National Chengchi University
國立政治大學

moda 數位發展部
Ministry of Digital Affairs

公部門人工智慧規範架構

計畫主持人：曾憲立 / 臺南大學行政管理學系 副教授

協同主持人：朱斌妤 / 政治大學公共行政學系 特聘教授

戴豪君 / 世新大學法律學院法律學系 副教授

許慧瑩 / 東吳大學法律學系 助理教授

顧　　問：蕭乃沂/ 政治大學公共行政學系 副教授

National Chengchi University
國立政治大學

moda 數位發展部
Ministry of Digital Affairs

目錄
Content

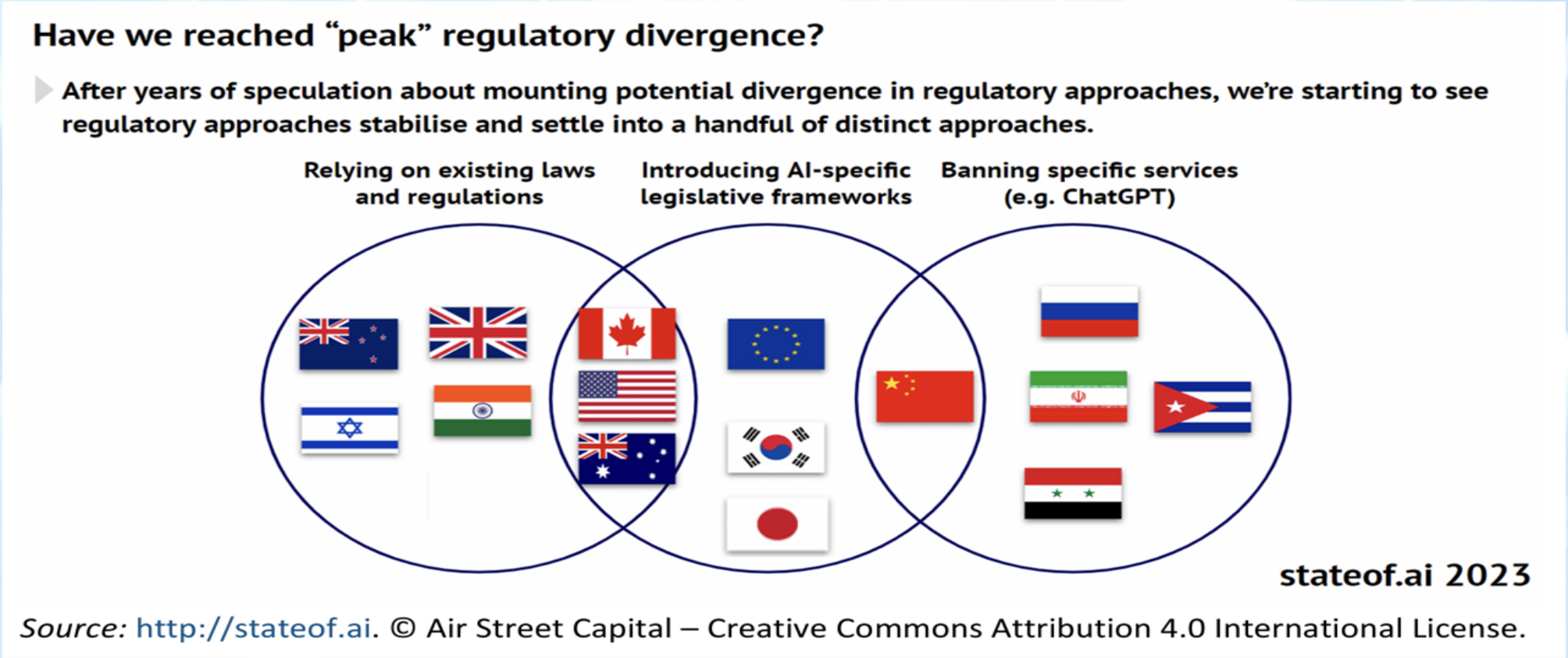
01

研究背景

- 研究背景及目標
- 文獻回顧

研究背景

- 隨AI技術的快速進步，世界各國均加速制定相關法規或行政命令，以因應人工智慧帶來資安防護、民眾隱私保護、數位權利、公部門數位轉型，與「不遺落任何人」的數位涵容社會。



研究背景 - 人工智慧在政府中的職能概述

治理功能	AI潛在用途
政策制定 （policy making）	1. 更快發現社會問題 2. 改善公共政策決策（並預估政策的潛在影響） 3. 監督政策的實施（並評估現有政策） 4. 加強公民對政策制定的參與
提供公共服務 （public services）	1. 改善組織的資訊服務 2. 改善向企業和民眾提供的公共服務 3. 發展創新的公共服務
內部管理 （internal management）	1. 改善人力資源配置 2. 改善公共組織的招募服務 3. 改善組織的財務管理 4. 改善詐欺和貪污的偵查 5. 改善維護 6. 改善公共採購流程 7. 改善組織網路安全

資料來源：van Noordt & Misuraca（2022）、朱斌妤（2023）

研究背景

- 人工智慧規範架構（AI Governance Framework）是一套指導原則和標準，用於規範人工智慧技術的開發、部署和使用，確保其符合倫理、安全、透明和責任等多方面的要求。

比較項目	AI規範架構 AI Governance Framework	AI使用指引 AI Usage Guidelines
層級和廣度	是宏觀層面的，涵蓋AI技術的整體治理和策略指導。	是微觀層面的，針對特定操作和應用場景提供具體的實踐建議。
目標和應用對象	為政策制定者提供高層次的數位治理指導。	為具體的AI技術用戶和操作人員提供具體的操作指南。
內容和詳細程度	內容廣泛，涵蓋倫理、透明性、問責、安全等多方面。	內容具體，側重於實際操作步驟和具體的合規要求。

