行政暨政策學報　第七十八期
民國113年6月　　第95-126頁
Public Administration & Policy, No. 78
June 2024　　pp. 95-126
https://doi.org/10.29865/PAP.202406_(78).0004

人工智慧風潮下公共服務制度設計策略與法規調適作爲之研究*

張濱璿**、陳敦源***、廖洲棚****、黃心怡*****、王千文******
張瀠心*******、陳郁函********

摘要

背景：政府運用人工智慧（AI）技術的數位轉型過程中，法規制度建立及風險控管措施，都需前瞻的政策規劃，值得全面探究思考。

方法：本文參採制度分析與發展框架（Institutional Analysis and Development framework, IAD）結合管制影響評估（Regulatory Impact Assessment, RIA）理論的內涵，藉由法學比較分析法彙整上開理論之所需要件，蒐集歐盟、OECD 等國際組織及美、英等先進國家，對應用 AI 於公共服務時，所應考量之問題、遭遇困難、制度解決方法與策略等進行彙整，尤其對人權保障所考量議題，以及政策執行機制操作方式深入整理，並進行國內外制度落差比較分析，討論可能的法規調適作爲。

結果：國外於 AI 監理層級單位的專責、領域的多元差異、共通執行流程，

* 本文係國立政治大學數位治理研究中心受國家發展委員會委託「公部門機器演算法應用之制度調適與路徑分析」（計畫編號：NDC-MIS-110-001）之研析報告成果中，關於法規調適部分之內容重新撰寫而成。感謝全體共同作者於研究過程中之參與。並感謝委託單位的支持與學報匿名審查人、編輯委員會對本文提供詳細的寶貴意見與修改建議。然本文文責仍由作者自負。本文係作者意見，不代表委託單位立場。
** 臺北醫學大學醫療暨生物科技法律研究所兼任助理教授、馬偕兒童醫院兼任主治醫師、昶騰法律事務所主持律師（通訊作者），電子郵件：brianchang.mdlaw@gmail.com。
*** 國立政治大學公共行政學系教授，電子郵件：donc@nccu.edu.tw。
**** 國立空中大學公共行政學系副教授，電子郵件：zpliao@mail.nou.edu.tw。
***** 荷蘭萊頓大學政府與全球事務學院公共行政系助理教授、國立臺灣大學政治學系與公共事務研究所兼任副教授，電子郵件：h.i.huang@fgga.leidenuniv.nl。
****** 淡江大學公共行政學系助理教授，電子郵件：cww@mail.tku.edu.tw。
******* 國立中正大學法律學系博士生。
******** 國立政治大學公共行政學系博士生。

收稿日期：民國 111 年 12 月 12 日。接受刊登日期：民國 112 年 7 月 24 日。

都有一定的成熟發展路徑，由初期的政策白皮書，發展至較爲明確的指引，最後成爲具體法規草案，可爲我國借鏡。經由比較，本文以四個原則面向分析：倫理導向、風險回饋、組織監管及循證決策，提出我國未來公共服務導入 AI 之法規調適作爲建議。

研究發現及建議：從倫理導向切入，各國共通性原則皆立基於「以人爲本」之人權保障；在風險層面，需以不同服務風險導向進行不同管制措施；組織層面，建議需分別有負責整合協調及監督治理之組織；決策方面，建議以倫理價值爲基礎之具體規範指引，使 AI 應用之發展保有彈性。

關鍵詞：制度設計、法規調適、人工智慧、公共服務、演算法

壹、前言

新興科技發展越趨蓬勃，人工智慧（以下簡稱 AI）廣泛運用於公部門服務中。在政府數位轉型過程中，法規制度建立及風險控管措施是極大挑戰。AI 過往以執行程式規則進行複雜任務的「規則基礎演算法」（rule-based algorithms）爲基礎，逐漸演進到機器學習演算法（machine learning algorithms）（European Commission, 2021a），訓練機器自動學習而不是「編寫程式」，這需要大量計算能力和大數據資料（big data），多數 AI 也同時使用兩者（European Commission, 2021b）。

目前多數公部門使用的 AI 並不複雜，但正確性與效率受到質疑，對公部門傳統究責制度也產生挑戰，若行政機關內部技術能力不足，將衍生民眾質疑運算結果。爲實現有意義的課責制度，對於實務和技術跨領域思考是有迫切需求（Engstrom et al., 2020）。政府的關鍵作用之一是制定和執行法規，確保 AI 的發展可以支持並且維繫民主的價值觀和制度，因此需建構與演算法相關之監理框架及道德標準，進一步進行法規調適，以建立可能不同於過去規制人類行爲而針對 AI 特性的制度體系。

貳、我國 AI 規範現況與問題

由於科技打破許多傳統人類社會的定義，例如地域、市場、新型態人權等，使得相關規範可能難以在現有規制基礎上進行修改或解釋，而可能需要創造新的遊戲規則。就我國公部門數位轉型過程中，如何解決監理制度與策略問題，需要深入探討。

一、我國 AI 規範概況

數位發展基礎來自資料分析，尤其公共服務將使用許多個人資料。目前臺灣就資料使用雖有「個人資料保護法」（以下簡稱個資法），但缺乏實際的監理機關，而由各目的事業主管機關自行適用及解釋，對於巨量資料使用亦無其他統一法規範。生醫研究可藉由事先取得檢體或資料捐贈者的概括同意，對合法申請設立的「人體生物資料庫」所保存之檢體與資料進行研究使用；但如中央健康保險署委託國家衛生研究院所推動「全民健康保險研究資料庫」之建置，由於後續研究已超出原保費申報與稽核使用目的外之次級使用，迭生爭議。憲法法庭於 2022 年 8 月 12 日就健保資料庫所衍生個人資料使用之憲法爭議作成 111 年憲判字第 13 號判決，對於政府遲未有負責個資法執行之主管機關要求限期改正，可見我國對於資料監理仍有不足。

我國政府將 2017 年宣誓爲「臺灣 AI 元年」，相繼於同年 8 月推出「AI 科研戰

略」，隔年再推動「臺灣 AI 行動計畫」（行政院，2019）；2019 年 9 月科技部（現爲國科會）與其轄下四個 AI 創新研究中心共同發布「AI 科研發展指引」，期以「以人爲本」、「永續發展」及「多元包容」三大核心價值，並提出共榮共利、公平性與非歧視性、自主權與控制權、安全性、個人隱私與數據治理、透明性與可追溯性、可解釋性，以及問責與溝通等八大指引，提醒科研人員於研究發展時應審愼關注（國科會，2019）。然此指引僅針對開發前階段研究，不適用於產品與服務上市或提供之後階段。

政府過去曾針對使用 AI 之公共服務數位轉型，委託制定「數位轉型 AI 育成手冊」（曾冠球等人，2019），作爲政府政策制定或作業程序的準則。至於較具體的 AI 運用規範，目前僅有金融領域「自動化投資理財顧問」也就是「理財機器人」，原由信託暨顧問商業同業公會於 2017 年制定「中華民國證券投資信託暨顧問商業同業公會證券投資顧問事業以自動化工具提供證券投資顧問服務（robo-advisor）作業要點」，其性質原僅爲民間團體針對會員的內部規範，但後因報請主管機關金融監督管理委員會函釋公告，作爲主管機關審查之標準，使其成爲具有外部效力之行政規則，爲我國目前唯一針對單一事業領域 AI 服務有法規範性質之規定，非針對 AI 發展一致性具體立法或準則，可見我國對於 AI 規範的缺乏。

二、我國規制面的問題

政府應用演算法進行數位創新與轉型雖令人期待，但也令人恐懼，因這些演算法仍受到人類價值觀的直接影響（Diakopoulos, 2015），若機器學習演算法受到人類不當指導，將可能造成個人和公共利益的侵害。從公共價值（public values）角度討論演算法，可看出公部門必須面對 AI 帶來各種價值挑戰（Veale et al., 2018），學者 Andrews（2019）就認爲，政府演算法的應用會帶來四種治理風險，包括「選擇誤差」（selection error）、「演算法違法」（algorithmic law-breaking）、「操控與博弈」（manipulation or gaming）與「演算法宣傳」（algorithmic propaganda）的問題，需政府的介入政策並且避免之。

若公共服務擬以更有效能方式貼近民眾需求並完成任務，新興科技的協助便極爲重要。私部門運用 AI 若發生爭議，可透過私法規範解決爭端。然而，公部門推動過程會面臨數位成熟度不一、組織不清楚、爭端解決與責任體制不明確等問題，若機關擬導入演算法於公共服務中，需系統性盤點現今人力、資源配置上應如何進行，並從前期規劃、中期發展、後期解決問題等面向，進行法規制度的調適。

