 6(1), 119-149.
- Reis, J., Santo, P. E., & Melão, N. (2019). *Artificial intelligence in government services: A systematic literature review*. Paper presented at the World conference on information systems and technologies.

- 
- Rhodes, R. A. W. (1996). The new governance: governing without government. *Political Studies*, 44(4), 652-667.
- Richardson, J., Schlegel, K., Hostmann, B., & McMurchy, N. (2008). Magic quadrant for business intelligence platforms. *Core research note G*, 154227.
- Rosa, J. M., & Ruiz, C. J. (2020). Reason vs. emotion in the Brexit campaign: How key political actors and their followers used Twitter. *First Monday*.
- Rubin, V. L. (2022). Manipulation in marketing, advertising, propaganda, and public relations. In *Misinformation and Disinformation: Detecting Fakes with the Eye and AI* (pp. 157-205): Springer.
- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, culture & society*, 30(4), 22-46.
- Russell, S., & Norvig, P. (2002). *Artificial intelligence: a modern approach*: Pearson Education, Inc.
- Russell, S. J. (2010). *Artificial intelligence a modern approach*: Pearson Education, Inc.
- Samoili, S., Cobo, M. L., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*: Publications Office of the European Union, Luxembourg.
- Sayes, E. (2014). Actor–Network Theory and methodology: Just what does it mean to say that nonhumans have agency? *Social Studies of Science*, 44(1), 134-149.
- Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). *What's next for ai ethics, policy, and governance? a global overview*. Paper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.
- Schmidt, P., & Biessmann, F. (2020). *Calibrating Human-AI Collaboration: Impact of*



Risk, Ambiguity and Transparency on Algorithmic Bias. Paper presented at the International Cross-Domain Conference for Machine Learning and Knowledge Extraction.

- Schmitt, L. (2022). Mapping global AI governance: a nascent regime in a fragmented landscape. *AI and Ethics*, 2(2), 303-314.
- Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *The journal of politics*, 52(2), 510-529.
- Schoemaker, P. J., & Tetlock, P. E. (2017). Building a more intelligent enterprise. *MIT Sloan Management Review*, 58(3), 28.
- Schwartz, R. D. (1989). Artificial intelligence as a sociological phenomenon. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, 179-202.
- Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G.-J., Elkins, A., . . . Randrup, N. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, 57(2), 103174.
- Semeraro, F., Griffiths, A., & Cangelosi, A. (2023). Human–robot collaboration and machine learning: A systematic review of recent research. *Robotics and Computer-Integrated Manufacturing*, 79, 102432.
- Sharma, G. D., Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures*, 2, 100004.
- Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*: MIT press.
- Simon, H. A. (1947). A Comment on "The Science of Public Administration". *Public Administration Review*, 7(3), 200-203.

Sismondo, S. (2010). *An introduction to science and technology studies* (Vol. 1): Wiley-Blackwell Chichester.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51(5), 991-1006.

Smidt, A., Balandin, S., Sigafos, J., & Reed, V. A. (2009). The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, 34(3), 266-274.

Stead, W. W. (2018). Clinical implications and challenges of artificial intelligence and deep learning. *Jama*, 320(11), 1107-1108.

Stoker, G. (1998). Governance as theory: five propositions. *International social science journal*, 50(155), 17-28.

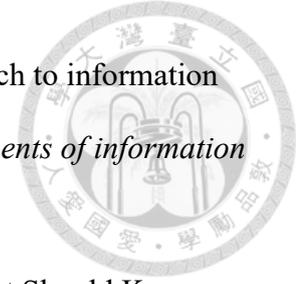
Subramaniam, N., Stewart, J., Ng, C., & Shulman, A. (2013). Understanding corporate governance in the Australian public sector: A social capital approach. *Accounting, Auditing & Accountability Journal*, 26(6), 946-977.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383.

Tam Cho, W. K. (2019). Technology-enabled coin flips for judging partisan gerrymandering. *S. Cal. L. Rev. Postscript*, 93, 11.

Tang, T., & Ho, A. T.-K. (2019). A path-dependence perspective on the adoption of Internet of Things: Evidence from early adopters of smart and connected sensors in the United States. *Government Information Quarterly*, 36(2), 321-332.

