《科際整合月刊》，2026 年 5 月，11(5)，1-25
Interdisciplinary Research Monthly. May 2026

高階公務人力的 AI 素養、媒體識讀對民主韌性與數位治理的功能

江岷欽*

摘要

隨著生成式人工智慧（Generative AI）技術快速發展，政府治理正面臨前所未有的挑戰與機會。特別是高階公務人力作為政策設計與風險應對的核心，其 AI 素養與媒體識讀能力已成為維繫民主韌性與數位治理正當性的關鍵變數。本文採用「循證政策」（Evidence-based Policymaking）與「敏捷治理」（Agile Governance）雙重途徑，透過政策病理學與成功典範對照的方式，分析台灣、新加坡、愛沙尼亞與芬蘭等國的治理實踐，並針對台灣現況提出三層次韌性建構架構與十項政策行動方案（Taiwan AI-10）。本文認為，唯有將 AI 技術治理制度化、文官訓練高階化與治理模式敏捷化，方能確保民主制度於 AI 時代之下持續運作、轉型與進化。

關鍵字：AI 素養、高階文官、敏捷治理、媒體識讀、民主韌性、系統性風險、政策沙盒

* 世新大學客座教授兼管理學院院長
本論文經兩位雙向匿名審查通過。收件：2026/3/19。同意刊登：2026/4/23。（本論文發表於 2025 年 12 月 16 日《2025 Talent X 公共服務與人才未來論壇》）

壹、演算法時代的治理危機與能力重構

隨著生成式人工智慧技術的發展，公共治理正經歷由單純行政工具應用至制度邏輯轉變的結構性變遷。當前行政實務面臨「技術決策」與「社會認知」的雙重治理挑戰：一方面，演算法自動化決策（ADM）之廣泛導入，對傳統科層體制的課責與問責機制產生顯著衝擊；另一方面，演算法驅動的資訊傳播機制，亦客觀上影響了民眾對數位治理正當性的評估基礎。

於此制度環境下，高階常任文官作為掌握行政裁量權與資源配置的核心代理人，其職能已從單一的政策執行，擴展至系統性風險控管與民主制度維繫。既有文獻多將「AI 素養」與「媒體識讀」視為獨立的分析單元；然而，政策實務上的治理失靈，往往肇因於技術邏輯（內部決策）與資訊環境（外部認知）兩者間整合機制的闕如。據此，如何彌合此一職能缺口，已成為維持民主韌性與行政法治之重要課題。

本文旨在剖析此一複合型治理情境，論證 AI 素養與媒體識讀之理論整合，並檢視其作為數位治理核心能力之機制。承此脈絡，下文將探討演算法時代下的治理危機與能力重構路徑，藉以建構公共行政體系回應技術變遷的理論分析框架。

一、 VUCA 環境下的數位利維坦與信任赤字

進入 21 世紀第三個十年，公共治理面臨一個結構性的典範轉移：從工業時代的穩態管制邏輯，過渡至數位時代的不穩定適應邏輯。這一過程發生於全球化碎裂、地緣政治張力升高、資訊空間高度操弄與科技倫理爭議交織的歷史背景中。在這樣一個以「易變、不確定、複雜與模糊」（VUCA）為特徵的後疫情時代中，政府能力的再定義已成為學界與政策實務的重要議題（United Nations, 2024）。

生成式人工智慧（Generative AI）、大型語言模型（Large Language Models, LLMs）、演算法決策機制（ADM）與即時數據流（Real-time Data Streams）等新興技術，正快速進入公共行政與社會治理的核心領域。這類技術強化了效率與預測力，卻也暴露出治理正當性、數據偏見、人權侵害與問責失靈等新型風險（Amnesty International, 2021）。在此情勢下，若決策者未能擁有足夠的 AI 素養與系統性風險意識，極可能導致所謂「數位利維坦」（Digital Leviathan）的浮現，即科技與行政結合成對社會契約與民主秩序的破壞性力量。

此類風險已在荷蘭托育補助演算法歧視案（Dutch childcare benefits scandal）、澳洲 Robodebt 自動債務系統失靈、以及英國 A-Level 評分演算法危機中具體顯現，導致系統性人權侵害與政策信任崩解（Royal Commission, 2023；IEA, 2024）。

二、 民主韌性的數位防線

本文欲釐清一項關鍵議題：在資訊操弄與演算法治理常態化的雙重壓力下，民主體制如何建構具有韌性與反脆弱性的數位治理機制？

所謂「民主韌性」（Democratic Resilience）是指：民主體制在面對內部極化、假訊息攻擊、政策錯配與外部混合戰等挑戰時，仍能維持核心價值、調適風險並持續創新（V-Dem, 2025）。此一能力仰賴兩項核心支柱：

(一)、 循證能力（Evidence-based Capacity）：能即時吸收環境變化中的數據與回饋，轉化為有效且正義的政策。

(二)、 敏捷反應能力（Agile Governance Capacity）：能於高風險不確定下，進行快速試錯與政策疊代。

然而，這兩項能力並非自然而然地出現在傳統官僚結構中，而需由具備 AI 素養與媒體識讀能力的高階文官所引導。因此，「高階公務人力」的職能轉型成為數位民主防線的關鍵樞紐（OECD, 2018）。

本文旨在透過理論整合與國際案例分析，試圖建構一套「融合循證治理與敏捷決策」的新型數位治理模式，並提出強化高階公務人力 AI 素養與媒體識讀的實作建議。其核心目的如下：