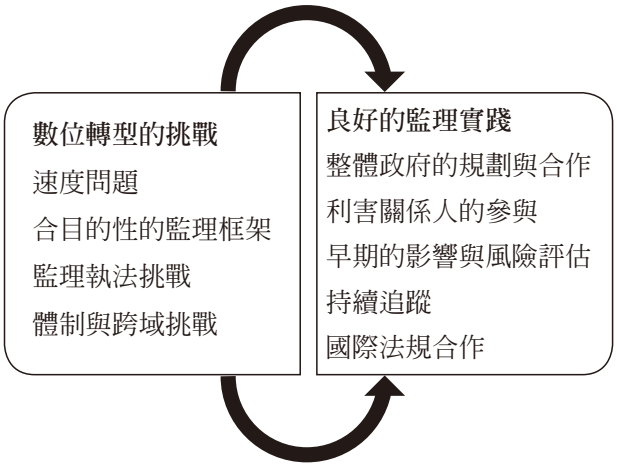
三、兼顧數位轉型與法規調適之思考

數位轉型運作是基於 AI 大量使用的結果；而 AI 運作又建立於背後演算法設計。因此在何種層級或階段進行規範、如何規範，都是挑戰。經濟合作暨發展組織（Organisation for Economic Cooperation and Development, 以下簡稱 OECD）曾提出在數位時代規制建立過程中，可能遇到的四項挑戰：（一）速度問題：法規制度建立或社會發展遠不及科技發展速度；（二）合目的性的監理框架：必須建立新數位秩序的規制方式；（三）監理執法挑戰：數位轉型可能改變傳統責任體系與究責觀念；（四）體制和跨域挑戰：數位化會挑戰傳統政府體制和專業分工，而必須重新建立制度（OECD, 2019a）。

OECD 基於上開挑戰提出了法規調適的策略思維。數位轉型帶來新型態行爲互動模式，可能造成管制錯誤的風險，因此速度問題並非趕緊提出具體法規。面對體制的挑戰，必須以政府服務整體思維整合規劃，也需各國政府對話及國際法規合作解決跨域問題；政府需要納入更多利害關係人參與，並及早進行風險與影響評估（OECD, 2019a）。對於數位轉型挑戰及監理實踐策略之關係呈現如圖 1。

圖 1

數位轉型的挑戰與良好的監理實踐



資料來源：修改自 *Regulatory Effectiveness in the Era of Digitalization*, by OECD, 2019a, https://www.oecd.org/gov/regulatory-policy/Regulatory-effectiveness-in-the-era-of-digitalisation.pdf

法規之本質實爲政策產出之具體成果，因此本文之目的，即基於因應 AI 風潮下進行公共服務的數位轉型的現狀需求，探討面對數位科技快速進步、跨域應用、跨專業發展的特性時，如何整合公共政策之理論與觀點，建立公共服務於 AI 數位轉型下所需具體化之法律規制，進一步提出法規調適策略。

參、理論基礎與研究方法

正如英國牛津大學未來人類研究中心學者 Dafoe（2020）認爲，人類面對「AI 治理」問題（AI governance problem），應思考如何從全球層次設計規範、政策與制度，使 AI 發展可爲人類帶來正面結果。本文即基於目前我國關於 AI 演算法應用之現況漏洞，藉由國外法制經驗之探討，提出我國法規調適之建議方向、著眼要件與法規調適可行方法。本文從政府對 AI 管制角色出發，擷取「管制影響評估」（Regulatory Impact Assessment, 以下簡稱 RIA）的制度性框架與管制品質關聯性的理論，最終建構以諾貝爾經濟學獎 2009 年得主 Elinor Ostrom（1986）之「制度分析與發展框架」（Institutional Analysis and Development framework, 以下簡稱 IAD）爲基底的四層次制度分析，作爲本文論述架構，期能作爲我國 AI 法制建立的參考。

一、解決 AI 治理問題的政府角色理論

制度理論是討論人與其環境限制的學說，Ostrom（1986）所提出對資源配置的 IAD，主要討論在法規、資源、能力有限的狀態下，組織制度如何運作及發展的變化，以瞭解如何更有效地進行組織與管制對象的改革。從憲政、集體與個人選擇的不同層次，分別探討對於環境、社群、個別事件及決策之特徵，以進行制度建構，尋找法規、組織及個人思維的創新方法。IAD 的架構或概念被廣泛應用於資源與數位轉型的制度研究中，透過檢視資源特徵、利害關係人的行爲規則、成本及可能結果、行爲與結果關聯、個人自主權利，以及可能效益等進行整體分析考量（Hess & Ostrom, 2003）。

傳統討論政府在社會上扮演的角色，可從其政治經濟環境中認知。公共政策學者 Weimer 與 Vining（2017）曾提出「共通性政策分析」（generic policy analysis）框架，[1] 認爲社會資源配置藉由市場價格系統，可以得到最有效率的解決，但市場會有失靈的時刻，政府介入（government intervention）便成爲最重要的解方。介入正當性包括效率價值因市場失靈而產生福利淨損失（deadweight loss），或政府因其他包括公平、透明或不確定性等社會價值而需介入市場。因此，若把 AI 發展當作一個科技市場，AI 治理問題也可視爲政府介入市場發展正當性與時機的問題。

從管制觀點切入，政府可能因 AI 市場發展的自然獨占而進行介入，讓知識經濟成果非僅由少數個人或團體獨享，並防止資通安全、勞動市場、或科技創新的傷害

[1] 此政策分析理論架構雖流行，但非所有經濟學與公共政策分析學者都認同以市場失靈爲主要政策介入依據的理論，反對意見參 Zerbe 與 McCurdy（1999）。

（Aghion et al., 2018; Frey & Osborne, 2017）；另有學者認爲，爲保護效率之外的社會價值，政府應介入並遵循利害關係人參與和公開透明的要求，訂定 AI（或機器學習演算法）使用的倫理價值規範（Asaro, 2006; Boddington, 2017; Lin et al., 2011），這兩類政府介入角色即滿足前述「共通性政策分析」的理論方向。但也有學者反對政府介入 AI 發展，認爲將折損企業創新的能量（Brynjolfsson & Mitchell, 2017; Chui et al., 2018; Kitchin, 2017）。當然，更多學者認爲，政府應更「聰明」地管制，一方面應用政府管制角色創造 AI 應用與發展的公共價值環境（如透明、公平、中立性等），確保 AI 發展與民主政治價值不會產生衝突（Danaher et al., 2017），但也不應影響 AI 市場創新的效率價值運作（Lucas, 2017; Zevenbergen & van der Voort, 2016）。

二、研究理論基礎：管制影響評估（RIA）

由 AI 發展文獻脈絡，發現學界對政府介入態度是「必要但須小心」（Stockmann, 2023），必要是因「跨國數位巨獸」[2] 在數位市場產生自然壟斷的市場失靈問題，政府應介入管制；但公共行政領域傾向以公共價值（尤其 AI 科技黑暗面）作爲政府介入理由，從平衡市場效率與公共價值的「聰明管制」角度出發（Chen et al., 2023; de Sousa et al., 2019; Wirtz et al., 2020），亦即政府如何在適當時機與設計下介入數位市場，一方面追求市場效率，另方面追求其他公共價值，是政府設計政策與自我評估的目標。但此設計專業，應由何領域來主導？

事實上，早在 1970 年代起，美國聯邦政府即要求政府部門在進行管制行爲前，應進行 RIA（Kirkpatrick & Parker, 2004; Radaelli, 2004; 張其祿，2008），此爲公共政策分析專業中的「預評估」議題，讓政府決策更加理性與循證，也是公共行政面對 AI 發展的狂潮需要儘快深耕的領域。歐盟 2020 年 12 月就當時審查中的數位市場法（the Digital Market Act）草案，[3] 也提出影響評估報告，強調網路公司因「守門人平臺」（gatekeeper platforms）造成市場缺乏競爭的問題，因此各國政府以營造「有意義的競爭」（effective competition）作爲介入市場的正當性，讓數位市場更有效率；該報告也強調程序正當性，特別提到這些事前管制作爲是經利害關係人認可的公共行動（European Commission, 2020a）。本文雖未擬對目前臺灣 AI 市場的政府特定管制作爲進行影響評估，但希望能夠藉由適當的分析框架，將國外的管制經驗，透過比較

[2] 包括 Apple、Google、Amazon、Meta、Microsoft，全是美國公司。事實上，美國利用全球的市場獨占，當作地緣政治競爭武器的態勢也越來越明顯。

[3] 歐盟數位市場法已於 2022 年 11 月 1 日生效，並於 2023 年 5 月 2 日開始實施。

法學方法將其精華進行整理，以作爲臺灣未來對 AI 產業的管制作爲，進行制度性的先導評估，此亦爲利害關係人在制度環境內互動狀態（institutional interaction）的預評估（de Francesco et al., 2012）。

三、制度性框架分析之研究方法

未來臺灣對 AI 市場管制介入成功與否，取決於管制政策運作品質，以及政府制度性環境中權力運作樣態，是一種「國家運作的藝術」（the art of the state）（Renda, 2006）。學者 Renda（2006）曾應用五個東歐國家在制度化管制影響評估上的異同，討論評估品質與各國制度化差異間的可能關係，正是本文意圖要達成的研究目的。基於我國深度繼受於國外法制的發展軌跡，比較法學方法在國內法學研究廣泛使用，除對外國法進行介紹外，如何發揮國外法制經驗與趨勢，擷取作爲我國法制發展的參考意義，並連結所需解決的問題，是比較法學方法的核心（黃舒芃，2005）。亦即，法學研究領域之比較法學方法兼採社會科學領域之文獻分析法與比較研究法，以擷取各國法制間可採用之共通精神原則與比較差異；本文進一步參採擷取 IAD 與 RIA 理論架構所需之內涵，達到更具有系統性比較分析的目的。各國法體系有所不同，但藉由擷取他國解決問題的因應經驗，發揮比較之功能，從不同法秩序中所存在的共同問題，反思我國需採取的解決策略。

