公部門人工智慧應用實務解析

從場景評估到風險管理及AI治理的全方位實戰指南

Presentation by 黃承漢（**Michael Huang** ）
2026/8

黃承漢經理

| 經歷 |

- 資誠聯合會計師事務所
  資訊安全暨鑑識科技實驗室品質負責人
- 資誠聯合會計師事務所 風險及控制服務 經理
- 資誠聯合會計師事務所 風險及控制服務 資深顧問
- 資誠企業管理顧問股份有限公司　資深顧問
- 華碩電腦 IT運營專案課 資安工程師
- 曜揚科技 技術服務部組長
- 曜揚科技 技術服務部 系統工程師

| 專長 |

- 人工智慧管理系統(AIMS)輔導諮詢服務
- 資訊安全管理制度(ISMS)輔導諮詢服務
- 營運持續管理系統(BCMS) 輔導諮詢服務
- 個人資訊管理系統(PIMS)輔導諮詢服務
- 品質管理(QMS)輔導諮詢服務
- 因應網路歐盟韌性法案(CRA)
- 零信任架構
- 資安治理成熟度分析
- PCIDSS
- PQC後量子密碼學和SBOM解決方案
- 資通安全管理法應辦事項因應

| 專業資格和證照 |

- PMP(Project Management Professional)
  國際專案管理師認證2012~2026年
- EC-Council CHFI v8資安鑑識調查專家認證
- ISO 42001 人工智慧管理系統 Lead Auditor
- ISO 27001:2005，ISO 27001:2013, ISO 27001:2022資訊安全管理制度Lead Auditor
- BS 10012:2009和2017 個人資訊管理制度 Lead Auditor
- ISO 29100:2011 個人隱私保護管理隱私框架Lead Auditor
- ISO 27018:2014 雲端個人隱私資料管理制度Lead Auditor
- ISO 27701:2019 和2025個人資料隱私管理系統 Lead Auditor
- ISO 22301:2019 營運持續管理制度Lead Auditor
- ISO 17025 : 2017 資安實驗室主管驗證訓練合格
- SGS「資通系統防護基準–合規技術專員」證書
- ITIL v3 Foundation 認證
- Trend Certified Security Expert 趨勢認證資訊安全專家認證
- 微軟MCTS(Hyper-V 虛擬化解決方案)
- 微軟MCTS(Exchange 2010)
- 微軟MCTS(SharePoint 2010, Configuring)
- 中華數位SPAM SQR專業證照

Agenda

1 公部門AI策略啟航與導入實務

2 公部門AI治理與AI法規地圖

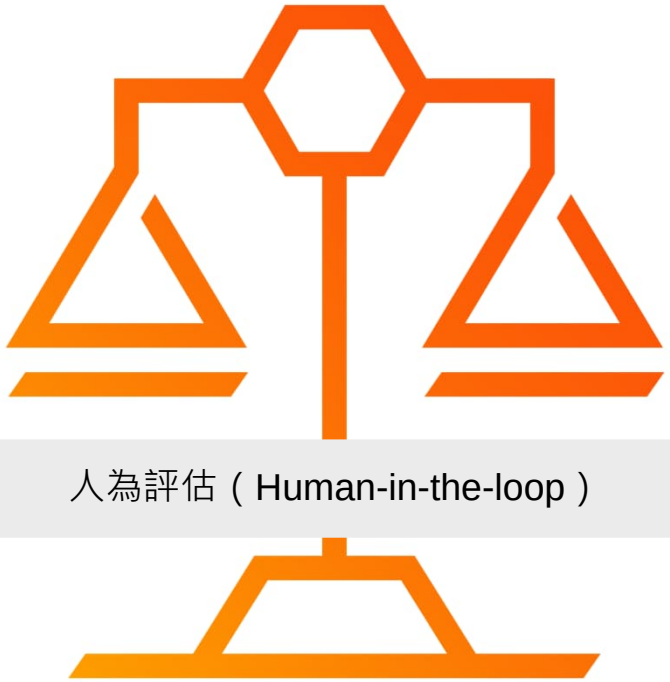
1

公部門AI策略啟航與導入實務

什麼是AI? 釐清期待與限制

AI的效能與優勢

- 減少工作時間
- 降低投入成本
- 提升重複性任務效率
- 協助做出更好判斷



人為評估（Human-in-the-loop）

AI的風險限制

- 輸出的不確定性
- 可能潛藏偏見與歧視
- 無法承擔最終決策責任

資安核心觀念

- 系統越智慧，人工評估的防線價值就越高。
- 完全由AI進行決策是一個極高風險的行為。

AI類別介紹

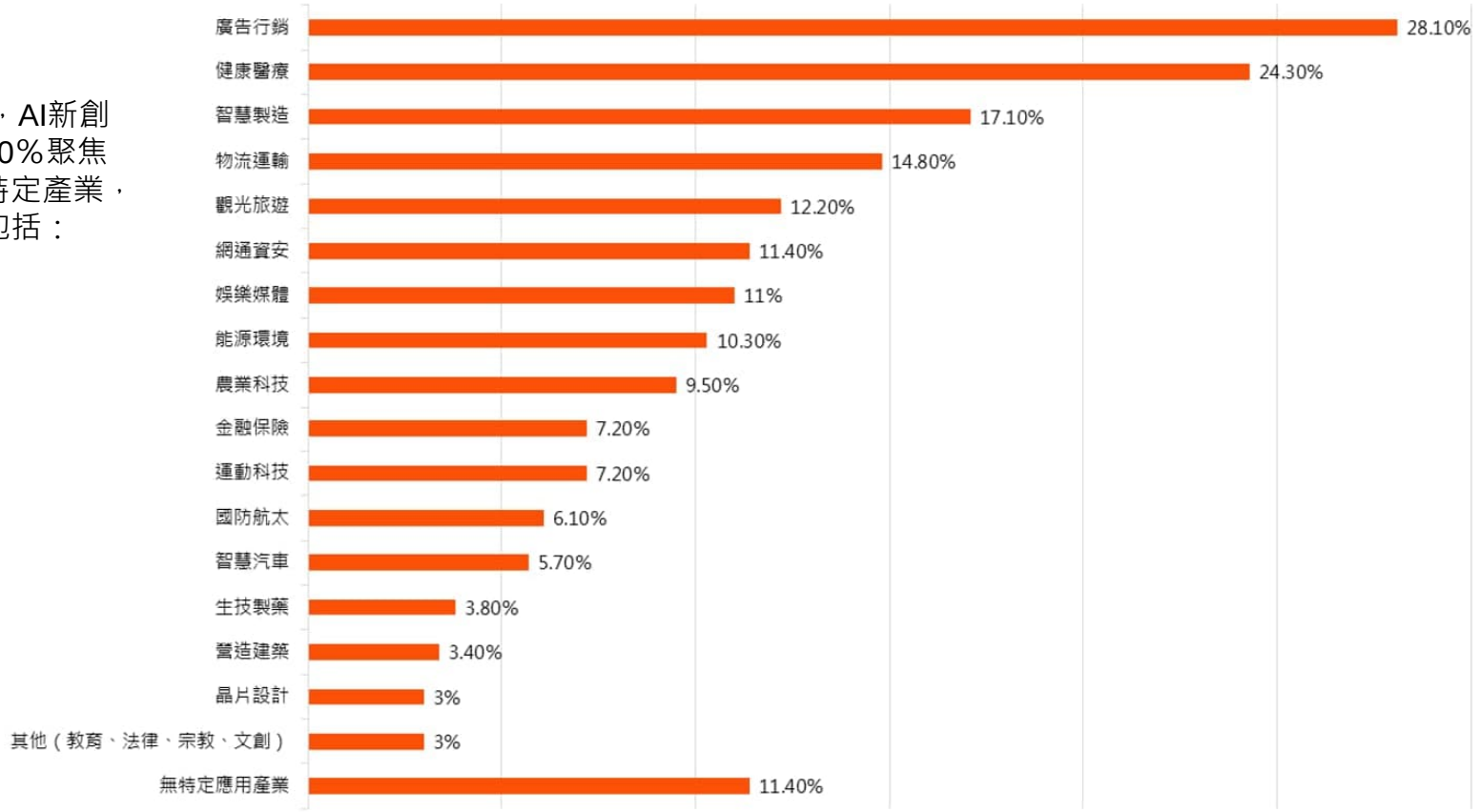

監督學習
非監督學習
半監督式學習
強化學習
簡介
使用時機
運作步驟
案例
使用有標籤的資料集來訓練模型
使用沒有標籤的資料集
結合有標籤和沒有標籤的資料來訓練模型
模型學習在特定環境中採取行動
有已分類的資料，希望 AI 系統能夠學習輸入與輸出之間的關係
不知道如何分類，想發掘不同分類的方法時
當標籤資料少，但無標籤資料充足時
任務需要在動態環境中進行交互式學習時
1. 標記訓練資料並定義
2. 資料用於演算法訓練
3. 訓練好模型後輸入新資料
1. 提供未標籤的資料
2. 演算法自行推斷資料的結構
3. 演算法提供分群結果
1. 標記部分資料
2. 標記資料用於訓練模型，未標記資料則增強模型性能
1. 演算法在環境中採取行動
2. 模型會接收到獎勵或懲罰
3. 演算法優化一系列的行為以最大化可獲得的獎勵
預測客服中心電話數量
房價預測
將使用數位服務的民眾分群
影像分類
遊戲玩家(AlphaGo)
自動駕駛汽車

在開發AI技術的產品或客戶領域，主要涵蓋那些產業或場景？

2025台灣新創生態圈大調查

根據《2025台灣新創生態圈大調查》，AI新創的產品或服務以「應用」為主體，且70%聚焦垂直應用領域，專注於將AI技術導入特定產業，且多集中於已具備數據累積的領域，包括：

- 行銷廣告：32%
- 健康醫療：30.4%
- 物流運輸：19.3%



資料來源：PwC（2025）

AI與產業整合案例

行銷廣告

媒體整合與搜尋引擎

AI能為龐大的內容資料庫提供更高效率的分類與資料管理，協助企業更精準的進行廣告投放與受眾定位，並進一步提升變現能力與整體營收

需克服的挑戰

- 資料量龐大且多為非結構化資料
- 如何在資訊過載的情況下，有效過濾雜訊、提取高價值訊息

健康醫療

診斷輔助

由AI驅動的診斷輔助系統，會以病患的個人病史作為基準，當系統偵測到與正常的樣本有偏差時，即可以標示出潛在的健康問題，並建議進一步檢查與治療。這類系統期待能夠輔助醫師診斷，同時在實際應用過程中持續累積資料數據，使模型不斷學習並提升準確度。

