國科會 114 年度
「邁向新世代前瞻人工智慧研究專案」計畫徵求公告

壹、專案背景

人工智慧(AI)技術近十年來發展飛速：2016 年 AlphaGo 打敗韓國圍棋冠軍李世乭(Sedol Lee)，2020 年 AlphaFold 根據蛋白質氨基酸序列準確預測其三維結構，2022 年底 ChatGPT 展現擬人的自然語言能力，席捲全球。這些突破性地進步，帶給人類社會前所未有的機會與挑戰。回顧 2024 年，OpenAI、Google、Anthropic 分別推出了多模態模型，能夠進行語音對話及圖像理解；歐盟頒布並開始施行首部 AI 法規：The EU AI Act；諾貝爾物理學獎頒給了 John J. Hopfield & Geoffrey E. Hinton，化學獎頒給了 David Baker、Demis Hassabis & John Jumper 結合領域與人工智慧研究的團隊。一方面 AI 技術推動了全球經濟的增長，然另一方面 OpenAI 解散了他們的 AI 倫理團隊，引發各界對 AI 負責任使用的擔憂。進入 2025 年，開源模型 DeepSeek v3 降低一次訓練成本至 560 萬美元，引起全球 AI 界廣泛的關注；而 DeepSeek R1 利用大規模的強化學習技術進行後訓練，在數學、代碼和推理任務的表現可媲美 OpenAI o1 模型，更是為全球科技市場投下不小的震撼彈。

人工智慧持續佔據了人們的討論焦點。根據 MIT 科技評論、時代、富比士雜誌報導，2025 年 AI 的重點趨勢包括自主 AI 系統、生成式影音及虛擬世界、能夠推理的大型語言模型、科學領域的 AI 大爆發、以及攸關 AI 治理與國防安全等議題。AI 將從簡單的客服對話系統轉向能夠自主完成任務的「代理人」系統，處理複雜的任務，如操作應用程式介面、編寫軟體。多模態 AI 將同時處理文本、圖像和影音，同時也能生成 3D 數位雙生虛擬世界加速實體世界自駕車、機器人的訓練模擬環境(例如 Nvidia Cosmos)。而愈來愈多代理人能自動化日常任務或提升業務運營效率，引發對 AI 治理、資料與模型安全的注重，特別需要防止偏見和保護隱私，同時從國家社會、經濟、安全的角度來看，更需要有自主的 AI 模型。

貳、專案目標

國科會(下稱本會)為打造臺灣成為 AI 智慧島、持續提升我國 AI 研究動能，以植基具國際能見度與對齊國家政策、社會需求為規劃基礎，延續 2018 年「臺灣 AI 行動計畫」(2018-2021 年)、及「臺灣 AI 行動計畫 2.0」(2023-2026 年)，特別規劃**「邁向新世代前瞻人工智慧研究專案」**(下稱本專案)：除引導研究計畫聚焦當前 AI 關鍵議題提升臺灣國際影響力，並考量百工百業在智慧轉型所需的 AI 基礎模型、演算法、與應用核心技術，朝高度自主、永續發展、安全強健、以及行動方向演進，協助 AI 科研及系統發展與培育 AI 人才，以鞏固臺灣 AI 核心技術競爭力。

本專案預期打造：

一、 保障國家的自主性、安全性及社會政治完整性的主權 AI。

二、 高效率低能耗的人工智慧技術與基礎模型，提昇 AI 的永續性。

三、 可信賴、強健、且具隱私保護的 AI 技術，確保系統的安全。

四、 具行動力、適應性、自然人機互動之 AI 行動代理，提升工作效率與產業價值。

一方面引導技術與人才發展朝向建構臺灣產業發展及落地應用所需之 AI 技術，同時深化 AI 基礎研究能量、專注關鍵技術的突破，透過自主創新、掌握產業需求，達到主權、永續、安全、並兼顧行動的科技智慧島願景，並促使我國擁有自主 AI 創新應用生態系，帶來實質的國際影響力。

參、專案說明

本次徵求計畫分為主題研究卓越計畫(下稱「主題計畫」)、以及突破性核心技術計畫(下稱「核心計畫」)。分別說明如下：

一、 主題研究卓越計畫

主題計畫的目標為發展主權、永續、安全、行動為主的 AI 技術，以推動臺灣前瞻 AI 技術、以及產業應用所需之科技、人才培育。主題計畫以整合型計畫為主要徵案方式。

(一) **主權 AI (Sovereign AI)**：發展自主可控的人工智慧技術以保障國家的自主性、安全性及社會政治完整性。

背景說明：AI 技術在關鍵基礎設施、國防安全、經濟決策及數位治理等領域扮演核心角色，倘若 AI 核心技術受制於國外，可能導致技術壟斷、資料安全隱憂，甚至影響國家戰略決策。因此，發展自主可控的 AI 生態系，包括自研 AI 模型、可信賴的資料基礎設施、及 AI 計算資源，已成為各國 AI 戰略的核心目標。此外，主權 AI 亦涉及技術標準的制定與國際規範的參與，確保臺灣在全球 AI 競局中擁有發言權，推動符合民主價值觀與社會利益的 AI 應用發展。透過強化 AI 技術自主性、資料治理、關鍵演算法掌控，主權 AI 不僅能提升國家科技競爭力，亦能推動符合倫理法制之負責任 AI，建立臺灣 AI 技術發展應用生態系。關鍵研究議題包括下列方向：

- 資料主權：建構人工智慧所需的關鍵資料與相關技術，包括資料的蒐集、儲存、管理及應用，以維護資料安全性、隱私權及國家利益，減少對外部資料來源的依賴。
- 演算法主權：發展自主可控且效能優異的核心演算法，掌握關鍵人工智慧技術，避免受制於外部團體，並提升技術自主性與競爭力。
- 模型主權：建立本土化的人工智慧模型，從基礎模型的訓練到應用的開

發皆由國內技術掌握，確保符合本國需求、倫理規範及產業發展方向，降低對外部技術的依賴，確保國家競爭優勢。

(二) **永續 AI (Sustainable AI)**：發展高效率低能耗的人工智慧技術與基礎模型，透過演算法與硬體設計以提昇 AI 的永續性。

背景說明：隨著人工智慧技術的快速發展與應用擴展，AI 模型的訓練與推理過程對計算資源與能源的需求也急遽增加。當今的大型 AI 模型(如深度學習與生成式 AI)往往依賴於高度集中的資料中心與計算基礎設施，其高能耗特性引發了對環境可持續性的關切。因此，如何發展高效率、低能耗的 AI 技術，並透過創新的演算法與硬體設計來提升 AI 的永續性，已成為全球關注的重要課題。關鍵研究議題包括下列方向：

- 節能高效：開發高效率低能耗的人工智慧技術與模型訓練和推論，減少碳足跡與降低能源消耗。
- 環保永續：設計符合環保和永續發展(例如：減少污染、支持綠色能源轉型)的人工智慧系統。