因此，依循類似的分析做法，本文將臺灣未來面對 AI 發展的政府管制作爲，與先進國家進行差異比較，進而提出未來臺灣 AI 管制政策建議。文本選取部分，先以國際組織著手，OECD 的法規調適策略指引，以及歐盟不斷提出 AI 相關規則（Regulation）與指令（Directive）指導會員國法規調適策略，可知國際組織早已預見數位轉型及數位市場經濟活動的需求；先進國家例如美、英，也積極提出各項 AI 應用監理的規範或指引。本文也選擇 Ostrom 最著名的 IAD 理論框架以輔助討論，過去此理論最主要是應用在公共資源（common-pool resources）的政策環境，Ostrom 說明該架構主要是「指認出一般制度環境中都出現之重要的結構變數，但其運作價值在不同制度規範下存在差異」的分析工作，[4] 並同樣可應用於政府對 AI 市場管制的制度環境分析。

在分析內容部分，參酌 Ostrom 等人的理論，將制度分析的框架應用於政府面對 AI 發展之「制度評價」（institutional assessment）（Williamson, 1980）目的

[4] 原文如下：“... an institutional framework should identify the major types of structural variables that are present to some extent in all institutional arrangements, but whose values differ from one type of institutional arrangement to another.”（Ostrom, 2011, p. 9）

上，進行 AI 管制政策「制度設計」（institutional design）的先導分析。另因比較法學進行實證影響評估有執行上之困難，因此藉由文本比較分析，參酌 RIA 的思維與要件，將分析概念範圍由大而小，區分價值（value）、系統（system）、組織（organization）到個人（individual）等層次，除逐層分析國外與我國法規範落差外，並將實務應用期待與需求作爲落差分析目標；再以 IAD 出發，以四個法制設計原則進行探討：最高價值層次的「倫理導向」，從以人爲本的人權保障，追求科技永續發展；系統層次的「風險回饋」，讓 AI 應用能因風險狀態，保有彈性發展與修正的可能；於組織層次的「組織監管」，需要能整合協調甚至有效監管的組織；個人層次的「循證決策」，於政策的制定與執行過程中，保持依循實證之原則。此層次的思維考量制度設計策略下，進一步從法律原則、風險等級、政府組織到個人互動等面向，提出法規調適的建議作爲（如表 1）。

表 1

制度設計策略與法規調適作爲面向

	制度設計策略原則	法規調適作爲
A：價值	倫理導向	法律原則建立與落實設計
B：系統	風險回饋	風險等級評估與回應設計
C：組織	組織監管	政府組織層級與任務設計
D：個人	循證決策	個人互動決策與制度設計

資料來源：作者自行整理。

肆、外國規範文本彙整

面對全球 AI 科技快速發展，各國際組織與先進國家早已看到 AI 可能帶來的問題，法規範也以漸進方式進行調整，從上位原則宣示，漸次進展到明確法律規則。本段將依文本選取之比較對象，從國際組織（OECD、歐盟）的指導方針出發，進一步整理美、英等較具體的政策或規範，作爲我國法規調適策略參考基礎。歐美國家法規範制定之腳步較爲領先，並成爲其他各國參採對象；亞洲國家雖新加坡、日本亦有框架性討論，但散在於各法規範或國家計畫中，截至 2023 年 7 月爲止，尚未見一致性之文件，因此本文以歐美作爲文獻探討之主要客體。

一、國際組織所揭示的AI原則

（一）經濟合作暨發展組織（OECD）

OECD 於 2019 年 5 月通過「AI 原則」（AI Principles）（OECD, 2019b），設定標準以促進使用具有創新性和可信賴性、並尊重人權和民主價值觀的 AI。其中有五項價值基礎原則：應注意包容性增長及持續發展和福祉、以人爲本的價值觀和公平並尊重法治人權和民主多元、透明性和可解釋性、穩定性和安全性、課責性；另建議執政者：應投資 AI 研發、培育 AI 數位生態系統（digital ecosystem）、塑造有利 AI 的政策環境、培養人才並爲勞動市場轉型做準備、促進可信賴 AI 的國際合作。

OECD 並於 2021 年 6 月就成員國導入 AI 原則，以及執行 AI 政策狀況出版報告（OECD, 2021）。報告中將 AI 政策的生命週期，分爲政策設計、執行、監測及國際與利害關係人合作等四個階段。另在治理組織與架構上，各國可採用現有部門、新設部門、成立專家小組、或採外部獨立諮詢和監督機構。法規架構部分，各國的政策制定者面對監理挑戰，從「觀望」到「實驗和學習」，再到徹底禁止等不同強度的監理，並建立軟性指引（guidance）或較硬性的成文法規。

（二）歐盟

1. 法規發展

初期歐盟法律並未針對 AI 制定具體法律框架，僅透過「一般資料保護條例」（General Data Protection Regulation, GDPR）和「執法指令」（Law Enforcement Directive）給予一般性的基本規範，並受「歐洲聯盟基本權利憲章」（the EU Charter of Fundamental Rights）平等原則約束，禁止任何形式的歧視。

近年則快速發展。2016 年起，歐洲議會開始進行機器應用涉及民事法律規則之討論（Civil Law Rules on Robotics）。「歐盟執委會」（European Commission）下設之「AI—高階專家工作小組」（high-level expert group on artificial intelligence, 以下簡稱 AI HLEG），於 2019 年發布「值得信賴的 AI 倫理指引」（Ethics Guidelines for Trustworthy AI，以下簡稱 AI 倫理指引）（European Commission AI HLEG, 2019），適用於所有受 AI 設計、開發、部署、實施、使用所影響之問題，自此「值得信賴的 AI」（trustworthy AI）成爲 AI 發展的中心思想。

其後歐洲議會於 2020 年發布「AI、機器人和相關技術的倫理框架」（Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies, 以下簡稱 AI 倫理框架）（European Parliament, 2020），歐盟執委會則於同年底發布「AI

白皮書」（White Paper on AI）（European Commission, 2020b）提出「卓越與信任的 AI 生態系統」願景；2021 年 4 月再提出「AI 法律調和規則草案」（Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence; Artificial Intelligence Act, 以下簡稱 AI 規則草案）（European Commission, 2021c），旨在平衡 AI 運用帶來的好處與潛在負面風險，此草案更進一步於 2023 年 6 月 14 日經歐洲議會通過爲「歐盟 AI 法案」（AI Act, 以下簡稱 AI 法案）。近期歐洲議會也在 2022 年 9 月提出了「AI 責任指令」草案（Proposal for an Artificial Intelligence Liability Directive）（European Commission, 2022），導入因果關係推定廠商責任的規定，積極保障使用者權利而更易獲得賠償。

2. 重點原則

(1) GDPR 之資料處理原則

在歐盟 GDPR 規範下，針對資料控制者處理個人資料時，應遵循之規則和原則包含：合法性、透明性、公平性、正確性、資料最小化、目的和存儲限制、保密性和課責。

(2) AI 倫理指引（2019）

歐盟執委會 2019 年的「AI 倫理指引」中，專家直接點明「可信賴度」（trustworthiness）是在開發、部署及使用 AI 系統之重要先決條件，並揭示「值得信賴的 AI」三項構成要素：「合法的」（lawful）、「合倫理的」（ethical）及「穩定的」（robust）。在此抽象概念下，值得信賴的基礎必須扣合「歐洲聯盟基本權利憲章」基本價值觀的保護，得到包含尊重自主、避免傷害、公平、可解釋性等四大原則；並提出七大要求：人爲介入與監管，技術穩定與安全，隱私與資料治理，透明性，多元性、非歧視性與公平性，社會與環境友善，課責性。落實方面，則應透過合適的「技術性」或「非技術性」方法，將各要求確實導入演算法系統中，才能達到值得信賴的 AI 目標。

(3) AI 白皮書（2020）

2020 年 AI 白皮書將政策方針主要分成兩區塊：a. 優質 AI 生態系統：仰賴跨部門、公私協力的共同行動；b. 值得信賴的生態系統：規範框架需在尊重公民價值觀與權利的同時，亦得快速、安全發展演算法技術。民眾面對 AI 決策，會因訊息不對稱而擔心權利和安全風險，業者也擔心法規監理制度的不確定性。因此監理架構必須從 AI「應用領域（scope）」出發，釐清「資料」與「演算法」這兩個主要元素的意義與規範，並釐清所應用領域的規範，訂定明確標準。