需克服的挑戰

- 醫療資料與敏感健康資訊的保護
- 人類生物結構的高度複雜性
- 新技術在病患、醫療機構與監管機關間的接受度與合規性

物流運輸

自駕卡車

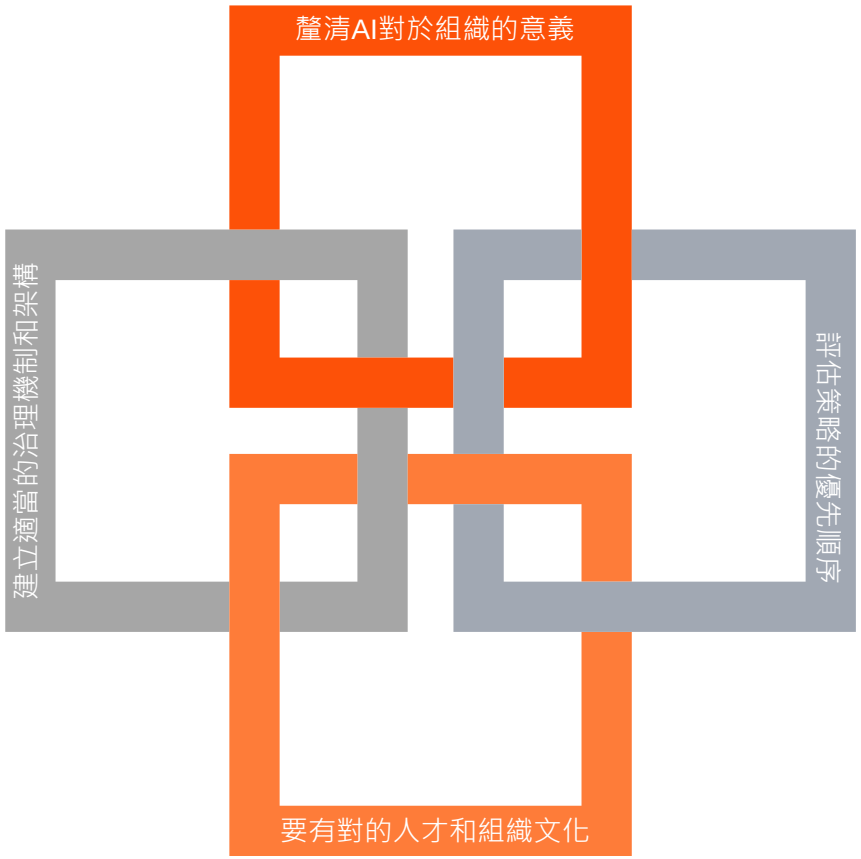
自駕卡車可以實現24小時全天候運行，降低人力成本的同時也可以透過提升資產的使用率來降低營運成本。這些變化，為AI新創帶來大量切入既有產業鏈的機會。

需克服的挑戰

- 自駕車隊相關技術仍在發展階段
- 法規、安全與社會接受度尚未成熟

資料來源：PwC（2025）

運用AI前應思考的四大議題



釐清AI對於企業的意義

應盤點產業內即將出現的技術發展與競爭壓力：何時發生、速度多快，如何因應。接著評估哪些營運痛點可透過自動化或AI技術來改善，現在已可使用的AI可能帶來哪些機會，以及中長期還有什麼新趨勢正在浮現。

評估策略的優先順序

關鍵問題包括：不同的AI解決方案，如何協助達成營運或成長目標？公司本身對變革的意願與準備度到什麼階段？想當早期採用者（early adopter）、快速跟進者（fast follower），還是保守跟隨者？

除了技術，更要有對的人才與組織文化

雖然現在投入AI看起來成本不低，但相關領域專家預期，隨著軟體愈來愈商品化，未來十年成本會逐步下降。當底層技術愈來愈標準化、商品化時，資料本身以及如何運用這些資料，將會成為最關鍵，且最有價值的資產

建立適當的治理機制與管控架構

信任與透明度是關鍵。以自駕車為例，AI等於要求人們把生命交給一台機器，對乘客和公共決策者來說，這是一個非常大的信任門檻。這種風險不僅存在於自駕車，通用於所有類型的AI應用。

資料來源：PwC（2025）

馬斯克 5 Step Algorithm –1

質疑每一項需求

- 每一個需求，都應該有人名
- 不要接受：組織一直這樣做、SOP規定、長官說、系統限制

01

刪除不必要的流程

- 如果你刪除的東西不到10%，代表你刪得不夠
- 很多流程是「表面必要、實際浪費」

02

簡化

- 將留下來的流程，盡可能簡單
- 越簡單，越容易執行與改善

03

加速

- 流程已經合理、必要、最簡後再開始追求速度
- 方法：平行處理、標準化、工具輔助、即時通知...

04

自動化

- 使用AI、Agent、Workflow、RPA、API等工具
- 讓流程「自動且持續地」產生價值

05

馬斯克 5 Step Algorithm –2

組織最常犯的錯誤

沒檢查流程

↓

直接買系統

↓

直接導入AI

↓

結果：
高速垃圾流程
(Fast Waste)

正確的順序

檢查流程

↓

刪除浪費

↓

簡化流程

↓

AI/自動化

↓

創造更大價值

Lululemon(高端運動休閒服飾品牌**)**

- 精簡產品線
- 刪除複雜流程
- 專注客戶體驗
- 成為高毛利品牌代表

TESLA

- 刪除冗於會議
- 簡化決策流程
- 高速迭代產品
- 營收20億→200億美元

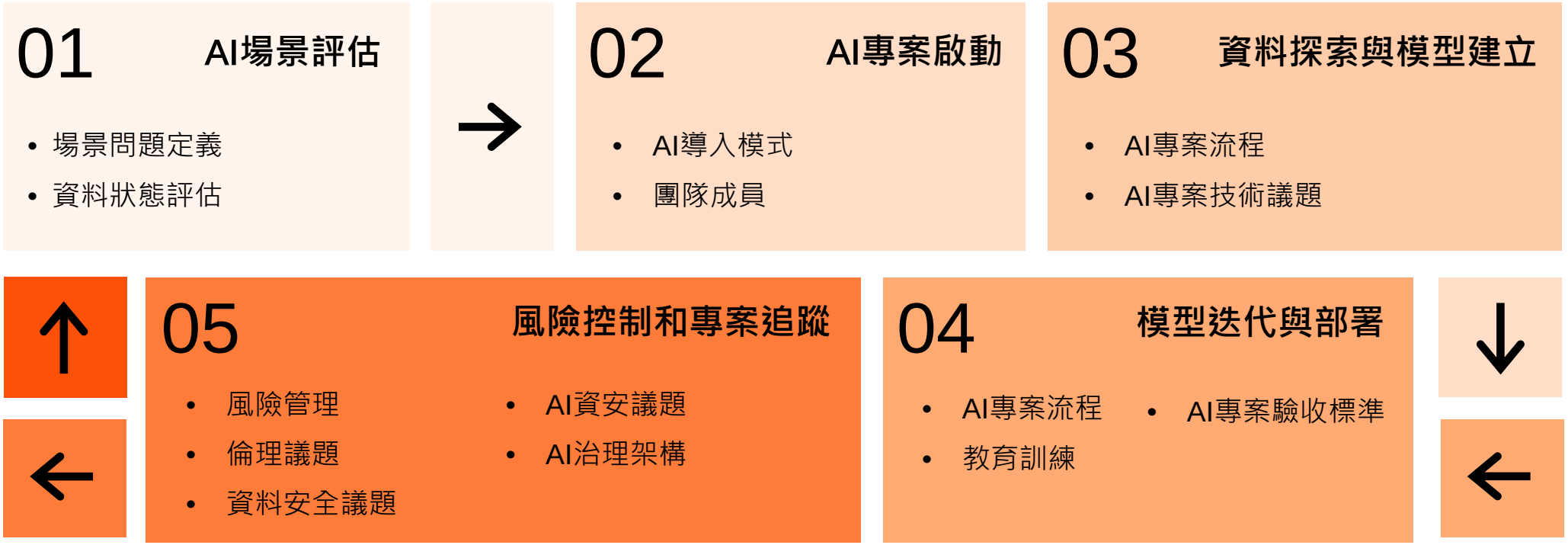
Alinea(三星米其林餐廳**)**

- 精算廚房動線
- 刪除不必要動作
- 流程及簡化
- 創造極致體驗

AI 導入

生命週期

AI 導入生命週期流程


01
AI場景評估
• 場景問題定義
• 資料狀態評估
02
AI專案啟動
• AI導入模式
• 團隊成員
03
資料探索與模型建立
• AI專案流程
• AI專案技術議題
04
模型迭代與部署
• AI專案流程
• 教育訓練
• AI專案驗收標準
05
風險控制和專案追蹤
• 風險管理
• 倫理議題
• 資料安全議題
• AI資安議題
• AI治理架構

【階段一】AI場景評估

從工作痛點出發，定義可被 AI 處理的問題

階段目標：明確專案目標和需求，並評估 AI 的適用性。

- 確定專案範圍和目標，避免「為導入 AI 而 AI」。
- 制定需求說明書，確保資料準備到位。

盤點痛點

先問「哪個流程痛、為何痛、痛在哪裡」，常見效益包含減少時間、降低成本、提升效率、促進人機協作與改善判斷。

轉成好問題

好問題應具備清楚目標、適當範圍、可行動性，以及可量化或可觀察的預期成果。

判斷 AI 適用性

優先選擇具備足夠資料、重複性任務、決策邏輯複雜，但能由專家驗證結果的場景。

【階段一】場景問題定義

場景評估的四個必答題：避免把 AI 放到不適合的流程

Q1：痛點是否明確？

是要縮短等待時間、減少人工審核、提升文件品質，還是輔助判斷？

Q2：資料是否足夠且可用？

是否有足夠樣本、標註、欄位一致性與合法可使用的資料來源？

Q3：成果能否被驗證？

是否能定義準確率、召回率、節省時間、服務滿意度或錯誤率？

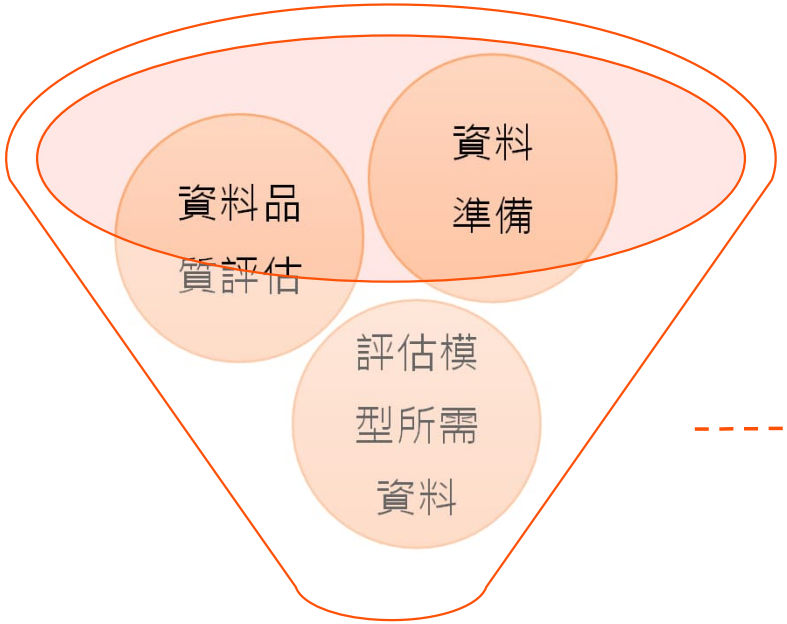
Q4：是否能被治理？

若 AI 產出錯誤、偏見或服務中斷，是否有人可監督、修正與負責？

【階段一】場景問題定義檢核清單

檢核項目	檢核結果	檢核說明
明確定義欲解決的問題		
本專案是否有適合透過 AI 解決的痛點？是什麼？	☐已完成 ☐未完成：_____ ☐不適用：_____	
您目前草擬的 AI 導入計畫，是否包含「清楚的目標」、「適當的問題範圍」、「可行動性」、「可量化或可觀察的預期目標」4 面向的條件？	☐已完成 ☐未完成：_____ ☐不適用：_____	
評估 AI 的應用可能		
您想以 AI 解決的痛點流程，是否符合「具備足夠資料」、「重複性任務」與「複雜的決策邏輯」的條件？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否邀請 AI 領域專家與熟悉機關場景的同仁，共同討論評估 AI 能否有效解決場景痛點？	☐已完成 ☐未完成：_____ ☐不適用：_____	
定義專案目標成效		
是否說明 AI 導入「關鍵指標」、「現況表現」與「設定目標」分別對應內容？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否以量化數字描述「現況表現」與「設定目標」？	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段一】資料狀態評估



