moda 數位發展部

公部門人工智慧應用參考手冊

Public Sector AI Playbook

文件修訂歷史

版本	變更內容摘要	頁數	提供日期
V1.0	初版訂定	95	115.01.28

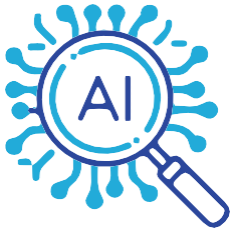
目錄

緣起

隨著人工智慧（Artificial Intelligence，以下內文簡稱 AI）發展應用越發多元，數位發展部也開始思考如何推動各機關積極運用 AI 技術，以對外提供民眾更便捷的數位服務、對內提升機關行政及業務運作效能，並且進一步落實數位涵容，確保所有人都能平等地受益於 AI 技術的發展。經研析，我們發現機關推動 AI 應用初期有三大痛點：

1. 機關同仁 AI 知識不足：

同仁對於 AI 概念的掌握度不足，無法了解 AI 的潛在應用機會；AI 應用於公務機關業務案例不足，難以類比到現有工作流程，從而限制了同仁對 AI 技術運用機會的想像。



2. 可參考的 AI 案例分散：

目前暫無符合機關需求之AI專案介紹且 AI 專案與傳統資訊專案存在差異，如何調整現有運作模式以順利地規劃管理 AI 專案成為一大挑戰。

3. 無 AI 專案實作導向指引：

現階段尚無供政府機關參考的明確實作面說明，已頒布者多為原則性、禁止性之 AI 規範，因此在實際 AI 導入階段缺乏操作指引。

故數位發展部撰寫本指引，期以簡要之說明，幫助各機關了解AI、促進對 AI 應用之想像，並作為 AI 導入階段時的參考手冊，提供機關 AI 專案流程指引，並在執行上提供具體且可參考之建議。

01 AI 概念介紹

章節摘要

- **AI 與生成式 AI 的差異**
- **AI 與生成式 AI 可以協助的任務類型，以及應用上存在的既有限制**
- **常見 AI 技術類型：監督式、非監督式、半監督式與強化學習**
- **國內外政府部門 AI 應用案例介紹**
- **AI 應用發展應遵循的原則**
- **AI 應用相關的法律與規範**

第一章旨在從科普的角度幫助機關同仁了解 AI 的基本知識。首先，討論 AI 的定義，闡明什麼是 AI 以及它的核心原理和功能。接著，探討 AI 的可能與限制，展示 AI 與生成式 AI（Generative Artificial Intelligence，以下內文簡稱生成式 AI）技術的發揮空間，同時也指出其存在的侷限性和風險。隨後，文章介紹 AI 的不同類別，幫助讀者理解各種 AI 技術的特點和應用範圍。

在應用案例部分，本文分享 AI 在公務業務優化、政府數位服務體驗優化與數位涵容精神的落實，提供具體的例子來展示 AI 的影響力。此外，此段落講述 AI 技術面臨的挑戰，如資料隱私、安全性和倫理問題。

最後，本文介紹與 AI 相關的法規和政策，幫助政府機關同仁了解當前法律環境下 AI 的發展方向，並符合法律以及既有規範。這些內容共同構成了一個全面而易懂的 AI 知識入門指南。

知識點：

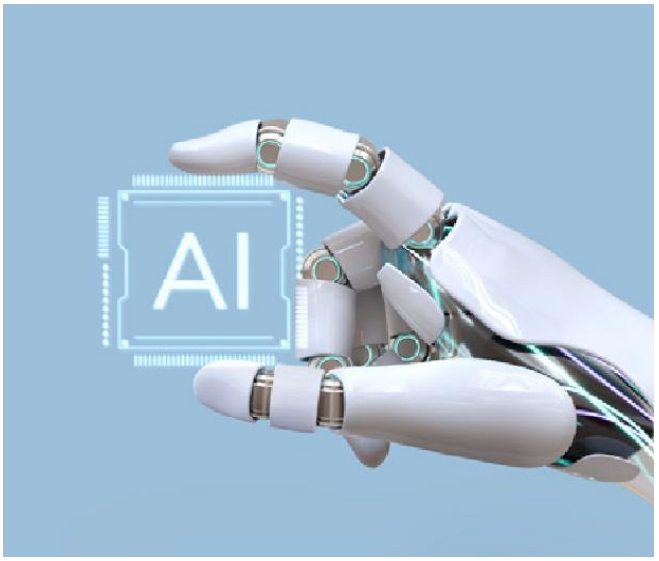
數位涵容 (Digital Inclusion)

運用科技縮減數位落差，確保有障礙人士及弱勢族群能享有與一般民眾相同品質的政府服務。

1.1： AI 定義

AI 是什麼？

AI 為 Artificial Intelligence 的縮寫，翻譯為人工智慧。隨著 AI 技術層面發展與應用案例增加，不同國家、學界以及應用產業都對 AI 範疇提出不同的定義。依據我國人工智慧基本法第三條：人工智慧指具自主運行能力之系統，該系統透過輸入或感測，經由機器學習及演算法，可為明確或隱含之目標實現預測、內容、建議或決策等影響實體或虛擬環境之產出。



* 來源：Freepik

什麼是生成式 AI ？

生成式AI，英文為Generative AI，縮寫為GAI。生成式 AI 並非全新產物，而是AI技術中的一個分支。與傳統 AI 模型不同，並不限於分類或預測，生成式 AI 能根據使用者輸入的提示，專注於生成多種形式的輸出，包括但不限於文字、圖片、音樂和影片等。 2022 年以生成式AI為核心的ChatGPT 公開後，大眾提高對於 AI 應用的關注，政府機關也開始思考運用生成式 AI 的可能性[1]。

知識點：

什麼是模型（Model）？

模型是一種用來簡化和模擬真實世界的工具。在機器學習中，模型是透過分析大量資料而建立起來的數學公式，這個公式可以用來進行預測或做出決定。

1. Generative AI Framework for HMG, GOV.UK （2024）

1.2：AI 的幫助與限制

AI 應用固然讓人充滿期待，但仍需了解 AI 應用的強項，以及現階段存在的限制。以下將以「AI 可協助的任務」與「AI 的限制」分別列示說明，也特別將生成式 AI 與廣義 AI的具體差異分別進行撰寫，提供後續服務發想時參考。

AI 可協助達成的任務

- 分析大量複雜的數據，分類、趨勢預測以及分群。
- 分析人類使用文字，解讀、產生以及處理文字資料。
- 辨識、追蹤和測量影片與圖片內容。
- 語音識別，將說話轉為可閱讀文字。

AI 無法達成的任務

- 進行全然的發想創新。
- 在沒有資料的情況下，提供好的分析結果。
- 在所有情況下，都提供最好的分析。

* 來源：KPMG 整理

AI 可協助達成的任務

1. 一次性分析大量複雜的資料，進行條件分類、趨勢預測以及分群。對應案例為群集分析(Clustering)，其為 AI 模型的一種，能將不同政府數位服務的使用者，進行相似**分群，找到需要相似類型服務的民眾。**
2. 分析人類使用文字，來解讀、產生以及處理文字資料。對應案例為自然語言處理(NLP)，**是進行翻譯、聊天機器人的基礎。**
3. 辨識影片與圖片內容，代替人眼去做辨識、追蹤和測量的任務。對應案例為電腦視覺（Computer Vision）領域，**能協助進行相片中的地址辨識，無須手動登打輸入。**
4. 語音識別技術，將說話內容轉為可閱讀的文字。對應案例為語音辨識（Speech Recognition）服務，可以將須公開之會議對話內容，轉成逐字稿。

AI 的限制

1. AI 須透過讀取資料進行預測生成，並非憑空進行發想創新。
2. AI 在沒有大量資料的情況下，無法提供合適的應用分析結果。
3. AI 無法在所有情境下，都提供最精準的分析。例如，並非所有 AI 對數字運算皆是以全運算方式進行，可能無法精確地回答數學問題。

生成式 AI 可協助達成的任務

- 更快的提供服務
- 提高日常工作效率
- 加強政府資訊的可取用性
- 更有效率的執行翻譯、重點摘要、名詞定義查找等任務

生成式 AI 無法達成的事項

- 很精確無誤的回應
- 具道德且無偏見
- 提供高專業度的建議
- 最新資訊答覆
- 記住過往所有對話紀錄

* 來源：KPMG 整理

生成式 AI 可以幫你達成的任務

1. 更快的提供服務：生成 AI 可以透過自然語言理解及數據分析，提供即時的客戶服務，處理常見問題，減輕人力負擔。
2. 提高日常工作效率：生成式 AI 可協助草擬信件與部分文件初稿，加速文件撰寫流程。
3. 快速取得現有政府資料：讓大量、不易快速理解的政府研究報告或網頁資訊，透過生成式 AI 工具作為新版索引工具，讓資料能被更多人查找使用，提升資料的可取用性。
4. 更有效率的執行翻譯、重點摘要、名詞定義等任務。

生成式 AI 的限制

1. 無法提供精確無誤的回應：生成式 AI 常有幻覺（Hallucination）產生，發生錯誤回覆或是編造資訊的情況，應再次檢核回應內容。
2. 潛在道德與特殊偏見問題：生成式 AI 的回應建立於訓練資料之上，根據訓練資料與提示（Prompt）指令，有可能產生具有偏見、冒犯以及不適當的用語產生。
3. 無法提供高專業度的建議：除非有用特定領域的資料做訓練，否則生成式 AI 受限於訓練資料，僅能提供一般性建議。另外，在法律、醫學以及其他要求極為精確之領域，不應將生成式 AI 當作專家的替代方案。
4. 可能無最新時效資訊：因其受限於訓練資料來源的限制，生成式 AI 可能未能包含最新資訊而導致回答有誤。
5. 無法記住過往所有對話紀錄：生成式 AI 產品可能在對話過程中，會遺忘先前對話的內容，並提供不連貫的答覆。
6. 無法有邏輯的解析生成式 AI 回覆內容：生成式 AI 是基於黑箱性質的神經網路應用，以致部分產出的對話、圖像與影音會無法明確考究。若將 AI 產出用於決策，在判別邏輯上會遇到挑戰。
7. 數字運算可能有誤：部分生成式 AI 為用來設計生成語言，而非進行精確數據計算，數字運算可能並非以全運算方式進行，因此模型在生成涉及計算的文本時，可能會產生錯誤。

知識點：

什麼是幻覺 (Hallucination)?

在 AI 領域，幻覺是指生成式 AI 模型產生看似真實，但其實不正確或誤導性的結果。導致錯誤的原因很多， 包括訓練資料不足、模型存在錯誤的假設或用於訓練模型的資料有偏差等。檢索增強生成 (Retrieval Augmented Generation, RAG) 是幻覺的一種解決方案，RAG 透過結合資訊檢索和文本生成的技術，利用外部知識庫來提高回答的準確性和資訊量。

目前 AI 應用還是以輔助特定任務為主，無法達成全能或近似於人類自學的應用，AI的產出都是基於既有資料以及部分人類干預的結果。在知識很少的情況下，也能自動完成各種跨領域任務的通用型 AI 仍在持續發展中。

AI 仍有其限制，故評估解決問題的方案時，AI 有時並非最好的選擇。AI 固然可以提升解決問題的效率，但有時傳統方法可能更快更精確，需透過謹慎評估以選擇適當解方。

1.3：AI 類別介紹

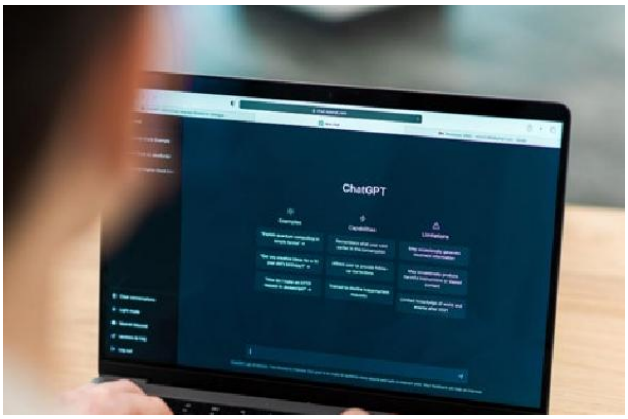
AI 是一個廣泛的技術，而機器學習是其中一個子領域。在此介紹常見四種類別的機器學習方法，能幫助閱讀者在初期對機器學習擁有全局視角，能根據資源限制，初步判斷可以使用機器學習模型類別。避免因機器學習模型多，不知從何選取作為解決方案的情況。有關模型選擇細節介紹，可參考 3.2 AI 專案流程章節。

* 來源：KPMG 整理

	監督學習 (Supervised Learning)	非監督學習 (Unsupervised Learning)	半監督式學習 (Semi-supervised Learning)	強化學習 (Reinforcement Learning)
簡介	使用有標籤的資料集來訓練模型，目標是學習輸入資料與輸出值（標籤）之間的關係	使用沒有標籤的資料集，目標是尋找資料中的結構或模式	結合有標籤和沒有標籤的資料來訓練模型，通常使用少量標籤資料和大量無標籤資料	模型學習在特定環境中採取行動，並在互動中將最大化獎勵作為訓練目標
使用時機	當你擁有已分類的資料，並希望 AI 系統能夠學習輸入與輸出之間的關係，以對新資料進行分類或預測時	當你不知道如何進行分類，或是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	當任務需要在動態環境中進行交互式學習時
運作步驟	1. 標記訓練資料並定義輸出變數 2. 將資料用於演算法訓練，學習輸入與輸出之間的關聯 3. 訓練好模型後輸入新資料，模型即可預測輸出變數	1. 給演算法提供未標籤的資料 2. 演算法自行推斷資料的結構 3. 演算法識別出類似屬性的資料群組，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 3. 訓練好的模型可開始運用於新範圍	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，目標是最大化可獲得的獎勵
案例	• 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（少量標籤影像、大量無標籤影像）	• 遊戲玩家(AlphaGo) • 自動駕駛汽車

模型技術概念解析

在了解完傳統 AI 的技術類型後，最可能有的疑問會是「那生成式 AI 會對應到前述哪個 AI 類別？」其實，廣義生成式 AI技術會同時應用非監督式學習、監督式學習與自監督式學習等技術。若聚焦於對話式生成式 AI 服務（類似於 2022 年推出的 ChatGPT 服務）討論，其初始訓練階段通常會使用非監督學習，能從大量無標籤數據中學習模式；而在模型微調過程中，經常會結合監督式微調和（或）結合強化學習技術。監督式微調通過使用有標籤數據，對模型進一步訓練，使生成式 AI 能更好地理解、執行具體任務和滿足使用者的需求。



*來源：Freepik

另一個常見的問題是「演算法與模型的差異為何」。簡單的說，演算法是針對輸入資料執行一系列運算或處理的過程，而模型是基於演算法與訓練過程的產物，能反映出演算法學習的成果。舉例來說，演算法從大量資料中學習到一套分類規則後，將其建立為模型，使用者後續輸入資料，模型將以同樣分類規則進行分類。

監督式與非監督式學習，差異點在於有無資料標籤，這會帶來甚麼樣的差異呢？資料有無標籤與模型應用相關，有標籤的資料代表資料中含有模型輸出的資訊；無標籤的資料代表資料本身未包含模型訓練後輸出的資訊。

知識點：

什麼是資料標籤？

資料標籤是一種能幫原始資料新增分類資料的資訊，資料標籤可以是數字或文字，目的是讓AI可以多一個能從中學習的分類。例如，在有很多不同品種貓狗的體重資料時，新增一個資料標籤標註該筆資料是貓是狗，將有助於AI做體重預測時，能納入此變數做為考量。

舉例來說，用包含收入的市民個人資料帶入迴歸模型，去尋找哪些個人資料與收入高度相關，是使用有標籤的資料進行預測，因為收入包含於其資料中。若是在一開始就未定義群體的情況下，想要透過數據規律，找出生活樣態接近的市民，則是使用無標籤資料。因為我們並未先設定特定市民樣態，代表既有資料中並不包含分類資料，之後的分類則為模型全新輸出的結果。

AI 的新進化：從「聰明助理」到「自主總管」

我們現在熟知的生成式 AI，像是 ChatGPT，它們最強大的地方是「創造內容」。可以把這類 AI 想像成一位知識淵博的「聰明助理」。它們的功能是：根據使用者的要求，快速產生文章、圖片、程式碼，或是一份詳細的旅遊攻略。它們是被動地一個口令一個動作，只專注於「產出內容」。

代理式 AI (Agentic AI)：賦予 AI「自主性」

AI 的下一步發展，就是讓它們從「助理」進化成一位「具備自主行動力的總管」，這就是所謂的代理式 AI (Agentic AI)。

「Agentic」這個詞的核心精神，強調的是 AI 所擁有的「自主性(Agency)」。它不再只是等待指令，而是能主動思考、規劃並執行複雜的多步驟任務，直到目標達成。這類 AI 的功能重點在於實際應用內容，自主執行複雜目標；行為模式是自主決策、目標導向，具備學習和適應性。舉例來說，生成式 AI 只能產出一份旅遊攻略，而代理式 AI 則能主動預定飯店、購買景點票券，並排定行程。

具備自主性的 AI 代理 (AI Agent)

當一個 AI 系統具備了這種「代理式能力(Agentic)」後，我們就稱它為 AI 代理(AI Agent)。AI 代理是一種能夠自主感知環境、記憶、規劃並執行任務以達成特定目標的人工智慧系統。它們能夠自主規劃，將一個大目標拆解成多個小步驟；能夠運用工具，知道何時需要上網搜尋或操作其他工具，並且具備自我修正能力，如果在第一個執行步驟失敗了，會自己反思並嘗試新的策略，直到任務成功。

市場趨勢

Gartner 研究報告指出，這類具備任務導向的 AI 代理 (AI Agent) 將在企業應用中爆炸性增長。預計到 2026 年底，將有 40% 的企業應用程式會整合 AI 代理，遠高於 2025 年不到 5% 的比例。這代表 AI 的未來趨勢，將是從一個「聰明的工具」轉變為一個能真正自動化工作、替人們處理複雜事務的「自主幫手」。

*來源：Freepik

以下為簡易範例，說明生成式 AI、AI 代理、代理式 AI 的具體差異：

類型	具體行動與能力舉例
生成式AI	反應式：根據預先定義的範本生成文件、程式碼，或是摘要複雜的網路資訊等功能。只能產生文本，無法執行動作。
代理式AI	高自主性：假設人類給定一個目標「在30天內完成95%工作量」代理式AI會執行： 1. 分解與協調：將任務分配給多個 AI 代理。 2. 動態調整、學習改進：若發現 A 代理速度慢於 B 代理，將會自主調用更多算力給 A 代理，或重新分配待處理的案件以確保總目標進度。

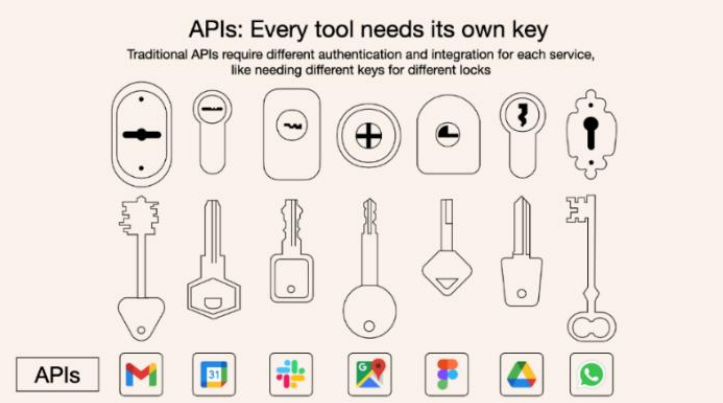
MCP：AI與應用程式的溝通協議

那麼，具體來說該如何使生成式 AI 變成 AI 代理呢？原先的生成式 AI 就像是擅長思考的「大腦」，沒有與「身體」結合，導致它不能靈活運用各種工具。AI 之所以未能連接工具，是因為這些工具說著各種不同的語言，導致 AI 和工具之間無法有效溝通。