(一)、 解構當前公共行政面臨的演算法治理危機；

(二)、 探討 AI 素養與媒體識讀如何成為民主韌性的制度性保障；

(三)、 建立一套結合循證與敏捷邏輯的治理分析架構；

(四)、 彙整多國失敗與成功案例，提出符合台灣國情之政策建議。

本文採用質性分析法，結合政策失靈（Policy Failure）與制度韌性（Institutional Resilience）分析框架，對照荷蘭、英國與澳洲之負面經驗，以及台灣、新加坡、愛沙尼亞與芬蘭之正向典範，從中萃取可行之策略模式。

三、 分析方法與理論依據

本研究主要採用文獻探討法，系統性蒐集並梳理國際組織（如 OECD、UNESCO）、各國政府智庫報告以及公共行政、數位治理與認知安全領域之核心學術文獻，藉以奠定本文之學理基礎。在資料分析與理論建構層面，本文則以雙元途徑分析法作為核心分析框架，融合下列兩項理論視角：

(一)、 循證治理途徑（Evidence-based Policymaking, EBP）：主張政策應以最佳可得證據為基礎，避免直覺與政治偏見決定政策方向（Bipartisan Policy Center, 2023）。

(二)、 敏捷治理理論（Agile Governance Theory）：借鑑軟體開發方法學，主張透過「快速原型」「持續反饋」「用戶參與」實現政策迭代（Deloitte, 2024）。

這兩者在數位治理中呈現高度互補性：循證提供邏輯嚴謹與風險評估，敏捷則補充現場回饋與實作可行性。融合後可催生出「敏捷循證治理」（Agile Evidence-based Governance）的新範式，特別適用於變動快速、技術滲透性高的治理情境。

四、 概念界定

本文首先針對研究主軸所涉及之核心概念進行操作型定義（如表 1），以確立分析之理論範疇。

表 1

研究核心概念表

概念	定義
高階公務人力	指具決策權之高階文官（Senior Civil Servants, SCS），涵蓋部會次長至司處主管。
AI 素養	包含理解演算法邏輯、評估 AI 系統風險、進行倫理審視與操作 AI 工具的能力。
媒體識讀	涵蓋資訊鑑別、假訊息辨識、敘事分析與戰略溝通能力，並具備對社群演算法的反身理解。
數位治理	指政府運用數據、AI、數位平台進行政策制定、服務交付與公民參與的整體過程。
民主韌性	民主體制在高風險環境中維持合法性、公正性與持續創新的能力。

資料來源：本研究整理。

貳、高階公務員的數位核心職能圖譜

一、 由數位治理邁向認知治理

面對高風險與資訊過載的治理場域，傳統官僚體制與科層邏輯已逐漸失靈。以 AI 與資訊科技為基礎的新型治理模式正在興起，不僅仰賴數位工具操作力，更需結合策略前瞻、倫理思辨、與媒體鑑別力。因此，本章從「數位職能」與「認知韌性」兩大理論基礎出發，建構一套高階公務人力的核心職能圖譜，做為後續章節案例評析與政策建議的邏輯基礎。

二、 國際機構治理素養架構之比較與應用

(一)、OECD：數位公務員能力三面向模型

OECD（2023）針對高階文官提出三項數位治理核心能力：

1. 數位規劃與設計（Digital Planning and Design）
   (1) 涵蓋系統思維、科技預測、平台化政府邏輯（Government as a Platform）；
   (2) 強調科技與政策交會處的「預警式決策」。
2. 數據治理與倫理（Data Use and Governance）
   (1) 不僅為技術操作，更關注資料治理、數據倫理（Data Ethics）與 AI 偏見覺察；
   (2) 涉及對《通用資料保護規則》（GDPR）與自動化決策（ADM）的理解與應對。
3. 數位管理與敏捷執行（Digital Management and Execution）
   (1) 包含敏捷專案管理（Agile PM）、服務設計思維（Service Design Thinking）；
   (2) 採購方式從「規格導向」轉向「成果導向」的政策契約設計。

此三面向共同構成數位治理能力之「策略－數據－執行」閉環體系（OECD, 2023）。

(二)、UNESCO：AI 素養與 ROAM-X 原則

UNESCO（2022）則將公務員的 AI 素養視為民主治理的基礎安全能力，並提出「ROAM-X 原則」做為全球數位治理基準：

1. Rights（人權）
2. Openness（開放性）
3. Accessibility（可及性）
4. Multi-Stakeholder Participation（多方參與）
5. Cross-cutting indicators（跨域衡量）

公務員必須具備「技術－社會－倫理」三軸整合判斷力，能進行「人權影響評估」（Human Rights Impact Assessment, HRIA），確保 AI 技術不被濫用以強化社會偏見或差別待遇（UNESCO, 2023）。

三、 媒體識讀能力的戰略升級：從媒介理解到認知防衛

媒體識讀（Media Literacy）之核心內涵，係指行為人主動對媒體資訊進行取得、分析、評估、批判及創造之綜合能力。於數位治理範疇中，其核心目的在於強化決策者對訊息真實性之辨識，並藉由對資訊產製意圖之判準，建立具備反思性之資訊處理程序。隨著生成式人工智慧對資訊生態之介入，此種個體層次的資訊素養，已面臨向制度層次轉型的迫切需求。

在資訊戰與混合戰日益成為國際政治常態的背景下，媒體識讀已從一般資訊素養升級為國家級認知防衛機制（Cognitive Defense Infrastructure）。此概念源自北歐國家如芬蘭與瑞典之「總體防衛」（Total Defense）戰略。

