・公共行政學報・　第六十五期
〈書評〉
民112年9月　頁119-132
國立政治大學公共行政學系
Doi: 10.30409/JPA.202309_(65).0004

公部門人工智慧治理：從倫理道德觀點出發

王禕梵*

書　名：*Ethical Governance of Artificial Intelligence in the Public Sector*
《公部門人工智慧道德治理》
作　者：Liza Ireni-Saban and Maya Sherman
出版者：Routledge
出版年：2022
頁　數：114
ISBN：9781003106678

壹、前言

在科技日新月異發展的浪潮中，各國政府逐漸開始利用「人工智慧」（Artificial Intelligence, 以下簡稱 AI）處理政府業務以及提供公共服務。人工智慧雖然可以提升政府的效率以及效能，但也同時威脅許多公共價值，例如降低決策過程透明度與課責性、侵害隱私權以及加深種族歧視等（Chen et al., 2023; Zuiderwijk

非送審類文章。

* 王禕梵（Yi-Fan Wang）為美國內布拉斯加大學奧馬哈分校公共行政學系博士候選人。email: molinawang@gmail.com。

et al., 2021），我國亦有相關討論，黃心怡等人（2021）指出政府使用人工智慧的效應並非全然正面，同時也伴隨違背民主價值與倫理道德的風險，此外，李翠萍等人（2022）認為歷史資料的偏差，可能會使人工智慧在決策過程中產生非意圖性歧視，損害弱勢族群權益，據上所述，當政府使用人工智慧處理業務時，必須重視該科技背後可能帶來風險。

在許多人工智慧的風險中，尤以倫理道德議題最受關注。在美國的脈絡下，人工智慧可被運用到許多公共政策中，更帶來許多倫理道德上的威脅，特別是犯罪防治政策中，政府運用以人工智慧為基礎的「預測性警察活動」（Predictive Policing）來降低犯罪事件發生，AI 可透過大數據以及演算法，預測某些地區或是族群的犯罪風險，提早配置警力降低犯罪產生的可能與損失（Bullock et al., 2020），然而，美國過去常因種族歧視，導致警察較容易取締與逮捕少數族裔，致使犯罪資料庫過度呈現少數族裔的犯罪行為，若人工智慧以此資料做成預測結果，亦可能複製過去種族歧視造成的負面效應（Fountain, 2022），此種人工智慧帶來的倫理道德風險，除嚴重侵害人權外，更使種族不平等更加惡化，因此，許多學者開始討論如何制定倫理道德的治理框架與綱要，讓 AI 的應用能夠符合公平正義（Young et al., 2021; Wirtz et al., 2020），同樣地，本文此次評論的專書論著《公部門人工智慧道德治理》也以此為出發點，據以論述如何改善人工智慧的倫理道德治理。

該書由 Liza Ireni-Saban 與 Maya Sherman 合力完成，Liza Ireni-Saban 為以色列「勞德政府學院」（Lauder School of Government）資深講師，主要關注公共政策與服務中的倫理道德層面，並將此概念應用到人工智慧以及災害防治中，另外一位作者 Maya Sherman 則是英國「全球數據公司」（GlobalData PLC ）的資深人工智慧分析師，她主要研究網路安全、人工智慧公平性以及種族平等議題，兩位作者依照各自專業，共同寫作此本論著，探討人工智慧之道德倫理、分類學與應用技術、責任型科技創新倫理架構、多元交織性與倫理道德以及全球化浪潮下應用等議題。

貳、專書重點論述

該書主要目的是以「慎思型道德」（Prudential Ethics）為基礎，並結合「多元交織性」（Intersectionality），發展分析公部門人工智慧應用所面臨的倫理道德問題，同時藉此提供合乎倫理道德的治理框架，進一步提升 AI 的決策品質。該書認