確保資料現況在制定
AI 解決方案時一併考慮

資料準備

- 盤點導入場景所涵蓋資料範疇。
- 確定所需的資料種類及其來源

資料品質評估

- 完整性（Completeness）
- 準確性（Accuracy）
- 可信度（Veracity）
- 即時性（Timeliness）

評估模型所需資料

- 預訓練模型（Pre-trained Model）
- 遷移學習（Transfer learning）
- 自行訓練（Training from Scratch）

【階段一 】資料狀態評估檢核清單

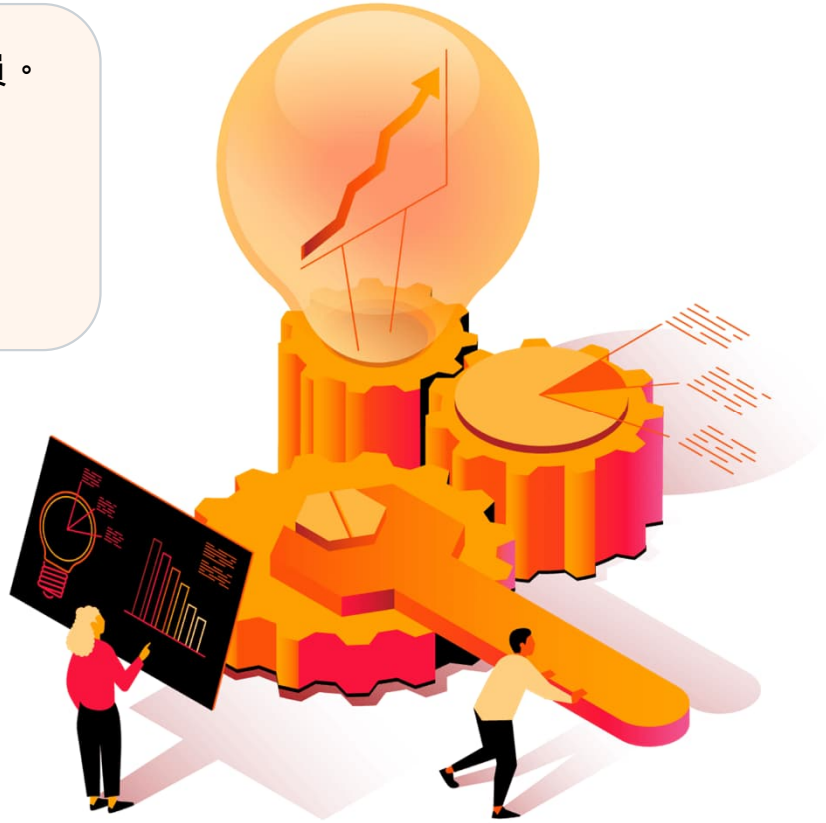
檢核項目	檢核結果	檢核說明
資料準備		
依照專案問題範疇，與哪些現有資料相關？是否有可取用的適當權限？是否有 AI 導入場景需要，但現階段無法取得資料的情況發生？是否有機會取得？或是存在可行的替代資料來源？	☐已完成 ☐未完成：_____ ☐不適用：_____	
訓練資料之蒐集及使用應排除侵害個資及智慧財產權的相關資料。 （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
資料品質		
資料來源是否真實可信？	☐已完成 ☐未完成：_____ ☐不適用：_____	
資料是否有時效性？所使用資料是否時更新？	☐已完成 ☐未完成：_____ ☐不適用：_____	
資料數量		
是否考量 AI 導入場景目標與模型使用資料的社會背景，並發現可能產生的潛在偏誤以及解決方案？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段二】AI專案啟動

階段目標：採取合適組織的內部開發或採購模式，建立 AI 專案團隊成員。

- 決定適合的開發模式，如自行建立或外包。
- 準備 AI 訓練所需基礎設施和資料。
- 確認專案團隊成員符合專案需求。

- AI導入模式
- AI 專案的專案團隊組成



【階段二】AI導入模式

既有系統延伸

若既有模組可滿足需求，不一定要啟動新 AI 專案。

機關自行建置

適合資料與能力掌握度高、需高度客製或長期維運者。

採購商業服務模組

適合需求通用、導入速度重要、可接受標準化功能。

研究單位合作

適合需求明確，但內部缺乏建模、工程或部署能力。

廠商建置

適合探索型、創新型或技術不確定性高的議題。

導入模式的核心判斷：需求成熟度 × 資料敏感度 × 內部能力 × 時程與維運責任。

【階段二】常見AI導入模式比較

	優點	缺點
既有系統	對於現有資源的盤點減少非必要的服務開發。	有時機關現有系統已無法滿足需求。
機關自行建置	機關能更貼近自身的業務需求，客製化開發所需的 AI 系統和應用。同時培養內部的 AI 技術人才，作為組織後續擴大 AI 的基礎。	需要大量的初期投入，包括人力、資金和基礎設施，風險較高。
採購商業服務模組	提供發展成熟且需求明確的 AI 應用，相較於自行建設成本更低且可快速部署。	提供發展成熟且需求明確的 AI 應用，相較於自行建設成本更低且可快速部署。
研究單位合作	機關可以根據自身需求靈活調整需求。機關可以掌握更多的自主權，減少對外部廠商的依賴。	需要較高的初期投入和時間成本，雙方須協調成果分配。
外部廠商建置	能大幅減輕機關的內部建置壓力。降低自行建設的成本和風險。	靈活性相對較低，可能有對廠商的依賴性過強、智慧財產權與資料安全等風險。機關需要建立健全的外包管理機制。

【階段二】AI專案常見所需團隊成員–1

專案經理

負責工作

- 負責專案規劃和組織
- 監督進度和資源分配
- 領導團隊進行協作

所需能力與經驗

- 具備AI 分析專案、AI 系統導入與資料分析專案管理執行經驗
- 能協助專案團隊與重要利害關係人做溝通
- 有分析資料、建立模型、解讀 AI 模型結果的經驗

資料分析師

負責工作

- 確定業務需求
- 提供領域知識
- 協助設定專案目標
- 解讀AI 分析結果，並提出業務建議

所需能力與經驗

- 具備專案領域知識經驗
- 擅長需求分析，能於前期進行需求訪談，協助專案團隊完成專案目標定義
- 對 AI 解決方案有高掌握度，能設定專案所需 AI 解決方案並向使用者說明對應成效

資料科學家

負責工作

- 進行資料清洗和特徵工程
- 建立和評估預測模
- 優化模型性能
- 檢核資料偏誤問題

所需能力與經驗

- 擁有機器學習和統計分析技能，像是 AI 建模、機器學習
- 熟練程式開發，常見使用語言包含 Python、R、 SQL

【階段二】AI專案常見所需團隊成員–2

資料工程師

負責工作

- 參與前期訪談
- 提供領域資訊

所需能力與經驗

- 具備大數據處理與 ETL 能力
- 資料庫管理技術
- 協助建立系統間的資料串接
- 具備大資料雲端平台與 API 管理經驗
- 具雲端算力與儲存所需資源規劃的經驗
- 具叢集運算架構經驗，熟悉 Hadoop 或 Spark
- 熟悉Python、Node.js 和 SQL 等程式語言

機器學習工程師

負責工作

- 模型開發部署
- 模型與演算法效能優化

所需能力與經驗

- 了解機器學習框架
- 熟悉預處理、特徵工程與數據清理等技術
- 了解機器學習演算法
- 具備模型部署、監控與優化的經驗
- 熟悉 Python、R 和 Java 等程式語言

【階段二】AI專案常見所需團隊成員–3

測試工程師

負責工作

- 系統效能測試
- 撰寫自動化測試腳本
- CI/CD 測試執行

所需能力與經驗

- 撰寫自動化測試腳本的經驗
- 熟悉自動化測試與追蹤管理測試工具
- 具測試執行經驗，了解熟悉 CI/ CD 工具，能夠配置和管理測試管道

領域專家

負責工作

- 參與前期訪談
- 提供領域資訊

所需能力與經驗

- 了解 AI 應用場景實際業務
- 後續 AI 服務實際使用者

技術諮詢專家

負責工作

- 提供 AI 專案諮詢指引
- 分享對應 AI 專案執行經驗

所需能力與經驗

- 具備大量AI 應用落地部署經驗
- 具AI 專案流程與相關實踐技術
- 熟悉環境部署工具，如：容器化技術、虛擬環境等

商業智慧分析師

負責工作

- 建立儀表板、報告、
- 資料視覺化報表
- 分析報表資訊提供分析洞察

所需能力與經驗

- 熟悉業務場景，能設立追蹤指標
- 熟悉 Power BI、Tableau等視覺化報表工具 軟體工程師、

【階段二】專案團隊成員檢核清單

檢核項目	檢核結果	檢核說明
是否評估機關已經配置充足人力？ (參照資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	
專案中是否有 1 位以上資料科學家，負責完成 AI 建模流程並確認模型效力？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否安排 AI 專案團隊與領域專家進行訪談，了解 AI 實際應用場景需求？	☐已完成 ☐未完成：_____ ☐不適用：_____	
測試工程師是否在專案啟動前進行完整系統測試，確保系統具備完善功能性和可靠性，並解決系統潛在問題？	☐已完成 ☐未完成：_____ ☐不適用：_____	
技術諮詢專家，是否具備 AI 專案實作經驗，能提供整體專案技術細節建議，並與專案小組進行執行問題討論。	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段三】資料探索與模型建立

階段目標：收集、清理和初步分析資料，接著選擇和訓練模型，進行初步測試和優化。

- 收集、清理和標註資料，確保資料安全。
- 選擇合適演算法，訓練和調整模型。
- 進行資料探索，識別資料特徵，確保資料適用性。
- 測試和評估不同模型，建立基礎模型。

資料蒐集

確認資料來源、權限、格式、欄位與是否涉及個資或機敏資料。

清理與標註

處理缺漏、重複、異常值，建立一致標註規則與資料安全控管。

探索與適用性

觀察資料分布、特徵、偏誤與可預測性，確認資料能支持目標。

【階段三】資料蒐集準備

1

資料前處理：

- 將原始資料轉換為可以作為 AI 訓練的模型資料。
- 針對敏感資料提前做去識別化的處理。

2

切分資料集：

- 將訓練資料切分為訓練集、驗證集與測試集。

3

建立資料使用管理機制：

- 建立對應版本管控機制。

【階段三】AI模型訓練

機器學習任務	簡介	案例
分類 (Classification)	將輸入的資料，依照欄位資訊識別不同的特徵和模式，再將所有資料區分為不同類別。	• 判定一批貨物是否需要接受邊境檢查。 • 判定一封電子郵件是否為垃圾郵件。
迴歸 (Regression)	用既有的變數資料，模型將尋找輸入資料與輸出資料之間的關聯，並預測輸出資料值。	• 根據房屋大小、位置或年齡等資訊預測市場價值。 • 預測城市中空氣污染物的濃度。
分群 (Clustering)	識別資料集中有相似資料點的資料群組。	• 將零售客戶分組，以找出有特定消費習慣的子群。 • 將智慧電表資料分群，以識別電器群組，並生成明細電費帳單。
降維 (Dimensionality Reduction)	在不影響資料關鍵資訊的前提下，將資料集的結構進行簡化，用更少資料展示原始資料集。	• 將人像照片去噪並進行圖像生成模型的訓練，使模型能夠根據低維的輸入照片生成逼真的人像照片。
排序 (Ranking)	訓練 AI 模型根據先前看到的列表對新資料進行排序	• 當使用者搜索網站時，按照相關性回傳關聯性高的搜索頁面。
生成 (Generative)	讓模型學習已知資料的模式，再以深度學習的技術產出全新的內容，包含文字、圖片、影片與聲音等。	• 與生成式AI 相關之大型語言模型提問，模型回覆問題解答。 • 對生成式 AI 下提示 (Prompt) 產出指定圖像。