(一)、媒體識讀的三層次模型

依據 European Commission（2024）與 Nordic Policy Centre（2023）之研究，本文整理出高階公務人力所需之三層次媒體識讀能力：

1. 內容層：真假訊息辨識、假新聞結構解讀；
2. 系統層：演算法理解、平台操弄機制分析；
3. 認知層：建構敘事框架（Narrative Framing）、反操弄策略。

此三層能力不僅對抗認知操弄，更可提升政策傳播的說服力與透明度。

(二)、台灣案例補強：認知安全即民主防線

台灣在應對資訊戰方面，將媒體識讀納入教育體系與政府應變流程。以 2020 年總統大選為例，政府透過快速打假機制、數位部戰情室與公民媒體合作，建立「快速、透明、真實」三原則（Belfer Center, 2020）。

四、 敏捷治理能力：數位時代的決策邏輯更新

除了識讀與倫理，高階公務人員更需掌握政策的「敏捷」邏輯，以因應變化快速的環境。

(一)、 敏捷治理四原則（取自 Agile Manifesto 公共治理應用版）

1. 以人民需求為核心：決策圍繞民眾反饋與現場回饋（User Feedback Loops）；
2. 迭代決策而非單一指令：接受「試錯即學習」，建立可回溯與修正制度；
3. 跨部門與跨域團隊協作：打破垂直決策與科層思維；
4. 快速實驗與原型開發（Prototype-first Policy）：以 MVP 模式優先測試小規模政策。

此原則已在新加坡 GovTech、英國 GDS（Government Digital Service）與芬蘭 AI 應用中廣泛落地（McKinsey, 2024；Deloitte, 2024）。

(二)、 「敏捷×循證」的融合邏輯

敏捷與循證並非對立：前者提供快速調適、後者確保政策正當性。兩者結合可形成「Evidence-informed Agile Governance」的新模型，其特徵（見表 2）：

表 2

傳統治理與敏捷循證治理模式之比較

面向	傳統治理	敏捷循證治理模式
決策依據	長期規劃、專家意見	用戶反饋、數據分析、快速原型
風險管理	規避風險（保守模式）	分散風險（小規模測試＋紅隊演練）
成本觀念	避免失敗	容許失敗、轉化為學習資產
實驗方法	RCT、政策評估（後設）	監理沙盒、政策實驗室（前瞻性）

資料來源：Sabel 與 Zeitlin, 2019

參、治理失靈的病理學與案例解析：AI 素養缺口與敏捷真空的系統性風險

一、 分析方法：政策病理學與反事實推理（Counterfactual Reasoning）

治理失敗不僅為政策挫敗，更是民主體制健康度的「徵候指標」。本章採用政策病理學（Policy Pathology）的概念，聚焦於技術系統介入公共治理時，所暴露出的制度性盲點與倫理裂縫。

此外，輔以反事實推理（What-if Scenarios）：若決策者具備 AI 素養與敏捷治理能力，是否能避免災難性後果？透過此法，不僅剖析失敗，也為第四章之成功模型提供對照基準。

二、 荷蘭：托育補助醜聞中的數據偏見與倫理真空

(一)、 事件概述

荷蘭稅務局自 2013 年起，導入自學演算法以預測兒童照顧補助的「詐領風險」，卻在缺乏透明度與外部監管下，產生大規模誤判。

(二)、 核心失敗機制

1. 種族化資料標註（Xenophobic Labeling）：系統將「雙重國籍」視為高風險標籤，導致數以萬計移民家庭遭受錯誤追訴與財務毀滅（Amnesty International, 2021）。
2. 系統封閉與責任錯位：高階主管過度信任 AI 輸出，缺乏「演算法審計」與「人機共決」制度設計，產生責任真空。
3. 反事實推論：若導入敏捷原型測試與演算法影響評估（AIA），問題可於早期發現與修正，避免社會信任斷裂。

(三)、 民主韌性受損指標

1. 信任斷裂指數上升（公民對政府誠信崩解）
2. 對技術治理的合法性高度懷疑
3. 移民政策與社會整合反向激化

三、 澳洲 Robodebt：知法犯法的自動化暴政

(一)、 事件概述

2015－2019 年，澳洲政府推出「Robodebt」自動債務比對系統，以演算法自動核對福利金與稅務資料，寄出數十萬張欠款通知。

(二)、 核心失敗機制

1. 統計邏輯誤用：系統以年度總收入平均分配至每週，無視非典型勞動者（如零工）的實際收入模式。

2. 高層知情卻推行（Willful Negligence）：根據 2023 年皇家委員會報告，官員在明知系統違法風險的情況下，仍強推政策，並壓制法律顧問意見（Royal Commission, 2023）。
3. 逆向舉證責任：政府將「證明自己無罪」的責任轉嫁給最弱勢公民，違背正當程序原則。

(三) 對治理結構的病理學影響

基於上述失靈機制之剖析，此一自動化政策對整體治理結構所造成之病理學衝擊，可進一步由法治、倫理與公眾參與三個面向歸納如下（見表 3）：

表 3

Robodebt 自動化政策對治理結構之病理學影響

面向	表現
法治原則	被演算法侵蝕（違憲卻持續執行）
倫理審查	欠缺道德風險評估與知情審議機制
民眾參與	無公聽或反饋通道，制度黑箱化

四、 英國 A-Level 評分危機：數據理性與個體正義的斷裂

(一)、 事件概述

2020 年 COVID-19 期間，英國 Ofqual 因取消考試，改由演算法預測高中畢業生 A-Level 成績。結果造成大量來自弱勢社區學生遭系統性降分，導致社會大規模抗議。