為傳統「義務論」（Deontology）與「效果論」（Consequentialism）對於道德的論述仍多著墨於外部規範與社會準則，尚無法全方位闡明與分析 AI 所帶來的複雜道德考量，因此必須提出更具有彈性、與應用脈絡相關、並能以個案為基礎的道德分析途徑，慎思性道德即可作為另一個途徑來分析人工智慧相關的道德議題與決策，此途徑關注如何整合個人利益、理性選擇與道德規範，亦即如何由自我內部約束出發，在追求個人利益時，亦重視利他性與倫理道德，政府若採用此途徑分析 AI 相關政策與應用，除可關注弱勢族群需求外，更可以思考隱藏在隱藏在「理性」決策過程背後的社會平等議題，例如是否有特定群體蒙受不平等待遇，或甚至被排除在政府服務之外，而此種思考路徑更與多元交織性理論相關，該理論認為不同的個人身分背景組合可能造成不同的歧視結果，換言之，個人可能因為種族、性別、文化與收入等因素，而受到不同程度的歧視與差別對待，造成歧視的因素可涉及多個層次與面向，若能整合慎思型道德與多元交織性用以分析人工智慧相關應用，則可將視角擴展至不同個人因其身份背景組合而受到的不平等對待，進而提出一個多層次的倫理道德架構來分析政府運用人工智慧面臨的挑戰。

該書第一章討論演算法倫理內涵，藉此提供後續分析人工智慧應用議題的基礎。該書作者認為 AI 可運用演算法分析大數據，亦可由此學習如何分析資料，提供人類決策的基礎，甚至在決策過程中取代人類，然而從倫理道德的角度，此種 AI 應用可能產生許多問題，首先，演算結果缺乏合理解釋，人工智慧雖可透過不同的統計模型產出分析結果，但有時變數間的因果關係並不明確，導致政府機關或是學術單位較難詮釋其數據；再者，低透明度與低可解釋性導致分析結果難以應用，當民眾無法理解 AI 的決策過程，對於該科技的信任度會大幅降低，致使政府難以順利推展以人工智慧為決策核心的政策；第三，AI 決策結果可能有所偏見，演算法有時僅反映特定群體的偏好，此外，政治與社會結構亦會造成 AI 系統發展與應用的過程中複製現存的社會偏見，例如種族以及社會經濟造成的刻板印象或甚至是歧見；第四，延續第三點，人工智慧除可能強化現存偏見外，更因為資料結構與特性，可能造成種族、性別或是其他身份組合的歧視；第五，人工智慧可能降低個人自主性，當 AI 能夠替個人篩選資訊並提供決策建議時，人類將無法觸及更多元的資訊，甚至過度依賴科技進行選擇；最後，人工智慧可能侵害個人隱私，AI 雖可整合不同資料庫並分析數據，但卻無法確保分析過程符合個人資料保護規範，導致個人資訊可能被錯誤使用。該書認為透過確立責任歸屬，能夠降低上述道德與倫理的風險，系統開發人員設計人工智慧時應思考不同族群的需求，並提高系統的

可追蹤性與解釋性，藉以接受不同社會團體的監督，提高該科技的倫理與道德保障。

該書於第二章介紹人工智慧的定義以及不同的演算技術。雖然人工智慧並非一個全新的產物，但各個學科以及專業領域對於其概念內涵仍有不同看法，迄今仍無統一的定義，該書作者爬梳許多文獻，整理出以下幾種較為重要的定義，首先，人工智慧是能夠改變人類社會的變革性技術；第二，該科技可重製人類的感知、論證以及行動能力；第三，AI 能夠針對問題提出不同的解決方案，並從當中選擇最佳解。但該書認為上述任一概念均無法完美定義人工智慧，而是需要在不同的情境下，比較機器與人類的相似與相異之處，才能夠較為合理地理解 AI，例如過去對於「智慧」的定義是能夠解決困難的問題，但隨著時代演進，「智慧」的定義更關注如何了解以及應用各項知識，而此定義則涵蓋許多認知功能，例如注意、記憶、語言、感知以及計畫等，這些能力則能夠進一步整合歸納資訊，並從中產生與選擇行動方案，而人工智慧則是要模仿上述人類具備的能力，以演算法為基礎收集以及分析資料。此外，亦有學者認為人工智慧不僅是純然的技術產物，而是需要與社會元素互動後，才能夠給予明確的定義，AI 也如人類一般，即便在自身知識與能力都未完全成熟時，都需要適應環境並生存。

另外，若從實務觀點來定義人工智慧，則更側重於該科技實際上的運作。「美國 2019 年國防授權法」（2019 US National Defense Authorization Act）則對 AI 提出許多定義，第一，人工智慧指任何能夠獨立在難以預測環境中執行任務的系統，此系統不需要過多的人為監督，甚至能夠從經驗與資料庫中學習且提升表現；再者，人工智慧泛指能夠模擬人類的感知、論證、計畫、學習、溝通以及行動能力的系統；第三，該科技能夠如人類行動與思考，包括認知結構以及神經網絡等功能；第四，AI 可涵蓋具備有機器學習能力的系統；第五，人工智慧需能夠以理性解決問題，此系統可應用於電腦軟體、系統、或機器人中，此外，美國軍隊則強調人工智慧在決策過程中分析與詮釋資訊的能力，歐盟則是認為人工智慧能夠分析資訊與環境，並作出理性決策以及採取獨立行動以完成特定任務，但政府部門仍需注意當人工智慧應用於軍需武器時，可能產生的道德爭議。