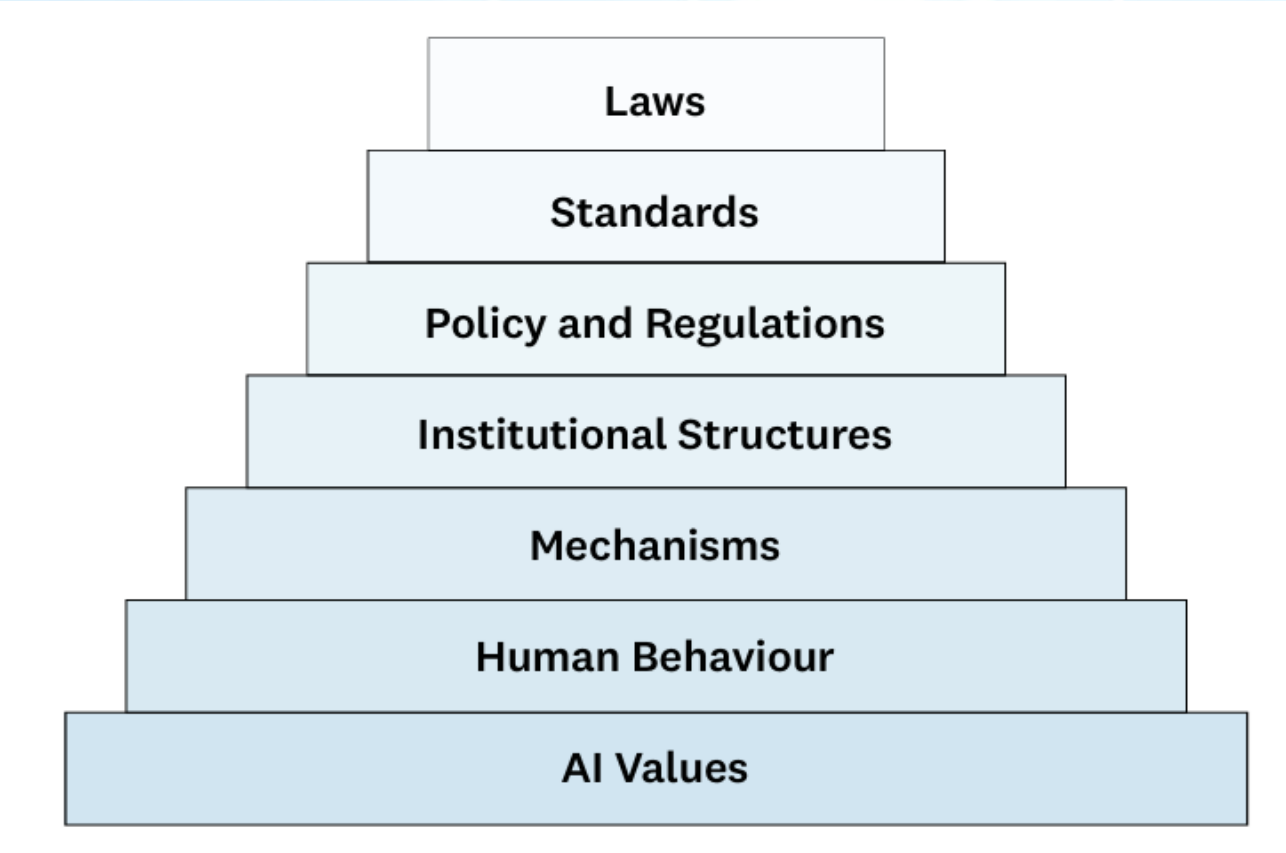
研究目標

- 研擬我國適用的AI規範架構，並依照我國國情篩選適用部份，做為我國政府運用AI時的高層次、整體性指導架構，以供各行政院所屬單位參考。
- 研究內容排除以下適用領域：軍事、國防、武器、研發、金融、醫療、媒體和平台內容監管。