傳統上，應用程式之間的溝通需要透過 API 的串接，但是 API 像是一個個不同的、精密的鎖，每個 API 都需要不同的鑰匙才能解開。例如，API 間的資料格式包括了 JSON、XML 等，驗證方式則有 API Key、OAuth等，不同的 API 需要使用不同的格式、驗證方式及錯誤處理機制，等於每串接一個 API，程式碼都要重寫一次。

為解決這個問題，美國 AI 公司 Anthropic 在 2024 年開發出了 MCP (Model Context Protocol, 模型上下文協議)。這是一套開源協議，定義了 AI 和工具之間的溝通規則，使彼此能順暢溝通，即讓 AI 代理能夠順利調用外部工具。

MCP 主要透過三個構件提供功能：Host（主機）、Client（客戶端）、Server (伺服器)。Host 負責理解使用者意圖，Server 最主要負責執行以及工具的使用，Client 則扮演兩者間的橋梁，若 Server 執行時發現使用者提供的資訊不夠充分，可以透過 Client 返回詢問使用者；Client同時也負責控制安全邊界，避免Server存取使用者並未許可存取的資料。



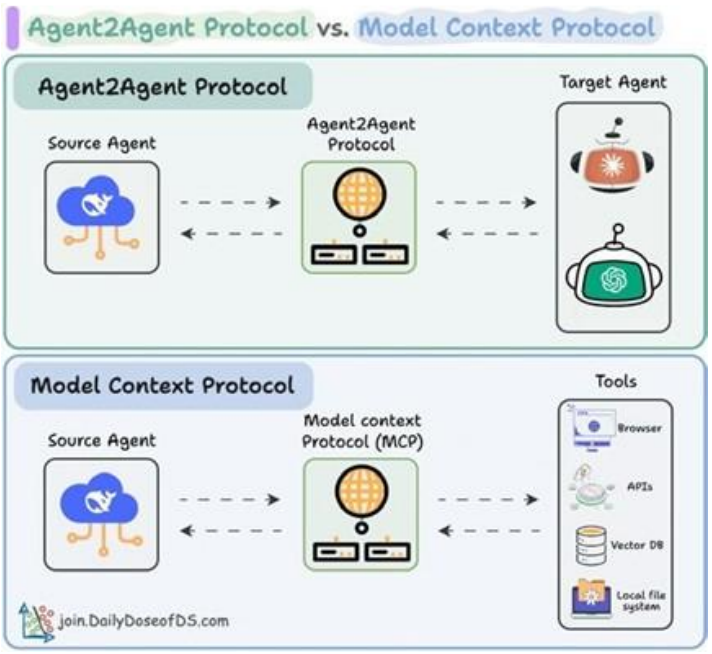
API的限制：像是一個個不同的、精密的鎖

*來源：Norah Sakal

A2A：AI代理間的溝通協議

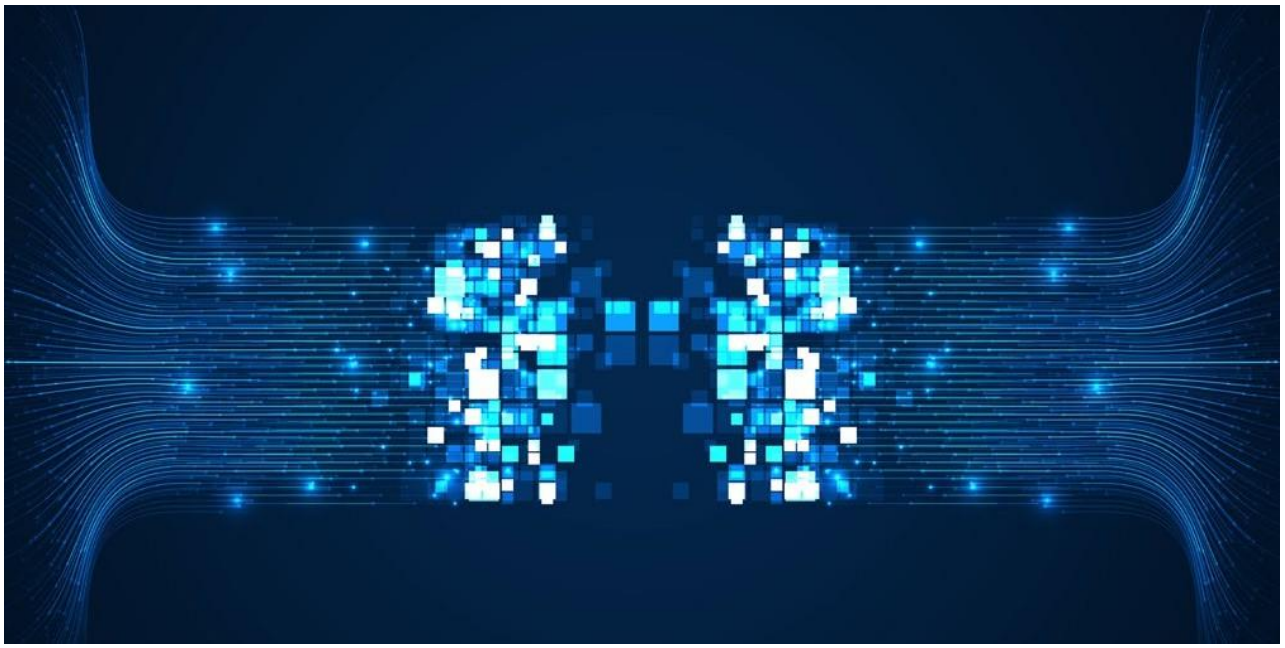
目前我們了解了 MCP 如何實現 AI 與外部工具的溝通，使 AI 代理可以順利地應用工具達成特定目的。但如果遇到龐雜的任務，單一代理無法完成，就需要透過多個代理共同執行任務。

不過，這些 AI 代理間由於語言不通，因此無法順利協作完成任務。例如，由Gemini驅動、專精數據分析的AI代理，無法與一個由OpenAI開發、擅長報告撰寫的代理溝通。如果要達成這件事情，必須針對每個不同代理撰寫膠水程式碼(Glue Code)來進行整合。



*來源：Avi Chawla

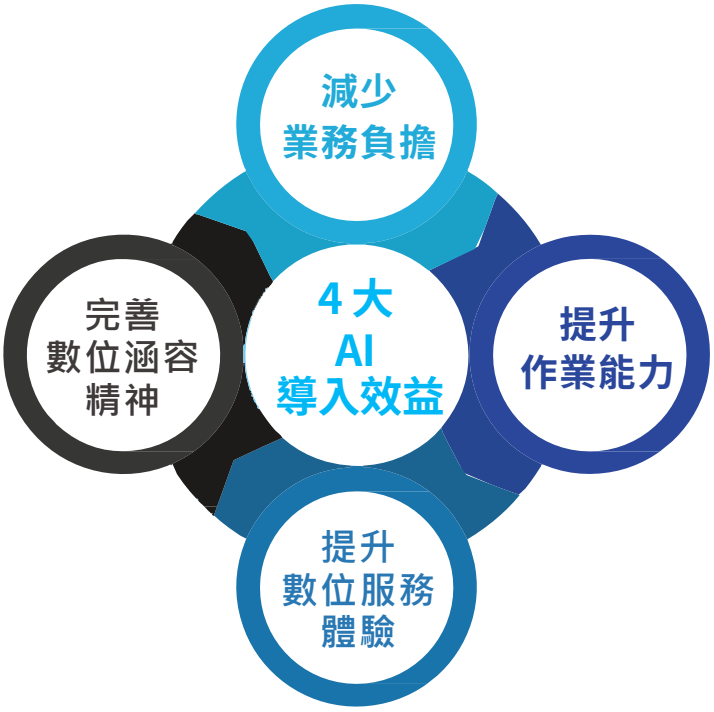
為了解決這個問題，Google在2025年推出A2A (Agent to Agent)協議，讓代理之間可以互相溝通，降低AI協作門檻，讓工作統籌功能更流暢。A2A技術專注在「Client-遠端」的溝通。MCP Host了解使用者的需求後，分析哪些代理適合執行這項任務，而後調動MCP Client與遠端的代理聯繫，達成跨代理的應用。



*來源：Freepik

1.4：AI 應用案例介紹

目前政府機關應用 AI 技術，依照場景可分為兩大類，分別是協助公務同仁以及提升民眾服務體驗。前者能帶來「減少業務負擔」及「提升作業能力」的效益；後者則可「提升數位服務體驗」，以及「完善數位涵容精神」以下介紹國內外近期實例，未來將會持續新增案例。



減少業務負擔

1. 以圖搜圖商標檢索系統[2]：

- 透過上傳商標圖片，快速尋找前 1,000 個已申請或是相似的商標，讓民眾評估是否進行商標申請。改變過往人工比對不易，導致倉促申請而遭退件的情況發生。
- AI 應用：透過 AI 影像辨識技術，分析上傳圖片與既有圖片的相似度。
- 單位：經濟部智慧財產局

2. 稅務智慧客服[3]：

- 提供稅務諮詢服務，民眾可於網頁客服視窗以簡要敘述提問稅務問題，無須等候真人客服，24小時皆可得到對應解答。
- AI 應用：能辨識簡易稅務相關語意，並提供對應文字回答。
- 單位：臺北市稅捐稽徵處

* 來源：Freepik

3. 法庭中文語音辨識系統[4]：

- 因應國民法官制度的需求，即時產出逐字稿筆錄，並整合司法文書編輯、數位錄音、錄影回放等功能。

2. 智慧局推出「以圖找圖」神器用AI檢索「圖形商標」（2024）
3. 北市稅處智能客服即日起上線（2019）
4. 法庭語音辨識系統上線正確率超過9成（2023）

- AI 應用：國、英語辨識模型準確率高達90%以上，並支援大部分法律專用詞彙，台語辨識準確率也已達85%以上。未來，司法院計畫持續提升模型辨識的正確率、語者分離精確度，以及相關功能的完善。
- 單位：司法院

提升作業能力

1.銀髮安居[5]：

- 集合政府與開放資料等超過 1.5 億筆跨機關資料，以機器學習方式找出「銀髮安居高度需求名冊」，供內政部「包租代管」及衛生福利部「老人照顧」等政策業管單位運用。
- AI 應用：透過戶籍地址、座標點位與衛生福利部長照、中低收入戶等資料，配合 6 大生活面向、18 項指標，使用「層級分析程序法」配予各指標權重，再透過機器學習的方式篩選出最需要協助的 1% 老人。
- 單位：內政部

2.AI 預測颱風強度[6]：

- 提供氣象人員客觀且自動化預測颱風強度、中心位置的方式，減少過去因偵測技術導致的主觀人為誤差。
- AI 應用：使用颱風衛星影像訓練 AI 模型，判斷近中心最大風速與中心氣壓的模型。
- 單位：交通部中央氣象署

* 來源：OpenArt.ai

3.MCP x AI Agent 應用，政府資料供民間使用[7]

- 過去，市場研究機構若要查詢各國統計資料，必須逐一進入各國官方網站並下載資料。然而，各網站的資料格式與圖表呈現方式差異極大，導致搜尋與整合過程面臨相當大的困難。為此，2025 年Google 推出 Data Commons MCP Server，橫跨多國的官方統計資料庫，內容涵蓋人口、教育、健康、經濟等領域，來源包括 OECD、WHO、世界銀行等可靠組織。
- AI 應用：本工具進一步透過知識圖標準化，將資料蒐集到產出簡易報告整合進一個流程完成。
- 案例：Google 與非營利組織 ONE Campaign 合作，開發出「The ONE Data Agent」，專門處理健康金融的資料。使用者能夠用自然語言在幾秒內完成數千萬筆資料的搜尋、下載經整理過的乾淨資料集，並串聯資料視覺化工具繪製出圖表。

5. 銀髮安居 , 內政部數位中心
6. 施政目標, 交通部中央氣象署 （2023）
7. Google: Introducing the Data Commons Model Context Protocol (MCP) Server: Streamlining Public Data Access for AI Developers (2025)

提升數位服務體驗

1.新加坡案例：AI 聊天機器人簡化民眾陳情流程[8]

- 問題背景：民眾在反映市政問題時，常因不知道該聯繫哪個政府機關、或無法騰出時間聯繫政府單位而造成延誤，導致問題無法即時解決，進而影響生活品質，分析其原因為傳統的陳情處理方式效率低下，無法滿足民眾需求。
- 使用案例說明：新加坡政府開發了一個託管在 VICA 平台上的 AI 聊天機器人，用來專門處理市政問題，讓民眾可以在熟悉的通訊平台（如 WhatsApp 和 Telegram）上報告問題。AI 聊天機器人會利用預測能力，引導使用者提供必要的問題細節，並自動將問題分配給相關機關。這大大提高了居民陳情的便捷性，加快了問題解決的速度。

* 來源：Freepik

- AI 應用：AI 聊天機器人是一種應用自然語言處理和機器學習技術的對話系統，能夠理解並回應使用者的語音或文字輸入。這種系統可以簡化陳情過程，自動引導使用者提供必要訊息，並將民眾的意見精確分配到相應機關進行處理。

2.新加坡案例：個人化求職推薦系統

- 問題背景：求職者要在眾多的線上職位資訊中尋找合適工作，則需要耗費大量時間和精力，也容易錯過適合自己的職位，影響了求職效率和效果。
- 使用案例說明：新加坡政府推出的 JumpStart 平台，專門提供針對勞動市場的 AI 服務，如工作或求職課程推薦。這個平台利用各種資料源（如技能、簡歷、使用者點擊流程、申請和職位描述）構建模型，根據求職者的技能和偏好提供個人化的工作推薦。而在新加坡政府實施這些新模型後，平台的求職申請增加了 9.1%，成功的求職配對增加了 21.5%。

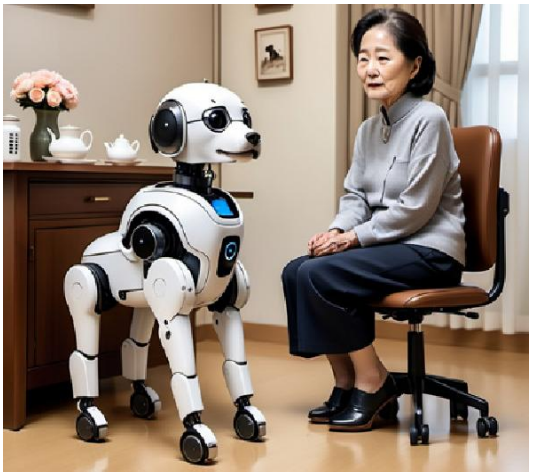
* 來源：Freepik

- AI 應用：求職推薦系統利用機器學習技術，根據使用者的技能、經歷和偏好，從大量職位資訊中推薦最適合的工作。這些系統能夠自動分析使用者資料並提供個人化建議，提高求職成功率。

8. AlphabotSG by VICA, Gov.sg （2023）

完善數位涵容精神

1. 南韓案例：伴侶機器人和護理機器人提升獨居老人的生活品質[9]

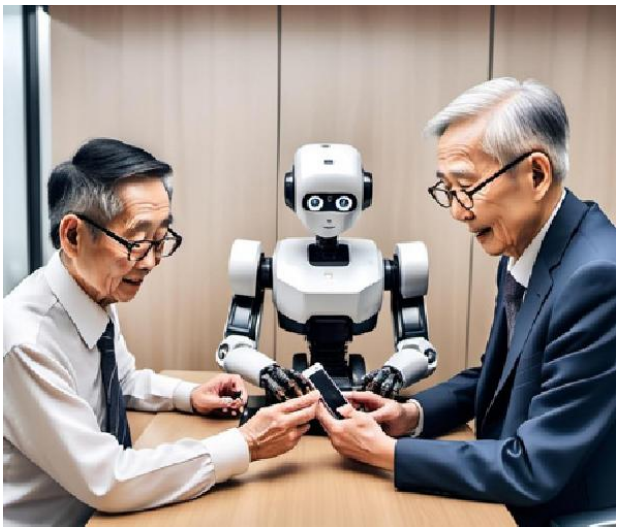


* 來源：OpenArt.ai

- 問題背景：獨居老人面臨心理和生理健康問題，傳統護理服務無法即時提供必要的支持。
- 使用案例說明：首爾市政府向獨居老人分發「伴侶機器狗」和安全管理設備， 2023 年已提供 430 隻機器狗，2024 年將再增加 50 隻。能透過機器狗監控老人的健康狀況，並在緊急情況下自動呼叫 119。
- AI 應用：伴侶機器人和安全管理設備通過語音指令和互動功能，提供情感支持和健康監控，提升老人的心理和生理健康。

2. 南韓案例：智慧機器人幫助老年人學習技術[10]

- 問題背景：許多老年人在使用智慧手機和數位自助式服務（如自助點餐機）時感到困難，這使他們無法充分利用現代數位服務，並感到與社會脫節，而傳統的教學方法效率低下，無法滿足老年人的需求。
- 使用案例說明：首爾市政府推出了一個 Liku 智能機器人教導老年人使用智慧手機和自助式服務的計畫。這個計畫在 17 個設施中為 3,000 名參與者提供培訓。Liku機器人利用語音指令和手勢，指導老年人逐步學習如何使用 KakaoTalk （南韓普及通訊軟體），幫助他們克服操作上的困難。根據試點計畫的結果，87% 的參與者對這項服務表示滿意，83% 的人希望繼續參與。



* 來源：OpenArt.ai

- AI 應用：智能機器人 Liku 利用語音指令、手勢和特別修改的智能手機，指導老年人學習如何使用通訊軟體。這些機器人通過與中央伺服器連接，能回答約 200 個問題，並提供當前天氣等資訊。Liku 擁有臉部識別和語音回應功能，能與使用者進行互動。

9. Elderly Care Becomes Smart in Seoul by Using Robots and AI Technology to Keep Seniors Healthy and Safe, Seoul Metropolitan Government （2024）
10.Robots help seniors learn to use technology in South Korea, United Press International （2020）

1.5：AI 應用原則

依據我國人工智慧基本法第四條：政府推動人工智慧之研發與應用，應在兼顧社會公益、數位平權、促進創新研發與強化國家競爭力之前提下，發展良善治理與基礎建設，並遵循下列原則：

1. **永續發展與福祉**：應兼顧社會公平及環境永續。提供適當之教育及培訓，降低可能之數位落差，使國民適應人工智慧帶來之變革。
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，並允許人類監督，落實以人為本並尊重法治及民主價值觀。
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，並採用資料最小化原則；同時在符合憲法隱私權保障之前提下，促進非敏感資料之開放及再利用。
4. **資安與安全**：人工智慧研發與應用過程，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**：人工智慧之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，進而提升人工智慧可信任度。
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，不應對特定群體造成歧視之結果。
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。

1.6：AI 應用涉及法規指引

截至 2025 年 12 月為止，有關機關提出之 AI 相關應用指引與規範整理如下，另 AI運用亦有可能涉及其他法規範，規劃時須一併納入考量。

法律規範

截至 2025 年 12 月為止，我國已制定人工智慧基本法作為上位原則。另機關於 AI 使用上，亦受相關法律規範。

1. 《人工智慧基本法》
2. 《個人資料保護法》
3. 《著作權法》
4. 《商標法》
5. 《專利法》
6. 《營業秘密法》

行政指導

1. 《人工智慧（AI）產品與系統評測參考指引（草案）》- 數位發展部數位產業署，於 2024 年 3 月發布。
2. 《行政院及所屬機關 (構) 使用生成式 AI 參考指引》- 國家科學及技術委員會，於 2023 年 10 月發布。
3. 《臺北市政府使用人工智慧作業指引》- 臺北市政府，於2024 年 09 月發布。