該書在第二章亦包涵人工智慧技術以及其演進階段。人工智慧可運用不同的「機器學習」（Machine Learning）技術分析資料，機器學習可分為「監督式」（Supervised）、「非監督式」（Unsupervised）以及「深度學習」（Deep Learning），監督式機器學習指演算法由系統開發者事先設計，所有分析皆依照人

類設計的參數進行，而非監督式機器學習則是指人工智慧能夠從資料中自主學習並且改善其表現，深度學習則是利用多層次的網絡來模擬人類大腦的決策過程，是機器學習中複雜程度最高的類型。此外，學者們也針對不同的人工智慧應用提出許多不同的名稱，例如「人類層級人工智慧」（Human-Level AI）、「通用人工智慧」（Artificial General Intelligence）、「狹義人工智慧」（Narrow Artificial Intelligence）以及「超級人工智慧」（Superior Artificial Intelligence）等，人工智慧仍在持續發展，且在部分領域表現已超越人類許多，但是當程式設計者開發與訓練 AI 時，仍可能會將現存的社會偏見帶入系統中，導致該系統的產出重製或是強化刻板印象或是歧視，例如聊天機器人可能會因為讀取過多帶有種族歧視的網路留言，而以此方式回應問題，又如人臉辨識系統也較辨認出特定族群的特色與臉部特徵等，據此，在該科技迅速發展的同時，如何處理相應的偏見或是道德議題，都需要進一步分析。

第三章則是探討責任型科技創新倫理架構。「責任型科技創新倫理架構」（Ethical Responsible Technology Innovation Framework）強調研發創新的程序必須透明且具互動，該程序中的行為者都須以倫理道德為基礎對彼此負責，此種創新模式能夠具有倫理可接受性、永續性以及社會合意性，該架構具備四個構面，第一，在責任型科技創新倫理架構中，「可預期性」（Anticipation）原則強調能夠具體預測系統特性與結果，並且在開發過程中強化公民參與以避免系統偏見；第二，「反思性」（Reflexibility）認為政府、系統開發者與民眾，能夠對於資料、演算法以及決策結果進行批判性思考，以找出可能的道德倫理風險；第三，「多元兼容性」（Inclusion）指出透過審議式民主或是焦點座談等方式，能夠納入多元的公共價值，同時滿足個人權益以及公眾利益；最後，「回應性」（Responsiveness）要求系統開發必須能夠回應利害關係人的需求與期待，並且提供充分的管道與資訊，讓利害關係人能夠充分表達其偏好。

該書作者認為責任型科技創新倫理架構可補足義務論與效果論在分析人工智慧應用時的不足，義務論強調個人需遵守程序上的道德倫理，效果論則是重視個人如何選擇對群體最有利的行動，但此二觀點都過於關注外在環境，忽略個人內部的自我制約與道德意識，因此，作者引入慎思型道德作為發展該責任創新架構的基礎，慎思型道德的發展可追溯至古希臘時代的「美德」（Virtue），該美德可由哲學角度或是實際脈絡進行思考，實際脈絡下的美德價值則被視為務實的智慧，而後被稱之為慎思性，蘇格拉底認為務實智慧是城邦以及社會存活的關鍵，柏拉圖認為平衡