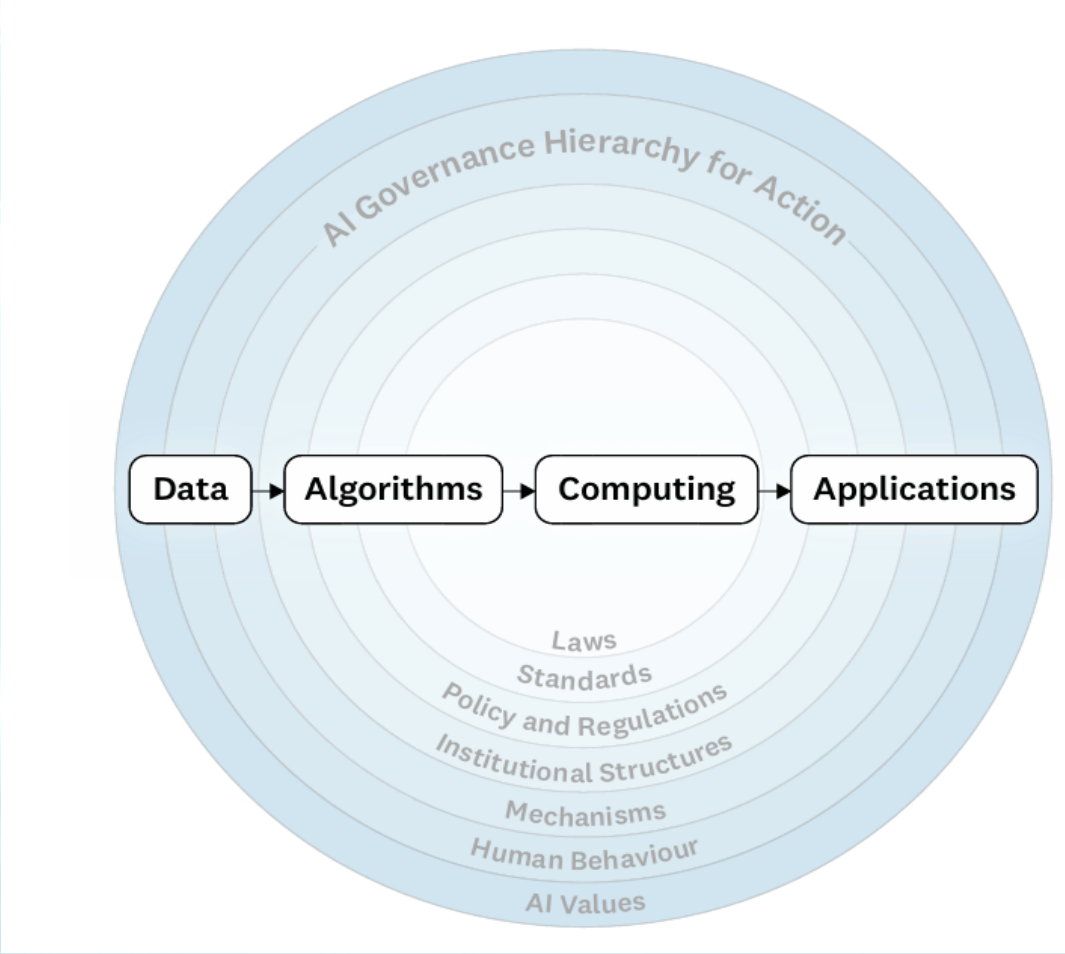
人工智慧治理-主要參考文獻

- **歐盟、美國、英國、澳洲**
- **國際組織/研究單位**
  - 《Governing AI for Humanity》（UN, 2024）
  - 《The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence》(MIT, 2024)
  - 政府AI應用五大分類及其應用(Engstrom *et al.*, 2020)
  - 經濟合作暨發展組織（OECD）
  - 七大工業國組織（G7）

《人工智慧治理框架》(UNU, 2024)



AI行動治理層級



AI治理模型

歐洲理事會（CoE）
人工智慧、人權、民主和法治架構公約

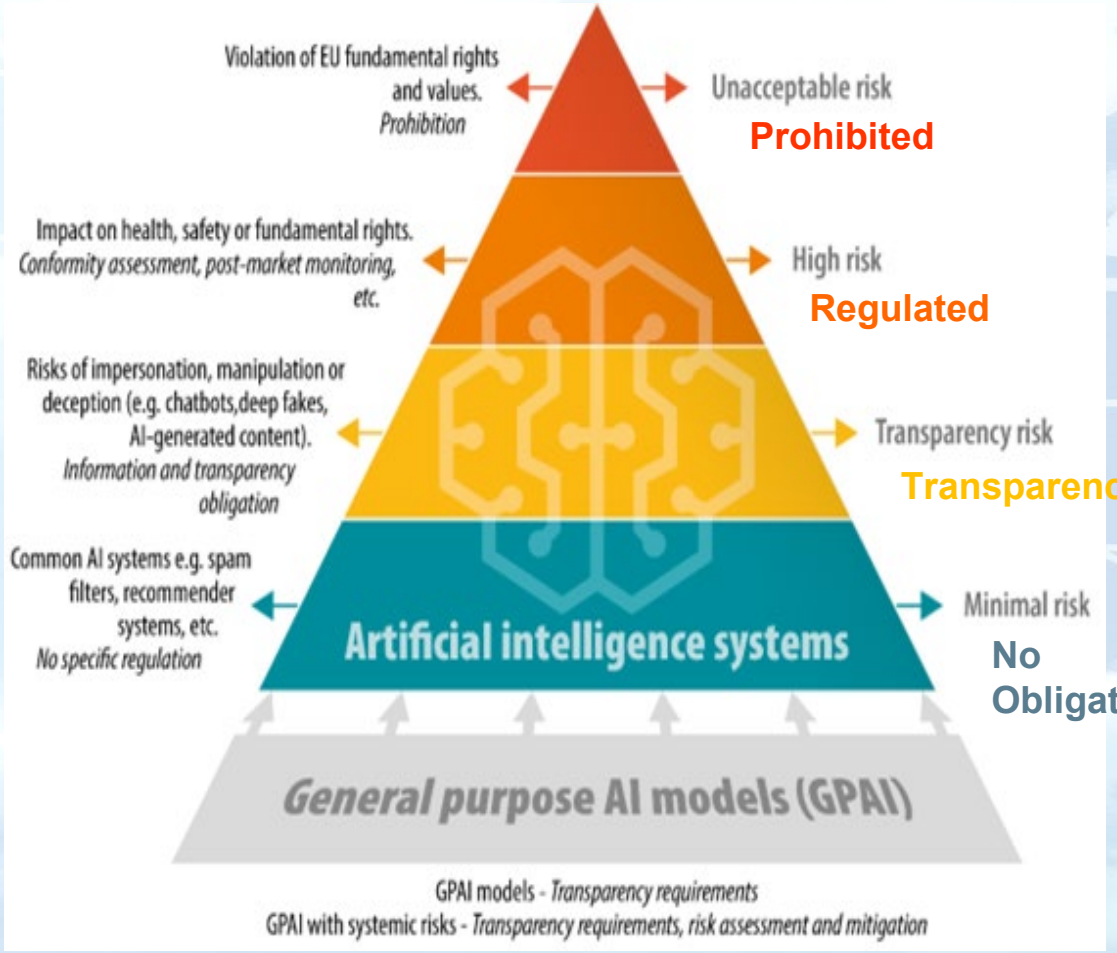
- 歐洲理事會（CoE）2023年12月18日公布關於人工智慧、人權、民主和法治架構公約草案（Draft Framework Convention on AI, Human Rights, Democracy, and Rule of Law），簡稱歐洲理事會人工智慧公約草案
- 2024年5月17日，歐洲理事會正式通過人工智慧公約，是第一個對簽署國具有法律約束力的人工智慧國際條約。 該公約將於 2024 年 9 月 5 日開放簽署。
- 主要目的
  - 確保AI系統的使用，須充分遵守人權、尊重民主運作並遵守法治，以維護歐洲人權基本價值
- 框架公約
  - 提供締約國一般性義務和原則，將具體細節留給締約國，透過國內立法的方式做進一步的補充