(4) AI 規則草案（2021）與 AI 法案（2023）

2021 年提出「AI 規則草案」，並於 2023 年通過爲「AI 法案」，是以「風險基

礎導向」（risk-based approach）制定，旨在平衡「AI 運用帶來的好處」與「AI 對個人或社會帶來的潛在負面風險」。依照 AI 系統對基本權利或價值所造成的風險程度，區分三種類型：a. 不可承受之風險（unacceptable risk）：原則上禁止使用；b. 高風險（high risk）：需遵循各種風險管控措施；c. 低度風險（low or minimal risk）：由系統提供者自願建立行爲準則（codes of conduct）。從風險管理、資料管制、技術要求、紀錄溯源與保存、資訊透明、用戶得理解、人爲監督，以及系統準確與穩定性、網路安全等面向，都獲得確保，也以附件（annex）方式列舉可能的「高風險」涉及公益之領域，包括自然人生物識別、重大基礎設施（水、電、瓦斯……）、社會資源分配、刑事執法行爲等。另要求資料控制者遵循資料治理和管理之規定，充分揭露系統運作設計和開發訊息透明性，使資料主體能知悉其個人資料正運用於該 AI，且完整、正確和清晰呈現相關訊息。

(5) AI 責任指令草案（2022）

由於 AI 運算結果是否爲「人」的行爲，一直是法律上討論的問題，也影響損害賠償責任認定。爲減輕因 AI 受損害者之舉證責任，簡化受害者的法律程序，導入「因果關係推定」（presumption of causality），由廠商負擔舉證責任；若爲涉及高風險的 AI，受害者可要求自廠商取得證據的權利，獲得程序性保障。

3. 監理架構概念

由歐盟演進可看出歐盟由執委會與議會共同並進，規範面從倫理指引到政策白皮書，經運作蒐集經驗後，再發展到提出立法程序的規則或法案；執行面則由歐盟執委會下設 AI HLEG，作爲 AI 發展的主導單位，並由各會員國進行執行。

二、個別國家規範發展

（一）美國

1. 法規發展

歐巴馬政府在 2014 年至 2016 年，由科技政策辦公室（Office of Science and Technology Policy）每年針對大數據與演算法發布相關報告。2016 年報告中強調「大數據是客觀的」這個假說前提是錯誤的，並提醒勿因過於依賴演算法，而忽略其歧視風險，有兩大原因：(1) 若演算法資料選取偏差，本身就是導致歧視的原因；(2) 若演算法設計不佳、限縮用戶自主、系統錯誤歧視的假設、資料代表性不足，最終產品可能會像「黑盒子」，技術運作過程會被操作者視爲機密或專有而無從檢視（Executive Office of the President USA, 2016）。

在 2019 年、2020 年則分別由川普總統頒布第 13859 號、第 13960 號行政命

令。2019 年發布以「維持美國在 AI 方面之領導地位」爲題的行政命令第 13859 號（Executive Order on Maintaining American Leadership in Artificial Intelligence, 以下簡稱 13859 命令）（The White House, 2019），稱爲美國 AI 倡議（the American Artificial Intelligence Initiative），也是國家 AI 政策；2020 年的行政命令第 13960 號（Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 以下簡稱 13960 命令）（The White House, 2020）也以「可信賴的 AI」爲題，提出適用於國安以外領域的一般性準則，以確保 AI 用途符合國家價值觀及公衆利益，進行跨部門合作、普及 AI 技術與知識、銜接現有法治框架、培育下一代 AI 人才。2020 年 1 月白宮科技政策辦公室再依據 13859 命令，與多個國家級委員會共同以備忘錄形式公告「AI 應用規範指引」（Guidance for Regulation of Artificial Intelligence Applications）（The White House's Office of Science and Technology Policy, 2020），提出十點政府機構規範私部門對 AI 技術應用的管理原則，供相關聯邦機構參考。

在立法部分，2019 年底衆議院通過的「AI 政府法案」（Artificial Intelligence Government Act）（116th Congress USA, 2019），要求聯邦政府各部門盡可能使用 AI，並設立能建議和促進聯邦政府開發 AI 創新用途的單位，但此法案未獲參議院通過。2020 年 3 月 12 日衆議院又提出「國家 AI 倡議法案」（National Artificial Intelligence Initiative Act of 2020）（116th Congress USA, 2020），是川普時代兩項行政命令後第一個具體化的 AI 法案，促進 AI 研究和機關間合作，並制定 AI 實踐標準。另由政府官員與來自學術界、政府和業界的 12 名成員，於 2021 年組成國家 AI 研究資源工作組（National Artificial Intelligence Research Resource Task Force, 以下簡稱 NAIRRTF），進行政策制定與規範研議之單位。

2. 重點原則

2019 年 13859 命令中提出五大原則及五個關鍵領域：研發、釋放 AI 資源、建立 AI 治理標準、培養 AI 勞動力，以及國際協作與保護，作爲美國維持其在 AI 研發與部署過程中所涵蓋的科學、技術及經濟等面向領導地位的重要指引。

2020 年 13960 命令中，提出在政府設計、開發、獲取和使用 AI 時，各機構應遵守的原則（The White House, 2020）：尊重國家對隱私、人權與自由的價值；利益須大於風險的目的和表現導向（purposeful and performance-driven）；準確、可靠、有效；安全、可靠且有彈性；可理解的；負責任和可追溯；定期監測；透明性；可課責的。

2020 年備忘錄形式的「AI 應用規範指引」，提出了十點一般性或特定領域之法規或非法規監理原則，並呼籲美國應領導國際社會制定 AI 技術標準。應注意的原則

包括：公眾信任、公眾參與、科學完整性和訊息品質、風險評估和管理、收益與成本等資源分配、使用彈性、公平與不歧視、揭露與透明度、資料安全、機構間協調等。這份文件可窺見促進創新的思維，對公部門導入演算法治理時具有參考價值（The White House's Office of Science and Technology Policy, 2020）。

3. 監理架構

美國一直在尋求開發 AI 的戰略和政策，對於保護公眾免受 AI 技術潛在風險、與鼓勵積極創新和競爭力間取得平衡。基於風險分級且顧及成本效益的 AI 監理方法，並在可能情況下優先考慮寬鬆非監理方式的指引，也要求各部門盤點 AI 使用的案例清單，作爲後續政策制定的基礎。

另爲 AI 政策執行和協調而成立 AI 專責委員會，除促進 AI 的美國優勢外，對私部門發展 AI 的法規監理原則，也確立必須兼顧促進 AI 創新應用、保障人民自由隱私、蒐集公眾意見增加信任、制定技術標準，以支持可靠、穩健與值得信賴的 AI 技術系統（The White House, 2019）。

（二）英國

1. 法規發展

英國政府科學辦公室（Government Office for Science, UK）於 2016 年提出一份關於演算法的報告，認爲政府應透過政策，提供 AI 相關產業發展及推展新技術的有利環境，對產品安全性、隱私權保護和從業人員進行倫理教育並制定完善的管理制度，以提高民眾信任、保護民眾隱私、降低科技風險。當政府將 AI 應用在行政管理和服務時，也須確定技術使用是否善盡相關義務，避免危害民眾權益，具體措施包括嚴密的個資保護措施、執法人員的科技倫理教育、部分技術內容公開、降低演算法偏差與歧視，以及明確的課責制度（GO-Science UK, 2016）。

2017 年 11 月英國商業、能源及產業策略部（Department for Business, Energy and Industrial Strategy, BEIS）發布「產業策略白皮書」（Industrial Strategy White Paper）（BEIS UK, 2017），將 AI 列爲未來產業發展的挑戰之一，並於 2018 年由英國商業、能源及產業策略部以及數位、文化、媒體暨體育部（Department for Digital, Culture, Media & Sport, DCMS）共同發布「AI 產業協議」（AI Sector Deal）（BEIS & DCMS UK, 2019），同時成立三個組織：AI 委員會（AI Council）、AI 辦公室（Office for AI）、數據倫理與創新中心（Centre for Data Ethics and Innovation），以尋求企業、科研及政府的三方合作；官方指引文件則分別發布「公部門 AI 應用指南」（A Guide to Using Artificial Intelligence in the Public Sector）（Office for Artificial

Intelligence UK, 2020）與「AI 路線圖」（AI Roadmap）（AI Council UK, 2021）。

2. 重點原則

英國 AI 委員會於 2020 年「公部門 AI 應用指南」中，提及 AI 應用應考慮 AI 所涉之倫理與安全，例如資料品質、公平性、課責性、隱私、解釋性與透明性，以及系統建立與維護成本。2021 年 1 月再提出獨立報告—「AI 路線圖」，回應業界、學界與社會關注，將發展重點分成四大項：研究、開發與新創；技能和多樣性；資料、基礎設施和公共信任；全國跨部門採用。

3. 監理架構

英國資訊專員辦公室（Information Commissioner's Office, 以下簡稱 ICO）於 2019 年 3 月提出「AI 稽核架構」（AI Auditing Framework），並以「治理與課責」與「AI 特定風險領域」爲架構的核心要素。前者指機構有責任採取措施以符合資料保護要求；後者則針對部分 AI 特定領域可能出現的潛在風險，以及爲了管理這些風險的適當措施（ICO UK, 2019）。由於 AI 應用會增加資料保護風險，且對資料提供者也可能帶來損害，因此使用資料者須重新評估現有的治理機制與風險管控是否有效。此架構深受歐盟 GDPR 影響，尤其是重要的課責性原則，ICO 並強調此框架是爲平衡與資料保護有關之公共利益與權利。