(三) **安全 AI (Safe AI)**：提升關鍵人工智慧系統的安全性，從而降低在人工智慧系統的間諜活動、破壞行為、運行風險、以及品質保證。

背景說明：人工智慧技術的普及與深化應用已改變現代社會的運作模式，從關鍵基礎設施、金融系統、醫療健康、智慧交通、國防安全到大型語言模型(LLM)與自動化決策系統，AI 正在驅動全球科技進步。然而，隨著 AI 技術滲透至關鍵領域，其安全性問題也日益受到關注，AI 系統可能成為間諜活動、惡意攻擊、系統破壞或運行風險的目標，進而影響國家安全、社會穩定與企業運營。本專案徵集提升 AI 安全性的前瞻研究，期望推動可信任、安全可靠、風險可控的 AI 技術，確保 AI 能夠在保障安全與隱私的前提下推動社會進步。關鍵研究議題包括下列方向：

- 強健的 AI 模型：具有抵禦能力以應對對抗性攻擊、模型破壞、或未經授權的存取。
- 可解釋性與可解讀性：具備透明與可理解決策邏輯的人工智慧系統。
- 隱私保護：運用差分隱私、同態加密、安全多方計算等技術，確保資料在人工智慧運行過程中的隱私性與安全性。
- AI 評量系統：Accountability、Fairness、Accuracy、Reliability、Transparency。

(四) **行動 AI (Agentic AI)**：發展具行動力、適應性、自然人機互動的 Agentic AI，智慧型代理系統。

背景說明：AI Agents 早在 1990 年代便已出現，但受限於當時的計算能力

與演算法的局限，真正能夠自主行動的 AI 系統始終難以實現。隨著 GPT 系列等模型在語言理解與生成上的突破，AI 系統已有多模態感知、規劃推理以及調用工具的能力，行動代理(AI Agents)在實體物理世界或數位虛擬環境中執行任務變成可能。行動代理的核心在於提升資訊系統的使用便利性，加速跨系統的流程自動化，減少人類在重複性任務上的時間成本，從而為企業與個人創造更高的生產力與創新性。然而如何提升推理與規劃能力，構建長期記憶與適應性，在決策自主性與可控性之間取得平衡，提高複雜流程的執行效率與準確性，以及跨域調用工具與多代理協作，提升對未知環境的泛化能力，都是 Agentic AI 的挑戰。同時，為了支援 Agentic AI 以及 Physical AI 的大腦開發，需有行動代理開發平台與工具鏈，尤其是實體世界需要空間智慧與世界模型，規劃任務執行策略、及具體行動方案。因此，未來行動 AI 的關鍵研究議題包括：

- 行動代理應用服務開發：開發具多模態感知、規劃、反思、工具調用，並能支援代理間協作，執行任務的行動代理系統(如 Robot、Softbot、No code RPA、GUI Automation)。
- 行動代理開發平台與工具鏈：建立行動代理的開發框架與系統(如 Cosmos、NeMo 等)，提供 AI 程式協作環境(如 Cursor 等)，及代理系統評估與測試工具。
- 空間智慧與世界模型建構技術：研發視覺語言模型驅動的推理模型，以支援行動規劃；建立支援空間推理與實體世界互動(Physical AI)的基礎模型，發展模擬環境中代理行為與物理交互的理論基礎。

二、 突破性核心技術計畫

為打造臺灣成為 AI 智慧島與全球 AI 前瞻核心技術的研發基地，本專案亦徵集四大主題以外的突破性核心技術研究，以追求顯著突破現有框架，具全球獨步或頂尖競爭力的創新技術。此類研究應能引領全球 AI 創新、挑戰現有技術框架(例如顛覆 Scaling Law)，產生典範轉移與實質影響力之研究。突破性核心技術計畫須明顯超越主題計畫範疇，提升我國在 AI 前瞻技術的國際地位與長遠影響力。核心技術計畫以個別型計畫為主要徵案方式。

背景說明：當前全球人工智慧研究正處於極快速發展的階段，然許多進展源自於對既有技術框架的深化與優化。隨著現有技術體系在提升模型效能方面逼近資料、計算資源的極限，如何突破這些框架、探索更具顛覆性的 AI 技術，成為全球頂尖 AI 研究機構關注的核心課題。例如不依賴擴大模型規模來提升效能，而是探索更高效、更具理論深度的演算法、架構、訓練方法。如近年來的稀疏建模、神經符號混合學習，以及以小數據驅動的大模型訓練策略，

皆顯示出改變 AI 發展方向的可能性。藉由突破性前瞻核心技術，推動更多具有國際影響力的研究成果、開源模型、與落地應用，進一步鞏固臺灣的 AI 戰略地位。

肆、計畫要求

一、 整體要求

(一)計畫所涉之主題研究與核心研究須符合本專案規劃之特定主題方向且具備引領全球 AI 之前沿技術，計畫書應說明所選擇挑戰關鍵研究議題的重要性與必要性。

(二)計畫須規劃全程「目標與關鍵成果」(OKR)，並訂定可協助整體目標達成之方法、可檢視之里程碑與各季度考核重點。

(三)計畫須自訂各季度考核之量化績效指標(如頂級 AI 會議及期刊論文、專利產出、獲得技轉或商轉金額、被學研社群引用次數及可能產生的影響)。

(四)計畫書內容之完整性、可行性與應用性(政府或產業相關組織之合作、產出工具或方法論之移轉與落地實踐，計畫之整合性及每季(年)預計達成目標等)；在研究可行性上需提出具體分年研究藍圖(roadmap)規劃。

二、 重點要求

(一)AI 技術：具備創新性、突破性或未來發展潛力的人工智慧技術，能夠推動 AI 領域的前沿研究，並塑造未來社會發展的關鍵動力。

(二)AI 治理：須配合執行本專案 AI 治理試行機制，進行可信賴 AI 評估(格式詳見附件 A 可信賴 AI 評估表，本會得視需要調整)，包括計畫所產出之資料集，須遵循前述(可信賴 AI)評估中「隱私與資料治理原則」自行管理，供後續相關研究參考。計畫須盤點訓練、驗證 AI 模型/系統所需之資料(格式詳見附件 B-1 訓練模型用之資料盤點表)。

(三)AI 模型及資料集共享：在合法授權再利用的基礎下，計畫團隊須就計畫產出之 AI 模型規劃共享模式(詳見附件 B-2 預計產出之 AI 模型管理與共享表)，若選擇公開授權之 AI 模型，須於本專案指定網站部署及共享(參見肆、「三、落地實踐要求」及「四、驗證與部署要求」)。建模後須提供產出資料集清單(詳見附件 B-3 計畫預計產出之資料集清單)。另，計畫所產出之資料集經本會認定具重要性、公共性或為發展主權 AI 所需者，本會得要求計畫團隊取得資料再利用授權並滿足其他再利用合法要件後，提供本會或本會授權之對象於必要範圍內使用。計畫團隊就計畫所蒐集或產出之資料集，得自行於確認具備再利用之適法性、經本會評估對關鍵核心技術開發與國家安全無不利影響後，對外提供共享。