歐洲理事會（CoE）
人工智慧、人權、民主和法治架構公約

https://www.holisticai.com/blog/europe-committee-artificial-intelligence-draft-framework-convention

EU - 人工智慧法 (AI Act)

- 為全球第一部規範人工智慧的法律
- 透過透明、課責與人類監督作為實現「歐洲數位權利和原則」的實例
- 以用途可能涉及的風險層級作為監管分級的理論基礎



資料來源：European Parliament, Artificial intelligence act Briefing, https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

EU - 人工智慧法 (AI Act)

- 依風險檢視技術應用之合法性、必要性，及與風險層級相對應之因應措施
- 提供者與部署者對應前述要求的義務，提供者與部署者還要在系統採用後持續進行上市後的監督，分享特殊嚴重事件的資訊
  - 提供者(provider):是指自然人或法人、公共機關、機構或其他團體，以有償或無償方式，開發人工智慧系統或通用人工智慧模型，或已開發人工智慧系統或通用人工智慧模型，並將其投放到市場（places it on the market），或以自己的名義或商標提供人工智慧系統服務（AIA Art.3 (3)）
  - 部署者（deployer）：是指自然人或法人、公共機關、機構或其他團體，在其授權下使用人工智慧系統，但個人非職業活動（a personal non-professional activity）中使用人工智慧系統的情況除外（AIA Art.3 (4)）

EU - 人工智慧法 (AI Act)

- **部署者人工智慧系統的基本權利影響評估(AIA Art. 27)**
- 部署除附件III第2點用於關鍵基礎設施人工智慧系統之外，受公法管轄機關構或提供公共服務的私人經營者（deployers that are bodies governed by public law, or are private entities providing public services），以及提供附件III第5點(b) (c)之評估自然人信用評等AI系統；人壽保險和健康保險自然人風險評估和定價AI系統之部署者。應進行使用該高風險AI系統基本權利衝擊影響評估（ Fundamental rights impact assessment for high-risk AI systems ）
- 評估包括說明部署者按照預期目的使用高風險人工智慧系統的過程、使用期限和頻率，以及可能受其使用影響的自然人和群體的類別，與可能受到的具體的損害風險；說明人類監督措施的執行；風險因應措施，包括內部管理和投訴機制的安排

EU - 人工智慧法 (AI Act)

- 要求公務機關（排除司法調查與犯罪偵查、邊境控管、移民難民或其他機關）使用高風險的AI系統，應先行至歐盟資料庫登錄

  - 風險控管制度的建置與落實
  - 訓練、驗證與測試資料與資料治理
  - 系統上市或使用前技術文件之整備
  - 系統運行紀錄之自動保存
  - 系統資訊之透明與提供
  - 人類行為的監督與介入
  - 系統運行的準確性、穩定性及資訊安全

美國 – AI監管的拉扯

- 拜登政府對AI治理呼應歐盟AIA採取嚴格的監管模式
  - 依循《推進美國AI法》與《第14110號安全可靠且值得信賴的人工智慧開發暨使用行政命令》，主要規制人工智慧的安全、可靠與可信的開發和使用，透過該行政命令推動AI技術之研發與應用，以因應新型態科技發展趨勢及降低AI對勞工、消費者、少數族群及國家安全造成之風險，進而確保AI技術及應用能促進人類福祉，同時避免公眾蒙受潛在威脅
  - 聯邦行政機關須遵循管理和預算辦公室（OMB）第M-24-10號《推進機構人工智慧使用的治理、創新和風險管理》備忘錄
    - 各機構採用AI提供更好的公共服務，同時對權利、公民自由和隱私的保護
    - 為機構制定新的要求和指導，包括建立切實有效的跨功能及跨機構協作以應對新的AI責任、管理AI風險與績效，以及通過創新採購促進具有競爭力的AI市場

美國 - AI監管的拉扯

- 2025年因政權移轉川普政府對AI治理調整為開放
  - 1月頒布Executive Order 14179《移除美國AI發展障礙行政命令》取消拜登政府時期針對AI技術的監管措施，強調AI應不受任何阻礙地推動政府效率與經濟成長
  - 於2025年4月頒布對AI採取輕度管理政策，支持與涵容創新，促進快速且負責任的AI使用、推動有效且高效的人工智慧收購，與應用AI為美國運作

美國 - AI監管的拉扯

- 政策須遵循管理和預算辦公室第M-25-21號《透過創新、治理和公眾信任加速聯邦政府對人工智慧的使用》備忘錄 (取代M-24-10)
  - 指示機構採用AI提供更好的公共服務，並保護權利、公民自由和隱私（以下節錄）
    - 各機構將釋出人工智慧採用成熟度評估，以更好地追蹤進展與需求
    - 引入單一「高影響力人工智慧」類別，用來追蹤需要進行更高程度審慎調查的人工智慧應用案例
    - 採用與創新提升為優先事項，同時增加社會和產業的透明度
    - 人工智慧的問責將與現行的政府IT使用過程相似，不創建新的批准層級
    - 最大化使用美國的人工智慧
    - 認可競爭的重要性，避免供應商依賴鎖定（vendor lock-in）
    - 將減少繁重的機構報告要求，完善採購流程，同時保護隱私並確保政府資料的合法使用
    - 對「高影響力人工智慧」實施最低風險管理措施
    - 確保其人工智慧的使用符合美國人民的利益