精神、欲望以及理性是道德思考的重要過程，亞里斯多德認為務實智慧是人類生存的重要憑據，道德教育可提昇人民對於道德慎思性的認識與執行（Ireni-Saban & Sherman, 2022, p. 54），此外，道德慎思性更與社會脈絡連動，道德內涵並非放諸四海皆準，而是會受到環境的影響，導致其實質內涵有所不同，休謨（David Hume）更認為道德可帶動人類思考作為或是不作為，在政治行動中，純然的理性無法完全解釋人類行為，必須帶入道德觀點，才能詮釋各種政治行動（Ireni-Saban & Sherman, 2022, p. 55），而伯克（Edmund Burke）則是強調慎思性可形塑政治上的道德倫理，該類道德倫理並非是舉世皆然或是純然抽象，而是具有脈絡特性，他同時也反對將政治視為科學理性主義以及功利主義的產物，功利主義認為所有個人利益均可視為群體利益的一環，但此論點卻無法解釋社會生活複雜性對於個人偏好的影響力（Ireni-Saban & Sherman, 2022, p. 56），該影響較難利用科學上的因果關係分析，自然就無法以功利主義來分析個人利益與群體利益的連結，據此，摩根索（Hans Morganthau）則是強調政治的神秘並非來自於工程師的科學理性，而是來自於政治家的智慧與道德（Ireni-Saban & Sherman, 2022, p. 56），此論點導致在討論倫理道德時，都可能需要暫時排除個人利益，但拜爾（Kurt Baier）認為若要以理性途徑探討倫理道德時，就需要闡述個人利益、理性選擇以及道德原則間的關係（Ireni-Saban & Sherman, 2022, p. 57），慎思型道德可提供另一個兼融上述三者的理論途徑，該理論認為道德上的「利己性」（Ethical Egoism）亦可視為個人利益的一種，雖然個人仍追求利益，但概念已與功利主義視角的詮釋有所不同。霍布斯（Thomas Hobbes）認為倫理道德本身即是一種社會規範，能夠促進社會群體的共同利益，當個人行為違反群體利益與社會道德價值時，就可能會受到群體的責難或處罰，因此當個人在評估是否採取行動時，會考慮其個人利益是否符合社會期待（Ireni-Saban & Sherman, 2022, p. 59），此時，個人已經將個人利益、理性選擇以及倫理道德三者全部納入評估標準中，但卡夫卡（George Kavka）認為即使個人以上述三者為基礎進行決策，但決策標準仍仰賴外部的社會制約，可能發生當行為者有足夠的資訊評估違反道德倫理產生的利益或是風險時，他們依然會破壞該社會制約，據此，他強調倫理道德不僅是外部的社會規範，更是個人的內部制約，個人可由自我的道德約束而採取符合倫理道德的行為（Ireni-Saban & Sherman, 2022, p. 60）。綜合上述論點，該書認為慎思型道德可作為發展責任型科技創新倫理架構的基礎，慎思型道德可讓公共管理者思考，各種數位科技的應用是否能夠保障弱勢群體的權益，呈現他們的需求，甚至是讓各個社會群體都能夠有機會參與人工智慧系

統的發展，使得人工智慧的應用能夠具有道德慎思性，而非隱含可能的社會偏見。綜上所述，作者在第三章探討許多道德倫理觀念以及許多哲學層次討論，用以支撐其所提出的責任型科技創新倫理架構。

第四章則是探討慎思型道德與多元交織性之間的關係，進而討論人工智慧應用的規範性框架。多元交織性的起源可追溯至 1960 年代，當時學界對於第二波女性主義提出批判，認為該主義的論點仍只關注白人或是優勢族群的權益，較少爭取同為女性的其他種族的權益，坎秀（Kimberle Crenshaw）則為第一位提出多元交織性的學者，他認為性別與種族的組合，可能會加深個人在社會的被邊緣化以及被歧視性，該論點強調社會平等為多面向與多層次的議題，當代的多元交織性理論已經視角擴展至移民身份、歷史以及社會階級如何與性別以及種族交互影響，產生不同的社會不平等現象，換言之，該不同背景產生的社會壓迫以及權力結構，都可能讓個人蒙受不同程度的歧視。多元交織性理論提供兼容多層次的分析途徑，以分析社會的複雜性對於不同個人的影響，該理論包括三種理論途徑，第一，「反類別複雜性」（Anticategorical Complexity）挑戰實證主義的化約論，認為社會現象相當複雜，並不能拆解為數個元素來探討其因果關係，應解構當前對於各社會群體的定義；再者，「跨類別複雜性」（Intercategorical Complexity）則是較關注個人身處不同的社會群體所面臨的多身份認同、面向以及權力結構；第三，「類別內複雜性」（Intracategorical Complexity）則是討論當個人的身份認同跨越傳統定義的群體時，其所面臨的處境。慎思型道德雖可整合個人利益、理性選擇以及倫理規範，但仍然缺乏完整的理論分析架構，但多元交織性理論可提供較為清晰的認識論，學者可依上述三種理論途徑，思考具體的方法論，換言之，慎思型道德提供對於人類社會生活道德內涵的本體論層次解釋，而多元交織性作為其認識論，思考社會互動與倫理道德間的關係，並藉此發展多元方法論來分析實存議題。