(四)國際影響力與產學合作：為促進科研、人才、治理的國際合作與接軌，並提升本專案團隊之國際能見度與影響力，本專案優先鼓勵在技術研發與突破、人才培育、產學合作等面向具有國際實質影響力與合作規劃之團隊。

三、 落地實踐要求

呼應國家發展政策，研究計畫除學術影響力外，需說明研究成果對臺灣社會、經濟、教育、環境等產生實際影響力及效益的做法及路徑，規劃具體的應用場域，合作對象例如：智慧雨林產業創生補助計畫、國研院智慧機器人研究中心、以及國家災害防救科技中心等，使其成果能夠轉化為實際應用。計畫團隊可選擇下列方式來落實研究成果：

(一)開源與技術共享

為促進技術交流與擴大影響力，研究成果可選擇開源至 GitHub 或 Hugging Face 等開源平台，並提供完整的程式碼、訓練資料說明及使用指南。

(二)創業或企業應用

若研究成果具備市場潛力，可考慮透過技術轉移或專利授權將技術導入既有企業，或選擇成立新創公司，進一步推動技術商品化。

四、 驗證與部署要求

根據所選應用方式，研究計畫需滿足相應的驗證要求：

(一)開源路徑

需確保 AI 模型可於本專案指定網站完整運行，以便公眾可對模型的有效性、可用性進行檢視，並利於技術成果快速擴散。

(二)企業應用路徑

可保留核心技術細節以保護智慧財產權，然需提供功能展示的簡化版本或案例說明，足以驗證技術可行性，並提供適合商業應用的部署規劃。

無論選擇何種路徑，計畫團隊均需提供充分證據證明方案的可行性與實用價值，確保研究成果能有效轉化為實際應用。

伍、計畫申請

一、 團隊組成及計畫經費

本專案接受整合型計畫及個別型計畫，整合型計畫鼓勵以跨領域、機關或單位合作模式組成研究團隊提出申請案，並優先考慮多年期計畫(114-117 年)。

(一)單一整合型研究計畫：

1. 總計畫主持人須將總計畫及子計畫彙整成一冊；子計畫數至少 3 個，且最多以不超過 5 個為原則；總計畫主持人須主持其中一項子計畫；各子

計畫主持人應實質參與研究。

2. 申請經費：以每件子計畫每年上限 500 萬元估算計畫總經費為原則，並將依審查結果決定補助金額。

(二)個別型研究計畫：

1. 將嚴格審查後擇優補助至多 10 件計畫，且得從缺。
2. 申請經費：以每年 500 萬元為上限，並將依審查結果決定補助金額。

(三)整合型計畫申請者應強化以下說明：

1. 整合型計畫須提出計畫團隊之領先能力與差異化評估：包括與全球及臺灣之現況比較，計畫團隊之能量及優勢等。
2. 計畫主持人之執行力。
3. 團隊成員分工與合作架構、關聯性、潛在優勢及跨領域、跨單位資源整合能力。

(四)核心計畫申請者，若計畫主持人在 AI 理論創新與技術發展上能顯著提升我國國際學術地位，具有下列三款重要優異實績之一，本專案將優先補助：

1. 曾獲國內外相關領域重要學術獎項，例如吳大猷獎、傑出研究獎等。
2. 於相關領域有極為傑出之研究表現，例如頂尖國際會議最佳論文獎、h-index 排名居前等。
3. 曾擔任相關領域之頂尖國際會議主席或主題演講者(Keynote Speaker)、或重要期刊主編(Editor-in-Chief)。

二、 申請資格

(一)須符合本會補助專題研究計畫作業要點之申請機構與計畫主持人及共同主持人之資格。

(二)整合型計畫之總計畫主持人(含共同主持人)及個別型計畫主持人限申請本專案計畫 1 件；整合型計畫之總計畫主持人不得擔任本專案其他計畫申請案之子計畫主持人。

三、 申請方式

(一)採線上申請，並選擇「專題類-隨到隨審計畫」；計畫類別為「一般策略專案計畫」；研究型別選擇「整合型」或「個別型」；計畫歸屬「前瞻處」；學門代碼為「P31」：申請「主題計畫」者請選擇「P31301221」；申請「核心計畫」則選擇「P31301222」。

(二)申請機構應依本會補助專題研究計畫作業要點之規定與格式，於 114 年 7 月 18 日(星期五)前線上提出計畫申請案，並備函送達本會，逾期不予受理(請彙整造冊後專案函送，並依本會收文日期為準)。

四、 計畫執行期間

全程至多 4 年。預計自 114 年 11 月 1 日至 118 年 10 月 31 日，計畫主持人須規劃多年期計畫，經審查通過者，核定補助二年；計畫執行至第二年期中時進行成果考核，並將依審查結果重新提送後續年度計畫書；本會可視情況調整作業時程。

五、 計畫書內容

(一)申請人應於申請書中研究計畫內容之首段明確勾選對應之徵求主題（請參考本專案說明），並依所選主題撰寫計畫書（CM03 格式範例如附件）

(二)整合型計畫書內容(CM03 表)以 50 頁為限，個別型計畫以 25 頁為限(均含圖、表，但不含參考文獻及附件)，包含研究目的、預計合作之公、私部門或機構以及與其工作內容規劃、研究方法、工作項目、預期成果、時程規劃、經費與人力分析等項目。

(三)整合型計畫須清楚說明計畫整體組織/執行架構，包括總計畫主持人、子計畫主持人，其他例如執行長、國內外研究人員等之業務分工與整合作法。

(四)核心計畫應具體說明理論基礎、技術藍圖、與預期影響。

(五)計畫團隊若屬跨機構，申請機構應就跨機構合作可能涉及之權利義務、經費使用、成果與智慧財產權(IP)歸屬、移轉或授權等相關事項，基於公平合理之原則，先行協調議定，並載明於計畫書中。

(六)配合推動多元、公平及包容(DEI)精神，推動負責任、可信賴 AI，計畫書須填寫「可信賴 AI 評估表」(附件 A)建模前部分，並說明採取何種作法以減少偏見與歧視造成的不對等，確保社會對 AI 發展的信任。

(七)計畫須盤點建模前預計使用之訓練資料集，請計畫團隊填寫「訓練模型用之資料盤點表」(附件 B-1)。

(八)在合法授權再利用的基礎下，計畫團隊須就計畫產出之 AI 模型規劃共享模式，依據性質團隊可選擇公開共享或有條件共享，請填寫「預計產出之 AI 模型管理與共享表」(附件 B-2)。

(九)為提供適量計算資源予本專案各研究計畫使用，本專案與國家高速網路與計算中心(下稱「國網中心」)，擬進行專屬計算資源合作案，請就研究之計算需求填列「計算資源需求申請表」(附件 C)，以利評估本專案之整體計算需求。

(十)國網中心另提供生成式 AI 應用服務開發平台(RAP)、可信賴雲平台(TRE)，鼓勵計畫團隊向國網中心(https://rap.genai.nchc.org.tw/poc)申請使用。