伍、法規範之分析比較

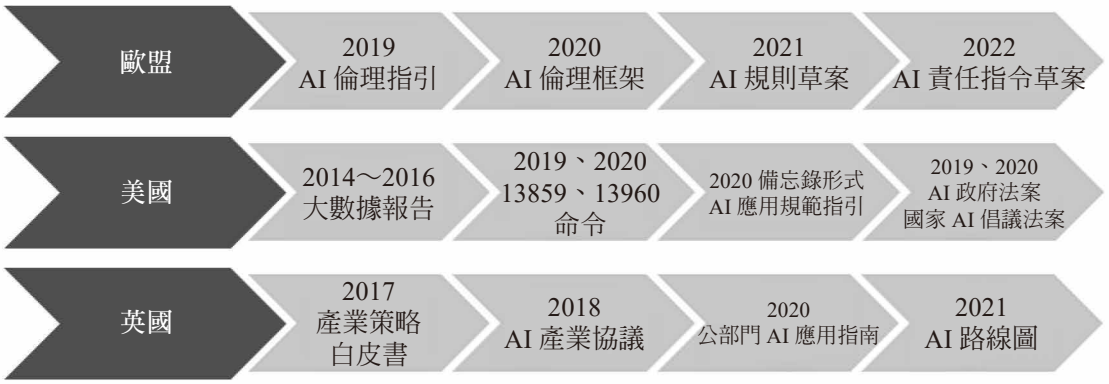
如前所述，本文透過比較法學方法，分析各國法規之特徵及與我國法規制度的落差，範圍由大而小區分價值、系統、組織到個人；各層次再依倫理價值導向、風險等級區分、政府組織監管、個人循證決策等四個制度面向進行分析探討。依循此層次順序的思維考量制度設計策略原則，進行法律原則建立與落實、風險等級評估與回應、政府組織層級與任務、個人互動決策與資訊等法規調適的策略設計，提出具體法規調適的建議作爲。

一、各國制度建立路徑的共同點

從以上蒐集各國對於 AI 的規範，發現各國在發展 AI 的過程中，均遇到 AI 角色爲何及適用領域之問題，也影響演算法如何規範的思考。AI 究竟是獨立主體、或是僅以處理較爲機械性、重複性的事務輔助人類決定，其角色會影響 AI 容錯與課責的思考。各國資料也顯示，目前多尚未有單一針對演算法應用管理法規，而是針對應用演算法之 AI 模式，逐步提出規範強度漸增的指引與原則。依本文所蒐集之資料，政

策規範發展較爲明確者有組織層級的歐盟，以及國家層級的美國、英國，至 2023 年 7 月之演進路徑整理如圖 2。

圖 2
歐、美、英 AI 規範演進比較



資料來源：作者自行繪製。

至於目前已明文的法規範部分，所重仍在於個人資料保護之法規，歐盟 GDPR 是目前最典型也最嚴格的規範，除了「當事人自主原則」影響各國法制思維，AI 政策還需考量以巨量資料進行機器學習演算分析所帶來的公共利益。因此從政策形成與制定而言，對演算法充滿高度不確定風險，外國經驗都會對 AI 在倫理（ethic）、法律（legal）、社會（social）等三個面向上（所謂的 ELSI）引發之相關問題或疑慮進行評估，再依風險程度決定下一步治理策略。

二、制度比較分析與發展

本文將國外相關規範文本內容，分別從價值、系統、組織、個人等四個層次進行分析。在倫理價值層面，最具體由歐盟所提出值得信賴的 AI 三要素：合法的、合倫理的、穩定的，以及四大原則：尊重自主、避免傷害、公平、可解釋性，都與 OECD 的倫理原則、美國 13960 命令及英國公部門 AI 應用指南所指出的價值原則相呼應，並進一步強調透明性、課責性等原則，以及隱私權保障的重要性。尤其個人資料的使用，目前國際的觀念似有從個人化權利的資料保護（data protection），逐漸過渡到群體式公益權利的資料治理（data governance），資料使用所帶來的利益，可能使私益保護的絕對自主部分退讓，因而需重塑可能的權利保障方式。

在系統層面，制度設計需自風險導向觀點設計不同管制強度之措施。尤其涉及基

本人權包含隱私、歧視等，須基於以人爲本的思維基礎，考量人民的利益、福祉與安全。例如歐盟 AI 規則草案列舉高風險等級領域，主要涉及具有基本權利可能侵害或造成歧視的特徵。英國 ICO 提出之 AI 稽核架構，也深受歐盟 GDPR 的影響，對 AI 自動決策涉及公平性、資料安全、公共利益、個人權利保障與行使等風險內容，作爲其核心治理的判斷要素。亦即，各國均認爲，監管措施應與風險評估結果一致，並以跨機構及跨技術領域之風險管理爲基礎。

在組織層面，OECD 整理各國政府組織運作方式，包含採現有部門運作、新設獨立部門、成立專家諮詢小組、或設立以提供意見爲主的獨立諮詢機構等方式。也因資料治理的思維，透過設立具有監理功能的資料倫理委員會，充分納入倫理考量以建立規範。個案部分，多採取綜合方式，除由原有科技、資料治理的單位繼續進行相關政策擬定與監理外，諮詢部分多以新設委員會或成立專案工作小組機構（辦公室）爲主。美國以採取設立 AI 專責委員會及成立 NAIRRTF，進行諮詢、監理與研究；英國則延續既有單位（即 ICO），並新成立 AI 委員會與 AI 辦公室、資料倫理與創新中心等單位進行運作。

在個人層面，著重循證決策，依實證基礎建立規範制度，以保障個人權利與互動的機制。歐盟 GDPR 除當事人自主原則外，也涉及對 AI 演算法之透明性要求，不分公私部門的資料控制者，均須主動向當事人揭露其個人資料是如何被處理及運用，因所謂 AI 透明度之精神，即在於該如何使資料主體知悉其個人資料非以傳統人爲而是機器學習方式所應用。其他可能的衡平方式，包括給予參與者回饋機制及動態同意的機會，或如英國衛生和社會照護法（the Health and Social Care Act 2012）賦予退出權（opt-out）的設計。

另特別就個人請求解釋權（right to explanation）部分，GDPR 未明文指出資料主體是否有權請求資料控制者，就如何做成個別自動化決策解釋其理由。歐盟會員國多認爲 GDPR 具有解釋權規定，無非是依據 GDPR 前言第 71 段提及，資料控制者應對資料主體提供保護機制（safeguard），包括要求資料控制者解釋如何透過演算法做成決策，此爲 GDPR 唯一明文提及「解釋權」之處；而歐洲議會 2020 年出版之「GDPR 對於 AI 之影響」（The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence）（STOA European Parliament, 2020）官方研究報告整理正反不同意見後，似也傾向支持解釋權的立場，但也擔心透明性問題可能引發對 AI 決策的不信任。

三、與我國之落差分析

由上可知，各國就 AI 發展之管制，早已發展多年，且循序漸進依研究報告、倫理作業指引、法律規範之步驟，已逐步提升至法規範管制強度；並依據倫理價值與政策原則、風險評估分級、組織規劃定位、規制方式制度，輔以各國既有監理機關與個人資料保護機制，循序漸進建構 AI 監理架構。

資料治理爲 AI 發展的初期步驟，我國已進入起步階段。現階段尚缺資料保護之主管機關，但行政院已著手籌備獨立機關「個人資料保護委員會」，可說已踏出第一步。而當公部門運用 AI 做行政決定時，因可能產生類似行政處分之效果，因此應透過要求資料控制者之主動揭露包含資料選擇在內的資訊，給予受影響之資料主體救濟途徑。對此歐盟即要求需建立機制，於前階段檢視資料蒐集是否帶有偏見，進而檢視設計所編寫之演算法、整體過程與參數。國科會亦於 2023 年 8 月公告「行政院及所屬機關（構）使用生成式 AI 參考指引」，規定製作機密文書禁止使用生成式 AI，也規定即便使用生成式 AI，最終判斷之決定仍須由業務承辦人進行，以確保公務機關之責任歸屬。

至於針對 AI 科技之具體規範，過去於我國較爲缺乏。前揭生成式 AI 參考指引，規範公部門使用既有之生成式 AI，並非對於 AI 演算法設計進行要求或監理規範；對於 AI 之開發，除國科會發布「AI 科研發展指引」針對科研階段給予一般性指引，但就應用領域只有民間同業團體針對「理財機器人」自行訂定內規性質的自動化投資顧問作業要點較爲具體，其內容與美國官方金融監管局（FINRA）數位投資工具研究報告建議之監管方式十分相近（王偉霖，2019）。該作業要點針對演算法監管有獨立之條文規範，強調演算法是自動化投資顧問服務系統之核心，反映業者對市場分析與研究之邏輯，其設計正確與否，影響運算結果，攸關客戶權益。故業者對系統所用之演算法，應進行有效監督與管理。此充分說明 AI 應用服務監管的核心在於演算法，因此該要點特別要求業者內部，對演算法應設立包括期初與定期審核的監理機制，演算法本身也是監督及告知的主要內容。但此要點屬民間團體內規，雖透過金管會公告而有法規範性質雛型，但並非具體立法或通案準則，故執行層面恐仍缺乏規制效力。