該書作者在該章節則提供具體的政策分析步驟。作者認為分析公部門 AI 應用可由包含以下階段，首先，探討人工智慧運用的政策領域與可能產生的道德風險；再者，分析應用個案中不同層次的資訊，包括各種質化與量化資料、資料管理與分析者的責任歸屬、當前政策與倫理性規範對於利害關係人的影響、以及各利害關係人在各層面的風險等；第三，提出可能的政策建議與解決方案，以提高人工智慧的倫理道德能力；第四，預評估可能政策方案在各個層面的影響，可特別關注利害關係人對於各種政策的觀點；最後，監督政策執行之結果，並隨時檢討執行成效。該書強調上開分析途徑，可回應慎思型道德，例如探討各群體中的個人意義、以較為

全面且具複雜性角度挑戰當前的不平等現象、以及分析政策是否能夠讓個人實踐其行動自由。此外，該書提出政府可分析人工智慧應用的三個層面，包含人權、集體社會價值以及互惠性，在人權層面，政府需評估人工智慧是否能夠保障基本權利以及個人尊嚴，於集體社會價值原則中，政府可評估 AI 對於各種社會價值的影響，而在互惠性中，政府可評估人工智慧的運用是否對於不同社會族群或是身份組合產生不同影響，是否出現歧視性或是不公平之政策效果。

第五章則是討論全球化時代的人工智慧應用案例。該書作者以跨國以及公私部門的案例討論 AI 對當代公共治理的影響，首先，作者以社會層面角度切入，探討人工智慧在「社會監視型」（Mass Surveillance）政策的應用，人工智慧可協助政府追蹤民眾的足跡與行動，並且可將收集到的資料進行大數據分析，提供政府預測民眾未來的行為，然而各國民眾對於該政策的接受度並不相同，具有文化上的差異，例如委內瑞拉的國民身分證系統可追蹤民眾在不同政府服務的使用行為，該國就有駭客侵入系統，將政治人物的資料刪除，藉此表達對於該政策的不滿，然而，中國的社會信用體系卻可順利執行，該體系會以監視型科技以及人臉辨識系統追蹤民眾在現實社會與社交媒體上的行為，並依照他們的行為給予評分，其中最明顯的案例就是將該科技應用於金融信用系統上，中國政府可透過該科技減少金融詐騙案件，該國民眾對於該系統的接受相當高，特別是社會經濟背景較高的國民，對於該政策的支持程度更高，該國政府也將此技術廣泛應用在對少數族群的社會控制上，例如可利用人工智慧監視新疆維吾爾族的漢化進度等。AI 雖可協助政府收集相關數據，但是，學者也對於該科技可能造成的社會不平等感到憂心，許多貪腐或是貧困程度較高的地區，少數或是弱勢族群面臨到因人工智慧而生的種族歧視問題更為嚴重，即便是在已開發國家，美國企業使用 AI 招募員工時，發現該科技對於女性的評估較為不友善，而金融信用演算法亦對於拉丁族裔或是黑人申請者產生歧視性對待。學者亦憂心該類政策可能在獨裁或是威權國家產生種族壓迫或是迫害，例如緬甸的羅興雅族可能會深受該科技之害，中國政府對於少數族群或是地方社群的控制，亦可能對其造成傷害，即便是在已開發國家，人工智慧的應用亦可能產生該類問題，例如美國國防部曾想利用該科技了解民眾在社交平台的行為，美國移民局也欲利用該科技追蹤移民的行動。上述案例皆闡明人工智慧雖具有整合多種資料的能力，但此技術也提高對於民眾隱私與其他權益傷害的隱憂。