產業指導

1. 《金融業運用人工智慧（AI）指引》- 金融監督管理委員會，於 2024 年 6 月發布。
2. 《文化藝術應用生成式AI指引》- 文化部，於2025年7月發布。
3. 《應用人工智慧/機器學習技術之醫療器材軟體預定變更控制計畫（Predetermined Change Control Plans, PCCP）申請要點暨撰寫說明指引》– 衛生福利部，於2024年9月發布。

政策白皮書

1. 《晶片驅動臺灣產業創新方案》- 行政院，於 2023 年 11 月發布。
2. 《金融業運用 AI 之核心原則與相關推動政策》- 金融監督管理委員會，於 2023 年10 月發布。
3. 《臺灣 AI 行動計畫 2.0》- 行政院，於 2023 年 2 月發布。

02 AI 服務評估

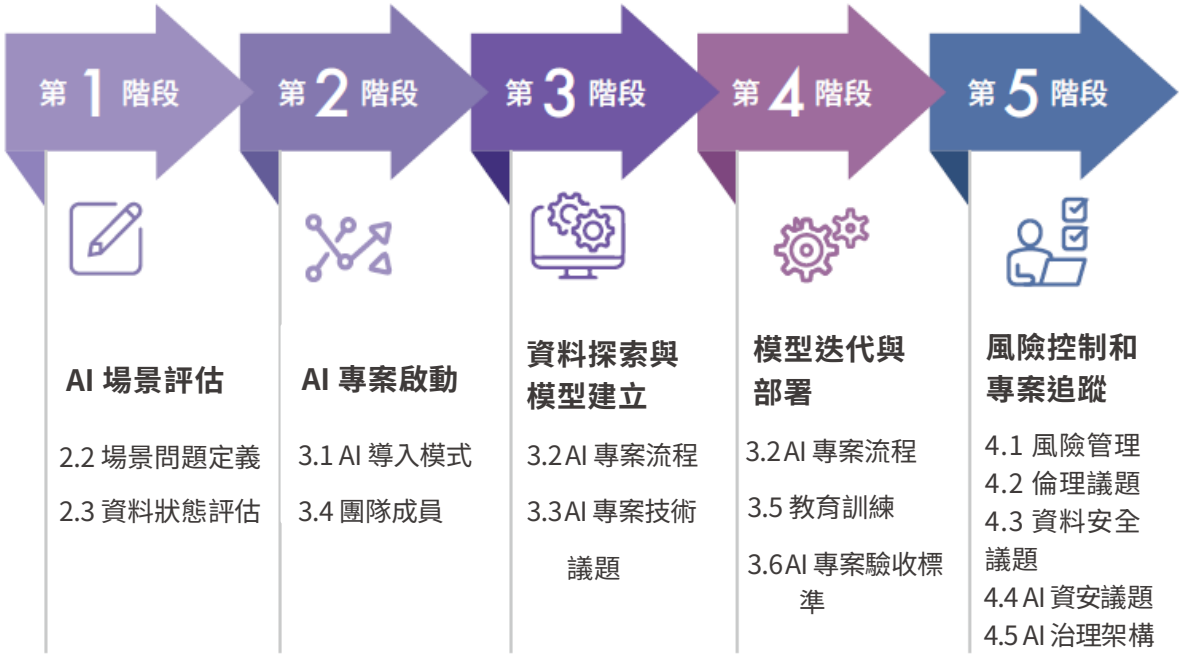
章節摘要

1. **AI 專案的完整生命週期**
2. **遇到問題時，如何去判斷 AI 是一個解決問題的好方法**
3. **如何評估 AI 專案所需的資料**
4. **有哪些 AI 相關的政府公開資源可以學習應用**

AI 作為新興資訊解決方案具有非常大的吸引力，但如何判斷 AI 是否為最適方案，以及機關本身是否具備導入 AI 的條件，均為決定是否使用 AI 的重要標準。以下段落將依序從場景問題定義、資料狀態評估、常見 AI 模型與對應應用場景三個段落，協助同仁評估 AI 是否為最合適的解決方案，判斷機關是否已準備好導入 AI 技術，並提供常見 AI 模型介紹。各段落皆有對應的「檢核清單」，供同仁於 AI 專案評估過程中，檢查是否已完成相應工作。

2.1 ：AI 導入生命週期流程

為確保 AI 導入能發揮效益，我們提供一套 AI 導入生命週期簡介。從 AI 應用場景辨識，到 AI 專案成立並進行建置等階段皆包含在其中，也建議機關可將此流程融入整體規劃中。我們將 AI 導入生命週期分為五大階段，並設立對應關鍵目標與對應任務。本節將提供一個完整 AI 導入流程全貌，詳細執行細節於後續章節進行詳細說明。



第 1 階段：AI 場景評估

AI 導入皆應從現況所面臨挑戰出發，清楚定義欲解決問題後，再評估 AI 是否為合適的解決方案，以避免組織落入「為導入 AI 而 AI」的情況，問題定義可參考 2.2 場景問題定義章節。定義好問題後，應評估 AI 是否為合適的解決方案，除了考量 AI 技術的應用可能外，單位資料準備度是否符合 AI 應用前提也會是關鍵，可參考 2.3 資料狀態評估章節。執行專案時，也應在專案初期，將風險管理納入考量，風險管理將於後續 4.1 章節進行闡述。

- **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。

第2階段：AI 專案啟動

在決定啟動 AI 專案後，機關應決定將自行建立或招商採購。若決定招商採購，也需評估將購買既有服務，或由廠商依需求建置對應內容。以上決定，可參考 3.1 AI 導入模式章節。另專案啟動後，應依照 AI 模型生命週期規劃對應管理權責，並找尋具備相關專業之人員加入專案。有關專案人員與職責，可參考 3.4 團隊成員章節。建議機關應於專案初期，選定 AI 專案小規模測試之場域，以利未來開發部署後的測試應用。這個階段的關鍵是確保您的基礎設施和資源準備就緒，以支援未來的 AI 開發和實施。

- **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。
  - 建立專案驗收標準。

第3階段：資料探索與模型建立

模型開發前，將進行資料準備與探索，目的在於確認數據品質和數據的可用性，可參考 3.2 AI 專案流程中的資料前處理段落。接著進入正式模型開發建置，在這個階段，機關將開始建立並選擇 AI 模型，對不同的 AI 演算法進行測試和評估。這個階段的目標是建立一個基礎模型，以便在後續階段進行進一步的優化和改進。有關基本模型與解決方案介紹，可參考 3.2 AI 專案流程章節。完成初版模型並進行迭代。在這個階段，機關需要不斷測試 AI模型效能，依照測試結果，將模型版本做迭代。

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

第4階段：模型迭代與部署

AI 模型須不斷測試迭代，且須確認其後續在部署環境中能有穩定表現。待模型表現穩定後，機關將建立監控模型性能之流程，並依需求將模型與系統架構融合，確保它能夠正常運行並產生預期的結果。同時，需要確保模型的性能和效果得到持續的監控和更新，且後續使用者能掌握如何解讀模型或是操作該 AI 服務。完成專案後，應依照最初擬定之目標，進行專案驗收檢核。可參考 3.3 AI 專案技術議題與 3.5 教育訓練與 3.6 AI 專案驗收標準章節。

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

第5階段：風險控制和專案追蹤

在 AI 專案從建置到維運的過程中，您需要持續關注和管理可能出現的系統有效性及專案管理問題，可參考 4.1 的風險管理章節。另外，隨 AI 技術演進，AI 服務將面臨道德面、資料安全與資安的挑戰，故可參考 4.2 倫理議題、4.3 資料安全議題與 4.4 AI 資安議題章節，了解並持續監控 AI 可能產生的風險。

- **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。
  - 監控和管理風險，包括道德和資料安全問題。
  - 進行專案追蹤和改進，確保長期運作順利。

除了從 AI 導入生命週期進行說明，本指引也針對 AI 專案特色與 AI 專案推動提供具體引導。由於 AI 專案與傳統機關資訊專案存在差異，執行上有特別需留意之處，故撰寫 3.7 AI 與傳統資訊專案的差異，針對常見資訊系統建置案與 AI 專案進行比較，列舉 AI 專案執行上的常見錯誤。另 AI 專案推動過程，以「擴大 AI 應用」為目標，故建置後之營運效益追蹤，以及如何確保 AI 建置資源具備共用性並保有擴充彈性，將於 4.5 AI 治理架構議題中探討，同時提供如何建立良好治理制度之建議，為後續 AI 應用擴展做好準備。此外，4.6 後續專案推動將說明三項可強化 AI 專案推動的具體措施。

2.2：場景問題定義

在 AI 導入的開始階段，明確定義問題是關鍵的第一步。以下提供三步驟的問題定義過程，以確保 AI 的應用能達成機關目標：

* 來源：Freepik

步驟 1：明確定義欲解決的問題

盤點工作中遇到的痛點：從具體的工作痛點出發，去連結如何以 AI 來解決這個痛點，用以確保後續 AI 應用能帶來具體效益。根據經驗，常見 AI 可改善的痛點包含以下一個或數個面向：

提出問題：接著請將痛點轉化為一個具體的問題，並非單純描述痛點。一個具體的問題建議包含「清楚的目標」、「適當的問題範圍」、「可行動性」以及「可量化或可觀察的預期目標」。

清楚的目標：

❎ 不好的提問：我們該如何改善政府服務？

✅ 好的提問：我們該如何以 AI 或其他數位工具，提高檢核文件的效率？

適當的問題範圍：

❎ 不好的提問：如何提升公務員的工作效率？

✅ 好的提問：如何以 AI 在合適的應用情境下，提升公文撰寫的效率？

可行動性：

❎ 不好的提問：我們如何以 AI 直接取代日常業務？

✅ 好的提問：我們如何和 AI 協作，加速會議紀錄產出效率？

可量化或可觀察的目標：

- ✗ 不好的提問：我們要如何以 AI 提升業務效率？
- ✓ 好的提問：我們要如何以 AI，將民眾的回覆等待時間減少兩天？

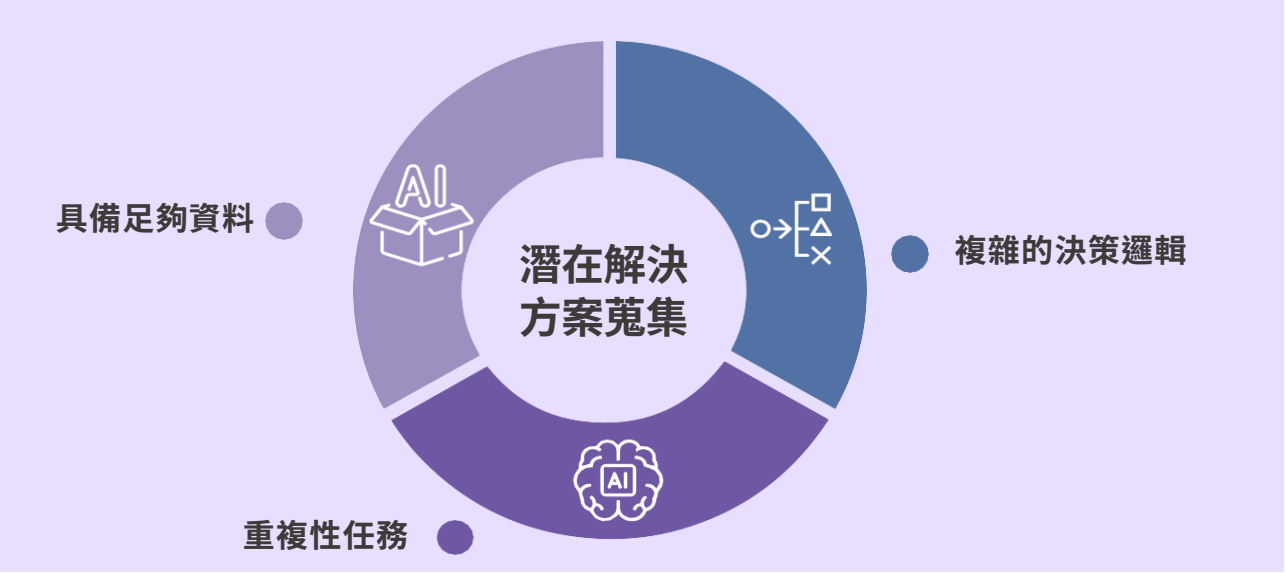
在導入規劃研擬初期，可能無法一次滿足所有好問題的條件，但在招標文件撰寫的過程中，承辦機關應對場景導入目標與預期效益有更高的掌握，以利後續方案選擇評估能有對應基準。

思考點：AI 導入可行性

在AI 專案的初期規劃階段，可試想：我們是否已釐清業務流程並劃分出各個階段？這有助於判斷哪些環節適合導入 AI，也可以初步評估導入的潛力，並提醒自己不要對模型有過高期待，特別是在資料不足的情況下，避免影響後續判斷。

步驟 2：潛在解決方案蒐集

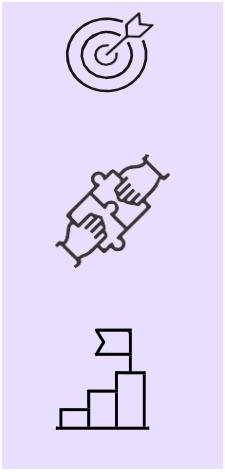
在找出痛點並明確問題後，應審視問題定義中的對應工作流程，判斷是否存在可應用 AI改善之面向。可參閱章節 1.4 AI 應用案例介紹與章節 3.2 AI 專案流程，尋找適合的 AI解決方案並評估其預期成效。若 AI 被納入最終解決方案，導入前通常具備以下條件：



1. **具備足夠資料：** 針對問題應用場景，已具備足夠可取用資料。
2. **重複性任務：**大範圍、量多且重複性高之任務。
3. **複雜的決策邏輯：**整體決策涉及大量複雜資訊比較，或尋找大量資料間的關聯規律，再依照分析的結果進行決策。

在潛在解決方案評估過程中，需留意切勿一味的追求 AI 導入，在進行解決方案選擇時，需一併考量傳統資訊解決方案（例如：RPA、傳統資料分析工具），並且需考慮現有資料基礎建設的成熟度、AI 方案與現有資訊系統整合，並留意後續擴大應用範圍的可能性。

步驟 3：定義導入目標成效



1. 關鍵指標：明確衡量改進工作流程的具體指標，如工時降低、人力精簡、AI 模型的精準度等。

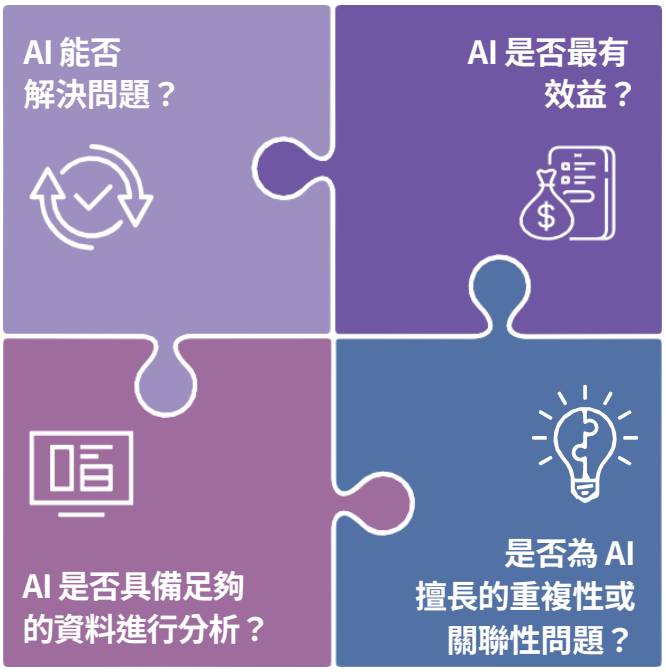
2. 掌握現況：了解機關目前的現況，作為導入 AI 成效之參考基準點，以利清楚地界定目前的工作狀態和改進幅度評估，這有助於後續評估AI解決方案的實際影響和效果。例如，目前每月花 50 小時執行某項任務，可做為導入 AI 後的比較基準。

3. 設定目標：定量設定導入期望達成的成果，這有助於設定所有利害關係人有正確的預期，並確保導入實施期間能逐步衡量是否成功達標。例如，每月減少 30% 的工時的專案目標。

四大思考面向

機關在依序完成問題定義、潛在解決方案蒐集以及導入目標設定等三步驟之後，建議採取以下四個面向來思考，AI 是否為最佳解決方案。

1. **AI 是否能解決機關所面臨問題？**
2. **AI 是否為效益最佳的解決方案？**
3. **是否具備足夠的資料進行分析？**
4. **想要解決的問題是否為重複性的任務，或需要了解大量資料中隱含的關聯性。**



完成「問題定義」、「潛在解決方案蒐集」與「導入目標設定」三步驟後，可以確保機關能掌握 AI 導入的目標、AI 方案的適配度與 AI 導入的常見績效指標。這將有助於機關進行後續專案需求說明書撰寫，並於採購階段選擇最適合的方案。而最後「四大思考面向」，能協助機關在方案選擇上，作為篩選評估的最終標準。

政府的 AI 應用尚處於實驗階段，AI 技術本身也仍在快速發展中，機關採行 AI 應用在短期內或許很難獲得具體效益。因此，建議機關在探索具不確定性的數位技術及其潛在效益時，優先進行小規模概念驗證，確認 AI 應用可達到預期效果。此外，若已存在性質類似且非使用 AI 技術的數位服務，機關應更加謹慎評估改用 AI 技術之必要性，避免為了 AI 而 AI，忽略其實際效益之情形。

思考點：目標與效益設定

在規劃 AI 專案時，可試想：我們希望帶來哪些具體效益？除了提升績效，也可以從減少人力與時間、提升工作能力、降低成本，或促進人機協作等角度來評估。另一方面，也別忘了思考模型需要多高的準確率，才能真正發揮效用，避免後續成果與期待落差太大。

2.2：場景問題定義檢核清單

編號	檢核項目	檢核結果	檢核說明
明確定義欲解決的問題			
2.2.1	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：_____	
2.2.2	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			
2.2.3	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：_____	
2.2.4	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：_____	

編號	檢核項目	檢核結果	檢核說明
定義專案目標成效			
2.2.5	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：_____	
2.2.6	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：_____	

2.3：資料狀態評估

在進行 AI 導入時，對資料的收集與評估是至關重要的。因為資料的品質直接影響到 AI模型的表現，也將進一步影響評估機關是否適合於現階段導入 AI 服務。在進行 AI 導入所需資料狀態評估時，可朝以下面向評估，應盡可能減少不確定，並減少所需資料不齊全的情況，以確保場景中的 AI 導入可順利進行。以下是評估資料狀態的幾個重要步驟和原則：

* 圖片來源：KPMG 整理

資料的準備

是否可使用所需資料，是 AI 專案導入的前置基礎。故在導入 AI 專案前，機關應盤點導入場景所涵蓋資料範疇，並確定所需的資料種類及其來源。除了內部資料集外，也可能包括內部資料庫整理整合新資料、公開資料集、規劃當次導入過程中蒐集、與合作機關介接等方法取得所需資料。但須留意資料準備為 AI 導入前提，必須預留時間與資源，並確保 AI 有可取得資料。待正式執行團隊確定後，在執行專案初期，機關應與執行團隊討論符合專案目標之合適資料提供範疇，包含來源、資料量與欄位細節。

資料品質的評估

確保可取得資料後，應對資料進行資料品質評估，以確保 AI 模型能有好的分析結果，減少 AI 模型預測不精確、偏誤結果產生。以下為幾個討論資料品質可參考之基本四大面向。

1. **完整性 (Completeness)：**檢查資料集是否涵蓋了 AI模型所有必要的欄位，了解資料是否有缺漏問題。
2. **準確性 (Accuracy) ：**評估資料集中的值，與真實情況是否一致。
3. **可信度 (Veracity)：**評估資料的可靠性，資料來源是否可信。
4. **即時性 (Timeliness)：**資料是否紀錄時間數值，並且確保有時效性之資料已更新到分析所需時間區間。