【階段三】AI模型訓練與迭代

1 選擇對應演算法

開發人員依照專案需求選擇對應演算法

- 自然語言處理（NLP）
- 電腦視覺
- 異常監控
- 時間序列分析
- 推薦系統

2 訓練與驗證模型

模型建立過程，包含需模型訓練與模型驗證兩階段。

- 訓練階段：準備好訓練集資料並選定要採用的演算法進行訓練，產出模型。
- 驗證階段：使用驗證集資料對模型進行驗證，每次進行驗證時，可針對模型的超參數或訓練方法進行調整，以不斷縮小模型產出結果與預期結果之間的差距。

3 評估模型

當模型在驗證階段接近理想狀態，就可以用測試集資料做測試。

- 模型訓練從選擇、訓練、驗證、評估四步驟，會需要不斷迭代進行。
- 整體開發與測試流程，將會不斷進行調整。整個模型建立過程，除了技術能力外，也須於最後確保 AI 模型的有效性和公平性。

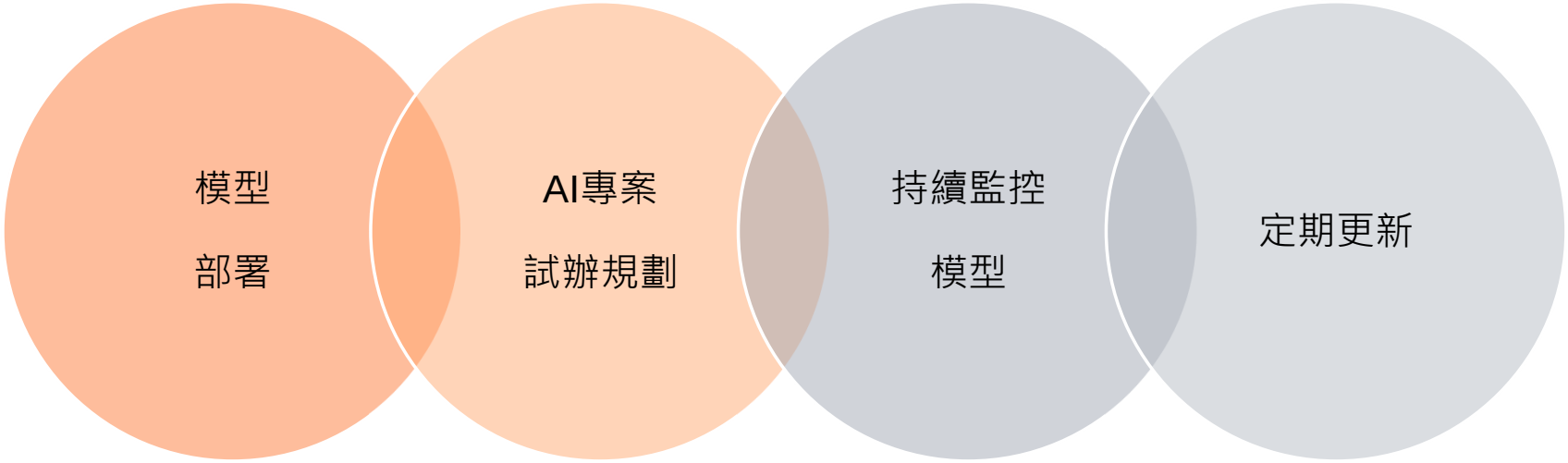
【階段三】AI 專案流程檢核清單

檢核項目	檢核結果	檢核說明
資料蒐集準備		
訓練資料是否完成前處理，了解執行團隊如何處理遺漏值、異常值、重複資料？	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 專案資料是否為手動資料標記預留足夠時間？	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 模型訓練與迭代		
是否了解 AI 專案團隊依照專案目標選擇合適演算法，並由 AI 專案團隊講解模型決策邏輯， 以及說明預測結果是否可解釋？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 專案團隊是否建立模型的評估指標， 並以評估指標衡量模型表現？	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段四】模型迭代與部署

階段目標：部署模型並進行反覆測試和優化，並進行教育訓練與驗收檢核。

- 部署模型，設計部署環境。
- 建立監控流程，確保模型運行良好。
- 反覆測試和迭代模型，確保穩定性。
- 參照驗收指標，確保 AI 專案成效。



【階段四】AI 專案流程檢核清單

檢核項目	檢核結果	檢核說明
AI 模型部署與監控		
AI 專案團隊是否規劃模型表現監控措施， 確保模型表現不佳時能第一時間進行追蹤調整？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否確保部署環境的硬體以及軟體配置滿足模型運行需求，並與訓練環境相容？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否配置監控工具以監控、記錄模型的運行性能（如延遲、吞吐量）、錯誤訊息和預測結果的準確性？	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 模型或系統是否有明確的描述，說明訓練資料、預期使用方式、使用的限制、效能或公平性指標等，供使用人員參考判斷？（參照行政院及所屬機關（構）使用生成式 AI 參考指引 ）	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 模型或系統是否能對預測結果進行解釋或分析， 供使用人員參考判斷？（參照行政院及所屬機關 （構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段四】教育訓練

共通性課程

1. AI服務及其功能介紹
   - 基本概念與技術、功能及適用範圍、使用方式
2. 應用場景與實際案例
   - 使用案例說明、實際操作、侷限性
3. AI系統的優化與監控
   - 如何解讀產出、內部相關規範、資料與隱私、負責任地使用AI
4. 疑難排解
   - 相關文件介紹、問題申報機制、聯絡窗口

差異化課程

系統管理者

1. 了解AI 建模原理、如何輸入資料與調整參數概念。
2. 能進行初步AI 系統故障排除，以及模型性能調整。
3. 了解系統資料使用範疇，能監控訓練資料的安全狀態。
4. 負責監控AI 資料清理前置準備，包含資料清理、標註。
5. 了解AI 系統與其他系統如何整合，或是擴大部署範圍。

系統使用者

1. 了解如何使用 AI 分析的結果，以及使用的侷限。
2. 了解如何辨識異常，以及對應回報機制。
3. 使用 AI 系統時，能遵守組織內資料使用規範。
4. 了解如何輸入和使用資料，確保 AI系統能獲最佳成果。
5. 掌握 AI 如何納入現有的工作流程並提高效率。

目的：確保使用者具備必要的相關知識與技能，建立人員對 AI 技術應用的信心，最大限度發揮投資價值。

【階段四】AI 教育訓練指引檢核清單

檢核項目	檢核結果	檢核說明
是否列舉本次受訓人員職務背景，客製化所需培訓內容？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否盤點受訓人員的 AI 系統應用場景，具體展示在該使用場景下如何使用 AI 系統？是否確保部署環境的硬體以及軟體配置滿足模型運行需求，並與訓練環境相容？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否說明 AI 系統之資安規範與資料應用權限，讓使用者確定能否了解輸入 AI 系統之資料會如何被使用、傳輸及保存？ （參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
承辦同仁是否具備 AI 基礎素養，有了解或執行過相關 AI 專案經驗？	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 模型或系統是否能對預測結果進行解釋或分析， 供使用人員參考判斷？ （參照行政院及所屬機關 （構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段四】AI專案驗收標準

設定模型性能標準

- 用途：確保模型在實際操作中的可靠性和有效性。
- 驗收標準：主要為機器學習在技術面表現的指標，

公平性指標

- 用途：衡量和評估模型在不同群體之間是否公正、平等地做出預測和決策的標準。
- 驗收標準：均等機會差距

系統穩定性

- 用途：評估 AI 系統在實際運行中的可靠性和效率
- 驗收標準：主要包括錯誤率、反應時間和資源利用率等指標。

使用者體驗標準

- 用途：評估 AI 系統是否滿足終端使用者的需求和期望。
- 驗收標準：主要包括系統使用便利性、回應速度與性能、功能符合需求程度和整體使用者滿意度。

【階段四】AI 專案驗收標準檢核清單

檢核項目	檢核結果	檢核說明
是否依照 AI 專案目標，考量模型性能標準、系統穩定性、使用者體驗與公平性等相關指標，選擇多種指標組合，作為驗收標準？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
對於生成式 AI 服務，是否將回應時間列為標準？生成式 AI 系統回覆時間將大幅影響使用者體驗，但生成時間與技術和資源高度相關，規劃時應納入評估。	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否確認機關公告之最新 AI 專案驗證規範，了解最新 AI 專案驗證方法與工具建議？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
針對此 AI 專案是否設立多元使用者體驗標準指標，包含但不限於系統使用便利性、回應速度與性能、功能符合需求程度、以及整體使用者滿意度。	☐已完成 ☐未完成：_____ ☐不適用：_____	
模型性能標準是否考慮實際應用場景各種因素，確保模型在實際操作中有效並易於後續維護？例如：程式碼清楚、模型文件完整、 metadata 完備等等。 (參照資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段五】風險控制與專案追蹤

階段目標：持續監控模型性能，進行優化和風險管理。

- 建立持續監控系統，定期更新模型。
- 進行專案追蹤和改進，確保長期運作順利。
- 監控和管理風險，包括道德和資料安全問題。

- AI專案的風險評估（建置前中後）
- AI專案面臨的議題

  →於後續章節說明

【階段五】風險管理檢核清單–建置前

檢核項目	檢核結果	檢核說明
建置前		
是否參考有關使用資料的相關規範，防止資料之蒐集與使用不符合法律、法規命令、行政指導（包含政府機構的公開指引）？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否於於專案前期，將會使用到的機關內部資料和外部資料嘗試做整合。再根據完整性 (Completeness)、準確性(Accuracy) 、 即時性 (Timeliness) 、真實性 (Veracity)與公平性 (Fairness) 的組合標準，來評估資料是否可使用？ （ 參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否瞭解並記錄 AI 模型之目的及用途，以及所使用的方法論。在建置前對模型進行測試，以確保其產出結果符合預期目的？	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段五】風險管理檢核清單–建置中

檢核項目	檢核結果	檢核說明
建置中		
是否針對模型的各個版本進行記錄，包含所選擇資訊的類型、來源、模型的輸出及預期用途？ （參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
若模型出現偏見或歧視的分析結果，是否規劃調整模型，確保您的模型的公平性和可解釋性，並且建立監控意外或偏差輸出的流程？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
專案開發團隊成員是否於專案早期投入準備，在資料科學家完成模型開發前，部署團隊應進行部署的測試規劃，以確保開發模型產出可順利整合？	☐已完成 ☐未完成：_____ ☐不適用：_____	

【階段五】風險管理檢核清單–建置後

檢核項目	檢核結果	檢核說明
建置後		
是否建立使用者發現偏見與歧視時的回報機制與管道，並預備開發團隊量能做模型調整？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
建立明確的管理框架：是否建立明確的權責劃分，定義機關、營運團隊、使用者對 AI 模型所需承擔的管理責任？	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否建立內部審查與監控機制，必要時可邀請不同領域專業人員參與相關評估過程？亦可建立獨立第三方之審查機制，並確保適當溝通管道讓外界在必要時得以瞭解相關風險或重要事項。 （參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	

數位發展部「人工智慧風險分類框架」–1

訂定依據

本框架由數位發展部依據《人工智慧基本法》第 16 條訂定，供目的事業主管機關訂定 AI 風險管理規範之用。

框架功能

協助各目的事業主管機關辨識與評估人工智慧應用可能涉及之影響範疇及風險程度，作為訂定以風險為基礎之管理規範及配套採行促進發展措施之程序。

適用範圍

主要適用於民用領域之 AI 應用，明確排除軍事用途。凡屬國防軍事目的之人工智慧系統研發、部署及應用，應依國防相關法令及規範辦理。

《人工智慧基本法》第十六條

- 數位發展部應參考國際標準或規範，推動與國際介接之人工智慧風險分類框架，並應協助各目的事業主管機關訂定以風險為基礎之管理規範。
- 各目的事業主管機關應視人工智慧應用風險管理之需要，循前項風險分類框架，訂定以風險為基礎之管理規範，並應協助相關產業自行訂定產業指引及行為規範。有第五條第一項所列之情事者，應依法令限制或禁止之。

數位發展部「人工智慧風險分類框架」–2

1 **盤點應用情境**

2 識別風險

3 評估風險

4 應對風險

盤點應用情境

- 蒐集並整理所管產業中，現有或預計導入之 AI 應用情境及相關資訊。
- AI應用情境盤點表

	項目	說明	備註
1.1	應用情景描述		詳述AI系統德具體用途、流程、目的
1.2	AI技術		說明可能使用的AI核心技術
1.3	利害關係人		誰會開發、部署；使用或受此系統影響?

