

成果發表一：我國政府推動AI為基礎的創新應用機制：兼論法規的落地與調適

公部門AI自動化決策 與責任治理

日期：2026年5月13日

報告人：東吳大學法律學系 許慧瑩 助理教授

email: hyhsu007@scu.edu.tw

大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

前言

背景

- AI進入行政決策
- 決策自動化，傳統責任架構受挑戰
- AI應用下公務人員責任與風險控管

研析目的

- 現行規範未明確連結風險與責任
- 指引與檢核工具缺乏法律拘束力
- 影響公務人員使用動機與動力

研究目標

- 釐清責任概念
- 公部門ADM之風險控管
- 責任導向制度設計

現行規範與治理挑戰

行政程序法

- §96、§97有關自動機器做成之行政處分，目前侷限於羈束性決定

人工智慧基本法

- 為政策框架，§ 4(7)、§17提及問責、負責等，規範風險分類，但未對應責任類型與減緩機制

公部門AI應用參考手冊 (2025)

- 為自願遵循指引，不具法拘束力

臺灣AI行動計畫2.0 (2023)

- 為AI政策綱領，促進創新面向為主

數發部AI產品與系統評測指引

- 偏向技術評測，責任連結不足

計畫訪談：公部門人工智慧採用現況

科技層面

(Technology Dimension)

數位資料積累與模型參數



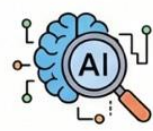
模型精準度
(Model Precision)

資料需進行數位化，並設定合理參數

優先應用於具比較利益之業務



人類範疇



AI範疇

優先使用在影像辨識、趨勢預測以及網站稽核

去識別化落實的限制



特殊個案仍有被辨識風險

組織層面

(Organization Dimension)

業務性質：重複性高者先行



水利監測



稅務查核等

耗費人力與重複性日常業務先行

跨組織資料分享與集中管理



面臨法律規範與責任歸屬困難時，可採行 AI 集中管理策略，進行跨局處整合以降低分享難度。

推動策略

模型部署		商用模型處理開放資料；機敏資料使用落地模型（地端）
激勵機制		建立科技預算與內部標章認證，鼓勵公務員放心創新
驗證模式		搭配概念驗證（POC），由公務機關主動搜出需求
開發重點		專才型 AI，針對特定業務設計小型代理人（Small Agent）

環境層面

(Environment Dimension)

適法性與法規調適



AI 採用必須符合政府資料保護法規與行政倫理，初期可透過「內部行政指引」先行調適。

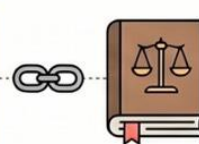
尊重場域原有法律規範



稅政



交通



在特定領域（如稅務、交通）應用 AI 時，仍須嚴格符合該領域原有的法律框架。

主管態度與政治責任



政務官承擔責任

主權語言模型 (Sovereign AI)



強調建立本土主權語言模型的必要性，以確保文化與法制環境的適配。

計畫訪談：公部門人工智慧採用風險因應機制

人類主導與決策核心

AI 僅作為輔助，決策權在人



公務人員擁有最終決定權，並可隨時推翻 AI 建議以回歸傳統決策模式。



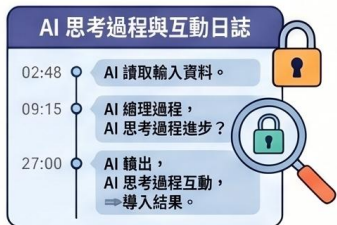
現行課責與避免衝擊策略

行政責任由公務人員承擔



無論是否使用 AI 輔助，擔負行政與法律責任的主體依舊是公務人員。

決策過程透明化與可追蹤



系統需記錄 AI 的思考過程與互動日誌，供人員在做最終判斷時參考。

人類擔任跨組織協調者



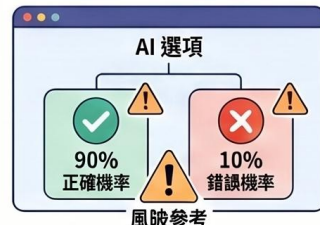
AI 目前仍無法處理複雜的跨部門決策，需由人類負責各單位間的協調與整合。

建立民眾溝通的緩衝介面



避免 AI 直接面對民眾，如市民熱線需由話務人員判斷 AI 資訊後再回覆。

附帶責任歸屬與機率聲明



AI 提供的選項中應闡明正確與錯誤機率，作為人員決策的風險參考。 7

問題意識

AI ADM 改變行政決策？

- 誰在決策？

責任仍停留在人？

- 誰要負責？

責任配置落差？

- 如何監督？

問題解決？

核 心 主 張

自動化風險 = 責任結構失衡

自動化不是卸責，而是責任轉移

治理重心應由「事後」究責與賠償移
至「導入前評估」

- 建立「基本權利影響評估」(FRIA)
- 建立「演算法影響評估」(AIA)