(二)、 核心失敗機制

1. 偏重學校歷史表現（Historical Overweighting）：忽略個人表現潛力，導致資源缺乏學校的優秀生遭「統計降級」。
2. 缺乏預警與回應（Slow Policy Feedback）：政策無迭代機制，直到民怨爆發才匆促修改，造成信任災難。
3. 誤用循證邏輯：使用過時統計模型與過度信任「客觀資料」，反而背離政策正義初衷（PMC, 2024）。

(三)、 政治與社會後果

1. 社會階級固化指數攀升。

2. 政策透明度與問責性受損。
3. 年輕世代對民主機制的不信任感增強。

五、 比較總結：AI 失靈的三種病理原型

經由上述各國政策失靈案例之剖析，本文將其關鍵病因與治理啟示彙整如下（見表 4），藉以識別系統性之素養缺口。

表 4

AI 失靈之關鍵病因與治理啟示比較總表

案例	關鍵病因	欠缺能力	治理啟示
荷蘭托育	偏見資料未被審計	AI 素養、監督機制	導入演算法審計機制與多方參與審議
澳洲 Robodebt	依法不依德，責任轉嫁	倫理判斷、敏捷撤案	制度需設有法律與倫理雙重閥門
英國 A-Level	統計推理壓倒個體正義	敏捷應變、社會溝通	即時反饋與公眾參與是避免制度危機的保險機制

資料來源：本研究整理

治理失靈並非僅為操作錯誤，更是「系統性素養缺口」的體現。當高階公務員未能掌握 AI 技術邏輯、倫理風險與媒體環境的結構性變化，系統將自動生成「不義的決策輸出」。

肆、成功的治理典範：敏捷與循證的數位治理實踐

一、 從失敗到韌性設計的轉向

與上一章的治理病理學相對，本章聚焦於那些成功整合 AI 技術、敏捷流程與民主價值的治理模式。這些國家不僅以技術為工具，更以數位信任（Digital Trust）為核心治理資本，展示「制度韌性設計」（Resilient-by-Design）在數位治理中的具體實踐。

本文以四個國家為案例主軸，代表四種治理取向（見表 5）：

表 5

成功治理典範之核心能力與特徵模式比較

國家	核心能力	特徵模式
台灣	協作式治理 + 敏捷沙盒	民主即服務（Democracy as a Service）
新加坡	技術官僚治理 + 全員 AI 素養	領導導向的數位轉型
愛沙尼亞	法治式數位信任架構	數據主權與平台國治理
芬蘭	全民媒體識讀 + 教育基礎防線	認知安全即國防

資料來源：本研究整理

二、 台灣：公民科技與政府平台的融合實驗

(一)、 2025 公務人才 AI 專案辦公室

台灣於 2025 年啟動「公務人才 AI 專案辦公室」（AITO），以三層架構強化文官 AI 能力：

1. 高階主管：治理戰略與倫理風險評估（如演算法審計報告解讀）；
2. 中階主管：Scrum 與敏捷治理實作；
3. 基層公務員：實用生成式 AI 工具與資料合規操作（Executive Yuan, 2025）。

結合行政院、數位發展部、國發會等多部會協作，AITO 成為「敏捷政府」的新型態架構單位。

(二)、 vTaiwan：數位審議與演算法共創

vTaiwan 平台透過 AI 工具（如 Pol.is）進行公眾意見聚類與共識建構，具體實踐「先審議、後沙盒、再立法」流程（FNF Taiwan, 2025）。其治理特徵（見表 6）：

表 6

vTaiwan 數位審議模式之治理特徵

元素	敘述
工具	使用 AI 聚合意見，非僅社群投票
流程設計	公民參與 → 政策沙盒 → 正式立法
韌性價值	促進信任、降低對立、擴大參與正當性

台灣此模式獲國際智庫稱為「亞洲的開源治理實驗室」（Democracy Technologies, 2025）。

三、 新加坡：由上而下的智慧國戰略與 AI 素養普及

(一)、 Smart Nation + GovTech 雙軌驅動

新加坡政府於 2020 年代持續以「智慧國」（Smart Nation）為國策主軸，並由 GovTech 執行「雙模 IT 模式」（Bimodal IT）：

1. 模式一：穩定核心 IT 基礎建設；
2. 模式二：敏捷實驗與用戶導向快速開發（如 Parking.sg 應用）(McKinsey, 2025)。

(二)、 全員 AI 素養政策與 Pair 助理系統

2025 年起，全體公務員需接受 AI 素養課程，核心訓練包括：

1. ChatGPT 類生成模型的倫理應用；
2. 結構化 Prompt Engineering；
3. 數據偏見與合規性知識（Straits Times, 2025）。

此外，GovTech 自建 AI 助理「Pair」，已成為 15 萬名文官的日常工作輔助工具，推動整體效能提升。

四、 愛沙尼亞：法治與區塊鏈確保數位信任

(一)、 X-Road 與數位主權架構

愛沙尼亞成功打造 X-Road 數據交換系統，覆蓋 99% 的公共服務。其創新特徵：

1. 所有數據存取皆以區塊鏈日誌記錄（Log Chain），可追蹤與問責；
2. 實現「司法就緒數位基礎設施」(Justice-Ready Digital Infrastructure)(GGI, 2025)。