Tao, C., Gao, J., & Wang, T. (2019). Testing and Quality Validation for AI Software—Perspectives, Issues, and Practices. *IEEE Access*, 7, 120164-120175.



Tatnall, A. (2003). Actor-network theory as a socio-technical approach to information systems research. In *Socio-technical and human cognition elements of information systems* (pp. 266-283): Igi Global.

TDS. (2021). 5 Types of Machine Learning Bias Every Data Scientist Should Know. Retrieved from <https://towardsdatascience.com/5-types-of-machine-learning-bias-every-data-science-should-know-efab28041d3f>

Thaler, R., & Sunstein, C. (2008). *Nudge: The gentle power of choice architecture*. New Haven, Conn.: Yale.

Thierer, A. D., Castillo O'Sullivan, A., & Russell, R. (2017). Artificial intelligence and public policy. *Mercatus Research Paper*.

Todd, P. M. (2007). How much information do we need? *European Journal of Operational Research*, 177(3), 1317-1332.

Trump, D. (2019). *Maintaining American Leadership in Artificial Intelligence*. Retrieved from <https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf>

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the turing test* (pp. 23-65): Springer.

UK. (2020). *Artificial Intelligence and Public Standards: report*. Retrieved from <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>

UK, G. (2017). *Government transformation strategy 2017 to 2020*. Retrieved from <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020>

UN. (1980). *Classification of the Functions of Government*: New York: United Nations,

Department of International Economic and Social Affairs.

UN. (2018). *UN E-Government Survey 2018*. Retrieved from

<https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2018>



Unver, A. (2018). Artificial intelligence, authoritarianism and the future of political systems. *EDAM Research Reports*.

Valle-Cruz, D., Alejandro Ruvalcaba-Gomez, E., Sandoval-Almazan, R., & Ignacio Criado, J. (2019). *A review of artificial intelligence in government and its potential from a public policy perspective*. Paper presented at the Proceedings of the 20th Annual International Conference on Digital Government Research.

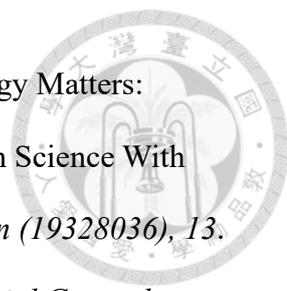
Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, 37(4), 101509.

van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security*, 23(4), 323-340.

van Bruxvoort, X., & van Keulen, M. (2021). Framework for assessing ethical aspects of algorithms and their encompassing socio-technical system. *Applied Sciences*, 11(23), 11187.

van Noordt, C., & Misuraca, G. (2022). Exploratory insights on artificial intelligence for government in Europe. *Social Science Computer Review*, 40(2), 426-444.

Veale, M. (2020). A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence. *European Journal of Risk Regulation*, 11(1), e1.

- 
- Waldherr, A., Geise, S., & Katzenbach, C. (2019). Because Technology Matters: Theorizing Interdependencies in Computational Communication Science With Actor-Network Theory. *International Journal of Communication (19328036)*, 13.
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1-37.
- Watson, L. (2018). Systematic epistemic rights violations in the media: A Brexit case study. *Social Epistemology*, 32(2), 88-102.
- Weber, M. (1978). *Economy and society: An outline of interpretive sociology* (Vol. 2): University of California press.
- Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco—Or was it something worse? *Australian Journal of Public Administration*, 80(2), 340-360.
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., . . . Schwartz, O. (2018). *AI now report 2018*: AI Now Institute at New York University New York.
- Whittlestone, J., Nyrop, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *London: Nuffield Foundation*.
- Wihlborg, E., Larsson, H., & Hedström, K. (2016). "The Computer Says No!"--A Case Study on Automated Decision-Making in Public Authorities. Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
- Williams, R., & Edge, D. (1996). The social shaping of technology. *Research policy*,

25(6), 865-899.

Wilson, C. (2022). Public engagement and AI: A values analysis of national strategies.

Government Information Quarterly, 39(1), 101652.

Winfield, A. F., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085.