大綱

背景與目的

- 為何公部門使用ADM需特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制、手冊草稿定位與後續方向

行政自動化與AI導入(1/2)

自動化決策 (Automated Decision-Making, ADM) : 透過電腦系統自動化行政決策程序的一部分或全部。

應用領域	傳統 ADM 案例	AI Based ADM 案例
稅務	自動核算稅額、逾期罰款 (如愛沙尼亞 e-Tax)	AI 預測逃漏稅風險、異常申報偵測
社會福利	社會福利資格條件自動審核 (收入門檻)	AI 評估失業給付詐欺風險 (如荷蘭SyRI)
交通執法	交通測速照相自動開罰 (如荷蘭自動測速與紅燈攝影)	AI 影像辨識違規 (闖紅燈、違停)
移民 / 簽證	移民簽證規則型黑名單比對	AI 審查入境風險評分 (如美國 ATLAS)
司法 / 矯正	前科紀錄自動查詢	AI 假釋風險評估 (如美國 COMPAS)
警政	警政自動比對通緝資料庫	預測性警務 (如 美國PredPol 熱點預測)
醫療給付	健保費用自動審核規則	AI 審查醫療申報異常 (防詐欺)

行政自動化與AI導入(2/2)

面向	傳統 ADM	AI Based ADM
透明度	規則公開，較易審查	黑箱問題，難以審查
可解釋性	高（邏輯較明確）	低（模型複雜）
責任歸屬	相對清楚	模糊（開發者？使用者？）
歧視風險	可能因規則設計偏誤	可能因訓練資料偏誤，且更難察覺
正當程序	較易符合說明義務	說明義務履行困難
規範密度	現行法多可涵蓋	可能需要專法或專章處理

AI Based ADM行政決策結構從以「人」為中心轉向「人機協作」，由資料、模型與系統共同形成決策

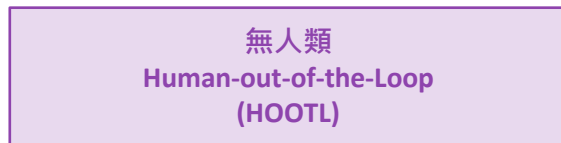
ADM自動化程度的光譜

以學者Ulrik B. U. Roehl 提出的自動化程度可分為A-F層級



← 低度自動化

高度自動化 →



ADM自動化程度與公務人員人機互動

	類型	說明
半 自 動 化	B類型自動化工具（取得與呈現資料）	<ul style="list-style-type: none">• 決策由公務人員與系統共享• 系統負責自動彙整、登錄與呈現相關資料，補充人工蒐集之資訊• 所有實質判斷仍由公務人員作成
	C類型程序輔助建議	<ul style="list-style-type: none">• 決策由公務人員與系統共享• 系統除資料呈現外，進一步提出後續應採取之程序步驟之建議• 最終決策仍掌握在公務人員手中
	D類型決策輔助建議	<ul style="list-style-type: none">• 決策由公務人員與系統共享• 系統不僅提供資料與程序建議，亦可提出可行方案或單一建議決定• 公務人員負責最終決策
全 自 動 化	E類型自動化（Automated）決策	<ul style="list-style-type: none">• 系統握有主要決策• 所有決策步驟均由系統自動執行，依固定且可預期的輸入輸出運作• 公務人員無須介入
	F類型自主（Autonomous）決策	<ul style="list-style-type: none">• 系統完全掌控決策• 行政決策在動態隱含的輸入輸出邏輯下運作，具高度自主性• 公務人員無須介入

大綱

- 為何公部門ADM特別關注責任問題？

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- **Responsibility**、**Accountability**、**Liability** 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

責任有三種...