英國 - AI監管之嘗試

- 保守黨嘗試以法律規範AI但未有結果
  - 2023年12月，Lord Chris Holmes提出的私人成員法案(Private member's bill)《人工智慧（監管）法案》進入下議院，在2024年5月因議會休會時未能繼續推進
  - 2024年9月新的法案提出，旨在規範在公共部門中使用AI系統進行「決策」過程，最終未有結果

- 保守黨對AI提出白皮書採取嚴格規範因政權轉換而未能接續
  - 《白皮書：支持創新的AI監管方法》提出5項關鍵人工智慧監管原則：1.安全性、穩定性和可靠性、2.透明度、3.公平性、4.責任和治理、5.可爭議性和救濟

英國 - AI監管之嘗試

- 2025年巴黎高峰會展現對AI監管採取觀望態度
- 2025年再度提出《人工智慧(規範)法案》但卻不被看好
  - 延伸監理還是協調輕度監管AI (政府支持創新)
    - 建立集中人工智慧監管機構
    - 延續5項關鍵的人工智慧監管原則
  - 從透明公開公眾參與
    - 首次對人工智慧開發者施加法律義務
    - 需要就人工智慧風險進行公共諮詢，並要求對第三方資料使用的透明度，特別是使用人工智慧訓練資料集時需獲得知情同意

澳洲 - AI治理指引或原則

- 未頒布直接規範AI的具體法律，僅提供自願遵循之指引或原則
  - 2019 年發布的《人工智慧倫理原則》
    - 針對人工智慧負責任的設計、開發和實施的八項原則
  - 2024年8月發布《自願性AI安全標準》（ Voluntary AI Safety Standard ）
    - 使用風險矩陣來確定風險的嚴重性

		Consequence				
		Insignificant	Minor	Moderate	Major	Severe
Likelihood	Almost certain	Medium	Medium	High	High	High
	Likely	Medium	Medium	Medium	High	High
	Possible	Low	Medium	Medium	High	High
	Unlikely	Low	Low	Medium	Medium	High
	Rare	Low	Low	Low	Medium	Medium

澳洲 - AI治理指引或原則

- 2024年9月發布因應高風險AI使用之強制性防護措施建議（Proposals paper for introducing mandatory guardrails for AI in high-risk settings）分為兩類：
  - 1. 人工智慧系統或通用人工智慧(General purpose AI, GPAI)模型之建議用途是否為已知或可得預見之高風險
  - 2. 高風險AI為適用於高階、能力強，存有無法預見的用途與風險的GPAI模型

澳洲 - AI治理指引或原則

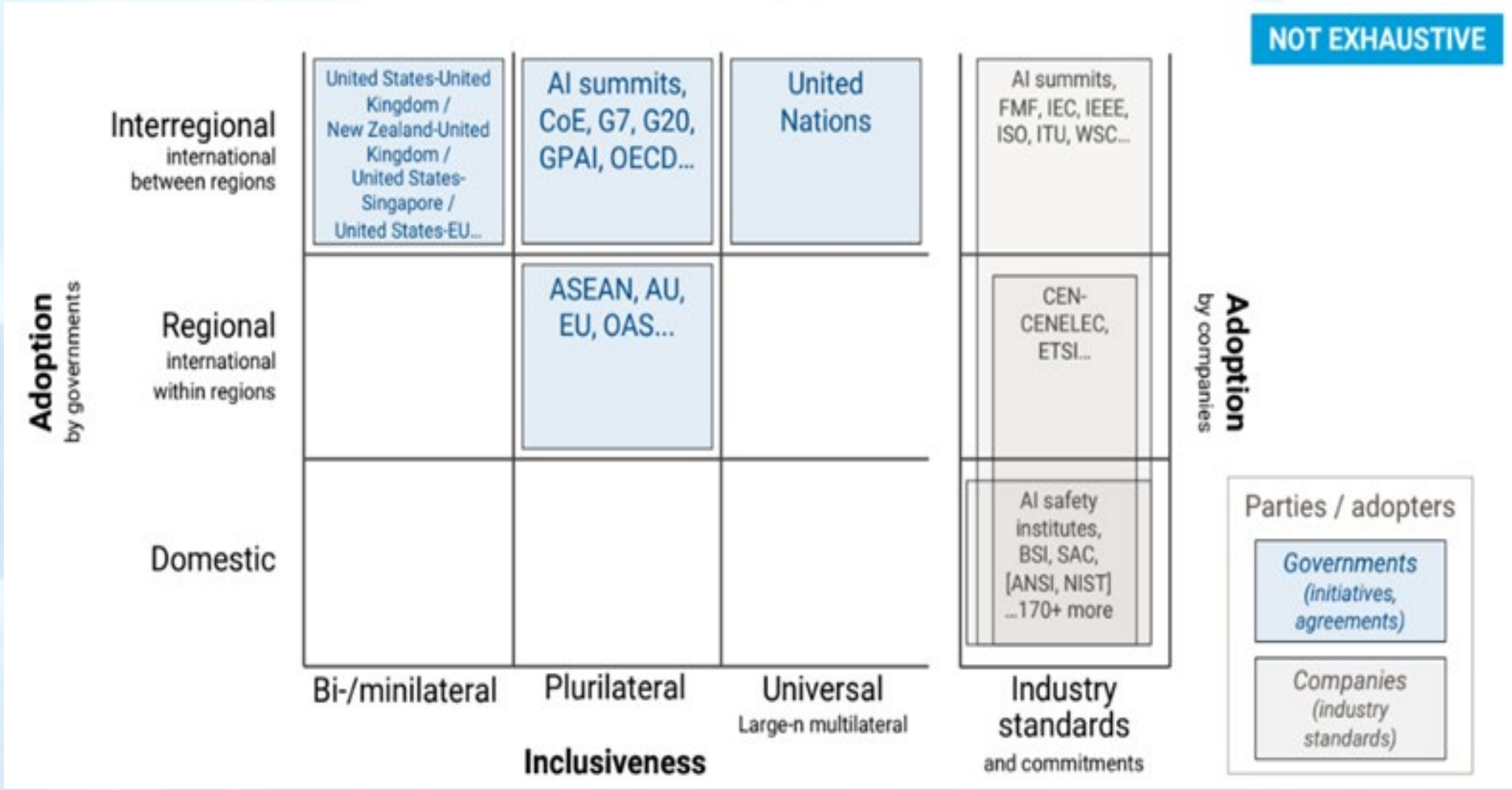
- 確保人工智慧測試、透明性與問責性的10項防護措施，涵蓋與其他組織的透明度、問責流程和人工智慧風險管理等，為組織提供實作指引，以便在利用人工智慧優勢的同時降低風險
  - 1. 建立、實施並公開問責流程
  - 2.建立並實施風險管理流程，以識別和減輕風險
  - 3.保護AI系統，並實施資料治理措施來管理資料品質和來源
  - 4.測試 AI 模型和系統以評估模型性能，並在部署後持續監控系統
  - 5.使 AI 系統能夠進行人工控制或干預，以實現有效的人類監督
  - (略)