(二)、 數據大使館與 e-Residency 計畫

愛沙尼亞在國外部署「數據大使館」（Data Embassies），一旦本土網路遭攻擊，資料與服務仍可運作。另推行 e-Residency，讓全球公民透過數位身份參與商業與治理流程，強化全球數位連結力。此模式證明：信任不是抽象概念，而是制度設計結果。

五、 芬蘭：媒體識讀作為韌性國策

(一)、 從幼教到國安：全民媒體識讀教育

芬蘭將媒體識讀納入國家安全戰略，自幼稚園起推動以下核心素養：

1. 資訊查證技術；
2. 敘事識別與反操作技巧；
3. 對演算法推薦系統的批判理解（Nordic Policy Centre, 2024）。

媒體識讀課程不僅強化民主防禦，更促進世代間的數位韌性傳承。

(二)、 整合式媒體韌性平台

芬蘭建置全國假訊息追蹤系統，結合媒體、學界與國安部門，每日協作分析錯假訊息，迅速回應輿情風暴，維持資訊場域清明。

六、綜合比較：成功治理典範的共同特徵

承前所述，透過對台灣、新加坡、愛沙尼亞與芬蘭之案例檢證，可歸納出成功數位治理模式在素養強化與制度設計上之共通特徵（見表 7）。

表 7

成功數位治理典範共同特徵比較表

面向	台灣	新加坡	愛沙尼亞	芬蘭
數位能力架構	AITO 三層訓練架構	全體公務員 AI 課程	數位主權法制化	全民媒體識讀教育
敏捷實踐	vTaiwan、政策沙盒	Pair 助理、敏捷開發	X-Road 實驗與升級	假訊息反制快速回應系統
韌性類型	公民參與與信任韌性	效率與技術韌性	法治與系統性韌性	認知與文化韌性
示範性政策	AI 十項行動、沙盒立法	Smart Nation 戰略	e-Residency 、Data Embassy	多語敘事教育與認知防線

資料來源：本研究整理

本文案例顯示，成功的數位治理不是單靠技術導入，而是透過制度設計 + 素養強化 + 多層韌性融合。這些國家在制度韌性上呈現「跨技術－跨文化－跨法制」的整合能量。

特別是台灣與愛沙尼亞展現出「數位信任是可以建構的」治理實驗，對中小型國家尤具示範意義。下一章將深入探討系統性風險治理與未來治理挑戰。

伍、AI 安全、前沿模型與系統性風險治理

一、 從效率工具轉向治理風險

隨著 GPT-4、Claude 3、Gemini Ultra 等前沿模型（Frontier Models）快速部署於政府、公營機構與社會治理場域，AI 正從「輔助工具」轉變為「系統性風險製造者」與「決策行為代理者」（AI as Actor）。這種質變引發對 AI 所帶來治理脆弱性（Governance Vulnerability）的高度關注。

AI 治理的核心挑戰不僅是模型本身的不確定性，更在於決策制度對風險辨識與干預能力的不足。因此，本章著重於建構「高階公務人員—制度—社會」三軸向的風險治理模型。

二、 AI 系統性風險類型總覽

在探討風險治理機制之前，有必要先針對前沿模型部署於公共領域時，可能引發之系統性風險進行結構化分類。根據 Stanford HAI（2024）、Center for AI Safety（2023）與 OECD.AI（2023）整理，目前公部門常見的 AI 系統性風險可分為五類（見表 8）：

表 8

公部門常見之 AI 系統性風險類型表

風險類型	說明
演算法偏見	對弱勢群體的不公平決策，常見於福利分配、信用評估等領域
模型幻覺與虛構	尤指大型語言模型生成「合理但錯誤」的內容，如虛構引用、政策數據錯植
黑箱決策	系統無法解釋其輸出，導致問責機制與公民監督失效
自動化擴散失控	未經人為審查即部署政策行動，自動產生惡性循環
民主弱化風險	AI 操縱訊息流向與情緒擴散，加深社會極化與政治操控

資料來源：本研究整理

這些風險的發生，常伴隨高階決策者對 AI 的過度信任（automation bias）與缺乏風險評估素養，導致治理體系本身變得更脆弱。

三、 國際治理機制發展動向

(一)、 歐盟 AI 法案（EU AI Act）

2024 年通過的《EU AI 法案》為目前全球最具前瞻性與法治化的 AI 管理架構，其分類標準如下：

1. 不可接受風險（Prohibited AI）：如社會信用評等、實時情緒偵測；
2. 高風險系統（High-risk AI）：如教育、司法、就業與公共服務等；
3. 一般風險（Limited-risk）：需標示 AI 系統使用事實；
4. 最小風險（Minimal-risk）：如 AI 助理、聊天機器人等。

此架構強調「預防原則（Precautionary Principle）」與「系統監督（Systemic Oversight）」，要求：

1. 預先風險評估（Risk Assessment）
2. 強制資料治理標準
3. 可解釋性與可回溯性機制
4. 第三方外部審計（External Red Teaming）

(二)、 美國 AI 行政命令（Executive Order on Safe, Secure, and Trustworthy AI）

拜登政府於 2023 年頒布 AI 安全行政命令，核心在於：

1. 前沿模型的「紅隊測試」強制化（Red Teaming）；
2. 國安機構需建立跨部會風險監控體系；
3. 公務系統部署 AI 前須提出「使用影響報告（Impact Statement）」(White House, 2023)。