	行為責任 Responsibility	課責 Accountability	法律責任 Liability
核心概念	基於職務所負之作為 與注意義務	行政行為的可追溯性 與說明義務	對損害承擔法律後果
時間向度	事前 (ex ante)	事後 (ex post)	事後 (ex post)
強制性	較低 (道德 / 角色義務)	中 (程序性義務)	最高 (法律強制)
關鍵問題	誰應該做什麼？	誰須說明為何如此做？	誰須承擔損害後果？
AI ADM 情境	誰應負起設計、監督、 部署 ADM 系統的職責？	誰須對演算法決策的 後果進行說明與接受 審查？	造成具體損害後，誰 須承擔法律賠償或被 懲罰？

自動化程度與責任解構

行為責任 Responsibility

- 職責邊界模糊
- 自動化程度越高，個人行為責任越趨向機關

課責 Accountability

- 課責不再僅仰賴公務人員盡職，與系統可解釋性亦密切相關
- 黑箱導致說明義務難以履行與監督

法律責任 Liability

- 民事刑事或懲戒，違法之法律後果（國賠等）因果關係難以認定
- E型責任重心移轉至機關
F型現行無合法空間

分類	自動化程度	人機互動	行為責任	課責	法律責任	備注
B	半自動	HITL - 完整介入	公務人員（高）	高		一般審查
C-D	半自動	HITL/HOTL 審查	公務人員 + 機關	中		審查AI建議
E	全自動	HOTL 事後監督	機關為主（低）	低		審查弱化
F	自主	HOOTL（無介入）	機關（不合法）	—	不合法	現行無合法空間

ADM下責任結構的轉變

傳統行政決策：三種責任的一致性

通常由同一人承擔、指向同一決策：

- 行為責任 (**Responsibility**)：誰作成決策，誰承擔職務上的作為義務
- 課責性 (**Accountability**)：決策者能否說明理由、接受監督與問責
- 法律責任 (**Liability**)：決策造成損害時，由誰承擔民刑事或賠償責任

ADM 的挑戰：三者的結構性脫鉤

自動化程度提高後，人類的實質控制能力隨之下降。

決策的形成從「由人判斷」轉變為「人機互動共同完成」，三種責任不再自然對應同一主體

- 行為責任**模糊化**：公務人員可能僅是系統的操作者或監督者，而非實質決策者
- 課責性**弱化**：演算法的黑箱特性使決策理由難以說明，程序透明度降低
- 法律責任**分散化**：損害發生時，責任可能落在開發者、採購機關或使用人員之間，歸屬不明

自動化不會使責任消除，責任重心從「作成決定」轉向「選用系統、設定參數與持續監督」

大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

荷蘭 DUO 演算法風險剖繪 (2012–2023)

Type E 全自動 | 教育補助詐領查核 | 涉及歧視性演算法 | 2023年停用 2024年政府道歉 · 啟動補償與重審

事實	爭點	結果與啟示
<ul style="list-style-type: none">• 年齡、就學層級、住址距離等指標自動標記高風險學生• 無人工介入，從申請到處分全自動 (System-level bureaucracy)• 追繳金額高達數千至逾萬歐元	<ul style="list-style-type: none">• 演算法指標代理種族/社經地位，構成間接歧視 (disparate impact)• 黑箱操作使學生無法理解「為何被選中」，剝奪程序保障• 98%的訴訟案件由具移民背景學生提起	<ul style="list-style-type: none">• 2023年被停止使用• PwC調查認定歧視• 政府公開道歉、補償、重審• 傳統一線裁量轉為「系統操作員」，公務人員責任質變

演算法代理變數 (**proxy variables**) 看似中立的指標隱含差別
需於系統設計前進行基本權利影響評估 (**FRIA**)

美國 COMPAS 再犯風險評估系統

Type B-D 混合 | 量刑輔助工具 | 黑箱商業軟體 | **State v. Loomis (2016)** | 仍在使用，明確要求法官不得以其建議作為量刑之唯一或決定性依據

系統架構與爭點

- 輸入靜態/動態因子，輸出1-10再犯風險
- 計算公式為營業秘密，被告無法檢視
- 預測準確率僅約.70，對黑人被告出現系統性不利模式（ProPublica分析）
- 統計公平指標之間存在「不可兼得」困境，無法以技術手段單獨解決
- 法院採有限容許加程序告誡折衷立場