數位發展部「人工智慧風險分類框架」–3

1 盤點應用情境

2 **識別風險**

3 評估風險

4 應對風險

識別風險

- 鎖定潛在風險類型，經識別為有的風險，即應進入下一階段評估其風險程度。
- AI風險類型表

AI 系統本身之技術設計缺陷	AI系統的安全漏洞與攻擊、缺乏透明性與可解釋性、AI行為偏離人類意圖與社會價值、AI 具有危險的能力、影響隱私與違反個人資料保護法規、侵害智慧財產權疑慮、 不公平的歧視或偏見、錯誤或誤導訊息
部署後操作及人機互動問題	過度依賴與不安全使用、生成違法內容、詐欺與深偽技術濫用、用於網路攻擊、AI 自主代理之授權外行為
社會結構與環境衝擊	企業及國家間競爭秩序失衡、權力集中與利益分配不公平、不平等加劇就業品質下降、人類在經濟與文化上之創作價值受損、環境傷害、認知作戰與資訊主權

數位發展部「人工智慧風險分類框架」–4

1 盤點應用情境

2 識別風險

3 **評估風險**

4 應對風險

評估風險

- 在完成風險識別後，目的事業主管機關應評估風險影響程度
- 以客觀固有風險為評估原則
- 評估是否有高風險 AI 應用
  - ✓ 對國家安全、人民基本權利、生命安全、財產保障、、社會秩序或生態環境可能造成嚴重危害，目的事業主管機關應認定該應用情境為人工智慧基本法第 17 條第 1 項所定之高風險 AI 應用。

數位發展部「人工智慧風險分類框架」–5

1 盤點應用情境

2 識別風險

3 評估風險

4 **應對風險**

應對風險

- 完備管理規範與措施。
- AI應用情境盤點表

	項目	說明	備註
1.1	應用情景描述		詳述AI系統德具體用途、流程、目的
1.2	AI技術		說明可能使用的AI核心技術
1.3	利害關係人		誰會開發、部署、使用或受此系統影響?

數位發展部「人工智慧風險分類框架」–6

風險識別、評估及應對表

識別風險			評估風險			應對措施				
			是否造成嚴重危害			現有措施涵蓋程度			新增措施	
風險項目	是否涉及	具體風險情境描述	不考慮現有措施下，是否可能造成嚴重危害	涉及之危害判准	評估說明	現行法規或行政措施名稱	現有措施類型	現有措施涵蓋程度	措施內容	說明
	◘是 ◘否 ◘不適用		◘是 ◘否 ◘不適用	◘國家安全 ◘基本權利 ◘財產保障 ◘社會秩序 ◘生態環境			促進發展 ◘產業標準與業者自律 ◘輔導合規資源 ◘其他 管理措施 ◘限制禁止 ◘事前許可 ◘透明度要求 ◘其他	◘充分 ◘部分 ◘不足 ◘尚無相關措施		措施類型： 風險緩解情形： 是否可能影響AI創新發展和應用：

AI 專案

技術議題

AI 專案技術議題

主要議題

軟體工具與環境

- 機關應依據 AI 應用目標和需求，選擇適合的程式語言。例如：
  1. C++：擁有高水準的性能和效率，成為遊戲類型 AI 應用的理想選擇。
  2. Java：具備易於調整、使用者友善介面，並且可以在大多數平台上使用特性。它適用於 AI 搜尋引擎演算法和大型 AI 專案。
  3. Python：目前主流選擇，初學者容易上手，具備許多模型建構套件。
  4. R：為預測分析和統計而開發，能進行資料前處理與統計分析。
- 程式開發工具與環境選擇上，AI 專案選擇的整合開發環境與傳統專案不同。

主要議題

數據資源管理

- 數據資源管理包含：
  1. 資料庫管理
  2. 確保資料儲存技術的可擴展性。
  3. 透過 ETL 工具，將資料從不同來源提取、轉換為標準格式並載入到目標資料庫。
  4. 確保資料的一致性和可用性。
  5. 何建立資料管道。
- 常見資料庫類型選擇：
  1. 關係型資料庫（RDBMS）
  2. NoSQL 資料庫
  3. 圖形資料庫

AI 專案技術議題

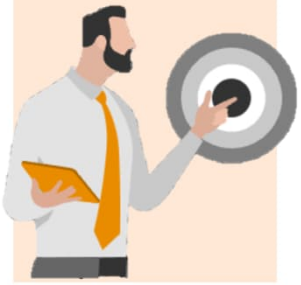
主要議題

硬體資源

- 地端環境
  - 優點：控制性高、無須依賴網路、遵守嚴格得隱私標準。
  - 缺點： 硬體資源需求、專門技術團隊維護和管理、擴展性有限。
- 雲端環境
  - 優點：可擴展性高、財務門檻低、雲端服務提供商開發工具與技術支援。
  - 缺點： 資料傳輸安全性、服務商限制、雲端服務不確定性。
- 雲地混合環境
  - 優點：最佳化資源運用、高齡活性、更高的控制與安全性。
  - 缺點： 複雜度增加、同步問題、潛在的更高成本。

主要議題

開發與部署工具



- API 管理，機關必須在使用 API 之前對其進行妥善評估需要了解使用 API 的條款和條件，並確保 AI 可取用資料規範的合理性，避免 AI 不當取用機密資料。
- 版本控制系統 (VCS)，版本控制系統幫助團隊追蹤程式碼和配置的變化，確保不同版本之間的相容性和穩定性。
- 容器化技術，容器化技術在 AI 專案中有助於簡化環境配置、提高部署效率和資源利用率。
- CI/CD 工具 ，CI/CD 在 AI 專案中加速模型更新和部署，確保程式碼變更的穩定性和品質。

AI 專案技術與環境議題檢核清單

檢核項目	檢核結果	檢核說明
使用 API 串接服務前，確認機關內部是否有關生成式 AI 與 API 的使用規範？（參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否考量所需資料型態、訓練資料如何存放、資料庫型態與專案的複雜度？	☐已完成 ☐未完成：_____ ☐不適用：_____	
部署環境會是否將影響到專案資料的存放位置方式與存放地點？（參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否了解潛在廠商的資料存放規範？（參照行政院及所屬機關（構）使用生成式 AI 參考指引、資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
使用雲端服務，是否考量其資料庫位置，或涉及個人資料之對應風險？（參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
若專案採取雲端部屬，是否已經確認機敏資料並且禁止上雲？（參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 專案的核心規劃團隊是否已到位，並確保專案範疇與痛點需求高度一致、分析洞察已納入目標設計，且資料架構已設計為可支援持續訓練、並能契合後續營運與既有系統整合的基礎設施？	☐已完成 ☐未完成：_____ ☐不適用：_____	

AI 與傳統資訊專案的差異

	傳統資訊專案	AI 專案
核心驅動	邏輯驅動（給定規則產出結果）	資料驅動（從資料中學習規則）
系統表現	穩定、100%可預測	具備不確定性、表現隨訓練與資料變化
開發過程	線性、需求凍結後開發	高度迭代、需要預留模型容錯與調整時間
驗收標準	功能是否上線、是否符合規格書	準確度、延遲、資料消耗等多維度指標

用驗收傳統資訊專案的思維來驗收AI，注定會面臨專案失敗與期望落差。

2

公部門AI治理與AI法規地圖

AI 議題

與AI治理

AI 專案面臨的議題

模型風險

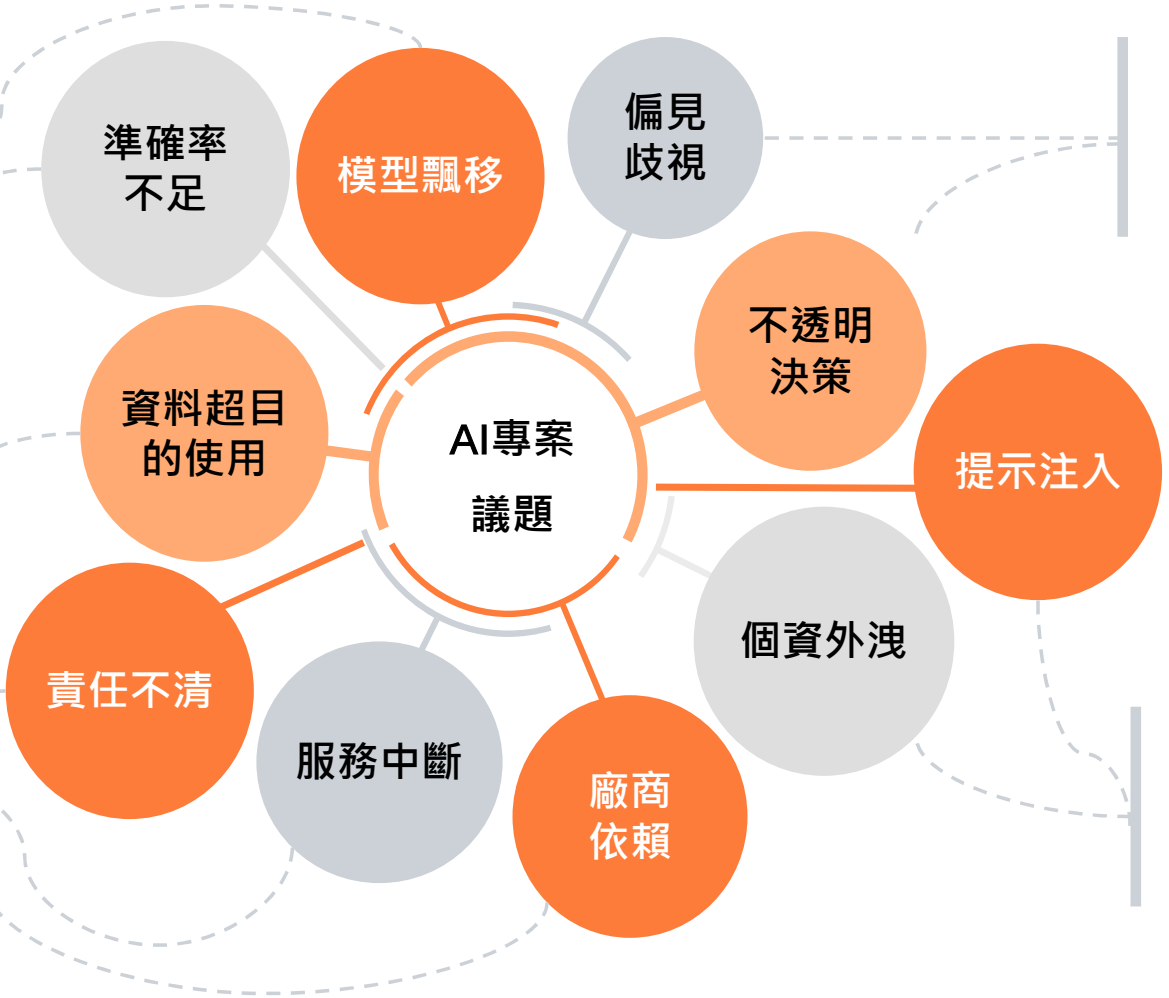
- 因應：設定驗收門檻、持續監控...
- 已於先前章節說明

資料安全議題

- 因應：資料最小化、去識別化...
- 將於此章節說明

AI治理架構

- 因應：備援機制、權責分工...
- 將於此章節說明



倫理議題

- 因應：公平性測試、可解釋性...
- 將於此章節說明

資訊安全議題

- 因應：減少AI系統授權權限...
- 將於此章節說明

AI倫理議題

01

公平性

如果使用個人資料或實際社會中的資料，應確保是否潛在傷害與偏見最小化。

02

完善AI專案權責機制

從完整 AI 專案生命週期確保對應責任義務如何劃分，在 AI 系統發生錯誤時能找尋對應負責人，並於長期未來推動 AI 審核機制。

03

對環境永續友善

機關應評估 AI 應用對環境造成的負面影響。AI 系統從模型訓練到應用的過程中，仰賴大量算力，將產生碳排放，或是資料存放的資料中心仰賴水資源做冷卻，甚至運算硬體裝置將使用稀有金屬。