UN - 國際人工智慧治理的指導原則和功能

- 原則 1：AI應該以包容的方式治理，為所有人的利益服務。
- 原則 2：AI必須在公共利益中進行治理。
- 原則 3：AI治理應該與資料治理及促進資料共同體相協調。
- 原則 4：AI治理必須是普遍的、網路化的，並根植於靈活的多方利害關係人合作中。
- 原則 5：AI治理應基於聯合國憲章、國際人權法和其他已商定的國際承諾，如永續發展目標（SDGs）。

資料來源:《Governing AI for Humanity》（UN, 2024）

UN - AI治理倡議來源與層級



資料來源:《Governing AI for Humanity》（UN, 2024）

UN – 風險分類

類別	風險類型
個人	人類尊嚴、價值或自主權（操控、欺騙等）、身體和心理完整性（健康、安全等）、生活機會（教育、就業、住房）、人權和公民自由（無罪推定權、公正審判權等）
政治與社會	歧視和不公平對待、不同身分的影響（兒童、老年人、殘障人士等）、國際和國內安全（自主武器、針對移民的警務等）、民主（選舉和信任）、資訊完整性（虛假訊息等）、法治（機構信任等）、文化多樣性和人際關係變化、社會凝聚力、價值觀和規範
經濟	權力集中、技術依賴、不平等的經濟機會、AI的過度使用、金融系統和關鍵基礎設施的穩定性、智慧財產權
環境	過度消耗能源、水和材料資源（包括稀有礦物和其他自然資源）

政府AI應用五大分類 (Engstrom *et al.*, 2020)

AI應用類型	應用類型	描述
執行 （enforcement）	1. 智慧識別流程	可辨識影像、影片、音訊或其他可偵測物理現象中的物體、人、地點、文字、情境和動作的過程。
	2. 審計和日誌記錄管理	記錄蒐集目的和來源，提供隨時影響特定操作、程序、事件或設備的活動序列的書面證據。
	3. 預測性執法流程	分析可用於預測和幫助預防未來潛在犯罪/錯誤/誤解的大量資訊的過程。
	4. 支援檢查流程	在主管機關介入之前用於識別不當行為或錯誤的支援流程（如檢查稅務狀況、註冊異常企業）。
管制分析與監控 （regulatory, analysis, and monitoring）	1. 資訊分析流程	資訊和數據分析是檢查、轉換和建模資訊過程。透過將資料轉換為可操作的知識（如儀表板）。
	2. 監控政策執行狀況	追蹤和評估政策實施的流程，以確保政策得到制定、認可和實施。
	3. 預測和規劃	基於預測模型的資源管理流程，以支援規劃。
裁決 （adjudication）	1. 就福利做出決定	用於做出有關批准、驗證或撤銷利益決策流程。
公共服務和參與 （public services and engagement）	1. 參與管理	加強與公民和企業的聯繫，建立信任關係。
	2. 資料分享管理	資料共享流程，考慮互通性和資料許可。
	3. 服務整合	整合管理多個服務供應商和資訊來源，以便為公民或其他組織提供新的客製化的特定服務。
	4. 服務個性化	考慮客戶（公民/企業/公務員）的需求提供客製化服務。
內部管理	1. 內部基本流程	為外部顧客創造價值的過程及對顧客（公民、企業）滿意度的影響。