治理啟示

- 裁量輔助（Type D）若公務人員未獨立判斷，實務形同全自動（Type E）
- 商業軟體的黑箱性質侵害當事人的可爭執性與理由說明
- 種族偏差源自歷史資料不平等的系統性再製
- 法官最終決定的制度設計，不足以消除演算法偏差的影響

採購商業ADM系統時，應於契約要求演算法透明
輔助建議（D型）仍需確保公務人員實質獨立判斷

波蘭 Syriusz 失業者自動化風險分類

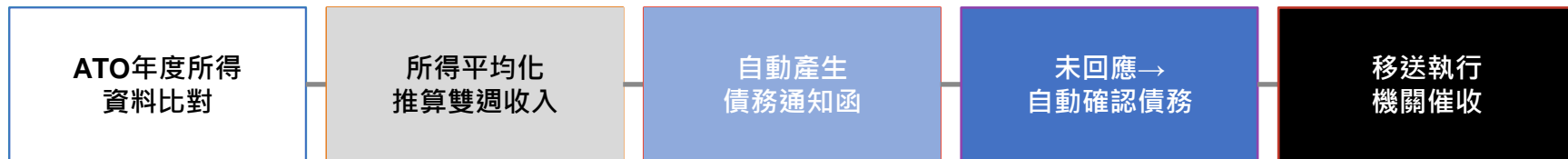
Type B / C / D 半自動化 | 公共就業服務 · 失業剖繪系統 | (2014–2019) 2019年違憲廢除

事實	爭點	結果與啟示
<ul style="list-style-type: none">• Syriusz系統：全國350就業中心導入，以登記 + 問卷分類失業者三類• 協助差異：高教育者13項、技能不足者29項、弱勢者多數無協助• 問題：公務員幾乎不推翻系統（僅1%）→ 實質全自動決策	<ul style="list-style-type: none">• 歧視風險：弱勢群體（ex 身障者、單親母親、農村居民）系統性被劃入第三類透明度不足：• 演算法保密，失業者無法知悉或要求人類介入• 個資處理未符合法律保留• 制度不一：各地執行差異大，侵害權利救濟	<ul style="list-style-type: none">• 2019年12月憲法法庭認定違憲，制度正式廢除• 人權專員提起訴訟：個資處理未具法律保留、侵害權利救濟• 形式保留人類覆核，實質已移轉為系統決策—裁量權空洞化

原為建議工具卻因公務人員幾乎不推翻系統決定，而實質成為全自動決策演算法黑箱 + 缺乏救濟管道 + 個資處理未具法律保留，三重缺陷導致違憲

澳洲 Robodebt (OCI) 社會福利債務追繳

Type E 全自動 | 社會安全給付 | 皇家調查委員會最終報告 (2023)



▲ 舉證責任倒置：公民須自證清白，否則視為確認

爭點一：違法的所得計算

以統計平均替代法律要求的「實際期間所得」，缺乏法律依據，被皇家委員會定性為「粗糙且殘酷的機制，既不公平也不合法」

爭點二：程序性負擔外部化

弱勢公民難以理解、回應與舉證，形成典型的「自動化決策將查證義務轉嫁公民」問題

爭點三：差別影響集中風險

核心違法在於所得平均化，但不利效果集中在收入不穩定族群，符合差別影響 (disparate impact) 模式

全自動化系統須有完整的使用者測試 (含弱勢族群) 、風險評估前置程序
在系統錯誤時，機關不得以系統計算作為免責理由

案例分析

半自動化案例：人類存在，但是否真正做決策？

- COMPAS: 風險分數輔助量刑，但黑箱與種族偏差引發正當程序爭議
- 波蘭失業者分類：系統建議名義下，人類幾乎不推翻系統結果
- 重點: 風險在「人類未實質介入」

全自動化案例：決策存在，但決策者消失？

- DUO: 系統自動標記高風險學生，引發歧視與補償重審
- Robodebt: 系統以平均所得自動推算債務，造成大規模錯誤追繳
- 重點: 風險在於「制度責任必須前移」