Winner, L. (2017). *Do artifacts have politics?* : Routledge.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596-615.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, 43(9), 818-829.

Wogu, I. A. P., Misra, S., Udoh, O. D., Agoha, B. C., Sholarin, M. A., & Ahuja, R. (2020). *Artificial Intelligence Politicking and Human Rights Violations in UK's Democracy: A Critical Appraisal of the Brexit Referendum*. Paper presented at the The International Conference on Recent Innovations in Computing.

Wong, P.-H. (2020). Democratizing algorithmic fairness. *Philosophy & Technology*, 33, 225-244.

Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823-832.

Wyatt, S. (2008). Technological determinism is dead; long live technological



- 
- determinism. *The handbook of science and technology studies*, 3, 165-180.
- Yeung, K. (2018). Algorithmic regulation: a critical interrogation. *Regulation & Governance*, 12(4), 505-523.
- Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*, 2(4), 301-313.
- Yu-Jen, T., Hsini, H., Tze-Luen, L., & Wen-Hsi, C. (2023). Expanding Governance Capabilities: The Experience of AI Implementation in Taiwan. *east asian policy*, 15, 44-62.
- Zeng, Y. (2020). Artificial Intelligence: An Accelerator for United Nations Sustainable Development Goals. Retrieved from <http://www.ai-for-sdgs.academy/ai-as-an-accelerator-for-sdgs>
- Zhang, B., & Dafoe, A. (2020). *US Public opinion on the governance of artificial intelligence*. Paper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.
- Zhang, H., Zhang, X., & Song, M. (2021). Deploying AI for New Product Development Success: By embracing and incorporating AI in all stages of NPD, companies can increase their success rate of NPD projects. *Research-Technology Management*, 64(5), 50-57.
- Zhang, L. (2020). Initiatives in AI Governance. Retrieved from <https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/5fb58df18fbd7f2b94b5b5cd/1605733874729/SRI+1+-+Initiatives+in+AI+Governance.pdf>
- Zohuri, B., Zadfathollah, R., Balgehshiri, S. K. M., & Paydar, A. Z. (2023). Artificial

Intelligence Driven Nuclear Reactor Instrumentation and Control Measurements
In-core and Out-core in Nuclear Power Plants. *Journal of Material Sciences &
Manufacturing Research*. SRC/JMSMR-162. DOI: doi. org/10.47363/JMSMR/2023
(4), 145, 2-5.

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial
intelligence in public governance: A systematic literature review and a research
agenda. *Government Information Quarterly*, 101577.

丁玉珍、林子倫. (2020). 人工智慧提升政府公共治理的應用潛力探討. *檔案半年
刊*, 19(2), 24-41.

方念萱. (2016). 媒介化, 行動化: 媒介化理論與行動者網絡理論的對話. *傳播文化
與政治*(4), 55-83.

王一旅. (2021). 深度學習 AI 醫療民事侵權責任之探討-以美國, 歐盟及我國法制
為中心. *輔仁法學*(61), 1-85.

王自雄、張腕純. (2020). 值得信賴的健康醫療 AI—從資料共享之治理論起. *法律
與生命科學*, 9(2), 1-28.

王志弘、高郁婷譯. (2023). *技術與社會理論(Technology and Social Theory)*: 群學
出版有限公司.

丘昌泰. (2008). *公共政策: 基礎篇*: 巨流圖書股份有限公司.

江伊嵐、陳奕廷、葉梓銓. (2020). 智慧型動態交通號誌控制系統-應用於台北市重
點壅塞區域. *土木水利*, 47(4), 20-27.

江明修. (1997). *公共行政學: 理論與社會實踐*: 台灣五南圖書出版股份有限公司.

江美虹、吳朝欽. (2015). 營業稅逃漏行為之研究—以留抵稅額為例. *財稅研究*, 44.