再者，該書亦由民主治理的角度探討人工智慧的影響。人工智慧亦高度挑戰民主價值，人工智慧可能產生操弄選民偏好的問題，社群媒體亦會使用演算法推送訊

息給使用者，但這些訊息多半基於流量與經濟考量，訊息的正確性仍有待商榷，若錯假訊息與政府相關，更可能傷害民眾對於民主制度的信心，人工智慧亦對於透明度與課責性產生負面影響，民眾通常較難理解演算法的運算邏輯，對於當中可能的偏見亦較難察覺，此種特性挑戰民主價值中的透明性，缺乏透明性亦難對該科技應用進行課責，此外，民眾亦較難分辨人工與 AI 產生訊息的差別，有心者可利用人工智慧生成錯假訊息，破壞民眾對於民主制度的信任，而此類訊息的產生與傳播速度相當快，可能改變選民的選舉行為，例如俄羅斯曾利用機器人在社群媒體對於西方民主國家進行攻擊。不過，人工智慧亦可辨別錯假訊息，政府亦可利用該科技偵測惡意產生的訊息與文字，降低假訊息對於民主政體的傷害，但該類應用亦可能侵害言論自由，且人工智慧有時無法分析較為複雜的語言架構以及其背後的文化脈絡，都可能影響其偵測效果。此外，AI 的「深偽技術」（Deepfake）導致許多公共價值受到挑戰。深偽技術指人工智慧可利用聲音與影像操弄人類感知，有心者可利用大數據訓練 AI 產生神經網絡，並且重置人類社會的規則與趨勢，系統開發者亦可利用演算法，讓人工智慧產生更為「可信」的錯假訊息，用以影響政治討論以及民眾信任。據此，人工智慧對於民主治理同時帶來正面與負面效果，其重點在於如何使用該科技改善民主制度以及降低錯假訊息對於制度信任的負面影響。

第三，該書從公共治理的角度分析人工智慧的應用。該書以新冠肺炎為案例，以公共治理視角分析，政府如何在新冠肺炎期間利用人工智慧控制疫情，例如 AI 可分析疫情趨勢，醫療機構亦可利用此科技對於病患進行治療，提高治癒率，但當政府利用人工智慧控管疫情時，亦可能產生侵犯人民隱私權的疑義，另外，因缺乏有關新冠肺炎的歷史數據，人工智慧的演算法尚未成熟，對於該疫情的控制仍未臻完善。作者認為政府必須利用適合的方式對於人工智慧進行治理，例如可透過跨國合作發展負責性 AI 系統，但各國政府亦須尊重文化差異，讓人工智慧運用可符合不同的文化脈絡，不同單位與團體對於該科技的想法亦有不同，公部門在利用該科技帶來的便利時，亦須重視其負面影響，方可運用可行且完善的治理方式，提升人工智慧在公部門的表現。

最後，該書雖未言明如何以責任型科技創新倫理架構分析實存的人工智慧應用，但本文仍試圖以作者第五章提供的個案與討論，回應該倫理架構。在第五章所提及的行為者，主要包括政府、私部門與民眾，但政府為當中最具影響力者，其可將人工智慧應用於各種公共事務中，然而，即便各國政府試圖制定法規來治理人工智慧，但該科技進展相當迅速，外部規範並無法全面提升治理品質，此時，責任型

科技創新倫理架構可彌補該落差，該架構重視如何以內部制約來平衡個人利益以及倫理規範，例如，政府雖可透過人工智慧進行社會控制以提升其利益，但政府亦重視民主價值與倫理規範，此即其內部制約，可影響其進行政策選擇之理性，除了慎思型道德的考量外，多元交織性亦可提供分析途徑，例如該書文中提到弱勢族群或是開發中國家民眾，更可能遭遇人工智慧的負面影響，此時政府就需思考各項人工智慧政策背後的邏輯，以及其對於不同身份組合產生的影響，甚至必須思考該科技是否創造過去未曾發現的負面問題，據此，本文認為該書提出的倫理架構，主要可處理在法律規範外的政府、私部門與人民之互動，憑藉該架構中的慎思型道德與多元交織性理論，可提供各行為者採取行動的準則，提升公共治理與民主制度的品質。

參、分析與評論

該書將道德治理視為公部門人工智慧應用的核心，並且結合慎思型道德、多元交織性以及責任型創新架構等概念，提出許多對於政府使用 AI 的建議，呼應當前美國學界關心「社會平等」（Social Equity）的趨勢。自 1970 年代「明諾布魯克會議」（Minnowbrook Conference）起，美國公共行政學界開始關注與重視社會平等議題，最初應用於代表性官僚研究（Gooden & Portillo, 2011），而後隨著資訊科技的發展，電子化政府與數位治理的研究亦開始重視社會平等的概念，包含「數位落差」（Digital Divide）、「數位識讀」（Digital Literacy）以及「行政排除」（Administrative Exclusion），數位落差強調個人對於科技近用性的落差，較為關注基礎建設以及網路設備等議題，而數位識讀則是指涉個人對於數位科技使用能力，即便民眾具備科技設備，但並不代表能夠順利地使用資訊科技，行政負擔則更與人工智慧的應用密切相關，當政府使用 AI 自動化處理部分業務時，某些族群可能因為系統設定，而被排除在使用自動化服務的客群之外，此時，該族群可能必須付出更多的行政成本，例如人力、金錢與時間等，去申請相關服務或文件，導致社會各族群間產生不平等的情形，該狀況則被稱為行政排除（Peeters & Widlak, 2018）。而該書更進一步探討道德在人工智慧治理的角色，基於政府組織與官僚系統具有一定程度的行政裁量權以及法規修正曠日費時的考量下，道德可能成為政府能夠合理使用 AI 科技的重要依憑，例如在法律未規範的情形下，政府可依照倫理道德等原則來設計人工智慧系統，確保社會中各族群能夠平等地使用政府服務，包