案例分析

決策形式有人類
實質已自動化

- HITL ≠ 實質控制

黑箱 + 無救濟

- 課責機制失效

責任仍在機關
但控制能力下降

- 責任錯位

大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

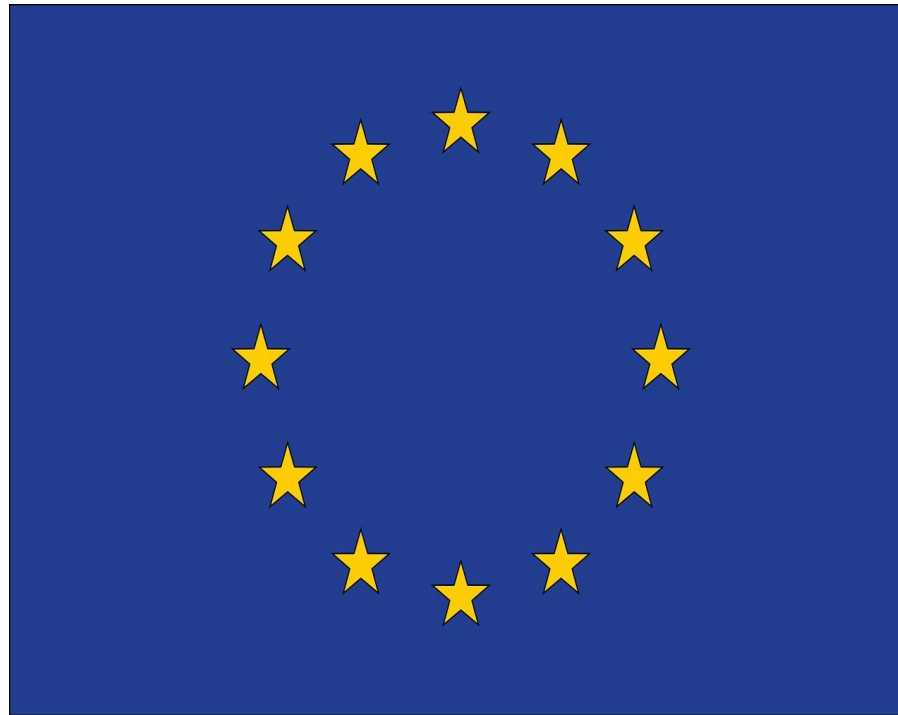
- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

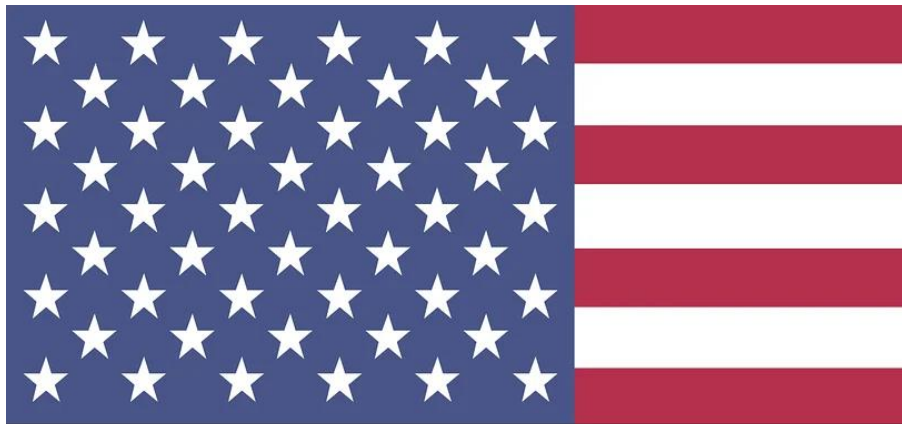
歐盟：高風險 + 權利保障

- 風險分級 (EU AI Act)
- 高風險系統：強制義務 (Art. 26)
 - 人類監督、資料品質、log 紀錄、系統登錄
- 關鍵原則
 - 決策仍歸屬機關、不得以外包規避責任、強調可解釋與救濟
- 核心：責任集中於機關