行政院. (2018). 臺灣 AI 行動計畫. Retrieved from

<https://digi.ey.gov.tw/File/4C622B6A10053DAD>



- 行政院. (2023). 臺灣 AI 行動計畫 2.0 (2023-2026 年). Retrieved from <https://digi.nstc.gov.tw/File/7C71629D702E2D89>
- 何建志. (2010). 台灣 H1N1 疫苗接種後不良反應事件之因果關係: 科學不確定性與法律舉證責任. *法律與生命科學*, 4(1), 1-22.
- 吳全峰、許慧瑩. (2018). 健保資料目的外利用之法律爭議——從去識別化作業工具談起. *Taiwan Law Review*(272).
- 吳定. (2017). *公共政策*: 五南圖書出版股份有限公司.
- 李怡青. (2012). 權力基礎理論: 權力定義與權力增生性. *中華心理學刊*, 54(2), 203-217.
- 李承嘉. (2005). 行動者網絡理論應用於鄉村發展之研究: 以九份聚落 1895-1945 年發展為例. *台灣地理學報*, 39, 1-30.
- 李承嘉、廖本全、戴政新. (2010). 地方發展的權力與行動分析: 治理性與行動者網絡理論觀點的比較. *台灣土地研究*, 13(1), 95-133.
- 李開復、王詠剛. (2017). *人工智慧來了: 天下文化*.
- 李聖傑. (2002). 因果關係的判斷在刑法中的思考. *追念黃建輝老師學術研討會---法學上之因果關係論文集*, 117-146.
- 汪林玲. (2014). 人民陳情制度之探討-以監察陳情制度為例. *中國行政評論*, 20, 81-114.
- 辛炳隆, 葉俊偉, 林金玲, & 陳秋蓉. (2023). 人與機械人協作趨勢與勞動權益保障之國際經驗研析. (勞動部).
- 周信輝、方世杰、李慶芳、蔡馥陞、劉亭蘭. (2022). 與機器共舞: 以交織理論探索人機協作的能動性. *中山管理評論*, 30(5), 857-897.
- 周信輝、蔡志豪. (2013). 網絡組織間關係治理的搭架: 行動者網絡理論的動態觀點. *臺大管理論叢*, 23(S1), 135-163.

- 
- 官美吟、彭立沛. (2018). 以行動者網路概念探討水梯田復育之治理模式. *農林學報*, 66(3, 4), 201-220.
- 林子倫. (2017). 大數據與健康研究之公民參與. *台灣醫學*, 21(1), 54-61.
- 林子倫、陳亮宇. (2009). 重返民主的政策科學—審議是政策分析概念意涵與途徑之探討.
- 林子倫、蔡甫昌. (2021). COVID-19 疫情時代的網路公民科技：從口罩地圖談起. *人文與社會科學簡訊*, 22 (4), 73-83.
- 林文源. (2007). 論行動者網絡理論的行動本體論. *科技醫療與社會*(4), 65-108.
- 林文源. (2014). *看不見的行動能力：從行動者網絡到位移理論*. 中央研究院社會學研究所.
- 林文源. (2017). 把疾病帶回來？病患實作中的多元疾病客體化. *台灣社會學*(33).
- 林水波、張世賢. (2006). *公共政策*. 五南圖書出版股份有限公司.
- 林守德. (2021). 揭密 AI 黑箱，開發者投入研究「可解釋的人工智慧」. Retrieved from <https://buzzorange.com/techorange/2021/02/08/explainable-ai-in-business/>
- 林金定、嚴嘉楓、陳美花. (2005). 質性研究方法：訪談模式與實施步驟分析. *身心障礙研究季刊*, 3(2), 122-136.
- 林彥廷、顏筱穎、張乃軒、林宏明、韓仁毓、楊國鑫、陳俊杉、鄭宏達、徐若堯. (2021). 應用 AI 學習技術於坡地崩塌預測分析-以高雄市小林村為例. *土木水利*, 48(2), 48-55.
- 林崇熙譯. (2004). 技術物有政治性嗎? In 傅. 吳嘉苓, 雷祥麟編 (Ed.), *科技渴望社會* (pp. 123-150): 群學出版社.
- 林葳均, 余佩儒, & 陳佳珍. (2021). 研析歐盟人機協作之研發策略. *經濟前瞻* (195), 73-79.
- 邱大昕. (2011). 為什麼馬殺雞？視障按摩歷史的行動網絡分析. *台灣社會研究季*