(十一)　本專案鼓勵計畫團隊與產業界合作，以縮短技術從學術研究到實際應用的落差。計畫團隊可與企業、政府機構或非營利組織建立合作關係，共同開發落地的 AI 解決方案，並進行實地測試與驗證。本專案亦鼓勵計畫團隊與國際學術機構或開源社群合作，以提升 AI 技術的實用性與全球影

響力。

陸、計畫審查與核定

一、 審查作業以書面審查與會議審查為原則；如有需要，將安排申請人簡報與答詢。

二、 確認計畫內容是否涵蓋並符合徵求主題與該主題重點內容，計畫書格式不符本專案徵案公告要求者，將不予送審。

三、 依申請案審查後推薦順序，擇優通知計畫主持人於線上提出「修正計畫書」；計畫主持人應依本會通知時間完成修正，逾期不予受理。

四、 審查項目包括完整性、技術創新性、可行性、團隊研究量能、AI 治理/安全。

五、 本會得核給整合型計畫之總計畫主持人研究主持費最高每月 50,000 元。整合型計畫之子計畫主持人及個別型計畫之主持人，本會得視計畫審查之結果，核給研究主持費最高每月 30,000 元。總計畫及子計畫主持人同時間僅得支領本專案 1 份研究主持費，若同時執行其他本會計畫，以最高額度計算，並得於計畫內採差額方式核給。

六、 經核定補助後，整合型計畫之總計畫主持人與個別型計畫之主持人列入本會專題研究計畫件數計算額度，子計畫主持人則不列入計算。

柒、其他注意事項

一、 本專案為政策導向型計畫，為強化計畫效益與成果，本會將對執行計畫定期進行檢視，計畫主持人及其團隊必須配合提供計畫執行進度與成果報告，並出席定期工作會議或各項審查會議；且本會得視業務需要，不定期請獲補助之計畫成員提供相關研究成果或資料，並有義務參加本會之學術應用推動活動以及配合本會相關國際合作及科普推廣活動。

二、 獲補助之計畫成員須配合參與國內外訪客相關交流與廣宣活動，如觀摩交流活動、成果發表、實務驗證案例等，並依本專案需求與審核結果，適時整合併入適當之整合型計畫。成果發表時計畫主持人須通知本會，以利成果詳實紀錄備查。本專案各研究計畫所產出之成果均依「政府科學技術研究發展成果歸屬及運用辦法」之規定實施。

三、 由本會籌組專家委員會，進行每年定期考核，並依據考核結果作為調整次年度經費之參據。若計畫年度成果暨可信賴 AI 自評(附件 A 建模中、建模後部分)經審議執行進度未達標準、預期成果無法達成或不受管考者，經考核會議討論後，可依照本會補助專題研究計畫作業要點第 23 點辦理計畫退場。

四、 執行本專案所需之博士級研究人員相關費用，請納入計畫經費中，不得另案依本會補助延攬客座科技人才作業要點第五點第一項第二款規定向本會申請博士級研究人員經費補助。

五、 本專案相關之簽約、撥款、延期與變更、經費報銷及報告繳交等皆依本會補助專題研究計畫作業要點、專題研究計畫經費處理原則、專題研究計畫補助合約書與執行同意書及其他有關規定辦理。

六、各年度所需經費如未獲立法院審議通過或經部分刪減，本會得依審議情形調整補助經費。

七、本計畫屬專案計畫，未獲補助案件恕不受理申覆。

八、本公告未盡事宜均依本會補助專題研究計畫作業要點及其他相關規定辦理。

九、說明會報名資訊請詳見報名網址：https://forms.gle/AZjBwXEaK1hsqbPR8

捌、聯絡資訊

一、 徵案召集人

張嘉惠教授
E-mail：chiahui@g.ncu.edu.tw
Phone：03-422-7151 ext. 35366

丁川康教授
E-mail：ckting@cs.nthu.edu.tw
Phone：03-571-5131 ext. 33756

二、 國科會前瞻及應用科技處

高雨瑄助理研究員
E-mail：yhkao@nstc.gov.tw
Phone：02-2737-7538

林滋梅研究員
E-mail：tm1lin@nstc.gov.tw
Phone：02-2737-7076

三、 計畫申請系統操作問題請洽資訊系統服務專線：02-2737-7590~7592

附件：

三、研究計畫內容(以中文或英文撰寫)：

(一) 專案類型

□主題研究卓越計畫：□主權 AI　□永續 AI　□安全 AI　□行動 AI (可複選)

□突破性核心技術計畫

(二) 研究計畫之背景。

(三) 研究方法、進行步驟及執行進度。

(四) 預期完成之工作項目及成果。

(五) 如為整合型研究計畫請就以上各點分別說明與其他子計畫之相關性。

(六) 核心計畫應具體說明理論基礎、技術藍圖、與預期影響。

(七) 請就計畫內容填寫徵案公告附件 A「可信賴評估」建模前部分、附件 B-1「訓練模型用之資料盤點表」、附件 B-2「預計產出之 AI 模型管理與共享表」、附件 C「計算資源需求申請表」。

(八) 請說明計畫之研究成果對台灣社會、經濟、教育、環境等產生實際影響力及效益的做法及路徑。

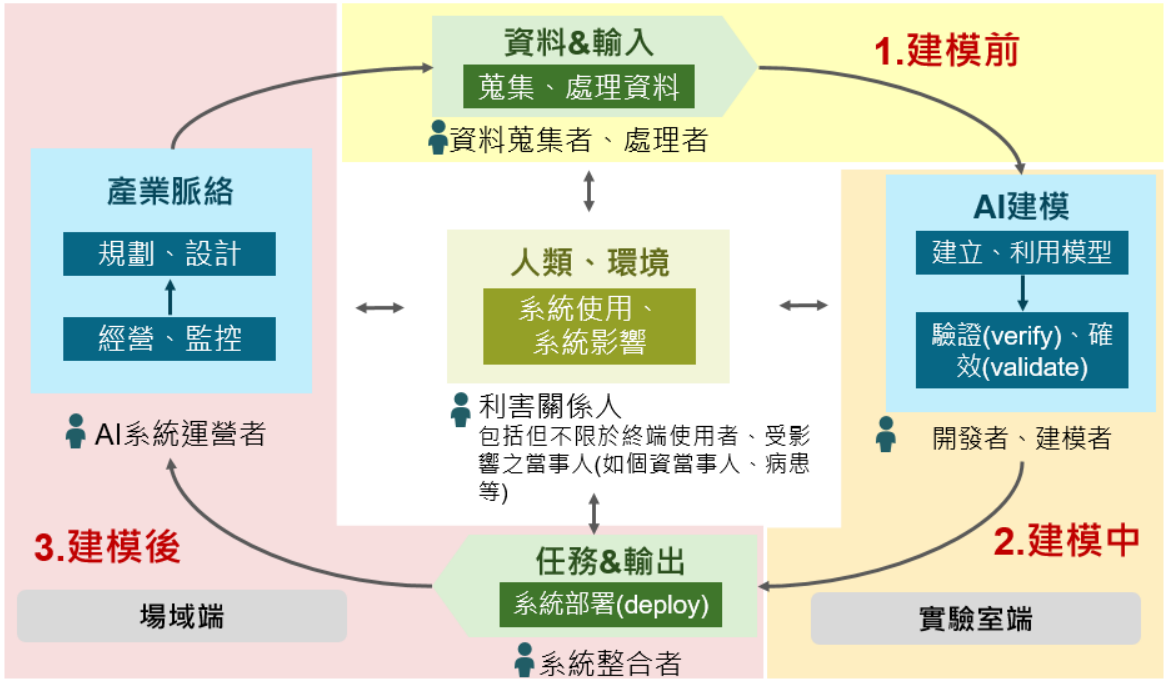
國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
可信賴 AI 評估表-建模前（徵案階段僅需填寫本表）