02

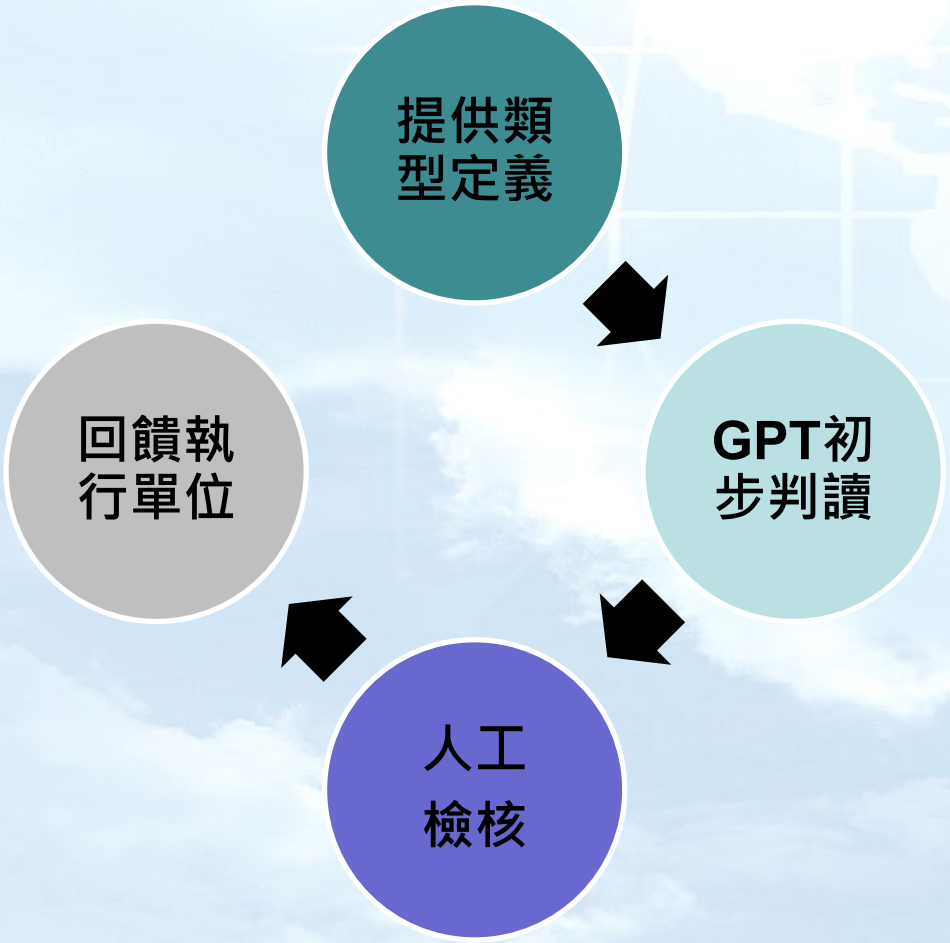
調查分析

- 訪談
- 次級資料分析

我國政府AI應用類型與場景

- 次級資料分析
  - DIGI+ 2023 – 40案
  - 服務品質獎(第1-6屆) – 32案
  - 台北市政府 – 28案
  - 內政部 – 90案
- 專家訪談
  - 資安專家、內政部、環境部、台北市政府、會計稽核

政府AI應用五大分類 – 人機協作方式



請你當我的助理，依我給你的這些資訊，結合你在網路上蒐集到的資訊，將我以下貼給你的案例，按照我說的指示，給我回應 (1)先初步判斷，該案例屬於五大類AI分類的哪一類(在AI應用盤點的文件中)；(2)將計畫置於該類後，摘錄重點，並告訴我該案例是用哪種AI技術；(3)依照報告內容文字描述，你覺得會有紅色標題的風險，請打勾，可以複選 ，紅色標題的風險為資料治理、人權保護、倫理、模型可解釋性、個人資料以及自動化決策風險六種。

政府AI應用五大分類 - DIGI+ (節錄)

AI應用類型	應用類型	案例名稱	使用AI技術	資料治理	人權保護	倫理	模型可解釋性	個人資料	自動化決策風險
執行	智慧識別流程	5G智慧警察行動服務研析 【內政部警政署】	影像辨識	V	V	V	V	V	
		AR頭盔前進搜救現場第一線，結合AI提昇救援效率 【內政部消防署】	影像辨識	V	V		V	V	
		建置公司登記文件影像自動分類，節省分類建檔成本 【經濟部商業發展署】	影像辨識、機器學習	V			V	V	
		車牌辨識結合AI，智慧勾稽異常清運行為 【環境部環境管理署】	影像辨識	V	V	V	V	V	V
		港區及聯外道路車牌辨識系統、船舶軌跡航行監控分析及新世代海巡偵防業務整合系統 【海洋委員會海巡署偵防分署】	影像辨識	V	V	V	V	V	V

政府AI應用五大分類 - 服務獎(節錄)

AI應用類型	應用類型	案例名稱	使用AI技術	資料治理	人權保護	倫理	模型可解釋性	個人資料	自動化決策風險
管制分析與監控	資訊分析流程	台北市政府工務局汙水管AI檢視	影像辨識	V			V		V
		台中市政府建設局自來水管「科技檢漏」	機器學習、動態影像分析辨識演算法	V			V		V
		基隆市消防局救災救護新應用「智慧消防2.0」	影像分析	V	V			V	V

政府AI應用五大分類 - 台北市政府(節錄)