評估模型需要的資料

歷經上方的資料品質評估，機關可了解可運用之資料量，並進一步思考可參考之 AI 模型方法有哪些。

思考點：資料準備與來源

啟動 AI 專案前，可先確認資料是否準備得夠完整。像是如果需要處理影像、文字或影片等原始資料，記得預留足夠的時間進行標記與處理；若資料來自不同單位，建議及早展開溝通與協調，避免影響後續進度。

1. **預訓練模型 (Pre-trained Model)**：許多常見的應用可以使用預訓練模型，像是圖像中的相同人物檢測或文件翻譯屬於此類。此訓練方式有使用條件限制，預訓練模型的訓練資料，需要與當下欲解決的問題一致，即可直接使用。在這種情況下，不需要收集訓練資料，只需收集一些資料來評估模型性能。
2. **遷移學習 (Transfer learning)**：可以從預訓練模型進行遷移學習來訓練新的模型。通常會使用遷移學習，該次訓練目標和要解決的問題，須與預訓練模型的訓練資料和目標類似，才能從同一個預訓練模型中提取部分特徵，或是針對目標微調預訓練模型。如同站在預訓練模型的肩膀上進行優化，這方法需較少的資料。
3. **自行訓練 (Training from Scratch)**：如果您從頭開始訓練 AI 模型，可能需要收集數百到數千萬筆資料。儘管如此，受益於少樣本學習（Few-Shot Learning）技術發展，機器學習的最新進展為可以使用少量樣本進行自行訓練。

有關 AI 模型採用之方案，機關可於 RFP 撰寫期間了解潛在方案的對應規劃合理性，是否考量機關現有資料品質現況，制定對應合理之 AI 解決方案。通過上述三個步驟，可以確保資料品質與數量現況在制定 AI 解決方案時一併考慮，進而提高最終 AI 應用的效能和可靠性。以上敘述已提供基礎的技術與數據評估架構，詳細模型選擇與對應資料評估，建議與資料科學家進行諮詢討論。

* 來源：Freepik

思考點：資料數量與品質

在評估模型訓練的可行性時，可試想：我們目前的資料量是否足夠？品質是否穩定？選擇技術方案時，也別忘了確認資料是否真的能支撐模型的需求，這樣才能讓後續的開發更順利。

2.3：資料狀態評估檢核清單

編號	檢核項目	檢核結果	檢核說明
資料準備			
2.3.1	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：_____	
2.3.2	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	**本案所需使用之AI訓練資料，其資料的完整性、準確性、真實性與即時性的評估結果。**		
2.3.3	資料來源是否真實可信？	☐已完成 ☐未完成：_____ ☐不適用：_____	
2.3.4	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			
2.3.5	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：_____	

2.4：政府公開 AI 資源

公開資料查詢

政府資料開放平台

AI 相關事務單位

1. 臺灣 AI 卓越中心
2. 行政院智慧國家推動小組 – AI 發展諮詢委員（為國家 AI 發展專業諮詢，參與成員皆為國內外產學研之 AI 知名專家）
3. AI 公務人才發展辦公室（2025 年 7 月正式成立營運，預計於 2026 至 2030 年逐步推展落實）

學習資源

1. 工研院產業學習網
2. 台灣人工智慧學校（e.g. Gen AI 素養認證實作班）
3. 臺灣人工智慧卓越中心（e.g. 資訊觀測分享）
4. 行政院人事行政總處公務人力發展學院資訊職能數位學習專區
5. AI 產業實戰應用人才淬煉計畫

AI 資源

1. TAIDE - 推動臺灣可信任生成式 AI 發展計畫
2. TryAI 政府 AI 實驗站
3. 製造業 AI 升級引擎
4. Taiwan AI RAP

03 AI 服務導入

章節摘要

- **了解政府 AI 專案有哪些導入方式**
- **政府 AI 專案的 3 個流程**
- **AI 專案相關的技術議題**
- **AI 專案的專案團隊組成與事前的教育訓練規劃**
- **AI 專案驗收指標如何制定**
- **AI 專案與傳統資訊專案差異**

在 AI 服務導入的段落中，我們將深入探討機關決定在特定場景導入 AI 服務，啟動一項 AI 專案後，如何透過以下流程有效地導入 AI 技術。首先，我們將介紹不同的 AI 導入模式，並討論機關如何判斷哪種模式最適合其需求。接下來，我們會詳細說明一個典型的 AI 專案流程細節，從需求分析到模型開發和測試，並探討其中涉及的技術問題，如資料準備、演算法選擇和性能優化。

當初版模型建立後，將進入模型部署階段。本章將介紹 AI 服務的部署模式，並說明 AI專案所需的團隊成員配置，以及機關內的教育訓練安排。最後，我們也會比較 AI 專案與傳統 IT 專案的差異，幫助機關同仁理解 AI 專案的獨特挑戰。

以上內容將做為機關執行 AI 專案時之初步背景知識，以利專案規劃與專案管理。但有關技術方案的選擇與判斷，仍需與資料科學專家進行討論。

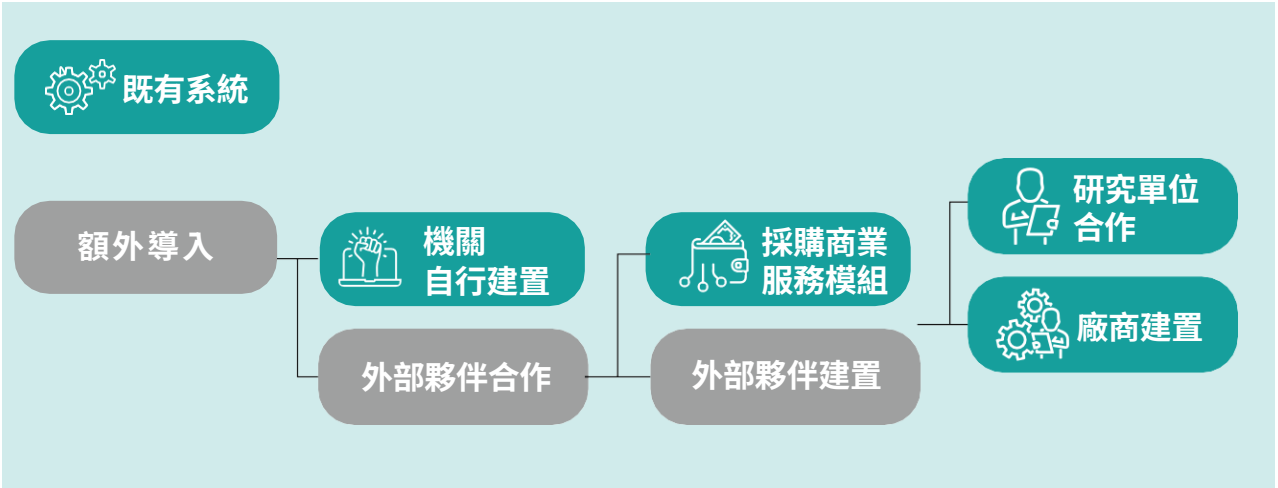
3.1：AI 導入模式

在釐清 AI 應用場景、定義問題與資料盤點後，即可開始進行 AI 專案的規劃與實施。AI 專案執行過程中，可依照專案現狀與組織資源，自行評估 AI 服務導入的方式。以下將說明幾個常見的 AI 服務導入模式介紹，以及選擇 AI 服務導入模式時，可考量之重要面向。

* 來源：Freepik

常見 AI 服務導入模式介紹

在進行 AI 專案時，大抵會有五種模式，本章節將針對以下五種導入模式分別說明。首先需要先確認，要使用既有系統或是額外建置。在選擇額外建置新系統後，應思考要由機關自行建置、外部合作建置或購買既有商用服務。選擇與外部合作建置，則可尋找外部合作廠商或研究單位。



是否要使用既有服務來滿足使用者需求？

如果本次需求與過去相似，或現有模組即能滿足，那麼不啟動一個新的 AI 專案，也是可考慮的問題解決方式。可透過盤點機關現有的 AI 專案資源，了解是否有與既有服務共用的可能。

選擇額外導入時，應自行建置或與商業夥伴合作？

* 來源：Freepik

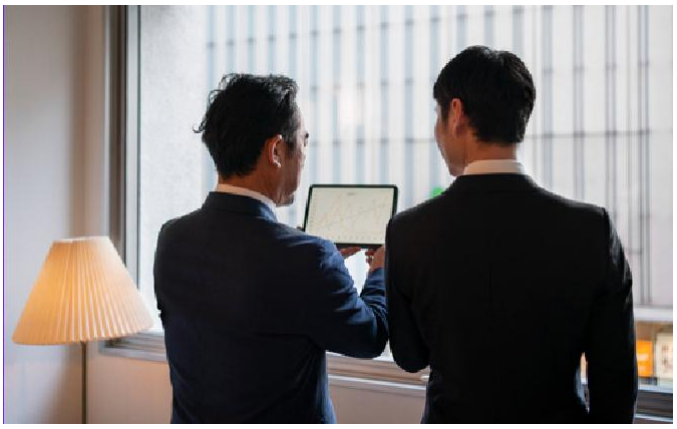
若採取自行建置，政府機關須自行組建一支專門的 AI 團隊，負責從頭開發和實施所需的 AI 解決方案。這種做法可以讓機關對 AI 技術有高度掌握，同時也能夠培養內部的 AI 技術人才，作為組織後續擴大 AI 的基礎。

通過自建 AI 團隊，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。團隊成員可以深入了解業務場景，並針對性地設計AI 模型和功能。這不僅能夠確保系統的適配性，也有利於持續優化和迭代。

但自建 AI 團隊需要大量的初期投入，包括人力、資金和基礎設施等。同時，機關還需要自行解決技術和管理方面的挑戰，例如人才招募、系統架構設計、營運維護等。因此，採用這種模式的機關需要做好充分的可行性評估和風險規避。

要選擇現成 AI 商業服務模組，還是找外部夥伴建置？

通常選擇現有商業服務模組，都因為目前需求是成熟且明確的 AI 應用，像是文字影像辨識。但若要使用現有的商業服務，應確保整體的使用流程是否流暢，以及現成 AI 商業模組是否能與現有系統架構進行整合。若需求為高度客製化，那尋找外部夥伴從頭開始建立 AI 服務會是另一種可行的選擇。



* 來源：Freepik

目前生成式 AI 存在多元的商業服務模組，使用者僅需登入即可開始使用，或是使用內建於既有的服務的 AI 商業模組。

登入式方案優點是多數使用者已經習慣透過此方式來使用 AI 服務，使用者只需要一個電子郵件地址即可註冊。缺點則是就機關而言，較難監控使用者的服務使用狀況，基本上僅能做規範勸導或直接禁止，來避免不當使用的情況。類似知名服務像是 OpenAI 的 ChatGPT、Google 的 Gemini、Microsoft 的 Bing 搜索服務 (截止 2024/06 資料)[11]。

11. 參考 statista 公告之 AI tool brand ranked, 選擇知名度高之 AI 工具

嵌入式生成式 AI 與既有產品結合的模式，多數能通過對話或提示（Prompt）形式來尋求問題解答或完成任務。但需留意嵌入式服務通常需建立在讀取原有服務的資料，包含文字、圖片、信件、資料庫、系統 log 等相關資料，機關需慎重管理嵌入式服務的資料訪問權限，也確保不會因保護資料隱私安全而過度限制導致工具的效益難以發揮。

應與研究單位合作，或是找外部廠商做建置？

1.研究單位合作

如果商業上可用的解決方案無法解決現有問題，或是有開源需求，機構可透過與研究單位共同合作，開發所需服務。

通過產學研發合作，機關可與研究機構的最新技術和專業人才合作，大幅提高 AI 系統的創新性和技術水平。研究機構提供 AI 算法、模型開發等核心技術支持，機關則負責提供資料庫、實際業務場景和需求反饋。雙方密切配合，共同推進 AI解決方案的設計和迭代。

這種合作模式有利於促進 AI 技術向實際應用的轉化，同時也能培養機關內部的 AI 研發能力。不過，也需要雙方協調建立健全的合作機制和成果分配制度。

2.外部廠商建置

在這種模式下，機關會將 AI 服務的開發和實施完全外包給專業的 AI 服務商。這樣可以快速獲得所需的 AI 服務，降低機關自行建設的成本和風險。同時，也可以充分利用服務商的技術、經驗和資源。

通過外包 AI 服務，機關可以專注於業務需求的定義和建置品質的控管，而不必過多地涉及技術細節。廠商負責系統的設計、開發、部署和維運，確保 AI 應用可以順利投入使用。這種模式的靈活性相對較低，但能大幅減輕機關的內部建置壓力。

不過，外包模式也存在一定的風險，例如對廠商的依賴性過強、智慧財產權與資料安全等。因此，機關需要建立健全的外包管理機制，確保服務品質和雙方利益都得到保障。

常見 AI 服務導入模式比較

	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。	較難監控使用者的服務使用狀況，基本上僅能做規範勸導或直接禁止。
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。

思考點：導入模式的選擇

在決定 AI 專案的執行方式前，可試想：我們是否已從技術成熟度、內部資源、導入方式（自建、商用或混合）及潛在風險等角度，做過整體評估？如果考慮使用現成模組，也別忘了確認它是否真的符合需求，並思考如何與現有系統順利整合。

3.1：AI 導入模式檢核清單

編號	檢核項目	檢核結果	檢核說明
AI 導入模式			
3.1.1	是否盤點機關內數位服務與 AI 服務資源，了解是否有現成解決方案或是可共用之開發資源。	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.1.2	是否評估研究單位與第三方廠商的技術能力與開源配合可能。（參照行政院及所屬機關（構）使用生成式 AI 參考指引、資通安全法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.1.3	若選擇現成生成式 AI 服務，需留意提供資料有可能被服務提供商直接再次使用，以及是否詳細了解現成生成式 AI 服務的使用規範。（參照行政院及所屬機關（構）使用生成式 AI 參考指引、資通安全法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.1.4	若要使用嵌入式生成式 AI 服務，需要了解服務需要取得哪些機關內部的資料權限，並評估提供相關資訊權限是否合宜。（參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	

3.2：AI 專案流程

AI 專案由於其模型建置訓練特性，其開發到部署流程較傳統資訊系統專案不同，此章節針對 AI 專案完整進行流程進行說明。

* 來源：Freepik

AI 專案開始前，需參照 2.2 章節之問題定義流程，依照領域知識與現有資料來明確定義 AI 模型的產出、應用方式並且確保 AI 能解決起始問題。在確定執行 AI 專案後，AI 專案從開發到部署可以分為三大階段，分別為「資料蒐集準備」、「AI 模型訓練與迭代」與「AI 模型部署與監控」。

第 1 階段：資料蒐集準備

1. **資料前處理：**在取得本次專案可使用資料後，開發團隊需要進行初步的資料探索，並在瞭解資料集可能存在的數據品質問題後，進行資料前處理，包含剔除雜訊資料與極端值、資料正規化、遺漏值處理等，將原始資料轉換為可以作為 AI 訓練的模型資料。另須考量資料的安全性與隱私性，並針對敏感資料提前做去識別化的處理。某些 AI訓練的資料前處理作業包含資料標記 (Data Labeling)。此項作業需耗費一定時間與人力，務必預留足夠專案時間。

2. **切分資料集：**在準備好可用資料後，應將訓練資料切分為訓練集、驗證集與測試集，才可進行正式 AI 模型訓練。AI 專案不會將所有資料皆投入於訓練中，會需要保留部分資料來評估模型表現。

在模型訓練過程中，理論上訓練、驗證，與測試三者應避免資料重複，需要完全切分，以確保模型評估的獨立性，否則將無法確認模型的真實學習表現。另外，切分資料集時須留意資料抽樣分布的公平性，避免造成偏誤。

然而當資料量極小時，完全避免重複並不切實際。可行的解法包括透過：

- 資料擴增(Data Augmentation)增加資料量
- 驗證與測試集上採取適度合併或權衡，以確保模型仍能進行基本評估。

需注意，這些指引屬於建議原則，實務上應依資料特性彈性調整，兼顧模型效能與評估合理性。

3.建立資料使用管理機制：應評估專案使用資料的安全性與隱私性，針對資料進行去識別化，並建立對應版本管控機制。

第2階段：AI 模型訓練與迭代

* 資料來源：KPMG 整理

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。

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，建議機關最終方案選擇過程，應與資料科學家做諮詢討論，以便選擇決策邏輯清楚且具可解釋性的方案。在此提供幾個常見 AI 技術與 AI 應用，同仁可依照期待解決的問題類型，去尋找對應要用到的模型，以利與技術人員討論。

* 資料來源：KPMG 整理

機器學習應用	簡介	案例
自然語言處理（NLP）	處理和分析自然語言，識別單詞、其含義、上下文和敘述。	• 文本背後的情緒分析 • 自動生成客戶電子郵件的回覆
電腦視覺	機器或程式模仿人類視覺的能力。	• 為自駕車識別道路標誌 • 用於自動化護照控制的臉部識別
異常監控	在資料集中找到異常的資料點。	• 識別使用者銀行帳戶中的欺詐活動
時間序列分析	了解資料隨時間的變化以進行預測和監控。	• 進行預算分析 • 預測經濟指標
推薦系統	預測使用者對給定項目的評價，以提出新建議。	• 根據使用者之前查看的文章推薦相關頁面

2. **訓練模型與驗證模型：**模型建立過程，包含需模型訓練與模型驗證兩階段。在訓練階段，首先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。完成初步模型訓練後，將會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（如學習率、正則化參數等）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。同時，需要注意避免過度擬合（over-fitting）和低度擬合（under- fitting）的問題。過度擬合是指模型在訓練集上表現良好，但在驗證集上表現較差，而低度擬合是指模型無法在訓練集上達到良好的性能。模型訓練與模型驗證的步驟，需要持續數次循環，以確保模型效能符合任務目標。

知識點：

什麼是超參數？

超參數 (Hyperparameter) 為訓練模型時用來控制訓練過程的變數。哪些是有效的超參數以及超參數的最佳值，可以手動或透過超參數演算法自動調校。

舉例來說，今天一個隨機森林的模型表現不佳，會嘗試調整的超參數包含樹的數量、每顆樹的訓練深度、樣本數據權重等不同的參數，來使模型表現變好。

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。根據結果，模型可能需要更多資料做測試，也可能已經適合正式使用。以上模型訓練從選擇、訓練、驗證、評估四步驟，會需要不斷迭代進行。通常整個流程中，會選擇多個演算法進行訓練，並比較訓練出的模型成果，再決定最終模型；在驗證模型階段，也有可能多次調整模型參數，以確保模型會有最佳表現。故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。

思考點：

資料與模型訓練

在模型訓練階段，可試想：我們是否已納入與分析目標高度相關的變數？也可以試著建立並比較多個模型的表現，從中挑選出最適合的方案，讓模型在部署前就具備更高的穩定性與可信度。

第3階段：AI 模型部署與監控

當模型經過訓練後，就可以在環境正式部署，並根據它「學到的」內容，依照新的輸入資料進行預測。當模型被部署進工作流程中，需確保其性能受到持續監控，確保預測品質符合需求。

1. **模型部署：**將訓練好的模型部署到生產環境中，需確保部署環境與訓練環境相容、將模型從訓練環境導出，轉換為適合部署環境的格式、將模型部署到預定的伺服器或雲端平台設置 API介面，並在部署環境中對模型做初步測試，使其能根據新的輸入數據進行預測。