前言

為避免 AI 發展對民主、自由、人權等基本價值造成難以逆轉的負面衝擊，國科會臺灣 AI 卓越中心（Taiwan AICoE）集結相關機構，共同建立「可信賴 AI 評估表」，將民主、自由、人權等基本價值融入整體 AI 系統生命週期各階段（如：訓練及測試資料之蒐集、模型建立、系統部署、系統經營及應用），增進研究人員及參與者對 AI 研發及應用相關倫理議題之敏感度，並促進研究人員、潛在 AI 使用者及可能受影響之人互動對話，及早消弭 AI 偏見與歧視，從而強化社會大眾對 AI 發展之信任。

本評估表先以 OECD「AI 建議」及國科會「AI 科研發展指引」作為基礎框架，列出七項可信賴 AI 基本原則，包括：透明與可解釋性、隱私與資料治理、公平與不歧視、人類自主、資安與安全、永續發展與福祉、問責，再依 AI 系統生命週期區分建模前、中、後（詳圖 1），預擬前述各項原則在不同生命週期階段，分別對應的評估問項及成熟度判斷準則。

建模前階段處於實驗室端，主要在蒐集與處理訓練、測試AI之資料。建模中階段亦處於實驗室端，涉及模型之選定、建立、驗證與確效。建模後階段則包括將 AI 進行系統整合、部署至應用場域及進行 AI 系統運營等。在徵案階段僅需填寫「建模前」相關問題（即本表）。此外，鑒於問項填答內容與「AI 開發情況及其所欲部署或應用之目的」息息相關，在七原則問項之前，亦設計「**基本資料表**」，作為計畫團隊填寫可信賴 AI 評估問項的前置作業，以便更準確判斷哪些問項為必要評估項目。



資料來源：OECD Framework for the Classification of AI Systems(2022)

圖 1. AI 系統生命週期各階段之情境及主要參與者

本評估表目的在「**引導團隊預先思考 AI 研發與應用相關倫理議題**」，評估 AI 模型或系統未來可能影響人權與民主價值之風險，促進團隊提前討論可行的因應做法，故**性質上並非法遵評分表(非所有問項答案都填「是」，就能得到高分)**。團隊填答時，**應依實際 AI 研發狀況，填寫是、否或不適用並附上原因或理由**。

基本資料表			
計畫名稱		填表人姓名	
		填表人於計畫中擔任之角色	
計畫主持人		填表日期	

項目		說明
1	AI 名稱	（以下簡稱本 AI）
2	性質	□ 1.現為 AI 模型，預計僅作為其他 AI 系統之部分元件 □ 2.現為 AI 模型，預計開發為 AI 系統 □ 3.現已為 AI 系統
3	計畫團隊基於什麼目的開發本 AI？本 AI 能夠解決什麼社會問題？	

<table>
<tr><td>4</td><td>本 AI 具有何種功能？</td><td colspan="3">□ 1.單純模仿人類行為，請簡要說明：
□ 2.尋找人類難以察覺的事物關聯性，請簡要說明：
□ 3.預測人類難以估算的結果，請簡要說明：
□ 4.其他：＿＿＿＿＿＿＿＿，請簡要說明：</td></tr>
<tr><td>5</td><td>針對本 AI，計畫團隊是否「參與」或「預計會參與」AI 生命週期的「建模後」階段（例如：AI 系統部署、AI 系統場域應用、AI 系統產品經營等）？</td><td colspan="3">□ 1.完全不參與或完全無預計參與；
□ 2.參與或預計會參與，包括（可複選）：
□(1)AI 系統部署 □(2)AI 系統場域應用 □(3)AI 系統產品經營</td></tr>
<tr><td>6</td><td>本 AI 開發過程中，是否需用到與人相關之資料（無論資料是否先經去識別化處理）</td><td colspan="3">□是
□否</td></tr>
<tr><td rowspan="7">7</td><td rowspan="7">所有參與本 AI 評估之人員
(表格不敷使用請自行增列)</td><td>姓名</td><td>職稱</td><td>在可信賴 AI 評估過程中所負責之事項</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

1.透明與可解釋性(Transparency & Explainability)

為增進社會公眾對 AI 研發及應用之信賴，AI 的發展應符合透明性及可解釋性。透明性與可解釋性彼此間具有目的與手段之關係。透明性是指讓利害關係人可從外部檢驗 AI 系統的資料、特徵(features)、模型、演算法、訓練方法與品質保證等過程，以幫助利害關係人衡量 AI 系統的開發及運作是否合乎所期待的價值。而可解釋性是達成透明性的手段之一，是指能說明 AI 之所以能達成預期部署或應用目的之理由，例如但不限於解釋 AI 產出結果或決策的因果關係、AI 產出結果或決策背後的科學知識、AI 產出結果或決策有效的正當理由。此外，為確保達到透明性與可解釋性，往往需要可追溯性(Traceability)原則作為輔助，以保存 AI 系統設計、開發、部署到應用等生命過程中的相關紀錄（例如但不限於訓練及測試 AI 所使用之資料、演算法或規則；所採行之訓練方法；驗證過程與歷次產出結果或決策等），以便受 AI 產出結果或決策影響之人提出挑戰、申訴或救濟時，能有跡可循。

AI 生命週期中各階段應關注事項：

1、 建模前：瞭解開發 AI 所用的資料

2、 建模中：解釋 AI 模型所產出之結果

3、 建模後（含系統部署至場域端經營及監控）：解釋 AI 系統所作之結果或決策

生命週期階段	對應問項	成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若「不適用」，請說明原因）	問項參考來源

建模前	1-1-1（資料品質）	您是否確保開發 AI（模型或系統）所用**資料（含訓練、測試資料）的品質、完整性，並評估符合預期部署背景或應用目的之代表性**？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化評估流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2# 技術穩健性與安全(準確性)、ISO/IEC 42001:2023 控制措施 A.7.4、A.7.6

2.隱私與資料治理(Privacy & Data Governance)