(三)、 G7「廣島 AI 原則」與 OECD.AI 框架

G7 於 2023 年推出「廣島 AI 原則」，主張：

1. 責任制（Accountability）
2. 人權導向（Human Rights-Based）
3. 安全預警（Safety by Design）

OECD.AI 平台則提供政府 AI 系統登錄與績效監測標準（OECD, 2023）。

四、 敏捷風險治理機制：沙盒、紅隊與韌性實驗場

高階公務人員需理解的不僅是 AI 工具，而是制度對風險的吸收、回應與適應能力。以下三項治理工具，是當代最重要的敏捷治理防線：

(一)、 政策沙盒（Regulatory Sandbox）

源自金融監理領域，近年擴展至數位治理與 AI 部署。特徵：

1. 小規模實驗 → 滾動調整 → 正式推行；
2. 開放公民與利害關係人參與；
3. 快速撤案機制。

台灣、英國與新加坡皆已在數據應用與智能系統導入過程中採用該制度。

(二)、 演算法紅隊（Algorithmic Red Teaming）

由美國國防部與 Anthropic、OpenAI 等合作推動，用以找出系統漏洞與意外行為。機制包含：

1. 有計畫地嘗試誘導錯誤輸出；
2. 測試多語種、邊緣案例與政策誤導性輸入；
3. 紀錄並回饋模型開發團隊與審計單位。

(三)、 風險韌性實驗場（Resilience Labs）

芬蘭與愛沙尼亞創設數位治理「實驗區」，模擬混合威脅情境，包括：

1. 假訊息爆發模擬；
2. AI 誤判政策評估；
3. 對數據失真做出政策反應時間測試。

五、 台灣的風險治理挑戰與轉型建議

(一)、 現況挑戰

對照國際間敏捷風險治理機制之發展趨勢，台灣當前於公務體系導入 AI 技術時，主要面臨以下四個面向之結構性挑戰與治理落差：

表 9

台灣公務體系導入 AI 之結構性挑戰

面向	問題描述
法制斷裂	數位與 AI 無專法、行政命令調適力不足
技術治理落差	地方與中央、部門間能力與資源高度不均
缺乏風險語言	高階官員無 AI 風險模型框架，決策語言貧乏
審計體制空白	無制度性紅隊或政策回饋迴路，反饋效能低

資料來源：本研究整理

(二)、 政策建議：建構三層次 AI 治理韌性系統

為具體回應上述之法制與技術治理落差，本文建議應從技術、制度與文化三大層次著手，建構一套具備反脆弱性之 AI 治理韌性系統（見表 9）：

表 9

建構 AI 治理韌性系統三層次架構

層次	核心機制	具體行動
技術層	模型可解釋性 / 演算法影響評估	建立 AI 審計單位、強制導入紅隊制度
制度層	法制與沙盒雙軌調整	設立 AI 專責機構與《AI 基本法》草案平台
文化層	高階素養 × 認知韌性	對高階文官推行 AI+認知安全模組訓練

資料來源：本研究整理

本文揭示，AI 的風險治理不只是科技問題，而是牽動整體行政制度設計、倫理監督與民主信任建構的關鍵挑戰。當前國際趨勢已從「技術中立」轉向「風險前提」，各國紛紛建立 AI 專法、審計制度與紅隊測試流程。

台灣若能善用敏捷治理工具、整合 AI 審計制度、並於高階文官中深化 AI 素養訓練，將有潛力打造亞洲首屈一指的「制度性 AI 韌性國家」。

陸、建構民主韌性與數位治理的戰略行動藍圖

一、 建構 AI 治理的三層次戰略架構

因應前述風險挑戰與治理落差，台灣應從「高階人力」的核心能力培育出發，向外延伸至制度韌性建構與法制調適，建立以下三層戰略架構：

層次一：個人素養強化（Cognitive Readiness）

首要之務在於個體層次，旨在透過核心職能與思維之重塑，全面強化各級公務人力面對複合風險之認知準備度（見表 10）：

表 10

AI 治理戰略層次一：個人認知準備度

行動目標	對象	策略與工具
強化 AI 素養	高階公務主管	高階 AI 執行課程（Decision-level AI Literacy）

媒體識讀即戰力	公部門全體人員	認知安全模組訓練（Cognitive Defense Modules）
敏捷思維內建決策習慣	各級文官	Scrum/Agile 公共治理適應訓練

資料來源：本研究整理

層次二：制度能力升級（Systemic Agility）

其次，於制度層次，為確保公務人員之素養得以發揮實質效用，必須同步推展政策監測機制與敏捷法制之結構性升級（見表 11）：

表 11

AI 治理戰略層次二：制度升級

行動目標	對應制度	策略與工具
設立 AI 政策監測機構	行政院層級的 AI 委員會	包含模型審查、資料治理標準、倫理風險回報
敏捷沙盒與紅隊法制化	數位部、金管會等主管機關	修法納入「政策演算法沙盒條款」、「演算法紅隊」
設計動態回饋政策機制	各部會政策迴圈	引入 Evidence-informed Agile Decision Loop 模型

資料來源：本研究整理

層次三：民主治理韌性（Governance Resilience）

最後，於宏觀之治理層次，為全面鞏固數位時代之公眾信任與民主防線，應積極推進跨域協作與社會整體認知防禦之重構（見表 12）：

表 12

AI 治理戰略層次三：民主治理韌性

行動目標	對象	策略與工具
建立「治理可信度指標系統」	國發會、審計部、監察院	結合 AI 使用公開性指標、政策透明指數、風險審核
公私協力治理模式再造	數位治理生態系（平台、媒體）	推動公共資料治理聯盟（Public Data Stewardship）
構建認知韌性公共衛生體系	教育部、文化部、國安單位	融合媒體素養、批判性思維與反資訊操弄教育課程