2. **AI 專案試辦規劃：**若機關於專案初期將 AI 系統的試辦納入流程，能有效測試 AI 專案帶來的幫助，以及可改進之處。建議機關應於模型在試辦環境部署前，進行系統試辦流程規劃，了解對機關衝擊並規劃應對方案。通常小規模試辦會衝擊到現行的機關流程，須於試辦前，向該場景合作承辦人，說明專案重要性，並共同針對可能造成之業務衝擊提出調整方案。

知識點：

什麼是資料飄移？

資料飄移(Data Distribution Shifts)指的是機器學習的一種現象，當資料隨著時間變動時，模型效能隨之下降。潛在原因多為後續新蒐集之資料的關鍵特徵，隨時間而有所變化，導致仰賴原有特徵的模型效能變差。

3. **持續監控模型：**模型部署後，須持續監控模型的性能和預測品質，注意輸入數據分布的變化（比如資料飄移），定期評估模型的預測準確性和其他性能指標以確保其穩定性和準確性。

4. **定期更新：**根據監控結果和新需求，對模型進行維護和更新，以保持其預測表現。因為這些技術仍處於快速發展階段，機關也需要注意最新的 AI 模型和演算法，如果出現新的 AI 模型或演算法，可能需要重新設計現有的 AI 應用。同時，若使用 AI 應用的業務出現與常規不同的變化趨勢，AI 模型也可能需要重新訓練，以適應新的業務需求。此外，定期評估 AI 模型是否符合業務目標至關重要，這樣可以確保持續的相關性，確保 AI 應用始終達成機關業務目標。

3.2：AI 專案流程檢核清單

編號	檢核項目	檢核結果	檢核說明
第 1 階段：資料蒐集準備			
3.2.1	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：_____	
3.2.2	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：_____	
第 2 階段：AI 模型訓練與迭代			
3.2.3	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：_____	
3.2.4	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：_____	

編號	檢核項目	檢核結果	檢核說明
第 3 階段：AI 模型部署與監控			
3.2.5	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：_____	
3.2.6	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：_____	
3.2.7	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：_____	
3.2.8	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：_____	
3.2.9	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：_____	

3.3：AI 專案技術議題

AI 導入專案，需要資料與運算資源兩項資源，故雲端經常作為 AI 專案的部署環境。又生成式 AI 發展迅速，眾多數位服務皆以 API 形式串接，打造創新的數位服務。此章節將就軟體工具與框架、數據資源管理、硬體資源、開發與部署工具四大段落進行說明。

軟體工具與環境

可建構 AI 應用的程式語言多樣，包括 C++、Java、Python 和 R 等，後兩種程式語言提供強大 AI 發展工具，如機器學習演算法程式庫以及資料視覺化程式庫。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。常見整合開發工具為 Jupyter Notebook、VS Code 與 PyCharm 做為開發工具，雲端選擇為 Google Colab。大數據處理則選擇 Apache Spark 和 Hadoop 技術，能更靈活處理大規模資料集。

數據資源管理在AI 專案中扮演關鍵角色，資料庫須能有效地儲存、組織和檢索大量數據，以支援模型訓練。一個良好的資料庫管理系統（DBMS）不僅能提高資料操作的效率，還能確保資料的安全性、完整性和一致性。常見資料庫類型選擇：

1. **關係型資料庫（RDBMS）：**如 MySQL、PostgreSQL和 Oracle，適用於結構化資料和複雜查詢，使用SQL 語言進行操作。
2. **NoSQL 資料庫：**如 MongoDB、Cassandra 和 Redis，適用於處理大規模非結構化或半結構化數據，提供高擴展性和靈活的模式設計。
3. **圖形資料庫：**如 Neo4j 和 Amazon Neptune，適用於處理圖形資料結構和複雜關係查詢，廣泛應用於社會網路分析和推薦系統。

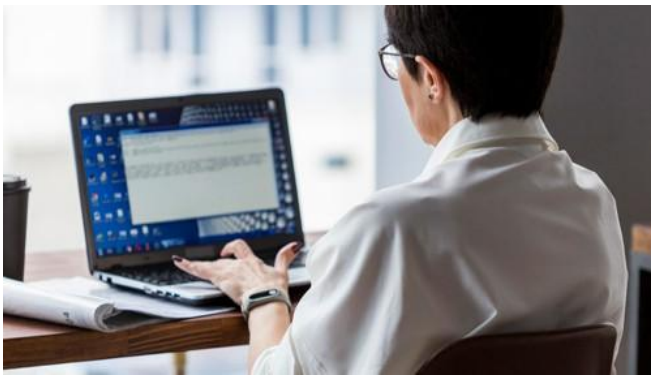
* 來源：Freepik

除資料庫管理外，數據資源管理也包含以下議題。如何確保資料儲存技術的可擴展性；如何透過 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中，確保資料的一致性和可用性；以及如何建立資料管道，透過一系列步驟持續處理整合大規模資料。

硬體資源

1.地端環境

對於傳統 AI 來說，前期原型設計和最小可行產品研究可以在個人或地端電腦上進行，適合在成本有限且運算資源需求較低的環境下進行。需要確保電腦性能足以支撐模型訓練和測試，通常與機構內的資料庫串接會較容易，能避免將機敏資訊傳輸到外部，但機構內仍應具備基礎的資料治理能力，防止未授權的存取與洩露。在此階段，對資源的需求較少，因此能以較低成本進行開發，但需注意未來擴展時的成本和資源需求。



* 來源：Freepik

2.雲端環境

雲服務提供了公共工具和付費 API 的功能，且會提供許多可用於建模測試的現成工具。除了能將運算上雲外，也提供基本的資料存取管理規範機制。常見可透過 AWS、 Microsoft 或 Google 企業帳戶註冊，來使用雲服務提供商的資料管理、AI 訓練與 LLM的 API 服務。雲端提供彈性的資源配置，能根據需求動態調整計算和儲存資源，從而提高開發效率。然需注意的是，若資料涉及機敏性資料，則需考量是否該上傳雲端，以避免資安問題。

3.雲地混合環境

雲地混合環境結合了地端和雲端的優勢，適合需要平衡性能、成本和安全性的 AI 開發和部署。此環境允許機關在地端基礎設施上處理敏感數據和進行初步開發，同時利用雲端的高性能計算資源來進行大規模訓練和數據分析。資料串接方面，雲地混合能夠在地端和雲端之間無縫傳輸數據。安全性則透過地端和雲端的雙重防護機制，包括地端數據保護和雲端安全措施，以及容器化技術配合運用來實現。雲地混合在成本上具有靈活性，可以根據實際需求動態調整資源配置，從而有效控制成本，同時保證高效能和數據安全。

* 來源：Freepik

AI 專案於雲端與地端部署的優缺點

	優點	缺點
雲端部署	1. **可擴展性高：**雲端部署允許機關根據運算資源需求的波動，快速擴大或縮小運算叢集的規模，以滿足不同時期的需求。 2. **財務門檻低：**不需要大量的初期投資，機關可以以較低的成本進入並使用雲端服務。 3. **開發工具和技術支援：**雲端服務提供商提供多種開發工具，例如可直接使用預訓練模型或是no code 平台，能加速專案進程，讓非技術背景者也有機會參與建立模型。	1. **資料傳輸安全性：**需要考慮在雲端環境中的資料傳輸安全性，確保敏感數據不被洩漏。另需留意資料存放之源頭機房位置，是否符合資料管理規範。 2. **服務商限制：**後續專案搬移會受到雲端服務商的一些使用限制，影響部署的靈活性和控制權。 3. **雲端費用不確定性：**雲端費用將作為後續模型運作支出，其費用將隨市場變化，需於專案開始前納入成本估算中。
地端部署	1. **控制性高：**地端部署允許企業多次使用 GPU 系統，無需考慮額外成本，從而更輕鬆地進行反覆嘗試和實驗。 2. **無需依賴網路：**地端系統運行不依賴於網路，因此在網路不穩定或中斷時仍能正常運行。 3. **遵守嚴格的隱私標準：**地端部署有助於符合醫療保健公司、人權組織及金融服務業等對資料主權及隱私權有嚴格標準的組織的要求，高敏感度的財務資訊或醫療紀錄等資料可以放在組織的防火牆內，保證數據的安全性和隱私性。	1. **硬體資源需求：**因無法依照需求變化快速取得 GPU 計算資源，需要預留更多的硬體資源來支持地端 AI 系統的運行和擴展。 2. **維護和管理：**需要專門的技術團隊進行硬體和軟體開發平台的維護和管理，增加了營運的複雜性。 3. **擴展性有限：**地端部署的擴展通常需要額外購買硬體設備，擴展速度較慢且靈活性不足。
混合雲部署	1. **最佳化資源利用：**可以將敏感數據或高負荷工作負載放置在地端系統中，其他業務運行於雲端，達到成本與效能的平衡。 2. **高靈活性：**能根據需求動態調整地端與雲端資源的分配，應對變動的運算需求。 3. **更高的控制與安全性：**敏感數據保留在本地，非敏感數據或應用程序可以運行在雲端，降低安全風險。	1. **複雜度增加：**管理和整合兩種不同的環境需要專業知識和額外的管理工具。 2. **同步問題：**在確保地端和雲端系統數據的一致性和同步性上，可能面臨挑戰。 3. **潛在的更高成本：**需要投資於地端基礎設施，同時也承擔雲端服務的持續費用。

開發與部署工具

1. API 管理

大多數大型生成式 AI 應用都有提供 API（應用程式介面，Application Programming Interface）。開發人員可以將生成式 AI 功能串接到解決方案中，大多需要註冊付費才可取得串接權限。通常這類型 API 開發仍需一定程度的客製化，避免生成資訊受誤導或浪費開發成本。此外，機關需要了解使用 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，像是經 API 將機關的資料自動傳輸給 AI 服務供應商。機關必須在使用 API 之前對其進行妥善評估。

* 來源：Freepik

2. 版本控制系統 (VCS)

版本控制系統 (Version Control System, VCS) 是管理程式碼變更、協作開發和維護程式碼歷史記錄的工具。在 AI 專案中，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。

- Git：最廣泛使用的分散式版本控制系統，支援分支和合併操作，常用平台包括 GitHub、GitLab 和 Bitbucket。
- SVN(Subversion)：集中式版本控制系統，適用於需要嚴格控制程式碼存取的專案中。

3. 容器化技術

容器化技術透過將應用程式及其所有程式庫依賴性打包到一個輕量級、可移植的容器中，確保在不同環境中一致運作。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。常見 AI 專案使用的容器化平台如下：

- Docker：最受歡迎的容器化平台，讓開發者可以建立、部署和運行容器化應用程式。
- Kubernetes：容器編排平台，管理和自動化容器的部署、擴充和操作。

4. CI/CD 工具

CI/CD 是一種自動化軟體開發流程的方法，旨在頻繁、可靠地建置、測試和部署軟體。

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。常見 AI 專案所使用的 CI/CD 工具如下：

- Jenkins：開源自動化伺服器，廣泛用於 CI/CD 管道的建置和管理。
- GitLab CI/CD：整合在 GitLab 中的 CI/CD 工具，支援從程式碼提交到部署的全流程自動化。
- Travis CI：與 GitHub 整合良好的 CI 服務，適用於開源專案。
- CircleCI：靈活的 CI/CD 平台，支援多種程式語言和環境。

思考點：系統整合與部署

在將 AI 模型導入實際環境前，可試想：資料串接是否穩定？資料來源的品質與上傳機制會不會影響模型表現？此外，也可以評估是否能整合過往專案中累積的技術資源與經驗，幫助提升整體部署效率，讓未來的 AI 推動更順利。

檢索增強生成 RAG

* 來源：Freepik

檢索增強生成，簡稱 RAG （Retrieval-Augmented Generation）是一種結合大型語言模型與外部資訊的技術。利用語言模型的生成能力和資料庫中儲存的資訊，來生成更真實和可靠的回應，經常被用於解決大型語言模型存在的幻覺問題。[12]

1. RAG 特色與使用時機

RAG 可以用使用者口語化的文字查詢資料庫，而 RAG 能夠從資料庫中檢索相關資料，並生成更準確的回應。相較重新訓練模型，RAG 技術更省時且低成本，不需要大量的計算資源來訓練模型。RAG 也提供了更透明和可解釋的生成過程，因為它能夠引用具體的資料來源，提升了回應的可信度。RAG 技術特別適合有以下需求的任務：

- 知識密集型：當任務需要處理大量專業知識或最新資訊時，RAG 能夠從資料庫中檢索相關資料，並生成準確的回應。例如，醫療診斷、法律諮詢和技術支持等領域。

[12] 生成式AI簡介與應用（下） – 國家發展委員會循證尋政

- 動態更新需求：如果任務需要頻繁更新數據，例如新聞報導或市場分析，RAG 可以確保生成的內容基於最新的資料，保持高相關性和時效性。
- 高準確性需求：在需要高準確性和可信度的情境下，RAG 能夠引用具體的數據來源，提供透明且可驗證的回應，這對於學術研究和專業報告非常重要。
- 節省成本和資源：相比於完全重新訓練大型語言模型，RAG 更加經濟實惠，因為它不需要大量的計算資源來微調模型，適合資源有限的情境。

2. RAG 工作流程

- 檢索：當使用者輸入提示 (prompt) 後，RAG 會根據提示在網路上或資料庫中檢索相關資料，取得可協助模型提供實用回應的資料。
- 擴增：檢索資料會與使用者提示一併輸入模型。
- 生成：模型會依據提示與資料生成真實、有效的回應。

這樣的結合使得模型能夠提供更具上下文和相關性的答案，並且能夠處理更專業和具體的問題。

3. RAG 的限制

- 資料清理：確保資料的品質和相關性是關鍵。低品質、結構混亂或不相關的資料會影響檢索與生成結果的準確性。因此，在將資料放入資料庫處理前，需要進行嚴格的篩選和清理。
- 生成品質：即使檢索到正確的資料，生成的回應也可能不夠準確或相關。這需要不斷調整生成模型的參數和提示詞，以提高生成品質。
- 維護更新：模型參考資料庫的資料生成回應，如果沒有及時更新資料，模型就可能參考過時資料回覆錯誤資訊，例如：法規變動、新流程。[13]

[13] Retrieval augmented generation: Keeping LLMs relevant and current - Stack Overflow

3.3：AI 專案技術與環境議題檢核清單

編號	檢核項目	檢核結果	檢核說明
3.3.1	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：_____	
3.3.2	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：_____	
3.3.3	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：_____	
3.3.4	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：_____	
3.3.5	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：_____	
3.3.6	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：_____	
3.3.7	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：_____	

3.4：團隊成員

AI 專案通常包含需求規劃、數據收集與處理、模型開發與訓練、系統整合與部署、測試與驗證、維護與更新，為了滿足以上需求，AI 專案需包含以下專案團隊成員，以確保專案能有良好品質。以下提供 AI 專案常見所需專案團隊角色。

專案經理

- 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

資料分析師

- 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

資料科學家

- 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

商業智慧分析師

- 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

資料工程師

- 參與前期訪談
- 提供領域資訊

領域專家

- 參與前期訪談
- 提供領域資訊

機器學習工程師

- 模型開發部署
- 模型與演算法效能優化

測試工程師

- 系統效能測試
- 撰寫自動化測試腳本
- CI/CD 測試執行

技術諮詢專家

- 提供 AI 專案諮詢指引
- 分享對應 AI 專案執行經驗

現有 AI 專案角色之能力定義多元，且相關職務也持續發展中，不同組織對於角色對應的工作職掌也有不同的定義。組成團隊時，可著重於專案所需之 AI 工作項目，是否與專案團隊的能力與經驗契合，以確保團隊具備執行 AI 專案所需之專業能力。

以下將以常見的 AI 工作範疇為基礎，介紹 AI 專案常見之成員以及其在團隊中負責的任務範疇，並以表格呈現所需的能力和經驗，以及常見之職稱作為參考。

AI 專案承辦之機關同仁若具 AI 專案相關經驗為佳，建議具備基本 AI 知識、數據管理能力以及資訊專案管理能力，以便與技術背景成員溝通，並掌握專案進展。另建議熟悉 AI相關法規與具備良好之溝通協調能力者為佳，考量 AI 專案涉及數據運用，仰賴與其他機關介接合作，具備良好法遵意識與溝通協調能力將有助於專案推行。

成員	負責工作	所需能力與經驗	常見業界職稱
專案經理 (Project Manager)	負責整體專案的規劃、執行和監控，確保專案按時完成並在預算範圍內。管理各成員的協作，協調內部和外部資源，解決專案中出現的問題。	• 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	專案經理
資料分析師 (Data Analyst)	負責收集、清理和處理各種數據來源，並運用統計分析和機器學習技術，發現數據中的趨勢。透過數據視覺化工具，資料分析師將複雜的數據轉化為易於理解的報告和圖表，幫助團隊和決策者深入了解業務現狀和潛在機會。	• 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	系統分析師、軟體工程師、專案管理人員
商業智慧分析師 (Business Intelligent Analyst)	商業智慧分析師負責設計和開發商業智慧解決方案，並通過構建和維護資數據基礎設施，確保數據的集中和一致。利用 BI 工具（如 Tableau、Power BI 等），商業智慧分析師建立互動式儀表板和報表，提供關鍵業務指標的即時監控和分析。	• 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、專案管理人員
資料科學家 (Data Scientist)	資料科學家主要負責設計和建立機器學習模型，進行資料分析和特徵工程。他們的工作包括從大量資料中發掘模式和洞察，並通過數據視覺化和統計方法提供決策者參考。	• 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	軟體工程師 AI 工程師

成員	負責工作	所需能力與經驗	常見業界職稱
資料工程師 (Data Engineer)	負責設計、構建和維護數據基礎設施，包括數據收集、儲存、處理和傳輸的系統。確保數據的品質和可用性，協助資料科學家和開發人員的工作。	• 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	軟體工程師 系統工程師 後端工程師
機器學習工程師 (Machine Learning Engineer)	機器學習工程師專注於開發模型與演算法，確保這些模型能夠在真實世界中高效運行。他們需要設計和建立可以處理大規模資料的系統，並確保模型的性能和穩定性。	• 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	軟體工程師 AI 工程師
測試工程師 (QA/Test Engineer)	負責系統測試，確保系統的功能性、性能和可靠性，識別和修復潛在的問題。	• 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	測試工程師
領域專家 (Domain Expert)	提供特定行業或領域的專業知識，幫助理解資料背景和業務需求，確保 AI 模型和解決方案的實用性和有效性。	• 了解 AI 應用場景實際業務 • 後續 AI 服務實際使用者	業務單位 承辦人
技術諮詢專家 (Consultant)	針對 AI 專案實際執行提供實踐經驗資訊，協助 AI 應用落地執行，確保技術細節正確與有效。	• 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	AI 工程師 軟體工程師

3.4：專案團隊成員檢核清單

編號	檢核項目	檢核結果	檢核說明
3.4.1	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：_____	
3.4.2	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：_____	
3.4.3	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：_____	
3.4.4	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：_____	
3.4.5	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：_____	

3.5：教育訓練

教育訓練的重要性

機關要能順利採用及實施新的 AI 技術，將依賴於針對部內人員的全面培訓。AI 服務之教育訓練應確保使用者具備必要的相關知識與技能，建立部內人員對 AI 技術應用的信心，最大限度發揮投資價值，並以最高標準為民眾提供服務。良好的培訓應達成如下目標：

1. 了解 AI 服務的潛在效益

AI 服務通常會引入新的工作流程、操作介面和互動方式。適當的培訓能夠幫助使用者系統性地理解 AI 系統，接納內部技術的轉型。

2. 掌握 AI 服務使用模式

AI 技術旨在通過自動化的方式簡化流程。若缺乏 AI 服務的基本知識將難以有效利用這些工具，導致成果不如理想。適當的培訓可以確保使用者能夠最大限度地發揮 AI 服務的全部功能，優化工作效率。