AI 的研發與資料之間具有密切關連，無論是訓練、驗證或測試 AI，皆須使用資料。研究人員及 AI 系統生命週期中的相關參與者，應採取良好的隱私保護及資料治理，以負責任的方式蒐集、處理、利用及共享資料，提升 AI 研發成果的準確度、公平性及可靠性。AI 訓練、驗證或測試資料的蒐集、處理、利用手段應正當合法，並避免在研發與應用過程中侵害他人隱私、智慧財產權或影響國家安全。蒐集個人資料及資料再利用之前，應顧及個人資料當事人之自主權，並在符合資料蒐集目的特定、資料最少化等原則下，處理、利用資料。研究人員及相關參與者，應建立適當的資料管理與安全維護措施，以達成良好的隱私及資料治理。

AI 生命週期中各階段應關注事項：

1、 建模前：AI 開發所需資料之蒐集、處理、利用（尤其是與人相關之資料）

2、 建模中：AI 模型建立過程與產出結果對隱私、營業秘密、核心科技發展與國家安全的影響

3、 建模後（含系統部署至場域端經營及監控）：AI 系統運作或應用對隱私、營業秘密、核心科技發展及國家安全的影響

生命週期階段	對應問項		成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	問項參考來源
建模前	2-1-1	為開發本 AI，您蒐集哪些資料？是否蒐集個人資料（先不論去識別化與否）？（若否，2-1-3、2-1-4 免答）	N/A	說明：	ISO/IEC 42001:2023 控制措施 A.7.2
	2-1-2	團隊蒐集前述資料的方式與管道為何？依據哪些法	N/A	說明：	EU Assessment List for Trustworthy

		令、規範或契約進行蒐集？			Artificial Intelligence (ALTAI) 7# 可歸責性(風險管理)、ISO/IEC 42001:2023 控制措施 A.7.3
	2-1-3（告知同意）[1]	除法律規定得免告知同意的情況外，您（或提供資料給您之人）是否向資	• 低度：完全未踐行 • 中度：部分或單次踐行	□是（成熟度為□中度/□高度），踐行方式說明如下： □否（成熟度為低度），理由說明如下：	EU Assessment List for Trustworthy Artificial

[1] 團隊若從開放資料(open data)平台蒐集資料，無須向資料當事人踐行告知同意，但需注意該開放資料平台資料授權問題，保留授權佐證資料。

		料當事人踐行告知同意？若有，如何踐行？若無，理由為何？	• 高度：已建立一致化踐行程序	□不適用，原因如下：	Intelligence (ALTAI) 3# 隱私和資料治理(資料治理)
	2-1-4（當事人行使自主意見）[2]	您是否提供資料當事人行使自主意願的管道（包括但不限於請求更正、停止蒐集處理利用、請求刪除資料、撤回同意、選擇退出(opt-out)）？若有，如何處理？若無，理由為何？	• 低度：未提供 • 中度：提供單次權利行使機會 • 高度：已建立程序化的當事人權利行使系統或管道	□是（成熟度為□中度/□高度），提供方式說明如下： □否（成熟度為低度），理由說明如下： □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 3# 隱私和資料治理(資料治理)

[2] 團隊若從開放資料(open data)平台蒐集資料，無須提供資料當事人權利行使管道，但需注意該開放資料平台資料授權問題，保留資料授權佐證資料。

	2-1-5（蒐集處理利用手段合法）	您是否確認資料蒐集、處理或利用手段合法，且未侵害個資隱私、智慧財產權、營業秘密或對國家安全及核心科技發展造成負面衝擊？	• 低度：未確認 • 中度：單次確認 • 高度：建立程序化確認機制	□是（成熟度為□中度/□高度），確認方式說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.7.3
	2-1-6（安全維護措施）	您是否及如何避免所蒐集或保存的資料被竊取、洩漏、竄改、毀損或造成其他侵害？採取什麼安全維護措施[3]？	• 低度：未防免 • 中度：已防免 • 高度：已防免並建立程序化安全維護機制	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2# 技術穩健性與安全(一般安全）、

[3] 安全維護措施包括但不限於配置管理人員；近用權限控管；資料及資產盤點；風險評估；個資去識別化處理；事故預防、通報及應變；資料及設備安全防護；資料及設備使用紀錄、軌跡資料及證據保存等。

					ISO/IEC 42001:2023 控制措施 A.7.5、A.7.6

3.人類自主(Human Autonomy)

AI 的應用是為輔助人類決策，不應導致脅迫、欺騙、操縱人類甚至取代人類。因此，AI 的開發及應用應循以人為本的原則，作為強化或補充人類認知、社會或文化技能，確保人類與 AI 系統交流的過程中，仍保有作出有意義的選擇權，並能保持充分而有效的自主性與控制權。AI 參與者應採取符合具體情況並與現有技術相符之機制和保障措施，透過建立相關監督機制以確保系統不會侵害人類自主性或是引發其他負面效果。

AI 生命週期中各階段應關注事項：

1、 建模前：瞭解 AI 開發目的

2、 建模中~建模後（含系統部署至場域端經營及監控）：

（1）瞭解 AI 操控人類之風險

（2）人類挑戰 AI 系統產出結果之可能性

生命週期階段	對應問項	成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	問項參考來源

建模前	3-1-1（研發目的）	您研發此 AI 之目的，是否在試圖了解或掌握個人的心理狀態或行為？ 例如：測謊、個人性格或行為側寫或預測、心理諮詢、聊天陪伴	N/A	□是，說明： □否	ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5

4.公平與不歧視(Fairness & Non-discrimination)

AI 的研發與應用，應避免延續或加劇對個人或群體之刻板印象、偏見或歧視。AI 參與者從開發、部署到應用 AI 的過程中，應關注 AI 產出結果是否在種族、膚色、民族、性別、性別認同、宗教、年齡、國籍、身心障礙、遺傳或其他分類上產生偏差，從而對個人或群體造成不公平待遇或歧視。

AI 生命週期中各階段應關注事項：

1、 建模前：

（1）瞭解開發 AI 所用資料

（2）確保 AI 相關參與者能認知到偏誤及歧視

2、 建模中：避免 AI 模型產出結果發生歧視或不公平

3、 建模後（含系統部署至場域端經營及監控）：避免及因應 AI 系統運作或應用結果所產生之不公平

生命週期階段	對應問項		成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若「不適用」，請說明原因）	問項參考來源
建模前	1-1-1（資料品質）已涵蓋	您是否確保開發 AI（模型或系統）所用**資料（含訓練、測試資料）的品質、完整性，並評估符合預期部署背景或應用目的之代表性**？	• 低度：無評估 • 中度：已評估 • 高度：已評估並建立程序化評估流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2# 技術穩健性與安全(準確性)、ISO/IEC 42001:2023 控制措施 A.7.4、A.7.6
	4-1-1（識別利害關係人）	您是否**識別** AI 模型（或系統）預期部署或應用目的之	• 低度：未識別 • 中度：已識別	□是（成熟度為□中度/□高度），識別方式說明如下：	EU Assessment List for Trustworthy Artificial