- 刊(83), 5-36.
- 施能傑. (2006). 文官體系能力與政府競爭力：策略性人力資源管理觀點. *東吳政治學報*, 22, 1-46.
- 柯于璋. (2013). 政府委託研究案代理問題之探討：一個結合賽局理論與代理人理論的研究取向. *行政暨政策學報*(57), 1-35.
- 洪新原、梁定澎、張嘉銘. (2005). 科技接受模式之彙總研究. *資訊管理學報*, 12(4), 211-234.
- 洪榮志、蔡志豪. (2011). 從行動者網絡理論看文創商品的展演：以安平劍獅的在地轉譯為例. *創業管理研究*, 6(4), 105-122.
- 孫本初. (2003). 政府績效管理的新思維. 收錄於*行政管理論文選輯 (第十七輯)*, 考試院銓敘部出版.
- 張文熙、王培得、陳秀宇、林哲豪、鄧宜芳、曾柏勳、平曉鷗. (2021). 運用人工智慧提升稅務服務之探討. *財稅研究*, 50, 93-115.
- 張春蓮. (2013). 加值型營業稅逃漏偵測之研究.
- 張國暉. (2017). 大型科技系統理論之評析：兼論探索移入台灣之系統的幾個構想. *科技醫療與社會*(24), 91-144.
- 張維安. (2001). 文字模式線上訪談的特質及其限制. *資訊社會研究*, 1, 279-297.
- 曹俊漢. (1983). 公共政策之評估及其在決策過程上的限制：運作模式與美國經驗之研究. *美國研究*, 13(1), 39-100.
- 許炳華. (2022). 美國選區劃分之合憲性爭議：以政治問題為核心. *東吳法律學報*, 34(1), 1-39.
- 許淑梅. (2019). 打造不遺漏任何人的世界—數位政府的挑戰與策略. Retrieved from <https://portal.stpi.narl.org.tw/index/article/10516>
- 陳昇璋、溫怡玲. (2019). 人工智慧在臺灣：產業轉型的契機與挑戰. (天下雜誌).



- 陳金貴. (2013). 治理之理論與發展. *公共治理季刊*, 1(1), 25-36.
- 陳柏良. (2021). AI 時代網路政治廣告之揭露義務: 以美國誠信廣告法草案為中心. *臺大法學論叢*, 50(3), 703-787.
- 陳敦源. (2009). 透明之下的課責: 台灣民主治理中官民信任關係的重建基礎. *文官制度季刊*, 1(2), 21-55.
- 陳敦源、朱斌妤、蕭乃沂、黃東益、廖洲棚、曾憲立. (2020). *政府數位轉型: 一本必讀的入門書*: 五南.
- 陳敦源、吳舜文、陳揚中、王光旭. (2019). 從工作要求資源模式檢視公共衛生護理人員職業倦怠因素: 以公共服務動機為調節變數. *台灣公共衛生雜誌*, 38(5), 478-497.
- 陳敦源、徐仁輝. (1999). 從權力授予概念看台灣的行政立法互動關係. *空大行政學報*, 10, 155-186.
- 陳敦源、黃建勳. (2019). 繁文縟節如何影響公共服務動機? 2011 年臺灣文官調查資料之分析. *文官制度*, 11(1), 35-71.
- 陳敦源、廖洲棚、黃心怡、張濱璿、王千文. (2022). *公部門機器演算法應用之制度調適與路徑分析*: 國家發展委員會.
- 陳敦源、劉宜君、蕭乃沂、林昭吟. (2011). 政策利害關係人指認的理論與實務: 以全民健保改革為例. *國家與社會*(10), 1-65.
- 陳敦源、簡鈺肆. (2018). 繁文縟節與組織績效: 以臺灣作為個案觀察. *文官制度*, 10(4), 25-60.
- 曾冠球、廖洲棚、黃心怡. (2019). *新興科技與公部門數位轉型: 個案採擷與模式建構*: 國家發展委員會.
- 黃心怡、陳敦源. (2023). 人工智慧協作下的公共行政研究: 對公部門組織議題的多層次反思. *政治科學論叢*(96), 139-178.