* 來源：Freepik

3. 降低錯誤使用者與管理者的風險

AI 技術伴隨著資料隱私及道德等方面的考量，例如因演算法偏誤而導致使用者因被誤導而進行錯誤決策，亦或是使用者將高度敏感性資料提供於 AI 系統。適當的培訓能夠幫助使用者具備迴避或應對各式 AI 風險的能力，也協助 AI 系統管理者能進行治理。

4. 促進 AI 素養與資料素養

AI 技術通常高度依賴資料。因此培訓內容也應著重於使用者的 AI 素養與資料素養，去理解資料的作用、解讀 AI 服務的邏輯，並根據其產出進行最終的決策。

教育訓練的前置作業

1. 培訓需求評估

❶確認學員

確認任何將與AI 服務進行互動之角色，包括決策者、使用者、管理者（IT人員）等，並將其列入須接受培訓之人員名單。根據不同 AI 服務的使用群體，調整相應的培訓內容及培訓方式。

❷現有知識與技能

評估待培訓人員當前的 AI 素養（AI Literacy）及整體數位科技熟練程度。可透過對內部員工的基本調查、非正式對談或技能評估進行評估作業。確定內部員工現有的 AI 知識與技能將有助於確定培訓範疇、重點培訓領域及培訓主題的優先順序。

❸盤點職責與任務

檢視培訓參與者的職位與職責，並確認 AI 服務將如何融入工作流程。此項分析將有助於訂定培訓過程中應涵蓋之應用場景教學，確保培訓的相關性及可應用性。

❹設定培訓目標

根據已確定的待培訓人員、其知識缺口及工作內容，明確界定培訓的預期成果。清晰的目標將有助於機關進行培訓內容的具體計畫。

- 系統管理者：著重如何監控與優化 AI 系統，與是否具備為使用者解惑的能力。
- 系統使用者：掌握如何使用 AI 系統加速作業效率，且確保使用時遵守內部規範。

思考點：人員培訓與素養思考

為了讓 AI 專案順利推動，可試想：參與的同仁是否具備基本的 AI 素養？了解專案流程、模型評估指標、資料倫理與模型維運等核心知識，能幫助他們更好地界定業務痛點、與技術團隊溝通，並掌握專案風險。同時，也別忘了盤點過往的實務經驗，將成功經驗與學習轉化為這次專案的助力。

2. 與 AI 服務廠商建立學習管道

在 AI 專案進行過程中，可要求 AI 服務供應商提供必要的教育培訓服務，且應針對使用者與管理者提供不同的教育訓練。這可能包括 AI 服務的相關文件說明，以及由廠商分享系統開發、部署及維護的專業知識。此外，在溝通過程中，應當與對方探討如何教導非技術人員 AI 服務的正確使用方式。

訓練內容

以下列舉正式導入 AI 服務前所需進行訓練之主題。具體訓練內容根據機關於教育訓練前置作業環節所得出之結論進行適當調整，提供對應課程：

1. 共通課程內容

❶ AI 服務及其功能介紹
- AI 服務的基本概念與技術
- AI 服務的功能及其適用範圍
- AI 服務的使用方式

❷ 應用場景與實際案例
- 使用案例說明
- AI 服務實際操作
- AI 服務的侷限性

❸ AI 系統的優化與監控
- 如何解讀 AI 的產出
- 機關AI 相關規範
- 資料與隱私
- 負責任地使用AI

❹ 疑難排解
- AI 服務相關文件介紹
- 問題申報機制
- 聯絡窗口

思考點：系統侷限與使用素養

為了讓 AI 系統被正確、安全地使用，可試想：我們是否已清楚說明 AI 的侷限性？像是模型能做什麼、不能做什麼，以及它的預測結果會受到哪些資料限制影響。同時，也別忘了提醒使用者，AI 是輔助決策的工具，最終判斷仍需仰賴人的專業與經驗。

2. 差異化課程內容

課程內容	系統管理者	系統使用者
技術深度	了解AI 建模原理、如何輸入資料與調整參數概念。	了解如何使用 AI 分析的結果，以及使用的侷限。
疑難排解	能進行初步AI 系統故障排除，以及模型性能調整。	了解如何辨識異常，以及對應回報機制。
資料安全	了解系統資料使用範疇，能監控訓練資料的安全狀態。	使用 AI 系統時，能遵守組織內資料使用規範。
資料管理	負責監控AI 資料清理前置準備，包含資料清理、標註。	了解如何輸入和使用資料，確保 AI 系統能獲最佳成果。
流程整合	了解AI 系統與其他系統如何整合，或是擴大部署範圍。	掌握 AI 如何納入現有的工作流程並提高效率。

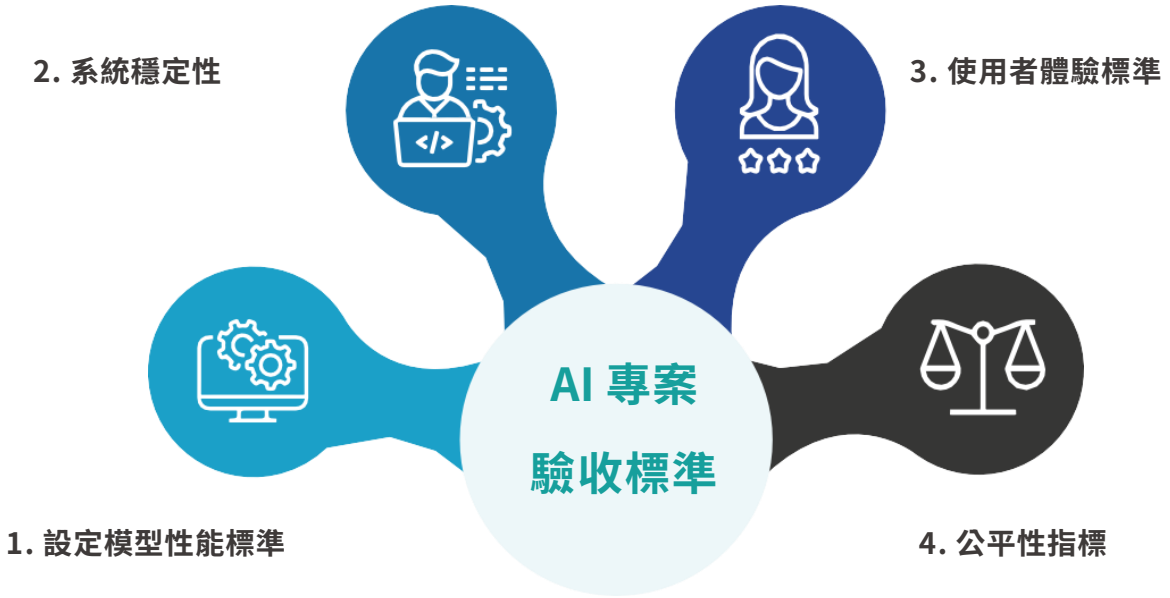
有關培訓課程內容，機關應與 AI 系統建置團隊共同規劃，確保學員於受訓後具備上述能力，以及面臨問題時能取得相關協助資源。

本章節探討之教育訓練目標為確保 AI 系統後續使用可達到目標，故以 AI 系統上線後之應用管理作為教育訓練主題。然而於 AI 專案建置過程中，承辦同仁也應具備基礎 AI 素養，以利於專案管理與風險控管。相關素養課程學習資訊，可參照章節 2.4 政府公開 AI資源中，有關學習資源段落說明。

3.5：AI 教育訓練指引檢核清單

編號	檢核項目	檢核結果	檢核說明
3.5.1	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：_____	
3.5.2	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？	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.5.3	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：_____	
3.5.4	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：_____	

3.6：AI 專案驗收標準



AI專案模型表現就技術面而言具有一定的不確定性，其最終驗收標準仍因專案性質與目標不同而存在差異。目前常見驗收模式會有四種，分別為設定模型性能標準、系統穩定度、使用者反饋與公平性指標。

設定模型性能標準

* 來源：Freepik

模型性能驗收標準主要為機器學習在技術面表現的指標，以常見 AI 的迴歸模型與分類模型為例進行舉例說明。迴歸模型評估指標包含均方誤差 (Mean Square Error)、絕對平均誤差 (Mean Absolute Error)、判定係數 (R-square) 與赤池資訊量準則（Akaike Information Criterion, AIC）。AI 分類模型評估包含準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall) 和 F1 分數 (F1 Score)。

並非所有模型皆適用以上2大種類型指標，實際需與技術專家共同討論決定。驗證這些標準的方法需要使用測試數據集進行模型評估，將驗證集資料輸入模型，並依照模型結果計算上述指標，並與預定標準做比較。

這些指標衡量傳統機器學習模型的預測表現。舉例來說，設定在一個預測網站是否為活躍使用者類別的分類模型，模型的準確率應達到特定百分比以上。有關指標的目標數值，將因專

案目標與欲解決的問題而有所差異。舉例來說，活躍使用者預測的精確度達 70%，也許已經具有一定的參考價值；而身分驗證相關的影像辨識，即使達 70% 仍無法取代要求近乎無誤的人工作業場景。

因此，除了與技術團隊討論外，專案團隊在理想情況下，可先行透過 POC 類型之專案，先行了解合理的模型性能標準為何，並納入評估現有訓練資源（包含技術與資料完整度）的齊備度，以及解決問題場景所需之技術標準，再設立對應的模型性能評估指標。

* 來源：Freepik

採用模型性能指標的優點是提供了量化指標做評估，能有客觀結果評斷。然而，指標設定有可能導致廠商未達驗收標準，而忽略了實際應用中的其他重要因素。例如，廠商可能專注於提高準確率以達到驗收標準，從而忽略了資料集中的少數群體。因此，在設計和評估模型性能指標時，必須全面考慮實際應用場景中的各種因素，平衡各種指標，確保模型在實際操作中的可靠性和有效性。

知識點：常見指標說明

迴歸模型常見指標

1. **均方誤差 (Mean Square Error)**：衡量模型預測結果與實際數據之間平均誤差大小的數字。
2. **絕對平均誤差 (Mean Absolute Error)**：衡量模型預測結果與實際數據之間平均誤差大小的數字。
3. **判定係數 (R-Square)**：計算出數值表示模型對數據變化的解釋程度。
4. **赤池資訊量準則 (Akaike Information Criterion)**：透過考慮模型對數據的擬合程度和參數數量，藉此來選擇最佳模型。

分類模型常見指標

1. **準確率 (Accuracy)**：正確預測比例，即 O 判斷 O，X 判斷為 X 的機率。
2. **精確率 (Precision)**：真陽性在預測為陽性中的比例，即被判斷且實際為 O，占所有判斷為 O 的比例為何。
3. **召回率 (Recall)**：真陽性在所有實際陽性中的比例，即判斷且實際為 O，佔所有真正為 O 的比例為何。
4. **F1 分數 (F1 Score)**：精確率和召回率的調和平均數，可看作為綜合指標。

系統穩定性

系統穩定性驗收標準主要包括錯誤率、反應時間和資源利用率等指標，以評估 AI 系統在實際運行中的可靠性和效率。考量現有 AI 應用服務，不乏生成式 AI 應用服務與 AI 模型應用平台建置的案例，皆應納入此面向之審查標準。例如，在一個生成式 AI 的問答機器人中，法律的回覆來源皆應來自指定資料庫，生成時間平均 5 秒以內，並且系統應能穩定運行。常見 AI 系統或平台建置，也可參考傳統系統建置的 SLA 標準。

使用者體驗標準

使用者體驗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，以評估 AI 系統是否滿足終端使用者的需求和期望。例如，機器學習 AI 儀表板的平台上，使用者滿意度應達到 80% 以上，並且使用者介面應該易於操作，以達成系統使用便利性的目的。驗證這些標準的方法包括通過問卷調查收集使用者反饋，進行使用者測試，並分析功能需求的實現情況。

* 來源：Freepik

公平性指標

AI 模型的公平性指標是用來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。這些指標旨在確保 AI 系統不會因為種族、性別、年齡、宗教或其他敏感屬性而產生偏見或歧視，從而保護所有使用者的權益。常見的狀況多指應用於比較不同種族或不同性別，在模型上的結果是否有重大差異。

以常見的公平性指標「均等機會差距」來舉例，此指標是用於衡量不同群體真陽性的機率是否相同，差異過大就代表有模型不公平的情況產生。舉例來說，有一個履歷篩選的 AI 模型用來判斷應徵者是否能獲得面試機會。若為公平的情況下，系統預測的準確率在男女群體間不應該有明顯差異。如果所有男性履歷的判定結果，通過檢測掃描且機器結果為正確的機率是 90%，而女性的機率為 70%，那這個系統在均等機會差距上就是不公平的。

實務作業中，為了達到公平性目標，AI 開發者會在模型訓練和測試過程中以用上述指標來檢測偏差，並通過調整數據、模型或訓練過程來減少偏差。

執行專案時，可視專案需求綜合選用上述 4 種指標，以避免執行團隊聚焦於單一指標，使專案成果不如預期。此外，除了前述四項指標，承辦人可依需求納入營運成本與資源效益、安全隱私、可擴展性等。在設計和評估模型性能指標時，必須全面考慮實際應用場景中的各種因素，平衡各種指標，確保模型在實際操作中的可靠性和有效性。

* 來源：Freepik

AI 產品的評測機制目前仍在持續發展中，可參考數發部數產署「AI 產品與系統評測中心」相關資訊。該中心的目標為建立國內AI 產品與系統評測體系，為國內的 AI 產品與系統提供評測服務。[14] 中心網站上提供了 AI 評測模擬測試題庫 [15] 供有需要的單位參考。目前，該中心 AI 測試實驗室處於試營運階段，服務暫時免費，並計劃於明年起研擬收費機制。[16]

思考點：驗收標準與效益衡量

在設定 AI 專案的驗收標準時，可試想：如果要用模型的表現來作為驗收依據，是否有參考人力或時間的基準？也建議與具備機器學習背景的專家討論，確認設定的標準在技術上是合理且可行的，這樣能幫助專案更順利地完成驗收。

14 成立背景-關於我們 | AIEC

15 下載專區 | AIEC

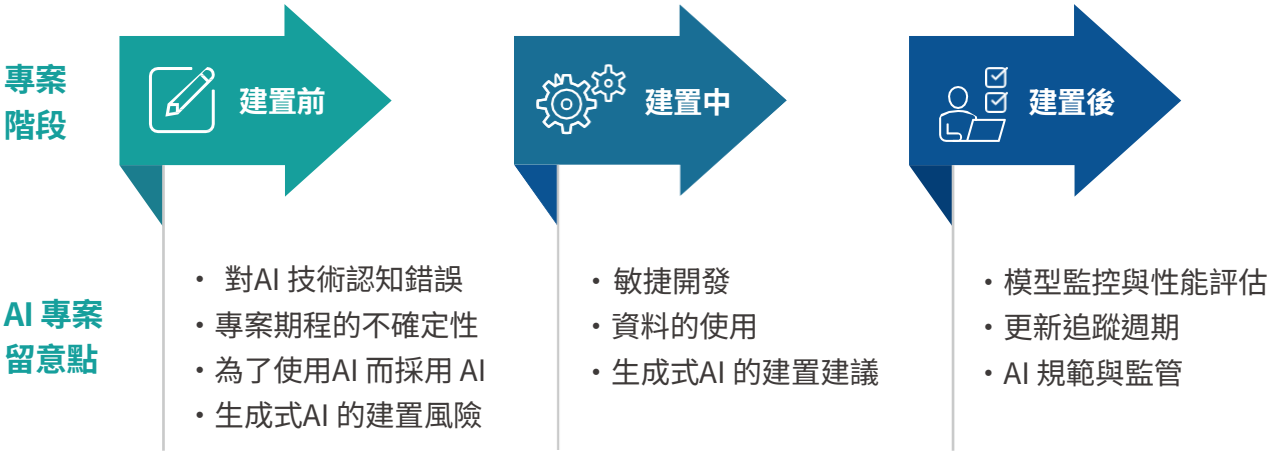
16 台灣首座AI測試實驗室來了！企業可免費預約「AI健檢」，工研院曝3大評估項目|數位時代 (2024)

3.6：AI 專案驗收標準檢核清單

編號	檢核項目	檢核結果	檢核說明
3.6.1	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：_____	
3.6.2	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：_____	
3.6.3	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：_____	
3.6.4	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：_____	
3.6.5	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：_____	

3.7：AI 與傳統資訊專案的差異

AI 專案與傳統資訊專案在本質上存在許多差異，需要機關在規劃、執行和後續營運階段特別留意。以下從專案的三大階段，介紹 AI 專案在執行過程中，其考量的面向與資訊專案有何不同，作為機關在推動 AI 專案前的參考：



階段	AI 專案與傳統資訊專案在專案執行過程中的差異
建置前	• **對AI 技術認知錯誤帶來不實的期望：**與傳統資訊科技不同， AI 作為近年快速發展的領域，除了技術飛速成長，其新穎程度也常使得人們對 AI 的技術能力有錯誤的判斷，進而造成不實的期望。政府機關可能因為缺乏 AI 相關知識及 AI 服務採購經驗，導致無法清楚表達 AI 專案所需的技術能力、訂定所要達成的目標與 AI 服務的合格標準，以及無法識別專案過程中技術面與需求面的實際需求。AI 專案負責人需要與專案的利害關係人進行深度溝通，讓大家對 AI 工具的預期效能有一致的理解。 • **專案期程的不確定性：**與傳統資訊專案相比，AI 專案的期程不容易預估與控制。使用者可能會認為 AI 只是一串程式碼，就像是一般的軟體，只需要設定好條件，AI 就會針對任務給出答案。然而，AI 模型的訓練包含複雜的資料處理與調校過程。高度的不確定性，導致團隊難以準確預估專案的期程。因此，AI 專案多採敏捷開發的形式，通過較短的開發循環，不斷迭代，快速應對變化多端的情境。 • **「為了使用 AI 而採用 AI」：**有些 AI 技術之所以被引進政府單位，只是因為專案中的成員希望能夠採用所謂的新技術。此外，AI 涵蓋了諸如「機器學習」與「深度學習」等技術，其中「深度學習」常採用的「黑箱模型」，可能導致部內人員在無從得知其確切運作的方式下，無法應用 AI 分析結果。政府機關不應該「為了使用 AI 而採用 AI」，而是須仔細思考相關的 AI 解決方案是否能夠派上用場，以及什麼 AI 技術符合對應的需求。 • **生成式AI 的建置風險：**若機關決定採用生成式 AI 的服務，則會額外衍生許多風險考量因素。例如，如何評估生成式 AI 解決方案的效能與限制，包含輸出的不確定性、生成回覆所需時間、回覆的精確度。也應盤點需使用之 AI 訓練資料，留意資料使用權限與資料存放是否符合機關規定。