		「**潛在使用者**」[4]及「**可能受影響之人或群體**」[5]？分別有哪些？	• 高度：已識別並建立程序化識別機制(SOP)	「潛在使用者」及「可能受影響之人或群體」如下： □否（成熟度為低度） □不適用，原因如下：	Intelligence (ALTAI) 6# 社會與環境福祉(對社會全體或民主之影響)、ISO/IEC 42001:2023 控制措施 A.5.4
	4-1-2（教育訓練）	您的團隊是否受過相關**教育訓練**、參加相關會議或蒐集資訊，以了解 AI 可能導致的偏誤[6]及歧視[7]？	• 低度：未參加亦不曾蒐集相關資訊 • 中度：偶爾參加或蒐集相關資訊	□是（成熟度為□中度/□高度），說明如下： □否（成熟度為低度）	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5# 多元

[4] 不限於終端使用者。

[5] 未來模型或系統部署至應用場域情境中，可能因 AI 受影響之人或群體。例如：輔助醫師診斷之 AI 系統，可能受影響之人或群體包括病患。需由團隊針對 AI 系統未來可能部署應用之情境，預先盤點。

[6] 如系統性偏誤、統計或運算偏誤、人類認知所造成的偏誤。

[7] 如 AI 系統依種族、膚色、民族、性別、性別認同、宗教、年齡、國籍、身心障礙、遺傳或其他分類所導致對人的不公正待遇或影響。

			• 高度：有針對模型特殊性進行專業訓練規劃，並針對訓練結果進行評量	□不適用，原因如下：	性、不歧視與公平(避免不公平的偏誤)、ISO/IEC 42001:2023 控制措施 A.4.6

5.資安與安全(Security & Safety)

AI 常見的安全威脅如：資料下毒(data poisoning)[8]、模型迴避(model evasion)[9]、模型逆向(model inversion)[10]等。AI 開發者及部署者，應確保 AI 在可預見的使用情境下，能正常、安全地運作。為此，需藉由辨識、防護、偵測、應變與矯正風險等方法，確保樣本蒐集、模型訓練、系統部署、系統運作環境及過程之安全，以維持 AI 產出結果之穩定性與可再現性，避免 AI 系統在實際場域應用時，因錯誤而對人類生命、身體、健康、財產造成損害。

AI 生命週期中各階段應關注事項

[8] 攻擊者竄改特定模型所用之樣本，有意影響訓練資料以操控模型預測結果，尤其發生在模型需透過網際網路持續從新資料中學習的情況。

[9] 攻擊者在模型輸入值回傳的過程中，對輸入值加入細微雜訊，以大幅改變模型的預測結果。

[10] 攻擊者利用機器學習系統提供的一些 API 來獲取模型的初步資訊，並藉這些初步資訊對模型進行逆向分析，以獲取模型內部的隱私資料。

1、 建模前：
（1）AI 開發資料蒐集管道之安全性
（2）AI 開發資料之異常偵測、應變與矯正
2、 建模中：AI 模型之安全性
3、 建模後（含系統部署至場域端經營及監控）：AI 系統之安全性

生命週期階段	對應問項		成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	問項參考來源
建模前	5-1-1（辨別來源）	您開發 AI 所需的資料中，是否含有生成資料(非從真實世界蒐集，利用生成式學習演算法生成之資料)？有哪些？	N/A	□是，使用之生成資料如下： □否	ISO/IEC 42001:2023 控制措施 A.4.3、A.7.2

	5-1-2（資料蒐集安全防護）	模型建立與訓練之前，您是否採取措施，對資料蒐集管道或來源進行安全防護[11]？採取哪些防護措施？	• 低度：未防護 • 中度：已防免 • 高度：已防免並建立程序化防護機制(SOP)	□是（成熟度為□中度/□高度），資安措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2# 技術穩健性與安全(抵禦攻擊之韌性和安全性)、ISO/IEC 42001:2023 控制措施 A.7.5
	5-1-3（資料異常偵測）	模型建立與訓練之前，您是否採取資	• 低度：未偵測 • 中度：已偵測	□是（成熟度為□中度/□高度），偵測方式說明如下： □否（成熟度為低度）	ISO/IEC 42001:2023 控制措施 A.7.2~A.7.5

[11] 例如從資料機密性、完整性、可用性等方面，採行資安防護措施，包括但不限於資料傳輸過程、儲存方式、利用權限等之防護。

		料異常[12]偵測措施？如何偵測？	• 高度：已偵測並建立程序化偵測機制(SOP)	□不適用，原因如下：	
	5-1-4（資料異常應變與矯正）	承上題，針對異常資料，您是否建立應變與矯正對策？	• 低度：未建立 • 中度：已建立 • 高度：已建立並能監督執行	□是（成熟度為□中度/□高度），對策設計說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.7.2~A.7.5

6.永續發展與福祉(Sustainable Development & Well-being)

AI 發展應維護人性尊嚴、人權與民主價值，並追求對人類及地球有益之結果，避免加劇人類及生存環境之危害。AI 開發者、部署者及使用者，應關注 AI 對弱勢群體的包容性，避免濫用 AI、使人類失去創造力與技能或對環境永續帶來負面影響。

AI 生命週期中各階段應關注事項：

1、 建模前：AI 開發所需資料的多元與代表性

[12] 例如：辨識或檢查有無資安與安全原則說明內容中的資料下毒(data poisoning)風險、資料遭竄改或毀損之情況。

2、 建模中：AI 模型產出結果對整體社會與環境的影響

3、 建模後（含系統部署至場域端經營及監控）：

（1）AI 系統包容性

（2）AI 系統運作及應用之整體社會及環境影響

生命週期階段	對應問項		成熟度判斷準則	評估結果與說明 （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	問項參考來源
建模前	1-1-1（資料品質）已涵蓋	您是否確保開發 AI（模型或系統）所用**資料（含訓練、測試資料）的品質、完整性，並評估符合預期部署背景或應用目的之代表性**？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化評估流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	

7.問責(Accountability)

問責原則橫向貫串前述六項原則，以確保AI從開發、部署至應用生命過程可追溯與可檢驗。AI 生命過程中的參與者，應針對前述六項原則之落實進行權責分工，負責任地管理開發 AI 所用之資料、模型與演算法，保存 AI 系統

設計、開發、部署到應用等生命過程中的相關紀錄，瞭解並確保 AI 開發、部署及應用遵循相關法律、規範或契約要求。而 AI 實際應用時所產出之結果或決策，將影響到利害關係人之權益，應給予受 AI 產出結果或決策影響之人提出挑戰、申訴或救濟之機會。

AI 生命週期中各階段應關注事項：

1、 建模前：可信賴 AI 生命週期管理與法遵責任

2、 建模中：AI 建模的管理責任

3、 建模後（含系統部署至場域端經營及監控）：AI 系統管理責任

生命週期階段	**對應問項**		**成熟度判斷準則**	**評估結果與說明** （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	**問項參考來源**
建模前	7-1-1（團隊成員角色與職責分派）	您是否在可信賴 AI 生命週期執行環節中[13]，針對前述各原則，定義團隊成員角色及相關職責，並指定合適	• 低度：無分派職責 • 中度：已分派職責 • 高度：已分派職責並建	□是（成熟度為□中度/□高度），分工說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 7# 可歸