由此可見，我國無論從政策價值提出、系統風險分析、政府組織規劃、個人保障制度等層面，仍有努力空間，亦爲本文擬嘗試提出策略的目的。

陸、法規調適策略之提出

參照 Ostrom 提出應用 IAD 架構進行制度分析與設計步驟，決定分析標的後，進

行現有領域中的行爲分析，經整合分析行爲交互作用模式後，提出分析的結果及建議（Polski & Ostrom, 2015）。因此藉由前揭分析比較的結果，就法規調適先提出思維層次，再進一步提出具體策略。

一、法規調適基礎價值思維

我國對 AI 規範體系尚在極爲初始階段，若依國外經驗，近程工作應先確立共同重點價值原則，作爲制度與政策規劃的基石，才能進一步指定或建立專責機關組織，訂定相關規範，建立政策方向。OECD 明確指出 AI 發展應具備以人爲本的價值觀，歐盟也強調尊重當事人自主爲基本出發點；透過自主原則的確立，美國進一步對於受影響的利害關係人就 AI 演算法應給予解釋使其理解如何受到影響，各國並強調應用結果的公平性與透明性，才能進行後續的課責，達到可信賴 AI 的目標。

但各國提出的各項並列之基礎原則，似仍有執行的先後順序。本文依據 AI 開發與應用之順序，思考在執行上可分爲三個階段層次（如圖 3）：[5]

（一）第一層爲告知程序與揭露義務的完備：屬一般個人資料保護的規範方式。由歐盟規範發展軌跡可知，資料使用爲 AI 開發與應用的初始步驟，當事人自主又爲資料使用的核心原則，踐行明確告知才能使當事人自主得以發揮。包括告知範圍、告知程序、揭露方式、揭露程度等事項都需有細節性界定。

（二）第二層爲可解釋性的考量：演算法如何使用或篩選各項資料及參數？演算法與機器學習使用資料產出結果之理由爲何？都可能影響機器學習判斷的結果與結論。因此在開發階段就應思考如何給予充分解釋，才能使受影響者或使用者對 AI 產生信任。

（三）第三層爲公平性或透明度之判斷：參數及過程均充分揭露與解釋後，何人有能力判斷公平性或透明度足夠？判斷之標準爲何？進而影響演算法與 AI 應如何具體管制或監督、後續救濟可能性與課責性等問題，基此建立整體監管的組織與法規架構。

---

[5] 感謝編輯委員提出本圖的順序似與歐美不同。本文參考包括 OECD（2019）AI 原則、歐盟（2019）AI 倫理指引、與美國（2020）13960 命令等文件所提出之倫理與價值原則。然而，歐美所提出之原則並無先後之分，作者則認爲於政策及法規執行上，仍有可行先後順序。因此，圖 3 爲作者認爲就臺灣法制尚未完備之現況，較佳之思維順序。

圖 3
法規調適思維層次

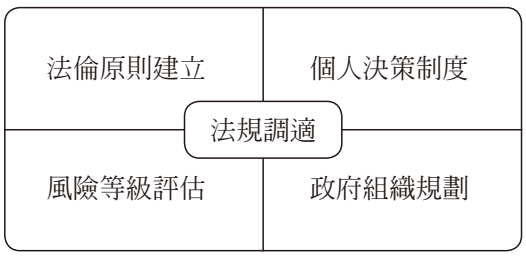
資料來源：作者自行繪製。

以上三個層次，以我國法規現況而言，目前還在努力完備第一層的告知與資料保護規範與監理。私部門運用演算法涉及公司內部商業或營業秘密等問題，歐盟認爲除非有相關法令，未必均能要求公開；但公部門受政府資訊公開之要求，問題應在於執行面能否確認已充分揭露。因此，當公共服務欲導入 AI 時，應建立透明的檢驗制度，對民眾進行揭露與告知，進一步訂定驗證演算法安全性與可靠性之標準，最後考量政府是否有能力執行對演算法透明監理並界定責任。AI 成熟奠基於前階段大量資料及參數之累積，從價值原則確立出發，建置完整之資料治理機制，才有可能完善後續衍生問題並課責；至於可解釋性與公平、透明，則可能在實際執行上有所重疊。

二、法規調適策略

依本文所規劃之制度設計策略與法規調適作爲之層次（表 1），並依倫理導向（價值）、風險回饋（系統）、組織監管（組織）及循證決策（個人）等 IAD 策略原則，分別以倫理法律原則建立與落實、風險等級評估與回應、政府組織層級與任務、個人互動決策與制度四個面向進行設計，作爲法規調適作爲之建議（如圖 4）：

圖 4
法規調適作爲分析面向



資料來源：作者自行繪製。

（一）法律倫理價值與政策原則

事實上 AI 學習過程源於人爲，演算法編寫將嚴重影響 AI 產出之結果。也因此各國際組織及各國均將倫理價值列爲首要事項，以追求可信賴的 AI 爲目標，以人爲中心、以人爲本、以人優先作爲思考核心。AI 系統必須合法、合倫理、穩定、正確且安全，並重視包括尊重自主、公平性、可解釋性、透明性、課責性、重視隱私等原則。倫理原則確立後，進一步於政策上採取有效的施行手段，達到介入監管、公衆參與，跨部門合作的目的。

抑有進者，公部門基於資訊公開的要求，更強調演算法及 AI 模型之可解釋性、透明性、可課責性等原則。私部門應用 AI 時，尚有主管機關得對私部門內控機制進行監督，然而公權力對個人與社會所產生之風險疑慮更甚於私部門。公務人員經民選或考試任用，具民主正當性及淘汰機制，但若應用 AI 系統產生自動化的行政決定，相對缺乏民主基礎。則在目前 AI 科技發展尚無法進行人類「價值判斷」之限制下，似難期待 AI 有能力分辨其決定是否侵害人民權利，亦難以課責或使人民提起救濟。據此，公部門使用演算法及 AI 模型是否能獲信賴並課責，均將成爲挑戰。

（二）風險評估分級

就事務性質本身，需不同強度方式監管，可依 AI 應用目的領域產生的風險強度爲據。無論歐盟「AI 法案」具體列舉高風險事務，或英國「AI 稽核架構」，都指出特定風險領域會涉及的問題，其風險衡量與倫理原則息息相關，包括偏見和歧視造成公平性問題、可解釋性、資料使用正確性、資料主體權利、決策過程審查衡平等事項，故高風險事項之監理強度便需嚴格。

以美國幫助法官評估被告再犯風險作爲量刑準據的 AI 系統 COMPAS system 爲例，其依被告問答、年齡、過往犯罪紀錄與類型等各項數據，推估被告再犯率。事實上 COMPAS 所仰賴的回溯性判斷做成前瞻性預測，易產生偏見（bias），白人獲得保釋機率往往較高。因而輿論開始檢討，這些數據意義爲何、如何判讀、預測是否合理等問題（Larson et al., 2016）。演算法模型若缺乏彈性，隨著時間及環境背景因素可能會失效，因此 AI 產出之結果仍偏向參考性質；政策改變也會使服務工作流程改變，造成原先 AI 模型產生偏離，也因此 AI 治理必須就演算法模型滾動檢視修正，以免影響風險評估。

（三）組織規劃定位

專責政府機構組織上，根據 OECD 建議，應思考區分協調組織及監理組織。在整合協調機關部分，OECD 建議以現有或新設獨立之部門進行；至於治理監督的機關，除以現有或新設獨立部門外，尚可以獨立之專家小組或新設獨立諮詢機構提供建議。除可提出政策方向之功能外，亦可進行案件之審查及個案之監理。

從美國、英國的經驗觀之，似多採取綜合方式，除由原有科技、資料治理的單位繼續進行相關政策擬定與監理外；諮詢部分多以新設委員會或成立專案工作小組（辦公室）爲主。美國除有 AI 專責的諮詢委員會外，並整合現有白宮幕僚組織與外部專家成爲 NAIRRTF；英國除原有的科學辦公室、ICO 外，另設立 AI 委員會、AI 辦公室等單位。

（四）個人互動決策制度與適用領域

基於人權保障的具體運作目標，各國多以概括到具體、拘束力由弱到強之步驟演進，此與一般法律體系由倫理原則走向具體法律之路徑類似。多先由政府進行 AI 相關研究的研究報告，在實證研究之基礎上，提出白皮書進行政策方向宣示，再進一步以相關指引的方式，提出對於 AI 服務審查應考量之重點；待有所共識後，再提出法規草案。