含是否具有管道以及使用技能，此外，該書作者提出的慎思型道德、多元交織性以及責任型創新架構，更能協助政府思考人工智慧應用的演算法與產出以及行政組織規範與慣習對於人民的影響，進而提升該科技在社會平等面向的表現。

該書提出的道德治理架構雖可提供許多指引與前瞻性建議，但仍有幾個層面略顯模糊，作者可進一步闡述。首先，該架構並未明確解釋道德治理如何在同「分析層次」（Level of Analysis）運用，該書作者在文中表示責任型創新架構主要可運用在制度層次的議題，該架構主要如何提升具備倫理價值的技術進步，其主要分析視角多半在宏觀層次，但書中許多建議與討論亦涉及個人與組織層次，此討論內涵似乎隱含該架構在中觀與微觀議題應用的可能性，畢竟慎思型道德可由個人內心出發，進而影響組織與制度運作，若作者能夠以不同分析層次闡述慎思型道德以及多元交織性對於人工智慧道德的影響，可讓論述能夠更為完整，例如多元交織性可能讓民眾使用人工智慧的可能性與便利性受到影響，民眾個人的身份背景可能帶來不同的數位識讀能力，而行政組織內的官僚組成，亦可能使人工智慧的系統設計有所不同，種族組成較為多元的單位，可能會更重視潛在的行政排除現象，而制度層次的規範與慣習，可能會影響政府機關與系統開發者的互動，進而形塑不同的人工智慧系統，因此，以不同分析層次作為基礎來討論道德治理架構的應用，可使該書提出的論點更具說服力與實用性。

再者，該書作者定義許多利害關係人，但卻未明確討論各該行為者在道德治理架構中的角色。在道德治理的論述中，該書認為政府、官僚、系統開發者與民眾為利害關係人，然而，各利害關係人如何能改善人工智慧的道德倫理層面卻未有明確說明，例如若是 AI 做出錯誤決策，是應由政府、官僚或是系統開發者擔負責任，而若是該系統開發有公民參與其中，其責任又應如何分配，都亟須說明，此外，若是政府將責任型道德治理作為目標，如何制定演算法設計的規則，如何讓民眾有意義地參與系統設計，而系統開發者與官僚如何避免本位主義與排除民眾的意見，提升人工智慧的倫理道德性，也須進一步討論，另外，若作者可利用不同分析層次來討論利害關係人角色，可使論述更為清晰。

最後，該書道德治理架構與公共價值間的連結仍須進一步闡述。該書提到部分公共價值，例如人工智慧的透明度、課責性、可解釋性、隱私權與信任度等，但該科技涉及的公共價值不僅如此，是否能夠納入更多價值的討論，另外，作者在書中有提到道德治理的不同操作步驟，而各該公共價值與各步驟的關聯性為何，例如 Chen 等人（2023）以不同的治理層面以及解決方案來討論人工智慧涉及的公共價