04

透明度與可解釋性

AI 系統運作過程、決策機制、資料使用狀況，以及上述行為對於使用者和社會之影響，機關能高度掌握，且使用者也對所使用 AI 服務有基本了解。

05

保持人為參與

若在 AI 應用上排除人為參與，前述之道德議題疑義發生時，將有可能被忽視，完全由 AI 進行決策會是一個高風險決定。

AI倫理議題

01

公平性

如果使用個人資料或實際社會中的資料，應確保是否潛在傷害與偏見最小化。

檢核項目：

1. 抽樣是否無偏見，以公平資料集進行模型訓練？
2. 是否使用合理的分析特徵，設計無偏見的分析架構與流程？
3. 是否審查輸出內容與決策建議的合理性，不以特定宗教、國籍、種族、性別、身心障礙、性傾向、政治傾向、年齡、文化等因素而產生有害偏見或歧視，以保護受眾權益？
4. 是否設計系統化的提示（Prompt）測試方式，針對不同性別、族群等產生測試指令，檢查是否存在偏見？
5. 是否建立回報系統，使用者可回報使用生成式 AI 產生的有害內容，機關並即時審視模型，並提出方案處理與改進？
6. 是否採取持續評估的方式，因應不斷變化的公平性考量與社會期望，與時俱進?

02

AI倫理議題

01

公平性

如果使用個人資料或實際社會中的資料，應確保是否潛在傷害與偏見最小化。

03

對環境永續友善

機關應評估 AI 應
系統從模型訓練
將產生碳排放
資源做冷卻 甚至運算

02

完善AI專案權責機制

從完整 AI 專案生命週期確保對應責任義務如何劃分，在 AI 系統發生錯誤時能找尋對應負責人，並於長期未來推動 AI 審核機制。

檢核項目：

1. 是否於開發、部署或使用 AI 時，遵守現行法律規定、政府機關之指導方針與國家政策，以及 AI 服務提供的相關使用條款，確保法令遵循，並符合法律以及既有規範？
2. 是否從開發到建置，明確界定 AI 生命週期中所有參與者的角色、功能及擔負之責任、管理機制與法律義務，並以書面或數位方式建立監控機制，賦予對應中高階管理人員監控職責、落實分層負責？
3. 機關若作為生成式 AI 的使用者，是否對使用生成式 AI 工具輔助日常工作（如撰寫電子郵件與報告）所產生的輸出承擔責任，並進行必要檢查？
4. 是否透過對 AI 系統從開發到營運各階段做審核驗證，已確保 AI 產出內容無發生可避免之偏誤？
5. 是否提供申訴與行動救濟管道，設立使用者回報機制，以即時處理相關問題？

AI倫理議題

檢核項目：

1. 是否優先使用可解釋性高的模型，同時注意生成式 AI 在解釋性方面的限制，適時調整解釋方式？
2. 是否建立評估與稽核機制，追蹤資料來源、設計決策、訓練過程等，確保各階段透明度？
3. 是否向利害關係人解釋 AI 模型運作與表現，確保於倫理面的影響可被接受？包含無歧視與傷害、合理的信任和背後流程設計。
4. 是否明確告知內容是由 AI 創作，並通知民眾何時與 AI 系統互動，增加透明度與信任度？
5. 是否於使用生成式 AI 創作內容或與民眾互動時，明確標示其使用情況，並盡可能標註由 AI 生成的內容，以增加透明度？

02

完善AI專案權責機制

從完整 AI 專案生命週期確保對應責任義務如何劃分，在 AI 系統發生錯誤時能找尋對應負責人，並於長期未來推動 AI 審核機制。

04

透明度與可解釋性

AI 系統運作過程、決策機制、資料使用狀況，以及上述行為對於使用者和社會之影響，機關能高度掌握，且使用者也對所使用 AI 服務有基本了解。

AI倫理議題

01

公平性

如果使用個人資料或實際社會中的資料，應確保是否潛在傷害與偏見最小化。

檢核項目：

1. 是否確認使用者應努力理解影響 AI 系統產出的因素，在諮詢 AI 系統前，先形成自身的觀點與組織立場，避免過度依賴 AI ？
2. 是否於使用生成式 AI 時，應有人工監督機制監控輸出，確保結果符合預期？

05

保持人為參與

若在 AI 應用上排除人為參與，前述之道德議題疑義發生時，將有可能被忽視，完全由 AI 進行決策會是一個高風險決定。

02

完善AI專案權責機制

從完整 AI 專案生命週期確保對應責任義務如何劃分，在 AI 系統發生錯誤時能找尋對應負責人，並於長期未來推動 AI 審核機制。

04

透明度與可解釋性

AI 系統運作過程、決策機制、資料使用狀況，以及上述行為對於使用者和社會之影響，機關能高度掌握，且使用者也對所使用 AI 服務有基本了解。

AI資料安全議題

合法性

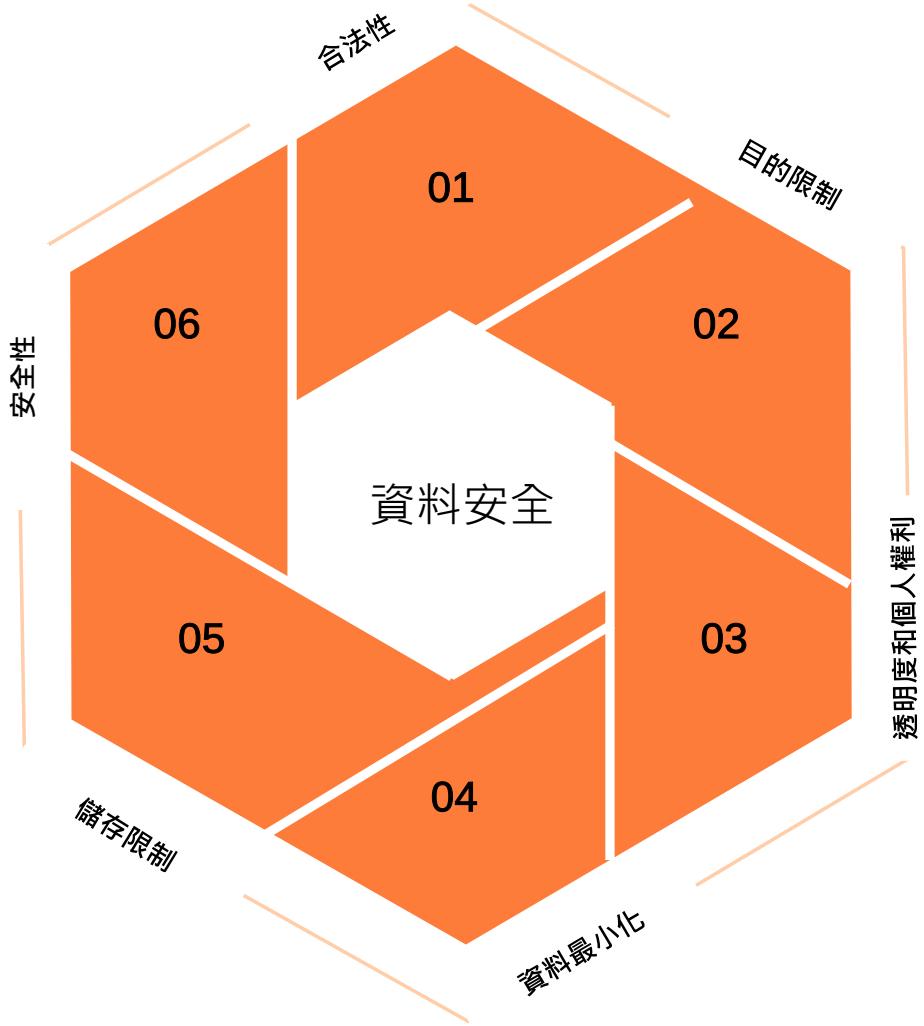
訓練模型使用機敏資料的行為，存在對應法律依據並依循資料保護或任何其他法規。

安全性

以合適技術和配套措施來保護機敏資料。

儲存限制

避免無使用目的情況下，長時間大量儲存機敏資料。



目的限制

取用機敏資料時，應用資料的範疇僅限核定的目標。

透明度和個人權利

應向資料主體說明，將如何使用他們的資料，並確保資料主體能選擇其資料不被蒐集、處理及利用之權利與機制。

資料最小化

以使用最少資料來達到 AI 訓練目標，規劃資料使用範疇。

資料安全議題檢核清單

檢核項目	檢核結果	檢核說明
AI 資料使用		
是否了解 AI 模型需要使用機敏資料的原因，並監控取得相關資料的流程？ (參照行政院及所屬機關(構)使用生成式 AI 參考指引)	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否確認以使用最少機敏資料為原則來訓練 AI 模型？ (參照行政院及所屬機關(構)使用生成式 AI 參考指引、資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否已確保 AI 模型在使用個人資料時，其使用方式、存放多久、取用單位等資訊皆為透明公開？ (參照行政院及所屬機關(構)使用生成式 AI 參考指引、資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否記錄 AI 系統訓練所使用的資料，以及 AI 系統可存取之資料範疇？ (參照資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	
AI 資料使用		
使用公開生成式 AI 服務（例如 ChatGPT），機關是否考量使用者可能存在不當使用行為，並設立相關對應機制，若僅良性勸說，是否可承受資訊外洩的風險？ (參照行政院及所屬機關(構)使用生成式 AI 參考指引、資通安全管理法及子法)	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否確保未在公開或 API 串接的生成式 AI 服務中提供個人訊息和機密資料？ (參照行政院及所屬機關(構)使用生成式 AI 參考指引)	☐已完成 ☐未完成：_____ ☐不適用：_____	

AI 資訊安全議題

- AI 模型仰賴大量資料，若建模過程中使用個人資料或機敏資料，面臨未經許可的存取風險或發生外洩事件，將帶來不良影響。
- 攻擊者可能透過注入惡意資料影響模型訓練結果。
- 利用反向工程，試圖竊取模型架構與模型內部資料等。

資訊安全議題檢核清單

檢核項目	檢核結果	檢核說明
內嵌式 AI 服務（例如 Gemini、 Copilot）所規範之資安條款，是否符合機關內部既有資安規範？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
是否留意內嵌式 AI 服務中，有未經資安驗證的 AI 外掛程式及其風險？ （參照資通安全管理法及子法 ）	☐已完成 ☐未完成：_____ ☐不適用：_____	
減少授權生成式 AI 無回復可能的任務的權限（像是信件寄送與修改紀錄），確保關鍵行為是否都是由人來檢核執行？ （參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	