資料來源：本研究整理

二、 關鍵政策建議十項行動方案（Taiwan AI-10）

依據前述之三層次戰略架構，本文具體提出以下十項政策行動方案（Taiwan AI-10），作為建構台灣 AI 治理韌性之行動藍圖：

表 13
台灣 AI 治理韌性行動方案

編號	行動名稱	說明
A1	高階公務 AI 執行力課程	為政務官與一級主管設計「策略 + 風險 + 應用」三軸課程
A2	AI 素養納入考試與升遷制度	結合文官考選與職涯發展，提升整體結構性 AI 能力
A3	媒體識讀x認知韌性模組	建立面對假訊息與深偽影像的防衛教育體系
A4	各部會 AI 審計專責窗口	成立「AI 應用監測員制度」，促進問責、透明與事前風險管理
A5	建立敏捷沙盒 × 紅隊聯測制度	法制化紅隊、政策回測、倫理模擬場景，導入制度型實驗文化
A6	政策生命週期透明儀表板	發展決策可視化平台，促進政策過程中的公眾監督與動態修正
A7	公私合作治理框架（PDG）	推動「Public Data Governance」夥伴關係平台
A8	數據治理標準與資料影響評估	全國性導入資料影響評估（Data Impact Assessment）
A9	建立 AI 風險應變地圖	對高風險政策領域建置模型錯誤回應場景與備援機制
A10	台灣 AI 民主治理觀測計畫	與國際接軌，發布年度《AI 與民主韌性白皮書》

資料來源：本研究整理

三、 對未來公務體制的轉型建議

(一)、 對人事制度的影響

1. 跨域職系設計：應增設「科技治理官（GovTech Officer）」與「資料倫理官（Data Ethics Officer）」職系；

2. 彈性編制試點：針對 AI 專案引入契約型、任務導向型公務角色（如政策駭客、技術評估師）；

3. 留才育才制度優化：提升科技公務員待遇，並設「數位留才基金」。

(二)、 對組織文化的要求

1. 從「風險避免文化」轉向「風險管理與學習文化」；
2. 鼓勵「快速實驗—失敗容許—快速迭代」組織流程；
3. 建立跨部會政策共享平台，促進制度記憶與知識傳承。

本文提出的政策建議體系，不僅針對 AI 技術應用，也直指當前文官體制與數位治理制度上的結構性挑戰。核心精神在於：素養是治理的起點，但制度才是韌性的底盤。

唯有將 AI 素養內化為文官的「治理語言」，並以敏捷法制與民主原則設計制度性防火牆，台灣方能於變動中的 AI 治理新世紀，維持其民主典範地位與治理信任資本。

柒、結論與建議

一、從風險到韌性的治理轉向

本文聚焦於「高階公務人力的 AI 素養與媒體識讀對民主韌性與數位治理的作用」，以循證途徑與敏捷理論為分析工具，從多國案例中提煉風險病理與成功典範，逐步構築台灣未來治理架構的藍圖。研究顯示：

(一)、 AI 技術的快速滲透，使高階文官面對治理風險之複雜度與速度顯著提升；

(二)、 缺乏 AI 素養與媒體識讀的治理系統，將傾向產生黑箱決策、歧視性輸出與民主脆弱化；

(三)、 成功治理案例（如台灣、新加坡、愛沙尼亞、芬蘭）皆展示出素養建設 × 敏捷制度 × 透明參與之交織設計；

(四)、 本土制度建議聚焦於三層韌性建構：個人素養 → 制度敏捷 → 治理正當性。

在這些基礎上，本文試圖構築出一個橫跨政策層級、風險類型與制度工具的整合型分析框架，有助於台灣未來面對更高階的 AI 模型與治理挑戰。

二、 研究建議

(一)、 理論方面

1. 敏捷 × 循證雙軌模型：融合 Agile Governance 與 Evidence-based Policymaking，創建「敏捷循證治理架構」。

2. AI 素養制度化模型：建構出以文官訓練為主軸、制度支撐為核心的 AI 素養推進機制。

3. 民主韌性新概念拓展：將「媒體識讀」「資訊操弄防禦」「技術透明度」納入民主韌性的新定義中。

(二)、 實務方面

1. 政策沙盒實作建議：建議各部會於導入自動化決策前，應建立具備反脆弱性之技術檢驗流程。具體而言，應提出 AI 紅隊測試、政策快速撤案機制與風險模擬實驗場等設計要點，藉由小規模實驗與滾動式評估，於前端阻絕系統性風險之擴散。

2. 公務人力培訓規劃：明確區分策略性 AI 能力、執行性操作技能與認知安全教育模組。在具體執行層面，建議優化現行國家文官學院「高階文官培訓飛躍方案」（Leadership Excellence Program）。目前該方案主要依據目標職務所需職能，規劃有「領導力」、「全球治理」、「公共政策」及「倫理價值與人文素養」四大核心模組課程外，應評估增設「媒體識讀對民主韌性與數位治理能力」專題訓練模組。此舉旨在將認知防衛之學理概念轉化為行政實踐，確保決策菁英具備維繫民主正當性之核心素養。

3. 制度改革藍圖：應從法制、組織、人事三面向推展可行之改革方案。例如推動「政策演算法沙盒」法制化、設立各部會 AI 審計專責窗口，並於人事體系中增設跨域之科技治理職系。透過上述結構性調整，方能實質促進跨部門協作整合，全面帶動國家數位治理韌性之升級。