除歐盟已通過 AI 法案外，目前相關倫理指引文件仍爲各國進行 AI 應用的指導原則與依據。至於 AI 適用之服務領域，經區分風險類型，並觀察現階段 AI 科技發展狀況，包含單純機械性事務領域、高資料密集性領域、受高度監管（例如金融業）之特許行業三類型，爲適合優先應用 AI 模型之領域。

至於 AI 應用結果會產生歧視或偏離之情形，在於 AI 透過演算法進行機器學習之過程中，接收各種數據及資料，因此應由專業人員進行定期維修、更新及報告，以合於法規調適之透明度與可解釋性要求。惟應特別注意的是，數據平臺的資料已非原始資料，可能帶有評價性，仍應審愼避免 AI 所導出之結果產生歧視。

由於 AI 應用漸廣，可能涉及人權侵害，因此對於 AI 監管朝向法律制定仍有其必要。參照外國規範制度演進的經驗，在法律制定以前，必須依循實證之基礎訂定政策方向與保障之人權價值，以操作指引或準則方式進行管制，並賦予資料使用者等一定程度之揭露義務，資料控制者亦須善盡告知義務，並給予當事人拒絕自動化決策之權利。意即當公部門欲利用 AI 完成或執行一定行爲時，需充分揭露並提供救濟可能。若與當事人所認定事實有所違背時，亦得挑戰預定之參數，以促使自動化決策或演算法的透明。

柒、結論

基於本文希望整合公共政策之理論與觀點，建立公共服務制度於 AI 數位轉型下所需法規調適策略之研究目的，本文總整對於外國制度經驗，以 RIA 的循證思考出發，使用比較法學之研究方法，擷取 IAD 與 RIA 理論所需之架構，從價值、系統、組織及個人的四個層次內涵，考量「倫理導向」、「風險回饋」、「組織監管」及「循證決策」等策略原則進行各國制度設計策略原則之分析後，再對應「倫理價值與政策原則」、「風險評估分級」、「組織規劃定位」、「個人互動決策與制度適用」四個層面釐清我國現況，並進而提出未來 AI 法規調適作爲之可行方向。

就制度設計策略原則之分析而言，「倫理導向」面向，各國大多是提出指引方針與處理原則，除歐盟近期的 AI 法案外，法律規範上仍較著重於個人資料的保護，立基於「以人爲本」的人權保障思考，再視需求增加細緻的原則，成爲人機協作下的互惠、互信、互賴的運作模式。

從「風險回饋」層面觀之，公共利益的最大化在未來需透過更多巨量資料的演算分析，因此，藉由風險導向觀點進行不同管制強度之措施，基於倫理價值基礎保障基本人權不受侵害，提高安全性和保護措施、更新演算法的監管方法、評估可接受的風險，進一步規劃落差分析方式取得循證基礎。

「政府組織」則爲落實原則的實踐者。OECD 提出分別的整合協調組織與治理監督的組織建議，現有國外組織多採取「跨部門組織」、「委員會」的型態，其重點爲進行整合與協調，並有權力可使政府各部門爲該運作提供協助，也須仰賴外部專家的參與與諮詢，使其可行。

「循證決策」面向爲倫理價值原則的具體落實。政府除思考未來政策制定、執行 AI 的目標及專責機關的設立與組成之外，也能有一套具體規範架構依循。各國多先採取軟性指引，包括策略擬定、應用指南、白皮書、備忘錄等，使 AI 應用過程中，仍保有彈性發展並隨時修正。

各國經驗可知在倫理原則基礎上，法規均以維護人權保障追求可信賴的 AI 爲目標，追求科技帶來的永續發展。經由法制落差的比較分析，本文以四個層次的法規調適途徑，提出目前我國如欲運用機器演算法於公部門中，應思量之問題與因應之道：

第一，倫理價值與政策原則層面：我國在建立規制方式制度時，應將倫理價值與政策原則列爲重要核心，可分爲三個層級：告知程序完備、透明性與可解釋性的考量、對公平性或透明度之判斷能力。

第二，風險評估分級層面：釐清各領域專業性及 AI 運用目的後，在共通原則上建立審查機制，以判斷於該領域中業務推動時的風險控管。依照風險程度區分其監管

力道，使前端的資料運用到後端損害救濟，得以受到規範保障。

第三，組織規劃定位層面：政府組織改造時程之速度一向無法趕上科技的日益更迭，各領域之專業性與多元性程度也高，整合與監理工作仍須配合各專業領域需求。因此以現有機關擇一負責，並與目的事業主管機關合作，作爲加速橫向連結與發展重點項目時的基礎，共同設計監理機制，應爲較可行之方式。

第四，個人互動決策制度與適用領域層面：我國目前首當其衝即是對於個人資料保護監理機制尚非完備，因此針對資料治理應有清楚的法規架構，先透過完整之資料治理規範，才有可能妥適處理後續 AI 衍生之問題並進行課責。

在全球化 AI 風潮下，由於公共服務涉及公眾及個人利益，因此規劃公共服務導入 AI 時，需要有較爲具體的制度設計策略與法規調適作爲。相較於各國的演進，我國的法規調適作爲明顯落後，除期待數位發展部加速推動以外，更需要各部會積極投入，以增進公共服務之品質以及公信力。

參考文獻

一、中文部分

王偉霖（2019）。理財機器人對我國金融及相關法制的衝擊與發展。**財金法學研究，2**（3），388-390。[Wang, W.-L.(2019). The development and impact of robot-advisor on financial industry and relevant regulations of Taiwan. *The Financial Law Review*, *2*(3), 388-390.]

行政院（2019）。**台灣 AI 行動計畫一掌握契機，全面啓動產業 AI 化**，8 月。https://www.ey.gov.tw/Page/5A8A0CB5B41DA11E/a8ec407c-6154-4c14-8f1e-d494ec2dbf23 [Executive Yuan (2019). *AI Taiwan action plan*, August.]

張其祿（2008）。法規（管制）影響評估理論與實務之初探。**研考雙月刊，32**（2），50-58。[Jang, C.-L. (2008). An introduction to the theory and practice of regulatory impact analysis. *RDEC*, *32*(2), 50-58.]

國科會（2019）。**科技部訂定「人工智慧科研發展指引」完善我國 AI 科研發展環境**，9 月。https://www.nstc.gov.tw/folksonomy/detail/dbf8da09-22be-4ef1-8294-8832fc6e8a26?l=ch [National Science and Technology Council (2019). *AI technology R&D guidelines*, September.]

黃舒芃（2005）。比較法作爲法學方法：以憲法領域之法比較爲例。**月旦法學雜誌，120**，183-198。[Huang, S.-P. (2005). Comparative law as a legal method: Taking legal comparison in the field of constitutional law as an example. *The Taiwan Law Review*, *120*, 183-198.]

曾冠球、廖洲棚、黃心怡、陳敦源、何宗武（2019）。**新興科技與公部門數位轉型：個案採擷與模式建構**（編號：NDC-MIS-108-003）。國家發展委員會。[Tseng, K.-C., Liao, Z.-P., Huang, H.-I., Chen, D.-Y., & Ho, T.-W. (2019). *Emerging technologies and digital transformation in the public sector: Case selection and model construction* (Project number: NDC-MIS-108-003). National Development Council.]

二、英文部分

116th Congress, USA (2019). *H.R.2575-AI in government act of 2020*. CONGRESS.GOV, December. https://www.congress.gov/bill/116th-congress/house-bill/2575/text

116th Congress, USA (2020). *H.R.6216-national artificial intelligence initiative act of 2020*. CONGRESS.GOV, March. https://www.congress.gov/bill/116th-congress/house-bill/6216

Aghion, P., Jones, B. F., & Jones, C. I. (2018). Artificial intelligence and economic growth. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 237-282). University of Chicago Press.

AI Council, UK (2021). *AI roadmap*. GOV.UK, January 6. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949539/AI_Council_AI_Roadmap.pdf

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and big data. *Public Administration*, *97*(2), 396-310.

Asaro, P. M. (2006). What should we want from a robot ethic? *The International Review of Information Ethics*, *6*, 9-16.

BEIS (Department for Business, Energy & Industrial Strategy), UK (2017). *Industrial strategy: Building a Britain fit for the future*. GOV.UK, November 27. https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future

BEIS (Department for Business, Energy and Industrial Strategy), & DCMS (Department for Digital, Culture, Media & Sport), UK (2019). *AI sector deal*. GOV.UK, May 21. https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal

Boddington, P. (2017). Does AI raise any distinctive ethical questions? In P. Boddington (Ed.), *Towards a code of ethics for artificial intelligence* (pp. 27-37). Springer.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

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

Chui, M., Manyika, J., & Miremadi, M. (2018). What AI can and can't do (yet) for your business. *Harvard Business Review*, *96*(1), 100-109.

Dafoe, A. (2020). *AI governance: Opportunity and theory of impact*. Effective Altruism Forum, September 17. https://forum.effectivealtruism.org/posts/42reWndoTEhFqu6T8/ai-governance-opportunity-and-theory-of-impact

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Morison, J., Murphy, M. H., O'Brolchain, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, *4*(2). https://doi.org/10.1177/2053951717726554

Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, *3*, 398-415.