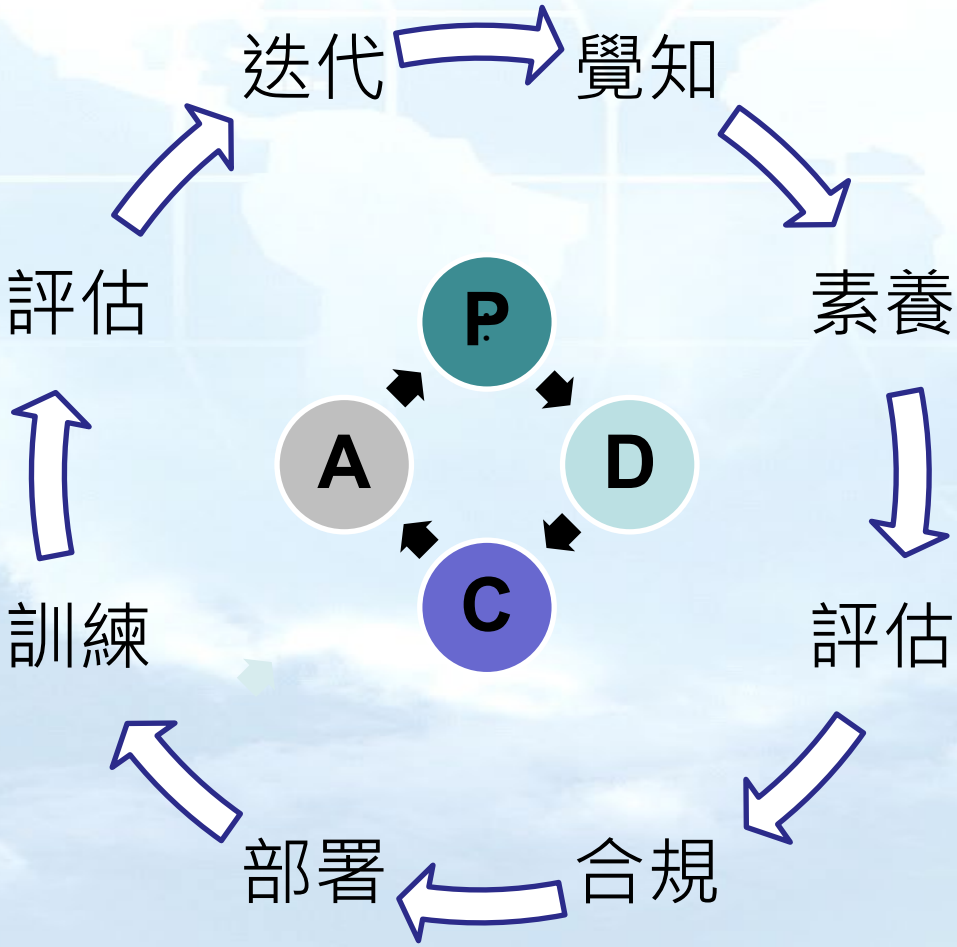
AI應用類型	應用類型	案例名稱	使用AI技術	資料治理	人權保護	倫理	模型可解釋性	個人資料	自動化決策風險
內部管理	內部管理流程	建置災害應變雲端協作平臺【台北市政府消防局】	機器學習、數據整合與視覺化	V			V		V
		決策輔助系統【臺北市立聯合醫院】	大數據分析	V	V	V	V	V	V
		護理資訊系統【台北市立聯合醫院】	大數據分析、ML、NLP	V	V	V	V	V	V

03

結論與建議

研究結論與建議

1.根據評估結果調整治理框架
2.持續更新風險管理措施
3.優化透明度與可解釋性實踐
4.融入政府服務流程並持續更新
5.持續進行人才培育與意識提升



1.確立 AI 治理架構與原則
2.建立風險管理機制
3.定義責任與課責機制
4.重視資料治理規劃
5.規劃透明度與可解釋性措施
6.規劃人才培育與意識提升
7.規劃負責任的 AI 創新與採購

1.監控 AI 系統的運行與影響
2.監控 AI 使用案例
3.評估風險管理措施的有效性
4.定期審視內部 AI 政策及治理方法
5.評估員工 AI 素養

1.部署 AI 治理與管理機制
2.實施風險緩解措施
3.執行透明公開
4.指定負責官員 (AOs)
5.促進跨職能與跨機構協作
6.建立內部註冊系統

其他關於治理的議題 1

- 政府單位間的6個落差
  - 算力落差（Computational Power Gap）
  - 資料資源落差（Data Resource Gap）
  - AI模型落差（AI Model Gap）
  - 人才落差（Talent Gap）
  - 法規與政策落差（Regulatory Gap）
  - 預算落差（Budget Gap）

其他關於治理的議題 2

- 監管、創新、服務面向的議題
  - 人工智慧素養
  - 主權AI
  - 假AI瓜分政府資源、真AI逃避政府監管
  - 大公司智財佈局阻礙創新
  - AI服務納入共同供應契約

其他關於治理的議題 3

- 創新實驗環境：人工智慧法草案第5條：為促人工智慧技術創新與永續發展，各目的事業主管機關得針對人工智慧創新產品或服務，建立或完備既有人工智慧研發與應用服務之創新實驗環境
- 我國目前依特定產業設計之無人載具科技創新實驗條例、金融科技發展與創新實驗條例等機制，如何面對各行各業適用AI之創新實驗需求
- 2024年12月歐盟啟動七項人工智慧工廠（AI Factories）計畫
  - 歐盟人工智慧法監理沙盒（AI regulatory sandbox）：是指由主管機關建立具體和受控的架構，為人工智慧系統的提供者或潛在提供者提供在監管監督下，依沙盒計劃在期限內開發、培訓、驗證和測試創新人工智慧系統的可能性（AIA Art.3 (55)）

歐盟數位主權EUROSTACK七大原則

- **主權與安全**：確保歐洲關鍵數位基礎設施處於歐洲管轄範圍內，並受到強大安全設計和隱私設計的保護
- **去專有化（De-proprietarization）和互通性**：促進跨開源、聯合技術堆疊的集成（an open-source, federated tech stack），同時減少對大型科技公司專有解決方案之依賴
- **永續性**：建立節能、資源彈性系統以滿足歐洲環境和氣候目標
- **資料為公共利益(Data as common good)**：將資料視為共享資源，以釋放創新，同時維護社會利益和基本權利
- **去中心化的主權基礎設施（Decentralized sovereign infrastructure）**：結合邊緣運算和集中式系統，提高效率和資料主權
- **涵容性治理**：確保法規協調一致、問責制以及短期彈性與長期自主權之間的平衡
- **強大的民主**：數位科技不會造成危害，並從根本上強化民主社會

後續研究建議

- 快速發展的技術與缺乏統一的全球治理框架
- 行政、法律、技術、產業跨領域的複雜性
- 主管機關與跨域合作機制的建立
- 公眾信任與接受度的不確定性
- 數位權利/數位涵容的落實

簡報結束
感謝聆聽 敬請指教

本研究成果不代表委託單位立場