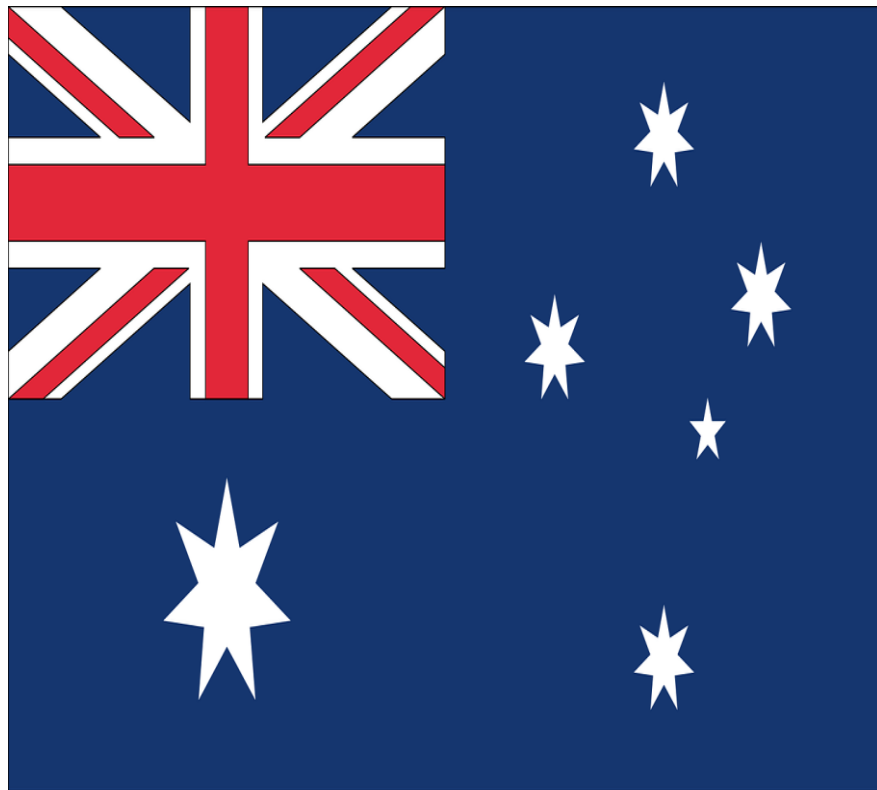
美國：分散 + 問責強化

- 無統一聯邦法
 - 以工具與政策為主：
 - GAO AI Accountability Framework
 - OMB備忘錄
- 州法發展
 - 偏誤審查 (bias audit)
 - 核心：高風險AI規範 (如Colorado AI Act)
 - 使用流程與證據降低責任風險



澳洲：行政法 + 最佳實務

- Robodebt事件 → 強化治理
- 核心要求：
 - 人類最終決定
 - 透明 + 可解釋
 - 不得委外卸責
- 工具：ADM Best Practice Guide (完整檢核表)
- 核心：行政法原則不因AI而改變，
流程治理
- 討論：2025年11月7日發布「自
動化決策法案草擬指引說明」



加拿大：影響評估（AIA）

- 強制適用：自動化決策指令（Directive on Automated Decision-Making, DADM）
- 風險分級：Level I-IV
- 對應義務：
 - 透明度、人為介入、外部審查
- 全生命週期管理：設計→部署→監測
- 核心：用評估制度化責任治理



ADM責任治理趨勢 (1/2)

- 三種治理模式
 - 權利導向 (EU) → 高風險 + 基本權保障
 - 分散治理 (US) → 透明 + 問責 + 風險管理
 - 程序治理 (AU / CA) → 指引 + 影響評估
- 共同趨勢
 - 不允許AI卸責
 - 強化：人類監督、可解釋性、稽核軌跡
 - 從「技術治理」走向「責任治理」

ADM責任治理趨勢 (2/2)

- **AI ADM 責任治理**
 - 責任分配 + 風險控制機制設計
 - 差異不在「要不要負責」，而在「如何讓責任可被證明與控制」
- **建議台灣應採**
 - 「評估制度 + 責任連結」的混合模式
(結合加拿大、澳洲 + EU權利保障)

項目	演算法影響評估 (AIA)
核心目的	評估AI風險
核心關注	模型風險、偏誤、監督與治理
評估對象	AI ADM系統
重點問題	是否適合導入AI
評估時點	導入前 + 運作中
內容	資料品質、可解釋性、人類監督
制度功能	風險治理與責任減緩
與責任關聯	建立可追溯治理與問責基礎