階段	AI 專案與傳統資訊專案在專案執行過程中的差異
 建置中	• **敏捷開發：**AI 專案通常具有長期探索過程，且需要進行大量的測試與調整。此外，AI 技術的發展速度，導致 AI 專案的需求有可能會在進程中隨時變化。專案團隊需要採用敏捷開發的專案管理方法論，在一次次的迭代中，快速嘗試不同的模型架構、參數及訓練方式，並因應專案需求的變動隨時進行目標的調整。 • **資料的使用：**AI 技術由資料所驅動。因此相較於傳統資訊專案，AI 專案在資料管理方面具有全新的挑戰。AI 模型的訓練通常使用大量且多樣化的資料。若想確保模型的表現，則需要花費許多時間資源針對資料品質進行管理。此外， AI 系統可能會在訓練過程中運用到具有機敏性的資料。專案團隊需要採取相應措施以保障資料的隱私與安全。 • **生成式AI 的建置建議：**生成式 AI 的特點在於能夠產生各種類型的新資料的能力，生成過程中也包括了使用者與生成式 AI 的互動。因此，進行生成式 AI 的建置時，可以針對使用者的輸入指令進行篩選，打造內容過濾機制，避免使用者輸入任何含有個人資訊或是違反組織規範的要求。此外，可通過 AI 浮水印的技術，在生成輸出以後添加標記，幫助使用者及大眾識別 AI 生成的內容。 • **仰賴機關與執行團隊協作：**傳統資訊系統建置案，依照明確需求規格即可進行相關開發建置作業，然而 AI 專案系統產出是否符合機關期待，需仰賴機關與執行團隊的討論與驗證。舉例來說，生成式 AI 的產出並無標準答案，如何確保一個運用生成式 AI 的聊天機器人的回覆符合機關期待，將有賴於建置階段的各個測試環節，由機關與執行團隊進行說明討論。
 建置後	• **模型監控與性能評估：**專案團隊在 AI 模型投入運用後，除了需要持續監控其表現，確保模型效能良好，也需要依據任務需求，定期優化模型的能力。這可能包括：抓取新資料、優化訓練流程、調整模型架構等。傳統系統建置 IT 專案往往更關注於功能與漏洞修復。 • **更新週期：**傳統 IT 專案在完成建置後，多為保固與營運，較少進行太大的變動，更新週期較長。而 AI 作為快速發展的領域，團隊需要持續關注或研究新的算法以及模型效能優化方式，回應外界針對 AI 技術不斷增加的需求。 • **AI 規範與監管：**無論是 AI 系統運行錯誤，或是使用者錯誤地使用 AI 技術，都有可能帶來嚴重的後果。團隊需要建立完善的 AI 治理機制。監管面向可以參考國際主要 AI規範，從安全性、可解釋性、公平性、透明性等方面，定期檢核模型表現。

有鑑於 AI 應用發展專案與傳統軟體專案之管理方式、重點差異巨大，傳統的軟體發展專案管理可能不適合使用於 AI 應用發展專案，建議機關仔細尋找合適的專案管理軟體和其他工具，以支援需要靈活性的 AI 應用發展專案。

思考點：最終確認與風險管理

當 AI 模型即將上線時，可以思考是否已完成必要的最終確認與風險管理，以提升系統的穩定性與可控性。例如，可以考慮是否已與所有利害關係人充分溝通，建立對 AI 效能與限制的共同理解，確保各方對模型的預期一致且務實。同時，也可以安排多輪擴大測試，模擬實際環境中的使用情境，進一步驗證模型在不同負載與邊界條件下的表現是否穩定。透過這些準備，有助於在模型上線後降低潛在風險，避免對業務運作造成不必要的衝擊。

3.7：AI 與傳統資訊專案的差異

編號	檢核項目	檢核結果	檢核說明
專案建置前			
3.7.1	是否評估生成式 AI 解決方案的效能與限制，包含輸出的不確定性、生成回覆所需時間、回覆的精確度？	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.7.2	是否對於 AI 專案的效能限制，預先設想規劃可行的配套措施？	☐已完成 ☐未完成：_____ ☐不適用：_____	
3.7.3	是否盤點需使用之 AI 訓練資料，留意資料使用權限與資料存放是否符合機關規定？（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：_____	
專案建置中			
3.7.4	模型表現將隨訓練階段不同有所變化，是否預留一定容錯時間？	☐已完成 ☐未完成：_____ ☐不適用：_____	

編號	檢核項目	檢核結果	檢核說明
專案建置中			
3.7.5	是否導入使用者的提示 (Prompt) 內容過濾機制，避免使用者輸入探詢隱私資料、含有個人資訊、打破審核機制（jailbreak) 的指令？（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：_____	
3.7.6	是否在生成式 AI 服務中，讓使用者明確知道自己在使用生成式 AI 服務，如透過介面提醒或顯示數位浮水印等方式呈現？（ 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：_____	
專案建置後			
3.7.7	是否建立模型監控和性能的評估機制，定期優化模型表現，且同步追蹤國際規範標準，確認應用符合法律與既有規範？（ 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：_____	
3.7.8	是否預設生成式 AI 解決方案有可能會提供錯誤資訊，並提前規劃系統上線時的配套措施？（ 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：_____	
3.7.9	是否設立生成式 AI 模型回覆品質設立檢核機制，並定期檢核。檢核項目可包含以下面向：回覆真實且無幻覺（Hallucination）發生、沒有攻擊性、沒有偏見（bias）、沒有過分歧視？（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：_____	

04 AI 營運管理

章節摘要

- **AI專案的風險評估**
- **AI專案面臨的倫理、資料安全與資訊安全議題**
- **AI治理相關議題**
- **如何加速擴大推動 AI 專案**

在 AI 營運管理章節中，我們將重心探討 AI 系統建置後，如何有效的進行管理營運。首先，我們會詳細說明 AI 專案在進行過程中面臨的風險議題，強調專案的潛在風險與防範建議。

接下來，此段落會說明 AI 專案所面臨的倫理問題，如決策透明度、公平性和偏見，這些都是保證 AI 符合法律和既有規範的關鍵。隨之而來的是資料安全議題，涵蓋如何保護訓練資料的機密性和完整性，以及在資料收集和使用過程中的合法性問題。資安議題也是不容忽視的部分，本章會介紹如何防範網路攻擊和資料洩漏，確保 AI 系統的安全性。

除此之外，我們還將探討 AI 治理的各個面向，包括機關如何制定清晰的政策和流程以確保 AI 系統的負責任使用。最後，我們將說明如何推廣 AI 於後續運用，幫助機關能有清楚的依循標準，能輕鬆為未來 AI 專案擴大推動做好準備。

接下來將以可採取的行動為主、AI 監管概念為輔，進行進一步的風險、倫理、資安、資料應用議題探討。

知識點：解決生成式 AI 偏見常見方法

為了解決生成式AI的偏見導致需要下架的風險，關鍵在於預先規劃與流程標準化，以維護服務連續性與使用者信任，有以下常見因應方法：

1. 服務備援 (Fallback)：建立回溯或人機協作備援機制，確保即使模型下架，服務亦能立即切換至安全模式或簡化流程，避免服務中斷。
2. 流程量化與標準化：制定標準作業流程，將錯誤分級，並透過模擬測試量化微調與重新上線所需的時間與成本，作為內部服務等級協議(SLA)的依據。
3. 維護信任：採取透明溝通策略，即時公告修正進度，強調調整是為提升服務的公正性與準確度，並告知預計的恢復時間範圍。

4.1：風險管理

AI 系統專案在執行過程中，需要留意各階段可能發生的潛在專案風險，並依照風險基礎原則加以管理。以下段落依照 AI 專案建置前中後三階段，說明對應防範舉措。在進行 AI 專案時，可依專案進行進程核對。

* 來源：Freepik

思考點：模型維運與驗證提醒

為了確保 AI 模型在長期運營中能維持其效能與業務價值，並及早發現資料漂移（Data Distribution Shifts）風險，可試想：是否已根據模型的關鍵性、應用場景的變化速度，以及訓練資料的時效性，來決定一套合理的、定期的模型驗證頻率？確保持續的驗證流程能夠有效確認模型在實際業務中的可用性與準確度。

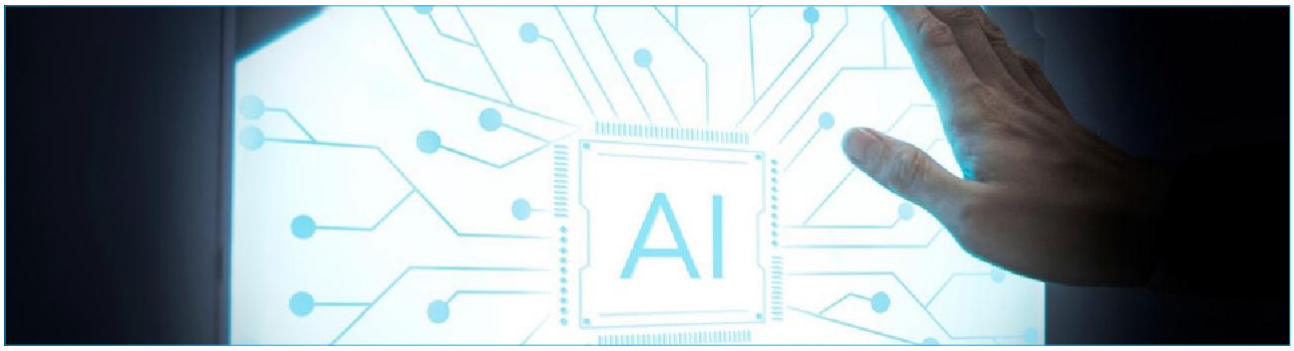
4.1：風險管理檢核清單

編號	檢核項目	檢核結果	檢核說明
建置前			
4.1.1	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、資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.1.2	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：_____	
4.1.3	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：_____	
建置中			
4.1.4	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、及來源、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：_____	

編號	檢核項目	檢核結果	檢核說明
建置中			
4.1.5	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：_____	
4.1.6	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：_____	
建置後			
4.1.7	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：_____	
4.1.8	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：_____	
4.1.9	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：_____	

4.2 倫理議題

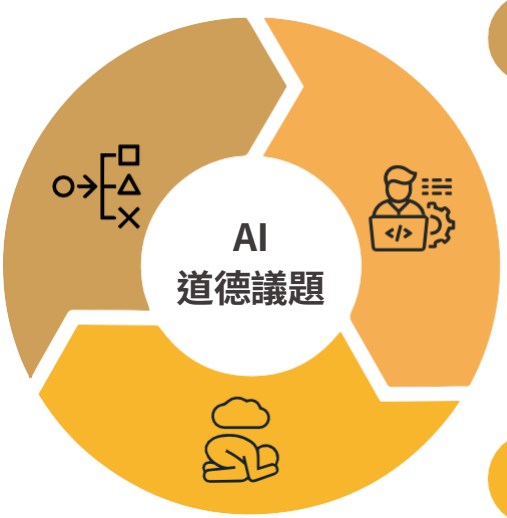
AI 倫理議題為在進行 AI 開發應用時，可能涉及之道德、法律及社會影響，涵蓋人類權利是否被影響、隱私性、公平性等議題。AI 系統應用將有可能對個人與社會造成重大衝擊，故確保 AI 系統應用不造成負面影響至關重要，開發團隊應將 AI 牽涉的倫理議題納入重點考量。



* 來源：Freepik

AI 倫理議題

多數 AI 系統開發者皆無意開發出危害社會或歧視的 AI 系統，但仍有可能衍生相關倫理議題，以下列舉之：



1. **以錯誤方式使用 AI 系統：**使用者以不正當意圖使用 AI 系統，使其產生非預期產出。例如對教育用之生成式 AI 服務詢問犯罪問題。

2. **開發者未能全面考量：**開發階段未能充分考量，產生漏洞。例如為擴大訓練資料提升 AI 效能，不經意取得不當資料進行訓練產生偏誤。

3. **未能預期的負面影響：**應用時未能考量到的負面影響，於系統上線時才被發現。例如使用 AI協助履歷篩選，發現自動將少數族裔名字當作關鍵變數來排除對應人選。

思考點：系統實境測試與倫理考量提醒

為了確保 AI 系統的設計是公正、公平且適用於所有目標使用者，可試想：是否已在系統測試期間主動納入不同背景、部門或使用情境的多元群體來收集回饋？我們如何利用這些實際情境中的數據和意見，來全面了解系統的表現，並確認其是否可能對特定群體產生潛在的偏見或不利影響？

以下參考英國官方圖靈研究院標準，在 AI 與生成式 AI 應用時，建立一個基礎倫理監管框架，提供給 AI 系統建置團隊做最基礎之參照，避免常見倫理議題，共有五點[17、18、19]：

1. **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。
2. **完善 AI 專案權責機制**：機關應可對 AI 系統影響負責，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。
3. **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。
4. **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。

17. Fast Track Principle, Research Gate （2021）Fast Track Principle, Research Gate （2021）
18. 人工智慧（AI）產品與系統評測參考指引（草案），數位發展部數位產業署（2024）
19. 金融業運用人工智慧（AI）指引，金融監督管理委員會（2024）

5. **保持人為參與：**除典型 AI 道德議題，也建議機關於 AI 工具使用過程中保持人為參與，包括前期模型訓練、測試、AI 工具使用、以及使用 AI 工具進行決策等。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。

思考點：AI 的人工評估提醒

決定導入 AI 系統之前，我們必須從倫理和價值創造的角度進行審慎評估，可試想：是否已透過人工、跨部門的方式，嚴謹評估所選定的 AI 適用情境，確保該應用能實質創造明確的公共利益與社會效益？我們如何全面識別並避免因 AI 導入而可能產生或放大的不必要風險（例如對服務品質的影響、倫理爭議或資源浪費），從而確保技術的應用是合理且負責任的？

4.2：倫理議題檢核清單

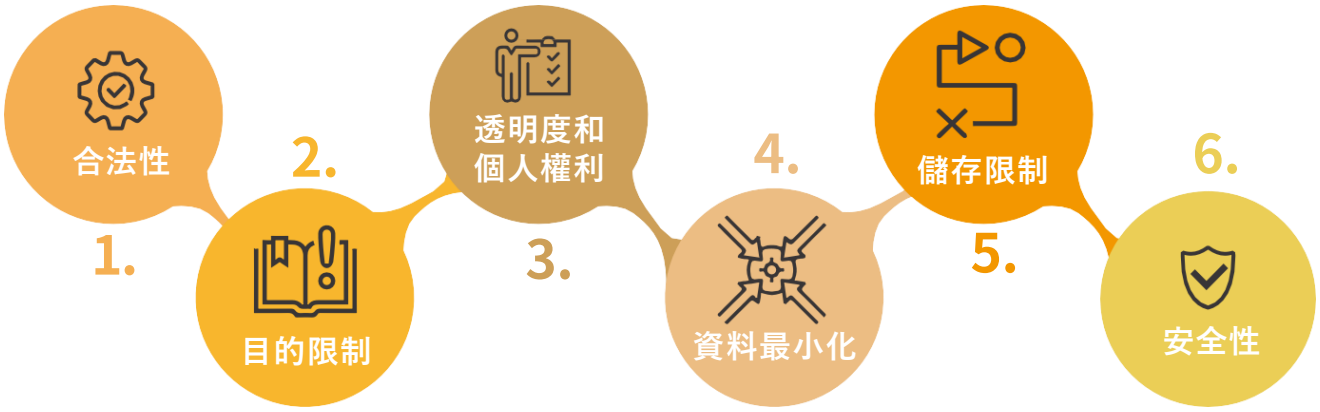
編號	檢核項目	檢核結果	檢核說明
公平性			
4.2.1	抽樣是否無偏見，以公平資料集進行模型訓練？（參照行政院及所屬機關（構）使用生成式AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.2.2	是否使用合理的分析特徵，設計無偏見的分析架構與流程？（參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.2.3	是否審查輸出內容與決策建議的合理性，不以特定宗教、國籍、種族、性別、身心障礙、性傾向、政治傾向、年齡、文化等因素而產生有害偏見或歧視，以保護受眾權益？（參照行政院及所屬機關（構）使用生成式 AI 參考指引）	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.2.4	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？（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：_____	
4.2.5	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？（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：_____	
4.2.6	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？（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：_____	

編號	檢核項目	檢核結果	檢核說明
當責與負責			
4.2.7	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？（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：_____	
4.2.8	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？	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.2.9	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？（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：_____	
4.2.10	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？（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：_____	
4.2.11	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？（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：_____	
透明度和可解釋性			
4.2.12	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？（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：_____	

編號	檢核項目	檢核結果	檢核說明
透明度和可解釋性			
4.2.13	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？（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：_____	
4.2.14	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。（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：_____	
4.2.15	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？（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：_____	
4.2.16	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？（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：_____	
保持人為參與			
4.2.17	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 ？（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：_____	
4.2.18	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？（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：_____	

4.3：資料安全議題

AI 系統在訓練和應用階段，將有可能使用到機敏資料，故機關在導入 AI 服務時，需要考慮如何保護與前處理。在進行模型資料訓練時，可參考以下資料使用原則：



1. **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。
2. **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。
3. **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。即便資料主體已經授權，仍有權撤回並請求機關刪除相關個人資料並不再運用。
4. **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。
5. **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。
6. **安全性：**以合適技術和配套措施來保護機敏資料。

識別資料來源非常重要，因為 AI 系統建置過程中，負責訓練資料存放管理的控制者將承擔個人資料管理責任。此外，若要在大型語言模型中，使用個人資料進行訓練，除了進行去識別化外，應了解並嚴格遵循資料保護相關法規。

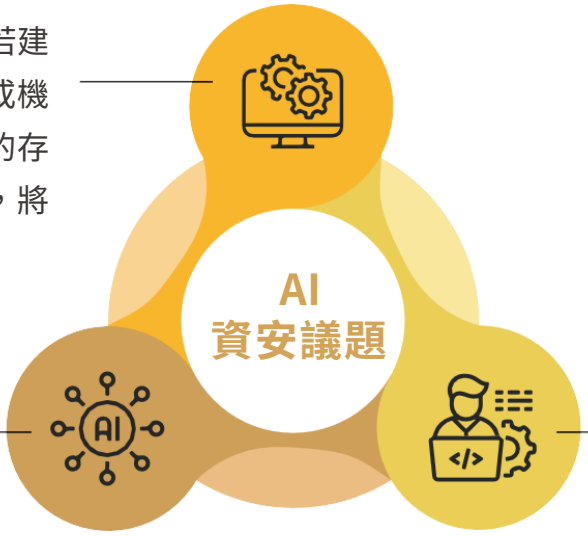
4.3：資料安全議題檢核清單

編號	檢核項目	檢核結果	檢核說明
AI 資料使用			
4.3.1	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：_____	
4.3.2	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：_____	
4.3.3	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：_____	
4.3.4	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：_____	
生成式 AI 應用			
4.3.5	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：_____	
4.3.6	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：_____	

4.4：AI 資安議題

AI 專案進行時，除了常見系統面臨資安風險外，也需關注 AI 專案將面臨的資安議題。包含但不限於：

1. 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。



2. 攻擊者可能透過注入惡意資料影響模型訓練結果。

3. 利用反向工程，試圖竊取模型架構與模型內部資料等。

另外，生成式 AI 現有商業服務存在的資安風險，機關導入時應留意以下面向：

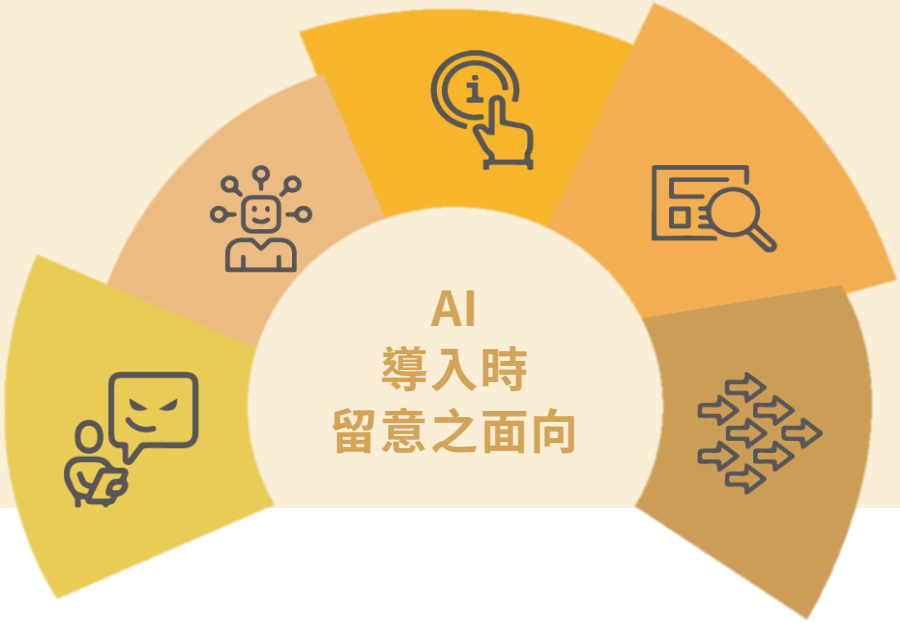
1. **提示注入威脅 (Prompt injection threats)：**不當使用生成式 AI 模型的提示 (Prompt) 所造成的資安漏洞。