AI治理架構

AI治理架構

建立管理制度

基礎 AI 系統管理制度至少需考慮以下面向：

- 誰是 AI 系統的使用者
- AI 系統建立目標
- 存在哪些利害關係人
- 如何檢查 AI 系統的使用紀錄
- 誰具有監督與糾正錯誤的權限與責任。

建立AI系統清單

清單應定期更新，包含以下詳細資訊：

- 每個系統的用途、使用情況和相關風險。
- 提供 AI 系統細節，如所使用資料、系統所有權、開發紀錄和關鍵日期等詳細資訊。
- 遵照組織規範和相關工具來維護準確和完整的清單。

建立治理委員會

- 機關宜設立 AI 治理委員會，或在現有的治理委員會中設置 AI 代表，定期且全面的 AI 治理議題討論。
- AI 治理委員會之成立，有助於機關整合 AI 專案相關資源、統一機關 AI 應用策略並共同研擬機關內 AI工具應用之規範。

公部門 AI 成功導入的三個結論

01

先定義問題，
再談 AI 技術。

02

先評估資料，
再建置模型。

03

先設計治理。
再擴大應用。

AI 專案的本質，是一套可被管理、可被驗證、可被問責的公共服務改善機制。

AI

法規地圖

各國人工智慧相關法規

歐盟

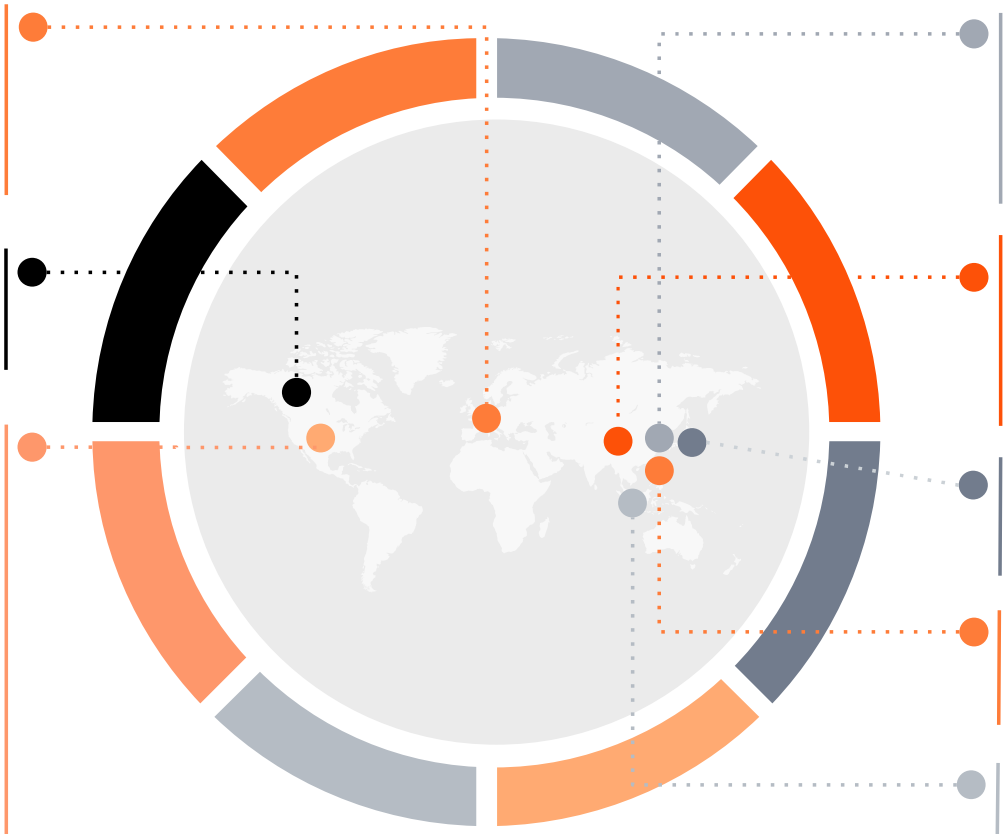
- 於2024年通過《人工智慧基本法》（EU AI Act）是全球首部針對人工智慧管制的全面性專法，於2024年8月正式生效。

加拿大

- 加拿大於2022年6月提出《人工智慧和數據法案》。

美國

- 美國參議院於2022年4月提出《演算法問責法案》，核心目標是防止演算法偏見，並強制科技公司提高自動化決策的透明度與可究責性。
- 美國國家標準與技術研究院（NIST）於2023年1月公布「人工智慧風險管理框架」，該框架提供相關資源，以協助組織與個人管理人工智慧風險，並促進可信賴的人工智慧（Trustworthy AI）之設計、開發與使用。



韓國

- 於2025年1月公布《人工智慧發展與建立信任基本法》，於2026年1月22日起生效，為繼歐盟人工智慧法之後第二部關注人工智慧的國家級立法。

中國

- 於2023年8月公布《生成式人工智慧服務管理暫行辦法》，要求服務提供者對生成內容負責，確保其符合社會主義核心價值觀。

日本

- 日本於2025年6月公布施行《人工智慧技術研究開發及應用促進法》。

臺灣

- 於2025年底正式三讀通過「人工智慧基本法」，並於2026年初正式施行。

新加坡

- 新加坡於2024年1月提出《生成式人工智慧（Generative AI）治理架構草案》。

關於臺灣《人工智慧基本法》

立法目的

促進以人為本之人工智慧研發與人工智慧產業發展，建構人工智慧安全應用環境，落實數位平權，保障人民基本權利，增進社會福祉，提升國人生活品質，促進社會國家之永續發展，維護國家文化價值及提升國際競爭力，並確保技術應用符合社會倫理

1. 基本原則

發展AI，應兼顧社會公益及數位平權發展，促進創新研發並強化國家競爭力，並避免AI應用產生的負面影響。

2. 評估驗證

由行政院設立國家人工智慧戰略特別委員會，訂定國家人工智慧發展綱領。

3. 風險管理

數發部需參考國際的法例、治理經驗，推動與國際介接的AI風險管理框架，由各目的事業主管機關訂定風險管理規範，協助各產業訂定產業指引及行為規範。

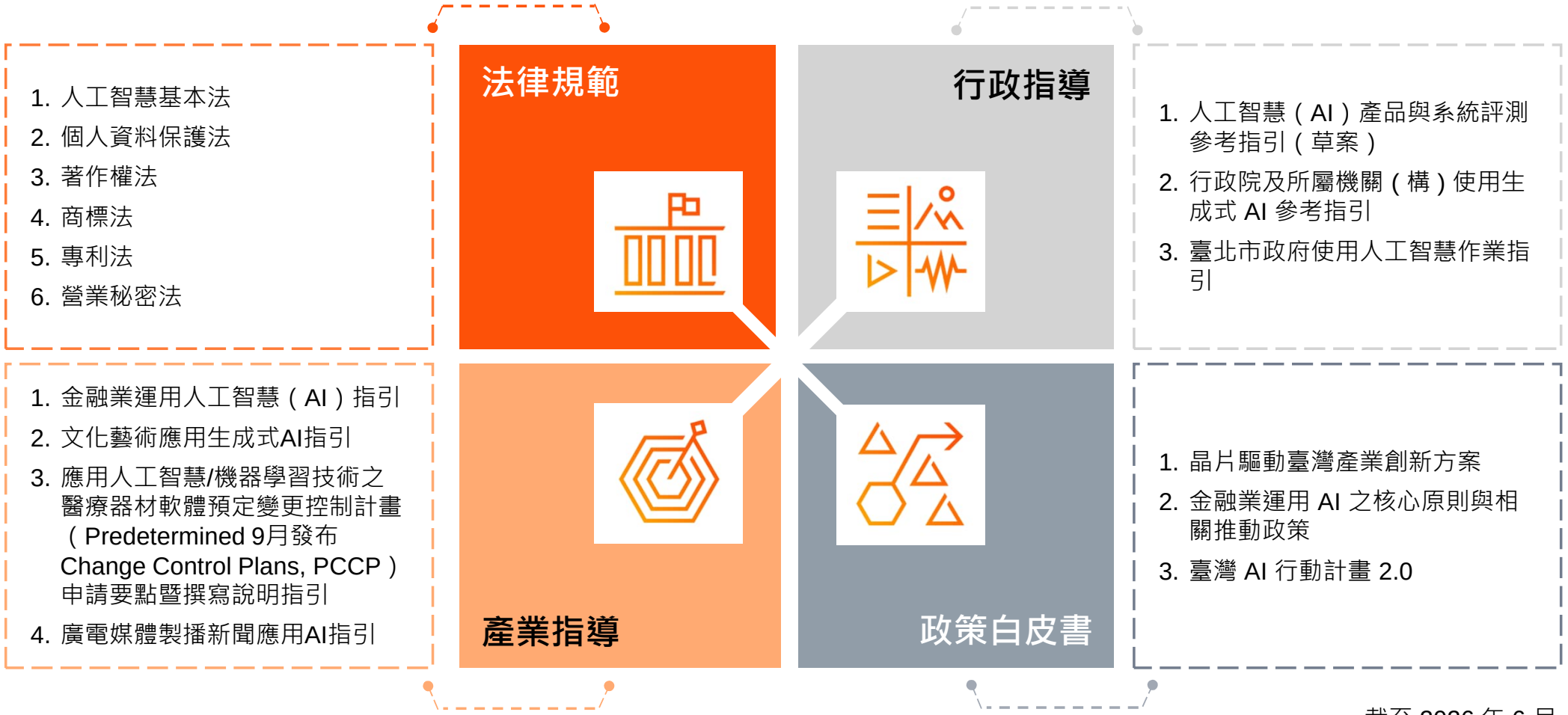
4. 資料開放

應推動資料開放、共享及再利用機制，定期檢視與調整相關法令及規範；為提升AI使用資料的品質與數量，確保訓練及產出結果足以展現國家多元文化價值與維護智慧財產權。

AI 應用原則–人工智慧基本法第四條

1. **永續發展與福祉**：應兼顧社會公平及環境永續。
2. **人類自主**：應以支持人類自主權、尊重人格權等人類基本權利與文化價值。
3. **隱私保護與資料治理**：應妥善保護個人資料隱私，尊重企業營業秘密，避免資料外洩風險。
4. **資安與安全**：應建立資安防護措施，防範安全威脅及攻擊，確保其系統之穩健性與安全性。
5. **透明與可解釋**：產出應做適當資訊揭露或標記，以利評估可能風險，並瞭解對相關權益之影響。
6. **公平與不歧視**：人工智慧研發與應用過程中，應盡可能避免演算法產生偏差及歧視等風險。
7. **問責**：應確保承擔相應之責任，包含內部治理責任及外部社會責任。

臺灣AI相關應用指引與規範



截至 2026 年 6 月

國際標準框架引領AI治理方向

這兩項標準的出現，標誌著AI治理正從抽象的原則倡議邁向具體的實踐規範。

比較面向	ISO / IEC 42001	ETSI EN 304 223
發布機構	國際標準化組織（ISO / IEC）	歐洲電信標準協會（ETSI）
發布時間	2023年	2024年
標準性質	AI管理系統標準	AI系統安全與可信度技術標準
適用範圍	所有開發、提供或使用AI系統的組織	AI系統開發者與服務提供者
核心焦點	AI系統全生命週期治理與風險管理	AI系統技術層面的安全性與可信度
主要要求	風險評估、影響分析、政策制定、持續改進	資料品質、模型穩健性、透明度、人機協作
認證與合規	ISO / IEC 42001為管理系統標準，可由第三方認證機構進行符合性評估	通常作為技術性或產品層面之合規 / 測試依據，需依歐盟或國家級合格評定確認合規
管理層級	組織管理系統層級（主要聚焦於組織治理與管理系統）	技術實作層級（主要聚焦於組織治理與管理系統）
法規連結	可支援各國AI監管要求	對應歐盟《人工智慧法案》
適用企業類型	各產業AI應用開發與導入企業	AI產品與服務開發商