(三)、 方法論建議

1. 本文採用「政策病理學 + 成功典範對照」的雙向邏輯，提供具行動力與預警力的比較式政策研究模型；

2. 引入「制度風險推理 + 反事實情境模擬」，有助於政策預測與制度重建能力強化。

三、 研究限制與未來方向

(一)、 研究限制

1. 本文以質性資料與案例研究為主，缺乏大型量化問卷或機器學習模擬資料支持；

2. 所蒐集之成功典範多來自政治穩定或治理集中型國家，對開放社會與地方自治型體系的應用仍需評估。

(二)、 未來研究方向

1. 建立 AI 治理成熟度模型指標，針對各級政府進行實證追蹤；

2. 探討 AI 與公共價值的關係，如公平、正義、包容性政策的演算法化；

3. 模擬深偽技術對選舉與政治信任的長期影響；

4. 推動跨國比較研究，將台灣經驗與歐盟、OECD 等治理模式進行對照研究。

未來十年，AI 將不僅重塑產業與勞動市場，更將徹底改寫民主治理的邏輯基底。面對來自深偽影像、演算法操弄與模型偏見的複合風險，我們必須重新回答：「誰來監督 AI？誰來治理治理者？」

答案不在單一法律或單一技術，而在於建構一套由人—制度—價值三者互構的治理生態系統。在此生態中，高階公務人力不再只是政策執行者，更是 AI 與民主之間的橋梁與守門人。

參考文獻

Amnesty International. (2021). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal*. https://www.amnesty.org/en/documents/eur35/3519/2021/en/

Bipartisan Policy Center. (2023). *Bipartisan Policy Center's exclusive toolkit for National Conference of State Legislators 2023*. https://bipartisanpolicy.org/bpc-ncsl-2023/

Center for AI Safety. (2023). *Statement on AI risk*. https://www.safe.ai/statement-on-ai-risk

Deloitte. (2024). *Government trends 2024: Government's newfound agility*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/government-public-sector-services/government-trends/2024/agile-government-is-imperative-for-public-sector.html

Democracy Technologies. (2025). *Taiwan: A regional laboratory of open governance*. https://www.democracy-technologies.org/

European Commission. (2024). *The EU Artificial Intelligence Act: Final legislative text*. https://digital-strategy.ec.europa.eu

Executive Yuan. (2025). *2025 台灣高階公務 AI 人才培育專案白皮書*。行政院人工智慧推動辦公室。

FNF Taiwan. (2025). *From vTaiwan to agile democracy: Civic tech and digital collaboration*. Friedrich Naumann Foundation.

GGI – Global Government Innovation. (2025). *Estonia's X-Road and digital sovereignty explained*. https://www.globalgovinnovation.org

McKinsey & Company. (2025). *Agile government: Lessons from Singapore's digital transformation*. https://www.mckinsey.com

Nordic Policy Centre. (2024). *Media literacy as cognitive defense: Finland's approach to hybrid threats*. https://www.nordicpolicy.org

OECD. (2023). *OECD framework for the classification of AI systems and AI incidents*. https://www.oecd.ai

PMC (Parliamentary Monitoring Centre). (2024). *A-level algorithm backlash: Policy failure or democratic deficit?* https://www.pmc.org.uk

Royal Commission into the Robodebt Scheme. (2023). *Final report: Findings and recommendations*. Government of Australia. https://robodebt.royalcommission.gov.au

Straits Times. (2025). *GovTech's 'Pair' assistant reaches all 150,000 civil servants*. https://www.straitstimes.com

UNESCO. (2022). *Artificial intelligence and digital transformation: Competencies for civil servants*. https://unesdoc.unesco.org/ark:/48223/pf0000384963

UNESCO. (2023). *Ethical impact assessment: A tool of the Recommendation on the Ethics of Artificial Intelligence*. https://doi.org/10.54678/YTSA7796

United Nations. (2024). *United Nations e-government survey 2024: Accelerating digital transformation for sustainable development*. Department of Economic and Social Affairs. https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2024

V-Dem Institute. (2025). *Democracy report 2025: 25 years of autocratization – Democracy trumped?* University of Gothenburg.

White House. (2023). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*. https://www.whitehouse.gov/briefing-room/

Democratic Resilience and Digital Governance: The Role of AI and Media Literacy Among Senior Civil Servants

Jiang Min Qin*

Abstract

As generative artificial intelligence (Generative AI) rapidly advances, public sector governance faces unprecedented opportunities and systemic risks. Senior civil servants, as the critical actors in policy design and risk response, must develop AI literacy and media competence to sustain democratic resilience and digital legitimacy. This study adopts a dual-framework approach, combining Evidence-Based Policymaking and Agile Governance theories. It uses a comparative model of "policy pathology" versus "successful exemplars," analyzing cases from Taiwan, Singapore, Estonia, and Finland. Based on this, the study proposes a three-tier resilience architecture and a Ten-Action Policy Package for Taiwan (Taiwan AI-10). The findings suggest that institutionalizing AI oversight, elevating executive-level AI training, and embedding agility into governance processes are essential to safeguarding democratic systems in the AI era.

Keywords: AI literacy, senior civil servants, agile governance, media literacy, democratic resilience, systemic risk, policy sandbox

* Visiting Professor and Dean of the School of Management, Shih Hsin University.
The paper was published under two double-blind reviews.
Received: March 19, 2026. Accepted: April 23, 2026.