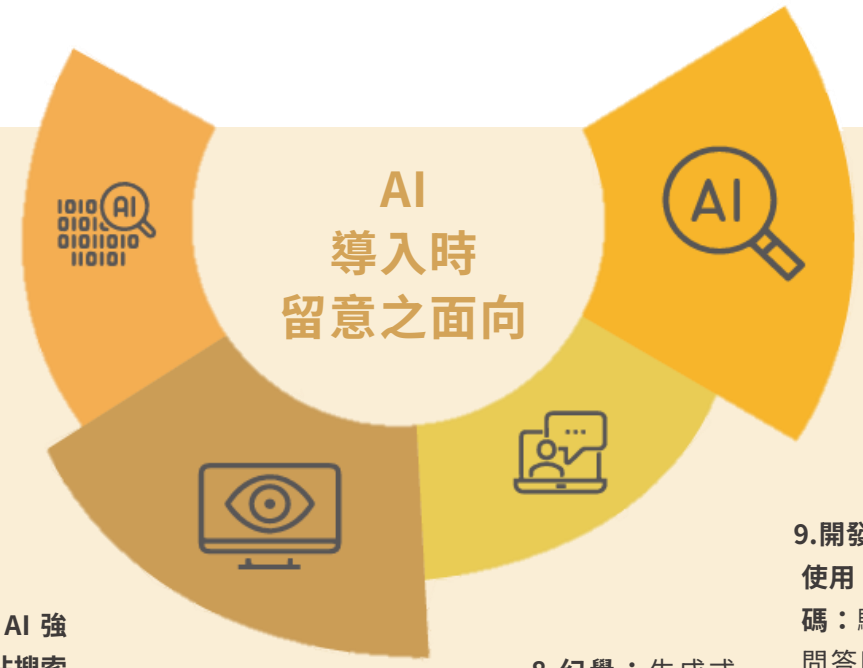
2. **政府網站上的生成式 AI 聊天機器人：**民眾詢問報稅連結，聊天機器人提供錯誤的連結給民眾，導致民眾被詐騙。

3. **以生成式 AI 強化現有政府網站的搜索功能：**有心人士可透過提示 (Prompt) 使查找內容出現有害資訊。

4. **私人生成式 AI 聊天機器人傳回建議的文件來源：**許多私人生成式 AI 會提供回應參照的參考資料來源連結，有心人士刻意於組織內資料庫上傳可疑的資料，使生成式 AI 提供此連結給其他尋找相關文件的使用者。

5. **資料外洩：**生成式 AI 的回覆洩露了敏感訊息，像是個人資料。





6.以生成式 AI 強化內部網站搜索功能：使用者透過內部生成式 AI 資料查找時，找到隱私或不符權限規範的資料。

7.生成式聊天機器人總結聊天對話：使用者詢問生成式 AI 一個問題涵蓋多個關鍵字，可能導致生成式 AI 判別語意時有所錯誤，提供不相干且侵害他人隱私回答訊息。

8.幻覺：生成式 AI 回應的訊息看似真實，但實際上是錯誤的。

9.開發人員未經審查便使用 LLM 產生的程式碼：駭客發現 LLM 在問答時會建議使用者安裝某一個不存在的軟體，便製作了一個同名的惡意軟體，使開發人員未留意便安裝了有害程式碼。

目前 AI 應用涉及資安議題多與資料外洩，以及全然相信生成式 AI 回覆，使有心人士有機可乘。使用者使用 AI 時，不可一味相信生成式 AI 結果，避免造成資安漏洞。可參考《資通安全管理法施行細則》與 1.6 AI 應用涉及之法規範中的資訊安全相關管理辦法。

思考點：API 串接安全與連結風險

與外部生成式 AI 服務進行 API 串接時，核心風險在於機敏資料的洩露，因為資料在傳輸和處理過程中存在潛在的暴露威脅；同時，AI 回覆可能夾帶未經驗證或惡意的外部連結，對使用者和機關系統構成嚴重的資安隱患。為有效應對，應採取雙重防護策略：

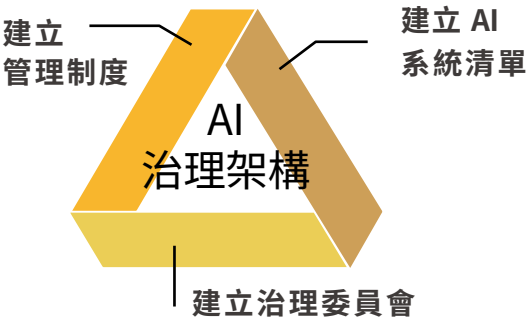
- 串接層面：實施隱私強化技術 (PET) 以降低資料敏感度
- 輸出端：須建立嚴格的內容檢核與連結過濾機制，並規定預設禁止或自動移除 AI 回覆中所有未經資安掃描與驗證的外部連結，確保系統和資料的安全。

4.4：資安議題檢核清單

編號	檢核項目	檢核結果	檢核說明
4.4.1	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：_____	
4.4.2	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：_____	
4.4.3	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：_____	

4.5：AI 治理架構

AI 治理涉及 AI 系統生命週期之風險管理、權益保障、資料隱私、資通安全、人員訓練、管理制度等事項，可透過法律、法規命令、行政規則、行政指導、自律規範、產業標準、使用契約、流程設計、審核制度、爭議處理等機制加以落實。



建立管理制度

機關使用 AI 應用時應建立必要的管理制度，特別在 AI 應用出現故障、異常，應該由誰負責處理。AI 應用出現問題的原因非常多樣，例如 AI 訓練資料存在偏誤、模型測試與驗證不足、AI 應用部署過程失敗、後續使用者不當使用等。機關評估是否採用 AI 解決方案時，應將管理制度納入評估項目之一。機關可規劃 AI 應用管理制度，列出 AI 設計、建構、營運與維護過程，由誰負責哪些工作，並在內部文件與流程中明確規劃。若管理制度無法讓利害關係人接納時，貿然發展 AI 應用發展過程卻無法管理，將導致失敗。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，以及誰具有監督與糾正錯誤的權限與責任。

建立AI 管理制度時，模擬問題發生時是否完成對應的管理規劃

1. 審核模型是否達成其目的，解決關鍵痛點。
2. 機關須建立 AI 系統的管理機制。
3. 需要制定 AI 系統的測試和監控框架。
4. 稽核演算法是否穩定、無偏見、公平且可解釋。
5. 回應與定期審視專案所使用資料，是否符合公民和使用者對其資料使用的期望。

建立AI系統清單

為掌握機關內 AI 的應用現況，機關應考慮建立 AI 應用清單，提供管理單位一個全面視角了解已部署 AI 系統。

清單有助於管理層和利害相關人，了解 AI 在各個計畫和專案中的使用範圍和規模，提供更好的監控，以及對任何使用 AI 進行決策的風險（如資料品質、模型準確性、偏見、安全漏洞和法規範）的認知。清單應定期更新，包含以下詳細資訊：

1. 描述每個系統的用途、使用情況和相關風險。
2. 提供 AI 系統細節，包括但不限於所使用資料、系統所有權、開發紀錄和關鍵日期等詳細資訊。
3. 遵照組織規範 （例如資料蒐集頻率、資料驗證機制、版本控制）、架構（例如資料存放與存取權限）和相關工具來維護準確和完整的清單。

建立治理委員會

定期且全面的 AI 治理議題討論，將有助於組織管理並強化後續的 AI 應用推動。委員會將作為 AI 治理框架的一部分，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，也可成立 AI 倫理委員會或分配對應職責角色於 AI 治理委員會中。

AI 治理委員會或委員會中的 AI 代表，將作為監督、管理和策略制定的角色，提供該機關AI發展管理以及後續推動方針，可以包含機關內高階主管、技術專家、人力資源專家、法律專家等。倫理委員會的主要職責是從倫理角度，評估組織內各種行動、計畫和決策的影響，重點關注公平性、透明度和隱私權等價值觀，可以包含倫理、行為科學、社會心理、人機互動、具產業領域知識之專家。

有關治理委員會設計與推動模式，目前仍在持續發展中。AI 治理委員會之成立，將有助於機關整合 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。

* 來源：Freepik

知識點

1.AI 治理委員會：指組織內為監督與指導所有 AI 相關活動而設立的跨部門高層次決策機構。其職責包括制定 AI 策略、評估倫理風險、確保法規遵循，並對重大 AI 專案進行最終審批。

2.AI 代表： 指在組織內 AI 治理架構中，領導 AI 治理委員會的最高階決策者。其擁有 AI 相關事務的最終決策權與執行權力，負責批准與監督所有重大 AI 策略、倫理規範及風險管理框架，是組織內 AI 戰略與治理的最高權威。

思考點：AI 治理提醒

為了確保 AI 專案從頂層設計上就符合治理、倫理與問責的要求，並獲得最高管理階層的承諾，可試想：是否已明確指派或成立高層級的組織結構（例如 AI 治理委員會或 AI 代表），來定義 AI 政策、分配開發與營運的角色責任？我們如何確保這個治理機制能有效納入跨部門、跨機關的利害關係人與外部倫理專家，以系統性地識別並處理 AI 應用中潛在的倫理爭議、公平性問題與法律風險？

建立 AI 審核機制

為了確保 AI 專案在規劃期能解決對應問題且不會帶來潛在風險，上線營運後能確保其表現與風險因素受良好監督，採取 AI 審核對 AI 專案來說非常重要。AI 審核需要分析這些 AI 演算法如何運作、AI 是否按預期運作以及運作過程中是否產生其他對社會危害。因此，完整之 AI 審核計畫涉及對人工智慧系統的設計、開發、部署和運行進行全面檢查，以確保其符合道德、法律和技術標準。

1. 規劃審核範圍和目標

- 訂立 AI 專案的 AI 審核具體目標：常見包含資料存放、演算法偏見、資安議題等。
- 確定 AI 審核的範圍與週期：需審核之系統、應用程式和流程，並規劃審核的固定週期。
- 確定 AI 審核的項目：基礎審核項目包含數據安全隱私、模型可解釋性與透明度、是否存在偏見、模型效能表現、模型是否符合法律與規範及模型對使用者可能帶來的潛在影響。可參考 1.5 章 AI 應用挑戰相關之議題。

2. 收集審核所需準備資料

- 收集有關審核範疇相關的系統檔案，如系統規格書、數據資料、訓練方法和算法說明。
- 獲取運行數據和使用紀錄，以及期間內使用者回報之 AI 系統問題。

3. 審核執行

- 審查資料：
  1. 檢查數據處理過程，確保數據收集、儲存和使用符合隱私和安全標準。
  2. 評估算法的公平性和透明性，檢查是否存在偏見和歧視。
  3. 分析系統的可解釋性，確保決策過程透明度與可追溯性。
- 測試驗證：
  1. 使用測試驗證 AI 系統的性能和可靠性。
  2. 進行模擬和壓力測試，檢查系統在不同情境下的表現。

4. 監控改進計畫

撰寫 AI 審核報告，總結發現的問題和風險。針對發生問題提出改進建議，並建立問題改善監控機制，確保問題解決且後續皆符合標準。

思考點：AI 審核機制設計提醒

為了持續監管 AI 系統，可試想：是否已設計一套組織內的 AI 審核機制，確保其能與系統運營同步進行？我們是否已諮詢第三方專家，以確保該機制在流程、專業度與即時性上足夠嚴謹？

目前 AI 審核機制仍在持續發展中，未有明確標準。相關審核計畫安排，稽核頻率與項目可依專案性質訂立。機關於設計 AI 審核機制時，可就 AI 系統於應用場景會面臨的風險議題進行審核項目設計，並結合相關 AI 服務驗證工具，以達 AI 審核定期監控 AI 系統表現與潛在風險議題防範的目標。以上措施將有助於 AI 系統長期發展，若於 AI 專案初期導入相關機制，將有利於後續穩健發展。

4.5：治理議題檢核清單

編號	檢核項目	檢核結果	檢核說明
4.5.1	是否於 AI 專案開始時，建立對應完整的管理制度，並隨系統上線、擴增等情境逐步彈性調整？	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.5.2	是否建立 AI 系統清單，掌握您機關內所有已部署 AI 系統？（參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	

4.6：後續專案推動

我們期待 AI 應用能在機關內快速穩健的發展，在完成機關內第一個 AI 系統導入後，可參考以下規劃，以利未來更多 AI 專案之推動。

實現現有 AI 專案成效

若第一個機關 AI 服務成效可有效體現，除了提高組織後續推動 AI 應用意願外，也會是後續機關學習的標竿，故確保現有 AI 專案獲得成功會是很重要的一件事。

* 來源：Freepik

❶從教育訓練提升服務使用體驗：

使用者對於 AI 應用的期待與想像，通常高過現有 AI 所能達成的狀態，多數人想像 AI 是一個全自動的服務，但實際上生成式 AI 是與人們協作的工具，且與 AI 協作需要額外學習熟悉。故如何讓使用者了解 AI 工具限制，並學會與 AI工具協作，會是提升使用者 AI 服務體驗重要的環節。AI 教育訓練內容應從實際應用場景出發，並陪伴使用者熟悉工具使用，以確保成功。

❷記錄專案成果並追蹤實際成效：

作業效率提升是 AI 工具的常見效益，機關應實際追蹤 AI 系統導入前後對於作業效率的改變資料。除了體現 AI 系統的成果外，也可依照此基準回顧是否達成建置前的目標。另機關可以在內部分享 AI 應用案例，例如如何更好的下生成式 AI 指令來完成工作，或是機關內推行 AI 專案的經驗等。類似討論都可以促進同仁對於 AI 應用的興趣，並發想更多應用可能。

完善組織資料治理

AI 訓練基於資料，而充足且高品質的資料仰賴良好的資料治理制度。資料治理並非一蹴可幾，除技術面管理整合資料並依循標準外，也仰賴機關利害關係人建立對於資料治理的共識。若 AI 應用納入品質不佳的資料，除 AI 表現不佳外，甚至可能阻礙服務推行。所以，機關若期待於未來導入更多 AI 應用，開始掌握組織內部具價值的關鍵資料狀態，並適時推動組織內資料治理制度，並考慮擴大可用資料來源，如文件、音訊、影像等其他非格式化資料，能避免資料到用時方恨少的遺憾。同時，良好資料管理機制將提升資料的可用性，成為跨部門 AI 訓練資源共享的基礎。

思考點：資料治理建議

若資源允許，建議深入了解組織的資料治理現況，並著手完善內部資料治理框架，這樣做將能大幅提高 AI 專案在資料取得、資料品質控管與應用的便利性與合規性。從長遠來看，這項工作能進一步建立跨部門資料存取的共用基礎設施，為機關後續所有 AI 應用與數據驅動的決策提供穩定且高效的基礎支持。

4.6：後續專案推動檢核清單

編號	檢核項目	檢核結果	檢核說明
4.6.1	是否了解現有 AI 服務使用體驗回饋內容項目，依照使用者場景滾動式提供教育訓練？	☐已完成 ☐未完成：_____ ☐不適用：_____	
4.6.2	是否開始盤點組織內部資料，推動資料治理標準，提升關鍵資料的資料品質？（參照資通安全管理法及子法）	☐已完成 ☐未完成：_____ ☐不適用：_____	

名詞解釋

編號	名詞	解釋
1	**準確率（Accuracy）**	在評估機器學習表現的混淆矩陣中，真陽性與真陰性佔全體樣本的比例。舉例來說，在一個以辨識出狗照片為目標的機器學習模型中，如果模型正確的辨識出 10 張狗的照片，以及 10 張不是狗的照片，而資料集中共有 40 張照片，則準確率是 50%。
2	**赤池資訊量準則（Akaike Information Criterion）**	透過考慮模型對數據的擬合程度和參數數量來選擇最佳模型。AIC 可協助找出參數數量和擬合數據間的平衡。
3	**資料飄移（Data Distribution Shift）**	為監督式學習的一種現象，當資料隨著時間變動時，模型效能隨之下降。
4	**深度學習（Deep Learning）**	深度學習是讓電腦模仿人腦學習，使用多層「神經元」來處理資料。這樣電腦能識別圖片、文字和聲音中的複雜模式，並自動完成像描述影像或語音轉文字等任務。
5	**數位涵容（Digital Inclusion）**	運用科技縮減數位落差，確保有障礙人士及弱勢族群能享有與一般民眾相同品質的政府服務。
6	**F1 分數（F1 Score）**	在評估機器學習表現的混淆矩陣中，精確率和召回率的調和平均數，可看為綜合指標。
7	**生成式 AI（Generative AI）**	生成式AI 是深度學習模型的一種應用，可根據訓練的資料產生文字、圖像、音訊和程式碼等內容。
8	**幻覺（Hallucination）**	在 AI 領域，幻覺指的是生成式 AI 模型產生看似真實，但其實不正確或誤導性的結果。導致錯誤的原因很多，包括訓練資料不足、模型存在錯誤的假設或用於訓練模型的資料有偏差等。
9	**超參數（Hyperparameter）**	超參數(Hyperparameter)為訓練模型時用來控制訓練過程的變數。哪些是有效的超參數以及超參數的最佳值，可以手動或透過超參數演算法自動調校。
10	**資料標籤（Labeling）**	資料標籤是一種幫原始資料新增的分類資訊，資料標籤可以是數字或文字，目的是讓 AI 可以多一個從中學習的分類。舉例來說，你有很多不同品種貓狗的體重資料，你新增一個資料標籤標註該筆資料是貓是狗，將有助於 AI 做體重預測時，能納入此變數做為考量。

編號	名詞	解釋
11	機器學習 (Machine Learning)	機器學習是讓電腦從大量資料中學習，然後自動做出預測或決定。舉例來說，就像教導小朋友認識貓和狗一樣，電腦也能根據學到的資料來辨識新出現的貓和狗。
12	絕對平均誤差 (Mean Absolute Error)	衡量模型預測結果與實際數據之間平均誤差大小的數字，通過將每個誤差取絕對值後取平均來計算，計算出的數值越小表示模型預測越準確。
13	均方誤差 (Mean Square Error)	衡量模型預測結果與實際數據之間平均誤差大小的數字，通過將每個誤差平方後取平均來計算，計算出的數值越小表示模型預測越準確。
14	模型 (Model)	模型是一種用來簡化和模擬真實世界的工具。在機器學習中，模型是透過分析大量資料建立起來的數學公式，這個公式可以用來進行預測或做出決定。
15	精確率 (Precision)	在評估機器學習表現的混淆矩陣中，在預測結果為正樣本之中，有多少比例實際上為正樣本。舉例來說，在一個以辨識出狗照片為目標的機器學習模型中，如果模型辨識出有 10 張狗的照片，而實際上這 10張照片中，只有 8 張真的是狗，另外兩張是貓，則精確率是 80%。
16	召回率 (Recall)	在評估機器學習表現的混淆矩陣中，真陽性在所有實際陽性中的比例。舉例來說，在一個以辨識出狗照片為目標的機器學習模型中，如果模型正確的辨識出 10 張狗的照片，而實際上狗的照片在資料集中有 20 張，則召回率是 50%。
17	檢索增強生成 (Retrieval Augmented Generation, RAG)	透過結合資訊檢索和文本生成的技術，利用外部知識庫來提高回答的準確性和資訊量，常用於應對大型語言模型的幻覺問題。

參考資料

Generative AI Framework for HMG, GOV.UK （2024）
Public Sector AI Playbook, Singapore Government Development Portal （2021）
A guide to using artificial intelligence in the public sector, GOV.UK （2020）
Crisp DM methodology, Smart Vision Europe

moda 數位發展部
Ministry of Digital Affairs

公部門人工智慧應用參考手冊