資料來源：PwC（2026）

ISO/IEC 42001 乃全球認可之 AIMS 驗證框架

可供驗證之國際標準

不僅是建議或原則，而是可透過獨立第三方稽核機構嚴格審查，獲得具有全球公信力證書的標準

能與既有 ISO 管理機制無縫整合

採用與 ISO/IEC 27001、ISO 9001 相同之系統化結構，可直接整合進現有管理體系中

與法令監管之精神保持一致

提供結構化管理系統框架，滿足各國法令規範，包含相對嚴苛之 EU AI Act，及臺灣人工智慧基本法等等

能與既有 ISO 管理機制無縫整合

可將 NIST RMF 作為內部建立 AIMS 的核心方法論，透過 ISO/IEC 42001 認證來驗證遵循成效

系統化治理

將 AI 治理從被動、零散的應對措施，轉變為主動、系統化的策略功能

風險管理

注入全生命週期的風險管理思維到 AI 創新流程中，確保安全可控

價值創造

確保每個 AI 專案都與營運策略緊密結合，創造可預測的價值

以 ISO 42001 對應 DJBIC CSA 負責任 AI 題組

將九大 AI 政策承諾與實踐方案轉化為 ISO 控制措施，作為得分利器

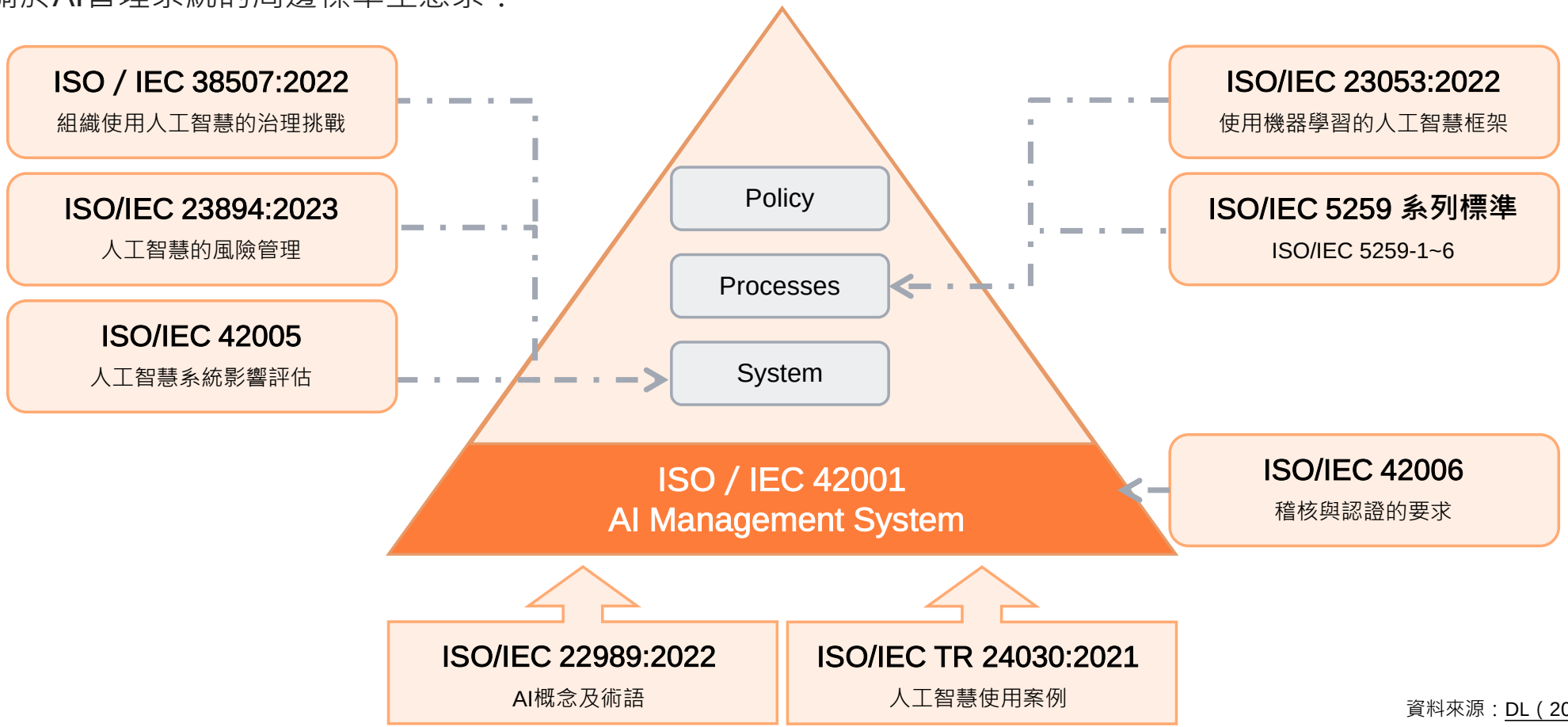
標普全球 ESG 評分（S&P Global ESG Scores Methodology）採 CSA 問卷評估，大幅提升「負責任 AI」門檻，要求公開 9 大政策承諾、實踐方案及外部驗證；取得 ISO 42001 驗證，是證明企業承諾已系統化、流程化且具第三方公信力之最佳佐證。

DJBIC CSA 負責任 AI 政策要求與 ISO 42001 對應（例舉參考）

CSA 政策	ISO 42001 條款與控制措施	例舉具體管控作為
尊重資料隱私	Annex A.7.3、Clause 4.1	AI 採購合約納入隱私條款，確保供應商符合 GDPR 與個資法
資訊安全防護	Annex A.10.3，結合 ISO 27001	AI 系統納入 ISMS 監控，防範模型遭竄改及資料外洩
避免潛在偏見	Annex A.9.3、A.5.4	導入第三方 AI 前，要求供應商提供公平性報告，建立內部抽查
保持真人介入	Annex A.9.2、A.6.2.6	嚴禁重大決策全自動化，系統流程圖明定真人覆核控制點
確保透明度與可解釋性	Annex A.8.2	使用者介面強制顯示「AI 生成」標籤，高風險系統提供決策邏輯說明
建立明確之責任歸屬	Clause 5.3、Annex A.3.2	指定高階主管負責 AI 營運，建立跨部門 RACI 矩陣
定義 AI 使用邊界	Annex A.9.4、Clause 4.3	制定《AI 工具員工使用規範》，嚴禁未授權高風險應用
具備低生態足跡	Annex A.5.5、Clause 4.1	採購 AI 服務納入供應商 PUE 值與綠電比例評選
不使用被禁止之系統	Clause 4.2、6.1.1	建立 AI 採購黑名單，禁用社會評分、未授權之生物辨識系統

《人工智慧管理系統》的標準生態系

目前ISO/IEC 42001標準雖然已經發行，但是與之相關的周邊標準，部分還在制定中。
關於AI管理系統的周邊標準生態系：



資料來源：DL（2024）

資誠獲得ISO 42001驗證 引領企業建立AI治理框架

資誠聯合會計師事務所（PwC Taiwan）於2026年5月25日正式取得「ISO / IEC 42001：2023 人工智慧管理系統（AIMS）」之國際標準驗證，成為台灣專業服務機構首家取得此國際標準驗證之組織。

對於本次獲證之戰略意義，資誠聯合會計師事務所數位資訊長張晉瑞表示：「人工智慧之發展已從單純之技術演進，躍升為董事會層級之營運風險考量。資誠本次取得ISO / IEC 42001：2023驗證，不僅是落實內部作業對資訊安全與模型運作品質之最高標準，更期望透過以身作則之實踐，為台灣市場建立可信賴之標竿。我們深信，唯有將嚴謹之治理框架深植於組織文化，方能在擁抱創新之際，穩健控管潛在風險，贏得利害關係人之長遠信任。」

針對企業落地實踐之挑戰，資誠智能風險管理諮詢公司董事林于翔指出：「面對各國日趨嚴格之監管趨勢，企業在推動AI專案時，往往面臨技術落地與法規遵循難以對應之困境。資誠顧問團隊協助客戶將抽象之國際規範，轉化為具體之營運流程。從跨系統協作介面之梳理、委外合約與內部文件之檢視，乃至於突發事件應變機制之建立，以客觀且中肯之視角，陪伴企業量身打造專屬之管理架構，進而全面強化組織之科技韌性。」

資誠將自身營運環境作為首要實踐場域，落實嚴謹之AI治理標準；藉由此次獲證，進一步印證資誠具備充分之實務量能，能協助台灣企業妥善應對複雜之技術風險，並穩健執行AIMS導入專案。

建立可擴展且合規的永續 AI 架構

資誠提供一套涵蓋 AI策略規劃、應用導入與治理建置的整合式顧問服務，從企業策路、業務需求、資料基礎、技術能力與風險治理等多元面向進行整體規劃，協助企業別AI 優先投入領域，並建立一套可擴展、易管理、合規範且能永續發展的AI 推動模式。

第一階段：AI成熟度分析

我們將從策略、應用、資料、開發與營運等關鍵構面，為企業進行全面性的AI能力與組織整備度評估，協助企業掌握現況、鑑別關鍵能力落差，並找出在資料、流程、技術與治理上的主要瓶頸，作為後續策略規劃與資源配置的堅實基礎。

第二階段：AI策略藍圖規劃

基於成熟度分析的結果，資誠將協助企業規劃未來三至五年的 AI發展方向，並依據商業價值、導入難度、資料成熟度、技術可行性與風險影響，擘劃出清晰的投資優先序位，從而辨識出短期內可收立竿見影之效的應用，以及中長期具高潛力的轉型契機。

第三階段：AI 應用試點導入

資誠將協助企業擇定一項兼具代表性與商業價值的應用場景作為試點，例如 AI智能助理、報告自動生成、市場資訊摘要或內部知識查詢。透過試點導入，可務實地驗證技術可行性、資料整合能力、模型輸出品質、使用者接受度與實際業務效益，並為後續規模化推廣累積寶貴經驗、奠定成功基石。

第四階段：AI 治理機制建置

我們協助企業建立完善的 AI 應用案例清冊、風險分級制度、審查流程、治理角色、責任分工、持續監控與回報機制，確保 AI 應用在導入與擴展的過程中，能做到「創新與風控並行」。
透過治理機制的建立，企業可在推動 AI 創新的同時，有效管理模型風險、資料風險、資安風險與監理合規要求。

核心服務涵蓋五大架構

1

商業及營運流程梳理

透過工作坊，在四至六週內完成企業現有數據資產的全面盤點現有 As-Is 流程、系統整合關係圖譜與優先強化建議，為後續架構設計提供可信的決策基礎。

2

數位平台建置

結合 Lakehouse 架構，搭配主流雲端平台（AWS、Azure、GCP）或混合雲環境，建置高效能、可擴展的企業數據中台及 戰情數據儀錶版，支援批次處理與即時串流雙模數據管道，確保 AI 模型能夠取用高品質的訓練與推論數據。

3

AI Agent 導入

部署客製化 AI Agent 群組，主動整合跨系統數據，執行異常預警、情境模擬與決策建議。應用場景涵蓋經營例會助理、成本異常偵測、供應鏈風險預警與跨系統自然語言查詢等，直接嵌入日常管理流程。

4

經營管理工作坊

提供客製化數據素養培訓課程、數據產品負責人認證計畫，以及持續性的數據治理健診服務，確保技術投資能夠轉化為長期組織能力，而非隨人員流動而斷鏈。

5

企業數據治理框架導入

以 DAMA-DMBOK 國際標準為基礎，結合台灣個資法、金融監理法規等在地合規要求，建立涵蓋元數據管理、數據血緣追蹤、數據目錄與存取權限控管的全方位治理體系，並提供角色責任矩陣（RACI）與治理流程 SOP 範本。

Thank you

pwc.com