Engstrom, D. F., Ho, D. E. Sharkey, C. M., & Cuéllar M.-F. (2020). *Government by algorithm: Artificial intelligence in Federal administrative agencies*. Administrative Conference of the United States (ACUS), February 19. https://www.acus.gov/document/government-algorithm-artificial-intelligence-federal-administrative-agencies

European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence) (2019). *Ethics guidelines for trustworthy AI.* European Commission, April 8. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission (2020a). *Impact assessment of the Digital Markets Act*, December 16. https://digital-strategy.ec.europa.eu/en/library/impact-assessment-digital-markets-act?fbclid=IwAR3flf3uGATqrQBz0YXMzajbdCMwY4mYJwTAgPXK0wW0z6-VrcUrH8vXu80

European Commission (2020b). *White paper on artificial intelligence: A European approach to excellence and trust*, February 19. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission (2021a). *Proposal for a regulation laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. EUR-Lex, April 21. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

European Commission (2021b). *Commission staff working document, impact assessment accompanying the proposal for a regulation of the European Parliament and of the Council, laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* EUR-Lex, April 21. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021SC0084

European Commission (2021c). *Commission staff working document, impact assessment accompanying the Proposal for a Regulation of the European Parliament and of the Council on machinery products.* EUR-Lex, April 21. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021SC0082

European Commission (2022). *Liability rules for artificial intelligence*, September 28. https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en

European Parliament (2020). *European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies* (2020/2012(INL)). EUR-Lex, October 20. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.pdf

Executive Office of the President, USA (2016). *Big data: A report on algorithmic systems, opportunity, and civil rights.* President Barack Obama the White House, May. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

de Francesco, F., Radaelli, C. M., & Troeger, V. E. (2012). Implementing regulatory innovations in Europe: The case of impact assessment. *Journal of European Public Policy*, *19*(4), 491-511.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254-280.

GO-Science (Government Office for Science), UK (2016). *Artificial intelligence: Opportunities and implications for the future of decision making*. GOV.UK, November 21. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

Hess, C., & Ostrom, E. (2003). Ideas, artifacts, and facilities: Information as a commonpool resource. *Law and Contemporary Problems*, *66*,111-146.

ICO (Information Commissioner's Office), UK (2019). *An overview of the auditing framework for artificial intelligence and its core components*, March. https://ico.org.uk/about-the-ico/news-and-events/ai-blog-an-overview-of-the-auditing-framework-for-artificial-intelligence-and-its-core-components/

Kirkpatrick, C., & Parker, D. (2004). Regulatory impact assessment and regulatory governance in developing countries. *Public Administration and Development: The International Journal of Management Research and Practice*, *24*(4), 333-344.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14-29.

Larson, J., Mattu, S. Kirchner, L., & Angwin, J. (2016). *How we analyzed the COMPAS recidivism algorithm.* Pro Public, May 23. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lin, P., Abney, K., & Bekey, G. (2011). Robot ethics: Mapping the issues for a mechanized world. *Artificial intelligence*, *175*(5-6), 942-949.

Lucas, H. C. Jr. (2017). Ethics and artificial intelligence: The moral compass of a machine. *IT Professional*, *19*(2), 4-8.

OECD (2019a). *Regulatory effectiveness in the era of digitalization*. OECD.

OECD (2019b). *Recommendation of the council on artificial intelligence*, May 22. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD (2021). *The state of implementation of the OECD AI principles: Insights from national AI policies*, June 18. https://www.oecd.org/digital/state-of-implementation-of-the-oecd-ai-principles-1cd40c44-en.htm

Office for Artificial Intelligence, UK (2020). *A guide to using artificial intelligence in the public sector*. GOV.UK, January 27. https://www.gov.uk/government/publications/a-guide-to-using-artificial-intelligence-in-the-public-sector

Ostrom, E. (1986). A method of institutional analysis. In F. X. Kaufman, G. Majone, & V. Ostrom (Eds.), *Guidance, control, and evaluation in the public sector* (pp. 495-510). de Gruyter.

Ostrom, E. (2011). Background on the institutional analysis and development framework. *Policy Studies Journal*, *39*(1), 7-27.

Polski, M. M., & Ostrom, E. (2015). An institutional framework for policy analysis and design. In Cole D. H., & M. D. McGinnis (Eds.), *Elinor Ostrom and the Bloomington School of Political Economy* (pp. 13-40). Lexington Books.

Radaelli, C. M. (2004). The diffusion of regulatory impact analysis: Best practice or lesson drawing? *European Journal of Political Research*, *43*(5), 723-747.

Renda, A. (2006). *Impact assessment in the EU: The state of the art and the art of the state*. Center for European Policy Studies.

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). https://doi.org/10.1016/j.giq.2019.07.004

Stockmann, D. (2023). Tech companies and the public interest: The role of the state in governing social media platforms. *Information, Communication & Society*, *26*(1), 1-15.

STOA (The Panel for the Future of Science and Technology), European Parliament (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. European Parliament, June 25. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530

The White House (2019). *Maintaining American leadership in artificial intelligence*. Federal Register, February 14. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

The White House (2020). *Promoting the use of trustworthy artificial intelligence in the Federal Government*. Federal Register, December 8. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

The White House's Office of Science and Technology Policy (2020). *Guidance for regulation of artificial intelligence applications*. The White House, January 6. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Veale, M., Van Kleek, M., & Binns, R. (2018). *Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making* [Conference presentation]. 2018 CHI Conference on Human Factors in Computing Systems, April 21-26, Montreal, QC, Canada. https://dl.acm.org/doi/10.1145/3173574.3174014

Weimer, D. L., & Vining, A. R. (2017). *Policy analysis: Concepts and practice* (6th ed.). Taylor & Francis.

Williamson, O. E. (1980). The organization of work: A comparative institutional assessment. *Journal of Economic Behavior and Organization*, *1*(1), 5-38.

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.

Zerbe R. O. Jr., & McCurdy, H. E. (1999). The failure of market failure. *Journal of Policy Analysis and Management*, *18*(4), 558-578.

Zevenbergen, B., & van der Voort, H. (2016). Applying the notion of algorithmic accountability to decision-making software in the public sector. *Government Information Quarterly*, *33*(1), 32-41.

Great Taste and Less Filling? A Review of Institutional Design Strategies and Regulatory Adjustment Action for Public Service through Artificial Intelligent Application

Brian Pin-Hsuan Chang[*] *Don-Yun Chen*[**] *Calvin Zhou-Peng Liao*[***]
Hsini Huang[****] *Chian-Wen Wang*[*****] *Ying-Hsin Chang*[******]
Yu-Han Chen[*******]

Abstract

Background: During the government digital transformation process, using artificial intelligence (AI) technology, the establishment of regulations and risk control measures all require forward-looking policy planning and worth to be fully explored and considered.

Methods: Starting from the theory of Institutional Analysis and Development framework (IAD) combined with Regulatory Impact Assessment (RIA), this study uses the approach of literature comparative analysis to review the experiences of international organizations and advanced countries. We will propose problems that should be considered, difficulties encountered, regulatory system solutions and strategies, etc. When it comes to applying AI in public services, we especially focusing on human rights

[*] Adjunct Assistant Professor, Graduate Institute of Health and Biotechnology Law, Taipei Medical University; Adjunct Attending Physician, MacKay Children's Hospital; Chairman, Attorney at Law, ChampTime BioMed and Law Firm. Email: brianpchang@gmail.com (Corresponding Author)
[**] Professor, Department of Public Administration, National Chengchi University. Email: donc@nccu.edu.tw
[***] Associate Professor, Department of Public Administration, National Open University. Email: zpliao@mail.nou.edu.tw
[****] Assistant Professor, Faculty of Governance and Global Affairs, Leiden University, the Netherlands; Affiliated Associate Professor, Department of Political Science, National Taiwan University. Email: h.i.huang@fgga.leidenuniv.nl
[*****] Assistant Professor, The Department of Public Administration, Tamkang University. Email: cww@mail.tku.edu.tw
[******] Doctoral Student, Department of Law, National Chung-Cheng University.
[*******] Doctoral Student, Department of Public Administration, National Chengchi University.

protection issues and the policy implementation mechanism.

Result: There are certain mature development paths for the AI supervision mechanisms, field diversity, and the executive processes in most of the developed countries. From the beginning policy white papers to relatively clear guidelines and finally a specific draft of regulations, the experiences of other countries can serve as our examples. In the comparative study, we applying the following four principles: ethical orientation, risk response, organization supervision and evidence-based policies. Finally, we propose methods to introduce AI in future public services.

Research findings and suggestions: From the perspective of ethics, the common principles of all countries are based on the idea of "people-centered" human rights protection. At the risk level, different control measures should be implemented according to different service risk orientations. From the organizational level, it is suggested that there should be separate organizations responsible for integration, coordination and supervision. In terms of decision-making, it is recommended to provide specific normative guidelines based on ethical values, so that the development of AI applications remains flexible.

Keywords: institutional design, regulatory adjustment, artificial intelligent, public service, algorithm