大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

影響層級分類框架治理要求

層級	類型	人機互動要求	風險特性	核心責任減緩
Level I 低影響	B型 半自動	可無即時人工，需可追溯	低度可逆，短暫影響	資料來源紀錄、系統log、通知民眾
Level II 中度影響	C型 程序輔助	人類參與品質保證，可介入	中度，可能可逆	偏誤檢測、公開決策邏輯、申訴管道
Level III 高影響	D型 決策輔助	人類必須審查輸出建議	高度，難以逆轉	人工覆核機制、稽核日誌軌跡、定期偏誤評估
Level IV 極高影響	E型 自動化決策	最終決策必須由人類作成	極高，不可逆轉	人類最終決定 (Human-in-command) 、獨立審查、退出計畫

ADM 生命週期責任控管

決定採用

- 基本權利影響評估 (FRIA) : 是否涉及基本權利、弱勢族群、差別影響
- 演算法影響評估: 風險層級分類
- 是否有低風險替代方案? 人工處理是否可行?

採購設計

- 契約中要求演算法透明化條款
- 資料來源、訓練資料偏誤審查
- 確保系統說明義務與可解釋性需求

部署測試

- 場景測試含弱勢使用者、邊緣案件
- 業務規則視覺化驗證 (非IT人員可理解)
- 通知設計: 民眾能否理解決策依據?

運作監督

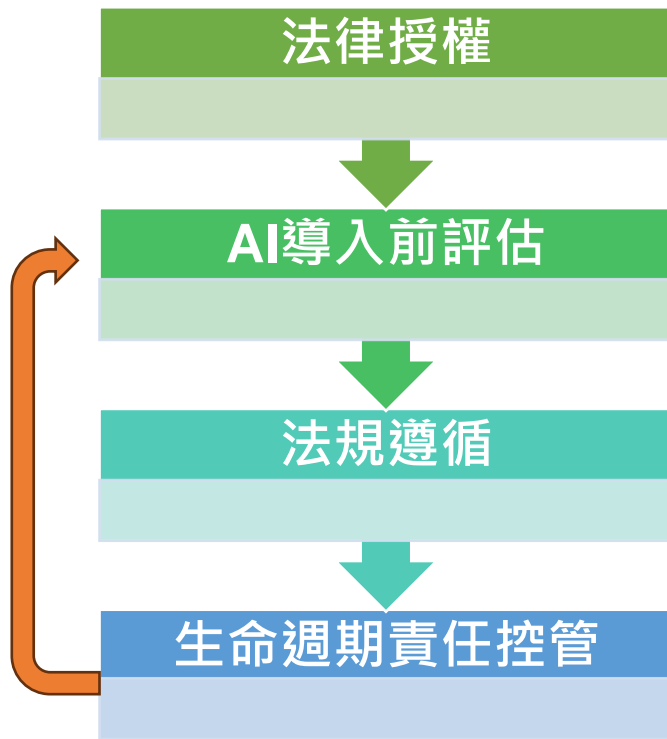
- 人工覆核機制與覆核頻率設定
- 決策日誌、稽核軌跡完整保存
- 申訴管道及錯誤更正程序

更新評估

- 定期重新評估模型偏誤與資料品質
- 法規或政策變更時同步更新業務規則
- 重大變更應重做AIA, 並知會相關單位

演算法影響評估之效果

- 聚焦於系統有多大風險→要如何治理？
 - 主要保護
 - 避免因系統缺陷未管控而產生的行政與民事責任
 - 舉證責任
 - 證明「已依風險等級採取對應治理措施」
 - 最大價值
 - 讓導入後的AI ADM 管理有制度依據可循