值，或許該書作者亦可分階段討論道德治理架構與人工治理價值的連結，會使論述更為清晰。此外，本文認為該書可以納入對於公部門人工智慧的信任以及該科技研發中對民眾的回應性，可提升整理論述品質，首先，民眾對於公部門人工智慧的信任度，可能會影響其使用政府 AI 服務的意願，甚至最終影響人民對於公共治理的信心，該書雖有探討信任度相關議題，但多著墨於人工智慧如何影響民主制度，但在公共行政相關研究中，不乏探討民眾對公部門人工應用信任度的研究（Aoki, 2020; Ingrams et al., 2022; Wang et al., 2023），民眾對該 AI 的信任度可能與其對於政府、開發商以及科技本身的信任有關，三者在倫理道德層面的制度設計與執行，例如政府是否對於人工智慧應用制定相應的法規或是準則、開發商是否能夠確保使用者的知情權與資料所有權、以及演算法本身是否能夠確保減少社會偏見或是產出公平合理的結果等，都可能影響民眾對其的信任度，而此信任度是否又不同的社會身份組合而有所不同，責任型科技創新倫理架構可在此議題中提供何種分析基礎，該書作者都可進一步論述；再者，公部門人工智慧研發與執行時對於民眾的回應性亦是重要議題，如該書所述，人工智慧的應用有時可能會符合特定族群的偏好，而非嘉惠各個社會族群，因此，本文認為在思考該科技的倫理道德時，仍需要考量該科技對於不同群體的回應性，在美國的脈絡下，聯邦政府在進行各項科技開發與建設時，時常會忽略美國原住民部落的需求，導致部落與整體社會的數位落差加劇（Korostelina & Barrett, 2023），因此，在開發與執行相關人工智慧時，就須納入與部落相關行為者的視角，讓 AI 能夠確實回應部落的需求，例如部落居民會利用不同的管道，例如網路、簡訊或是語音，向災害管理單位進行損失申報或是尋求協助，人工智慧的自動化系統設計就需能夠滿足民眾的需求，而這類民眾的需求通常須經「系統共同研發」（System Co-design）才能夠得知（Tsai et al., 2023），而在這類共同研發中，是否能夠確保少數族群的聲音能夠被納入，並且思考其文化背景脈絡可能對於使用偏好與經驗的影響，都是必須思考的道德議題，若該書作者能夠論述其道德架構與回應性的連結，更可提升該書在道德倫理與公共治理議題的討論深度，並可啟發更多後續研究對於人工智慧的倫理道德層面的分析與探討。

參考文獻

李翠萍、張竹宜、李晨綾（2022）。人工智慧在公共政策領域應用的非意圖歧視：系統性文獻綜述。**公共行政學報**，（63），1-49。[Lee, T.-P., Chang, C.-Y.,

& Lee, C.-L. (2022). Unintentional discrimination in application of artificial intelligence to public policies: A systematical article review. *Journal of Public Administration*, (63), 1-49.]

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。[Huang, H., Tseng, K.-C., Liao, Z.-P., & Chen, D.-Y. (2021). When AI joins the government: A reflection on AI application and public administration theory. *Journal of Civil Service*, *13*(2), 91-114.]

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, *37*(4), 101490.

Bullock, J., Young, M. M., & Wang, Y.-F. (2020). Artificial intelligence, bureaucratic form, and discretion in public service. *Information Polity*, *25*(4), 491-506.

Chen, Y.-C., Ahn, M. J., & Wang, Y.-F. (2023). Artificial intelligence and public values: Value impacts and governance in the public sector. *Sustainability*, *15*(6), 4796.

Fountain, J. E. (2022). The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. *Government Information Quarterly*, *39*(2), 101645.

Gooden, S., & Portillo, S. (2011). Advancing social equity in the Minnowbrook tradition. *Journal of Public Administration Research and Theory*, *21*, i61-i76.

Ingrams, A., Kaufmann, W., & Jacobs, D. (2022). In AI we trust? Citizen perceptions of AI in government decision making. *Policy & Internet*, *14*(2), 390-409.

Ireni-Saban, L., & Sherman, M. (2022). *Ethical governance of artificial intelligence in the public sector*. Routledge.

Korostelina, K. V., & Barrett, J. (2023). Bridging the digital divide for Native American tribes: Roadblocks to broadband and community resilience. *Policy & Internet*, *15*(3), 306-326.

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture–The case of the Dutch civil registry's master data management system. *Government Information Quarterly*, *35*(2), 175-183.

Tsai, C.-H., Rayi, P., Kadire, S., Wang, Y.-F., Krafka, S., Zendejas, E., & Chen, Y.-C. (2023). Co-Design Disaster Management Chatbot with Indigenous Communities. In *20th International Conference on Information Systems for Crisis Response and Management (ISCRAM), 2023*, 1-12.

Wang, Y.-F., Chen, Y.-C., & Chien, S.-Y. (2023). Citizens' intention to follow recommendations from a government-supported AI-enabled system. *Public*

Policy and Administration, https://doi.org/10.1177/09520767231176126.

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.

Young, M. M., Himmelreich, J., Bullock, J. B., & Kim, K. C. (2021). Artificial intelligence and administrative evil. *Perspectives on Public Management and Governance*, *4*(3), 244-258.

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*, *38*(3), 101577.