黃心怡、曾冠球、廖洲棚、陳敦源. (2021). 當人工智慧進入政府：公共行政理論對 AI 運用的反思. *文官制度*, 13(2), 91-114.

黃彥鈞. (2019). AI 將消滅大量職缺？唐鳳：輔助人類提升效率而非取代工作.

Retrieved from <https://technews.tw/2019/11/23/audrey-tang-ai-is-assistive-intelligence/>

黃琳、霍秉坤. (2016). 變革能動者在學校持續改進中之角色. *課程與教學*, 19(1), 135-170.

黃麗鈴. (2019). 從社會科學研究典範質量之辯探析混合研究. *臺灣教育評論月刊*, 8(6), 163-177.

楊弘任. (2012). 行動中的川流發電：小水力綠能技術創新的行動者網絡分析. *Taiwanese Sociology*(23), 51-99.

楊弘任. (2017). [養水種電] 的行動者網絡分析：地方政府，光電廠商與在地農漁民. *臺灣人類學刊*, 15(2), 45-96.

葉俊榮. (1999). 行政程序法的代理人觀點. *全國律師*, 3:4, 33-37.

葉俊榮. (2005). 提升政策執行力的挑戰與展望. *研考雙月刊*, 29(2), 3-16.

葉哲良、李育杰. (2022). *臺灣 AI 國力調查*: 國家科學及技術委員會.

賈文字. (2020). X-AI：智慧醫療長遠發展所需的關鍵. Retrieved from https://www.ailli.com.tw/ailli_detail/63.htm

雷祥麟. (2004). 直線進步或交引纏繞. In 吳嘉苓、傅大為、雷祥麟編 (Ed.), *科技渴望社會* (pp. 79-105).

廖元甫，趙偉成，楊智合，& 楊仲捷. (2020). AI 語音辨認技術應用於自動媒體節目字幕產製. *電工通訊季刊*, 1-15.

臺中市政府. (2022). *智慧科技治理空品成果發表會會議手冊*: 臺中市政府環保局.

劉育成. (2020). 人造社會性：從俗民方法學觀點初探人工智能如何作為一種社會



- 現象. *政治與社會哲學評論*(72), 1-55.
- 劉彥伯. (2019). 美國加州日前立法禁止機器人帳戶假裝為真人. *科技法律透析*, 31(4), 10-11.
- 劉瑞麟、王耀鐸、郭豐璋、鄭柏舜、董尚義、劉德添、陳學台、彭振聲. (2022). 臺北市 AI 智慧交控管理. *中華道路季刊*, 61, 23-40.
- 歐崇明、時文中、陳龍編譯. (2011). *人工智慧：現代方法(第三版)*: 全華圖書
- 潘慧玲. (2003). 社會科學研究典範的流變. *教育研究資訊*, 11(1), 111-143.
- 蔡允棟. (2002). 新治理與治理工具的選擇：政策設計的層次分析. *中國行政評論*, 11(2), 47-76.
- 鄭明政. (2020). 從 State v. Loomis 案件看 AI 應用於司法審判上的若干問題. *台日法政研究*(4), 165-178.
- 蕭全政. (1990). *政治與經濟的整合*: 臺北市: 桂冠.
- 蕭全政. (1991). 亞太經濟合作的發展. *政治科學論叢*(3), 1-18.
- 蕭全政. (1992). 法律與政治, 經濟間的辯證關係. *政治科學論叢*(4), 1-20.
- 賴維堯. (1994). 政策分類: 文獻探討與個案創新. *空大行政學報* 2, 59-70.
- 賴曉黎. (2012). 資通科技的工具面向— 從科技決定論談起. *資訊社會研究*(23), 1-35.
- 羅清俊. (2020). *公共政策：現象觀察與實務操作 (第二版)*: 揚智.
- 蘇彩足. (2013). 我國公共治理之挑戰與因應. *公共治理季刊*, 1(1), 52-60.
- 蘇彩足, 莊文忠, & 郭銘峰. (2017). *臺灣公共治理指標調查及公共治理相關議題研究*. 國家發展委員會
- 蘇經天. (2023). *新AI與新人類：學習、認知與生命的進化新路程*: 大塊文化.