大綱

背景與目的

- 為何公部門ADM特別關注責任問題？

ADM的定義與自動化程度的分類

- A-F 六種類型與人機互動模式

自動化程度與責任的結構

- Responsibility、Accountability、Liability 的轉變

國際應用案例與爭議解析

- 荷蘭DUO、美國COMPAS、波蘭Syriusz、澳洲Robodebt

主要國家規制現況

- 歐盟、美國、澳洲、加拿大的治理模式

影響層級與減緩機制

- 層級治理要求與查核清單框架

我國課題與建議

- 現行法制與後續方向

政策建議：制度設計方向

建議一 建立前置影響評估制度

借鑑加拿大AIA與荷蘭FRIA，要求高影響ADM系統在採購前完成演算法影響評估，並依風險層級設定最低治理要求

建議二 明確人類監督制度化規範

依自動化程度（A-E型）明定人類介入比例、覆核頻率與日誌保存要求，並將HITL/HOTL明確納入行政作業規範

建議三 責任與查核清單連結

公部門AI應用參考手冊應在現有查核清單外，增加各層級對應之責任類型（行政/民刑）與減緩措施說明，提升機關遵循意願

建議四 強化透明度與可爭執性

參考EU AI Act，要求ADM系統提供「通知義務」與「說明理由」機制，確保受影響民眾能有效行使異議與申訴權利

建議五 建立事後監督與究責機制

借鑑澳洲Ombudsman制度，建立獨立監督單位或跨部會審查機制，定期稽核重大ADM系統，並公開影響評估與治理報告

建議六 提升公務人員AI素養

將ADM風險意識與責任判斷納入公務人員訓練，特別是採購、部署與監督角色，確保其能進行實質獨立判斷而非淪為系統操作員

演算法核心的責任治理 (1/2)

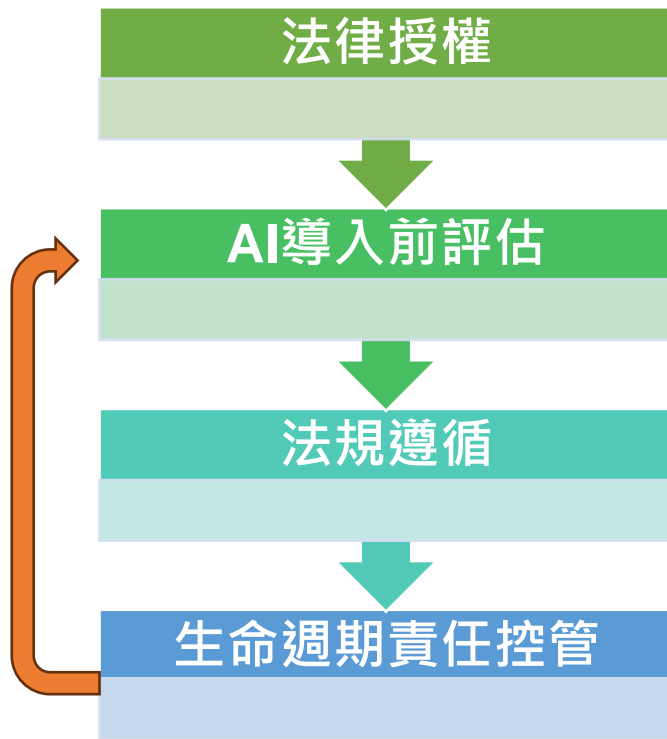
- **ADM導入公共行政，帶來三項結構性風險**
 - 黑箱風險：模型不可解釋，決策理由難以說明
 - 自動化偏誤：過度依賴AI輸出，實質降低人工審查品質
 - 裁量弱化：演算法標準化壓縮個案判斷空間
- **責任錯位與結構性轉移**
 - 自動化程度提高，責任重心轉移：承辦人員 → 機關 → 制度
 - AI實質影響決策，責任形式仍歸人類，人類實際控制能力卻持續下降 → 產生「責任錯位」
 - 責任並未消失，而是轉化為：評估（導入前審查）、監督（運作確保合規）、持續介入（異常時即時修正）

演算法核心的責任治理 (2/2)

- 治理重心前移

- 焦點從「部署後責任減緩」提前至「導入前評估是否採用AI」之「演算法影響評估」

- 建立已盡注意義務的舉證基礎
- 將責任從個人轉移至制度層面
- 降低訴訟和行政爭訟風險



責任治理還缺一塊

- **基本權利影響評估 (FRIA)**
 - 將基本權利保障，內化為AI導入的前置條件，在提升行政效能的同時，建立可問責、可追溯的治理



責任治理的完整性

	基本權利影響評估 (FRIA)	演算法影響評估 (AIA)
主要保護	避免侵害基本權利而產生的違憲/違法責任	避免因系統缺陷未管控而產生的行政與民事責任
舉證功能	證明「已評估比例原則與正當程序」	證明「已依風險等級採取對應治理措施」
最大價值	過濾不應導入的高侵害風險系統	讓導入後的管理有制度依據可循

第一步 決定是否採用 AI

法律授權確認

授權範圍涵蓋自動化嗎？

第二步 基本權利影響評估 (FRIA)

涉及基本權利嗎？

比例原則審查

替代方案

第三步：演算法影響評估 (AIA)

風險分級與治理義務

採購

部署

持續監督