[13] 例如：資料蒐集、處理、利用之管理；資訊安全維護；資料代表性評估；適用規範之遵循；AI 模型或系統產出結果之驗證、確效、監控等。

		的負責人？	立程序化機制以持續監督		責性(風險管理)、ISO/IEC 42001:2023 控制措施 A.3.2、A.10.2
	7-1-2（法遵盤點）	您是否盤點 AI 模型（或系統）所用的資料、演算法及研發領域所應關注或遵循的法律、規範或標準？有哪些？	• 低度：未盤點 • 中度：已盤點 • 高度：已盤點並建立程序化機制，能適時更新盤點結果	□是（成熟度為□中度/□高度），盤點方式說明如下： 主要資料、演算法、相關法律、規範、標準如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.4.3

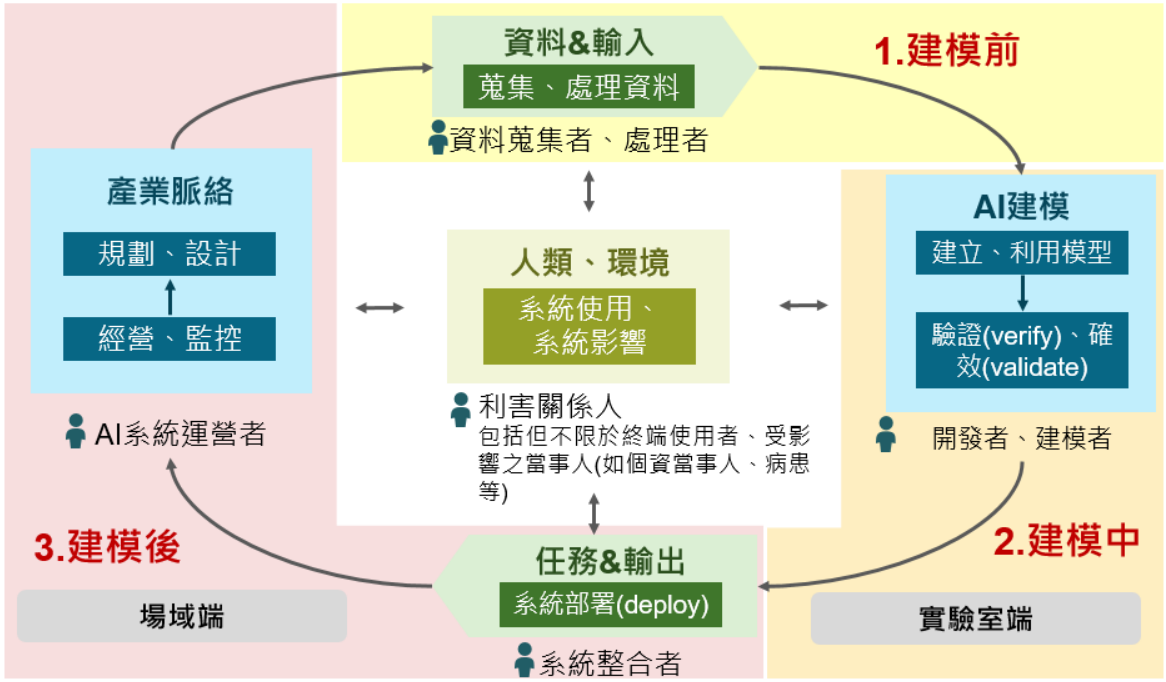
國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
可信賴 AI 評估表-建模中（徵案階段僅供計畫團隊參考，<u>不需</u>填寫）

前言

為避免 AI 發展對民主、自由、人權等基本價值造成難以逆轉的負面衝擊，國科會臺灣 AI 卓越中心（Taiwan AICoE）集結 AI 專案各計畫團隊之力，共同建立「可信賴 AI 評估表」，將民主、自由、人權等基本價值融入整體 AI 系統生命週期各階段（如：訓練及測試資料之蒐集、模型建立、系統部署、系統經營及應用），增進研究人員及參與者對 AI 研發及應用相關倫理議題之敏感度，並促進研究人員、潛在 AI 使用者及可能受影響之人互動對話，及早消弭 AI 偏見與歧視，從而強化社會大眾對 AI 發展之信任。

本評估表先以 OECD「AI 建議」及國科會「AI 科研發展指引」作為基礎框架，列出七項可信賴 AI 基本原則，包括：透明與可解釋性、隱私與資料治理、公平與不歧視、人類自主、資安與安全、永續發展與福祉、問責，再依 AI 系統生命週期區分建模前、中、後（詳圖 1），預擬前述各項原則在不同生命週期階段，分別對應的評估問項及成熟度判斷準則。

建模前階段處於實驗室端，主要在蒐集與處理訓練、測試AI之資料。建模中階段亦處於實驗室端，涉及模型之選定、建立、驗證與確效。建模後階段則包括將 AI 進行系統整合、部署至應用場域及進行 AI 系統運營等。此外，鑒於問項填答內容與「AI 開發情況及其所欲部署或應用之目的」息息相關，在七原則問項之前，亦設計「**基本資料表**」，作為計畫團隊填寫可信賴 AI 評估問項的前置作業，以便更準確判斷哪些問項為必要評估項目，並了解計畫團隊自評結果是否合乎 AI 所欲部署或應用之目的。



資料來源：OECD Framework for the Classification of AI Systems(2022)

圖 1. AI 系統生命週期各階段之情境及主要參與者

本評估表目的在「**引導團隊預先思考 AI 研發與應用相關倫理議題**」，評估 AI 模型或系統未來可能影響人權與民主價值之風險，促進團隊提前討論可行的因應做法，故**性質上並非法遵評分表(非所有問項答案都填「是」，就能得到高分)**。團隊填答時，**應依實際 AI 研發狀況，填寫是、否或不適用並附上原因或理由**。

基本資料表			
計畫名稱		填表人姓名	
		填表人於計畫中擔任之角色	
計畫主持人		填表日期	

項目		**說明**
1	AI 名稱	（以下簡稱本 AI）
2	性質	□ 1.現為 AI 模型，預計僅作為其他 AI 系統之部分元件 □ 2.現為 AI 模型，預計開發為 AI 系統 □ 3.現已為 AI 系統
3	計畫團隊基於什麼目的開發本 AI？本 AI 能夠解決什麼社會問題？	

<table>
<tr><td>4</td><td>本 AI 具有何種功能？</td><td colspan="3">□ 1.單純模仿人類行為，請簡要說明：
□ 2.尋找人類難以察覺的事物關聯性，請簡要說明：
□ 3.預測人類難以估算的結果，請簡要說明：
□ 4.其他：＿＿＿＿＿＿＿＿，請簡要說明：</td></tr>
<tr><td>5</td><td>針對本 AI，計畫團隊是否「參與」或「預計會參與」AI 生命週期的「建模後」階段（例如：AI 系統部署、AI 系統場域應用、AI 系統產品經營等）？</td><td colspan="3">□ 1.完全不參與或完全無預計參與；
□ 2.參與或預計會參與，包括（可複選）：
□(1)AI 系統部署 □(2)AI 系統場域應用 □(3)AI 系統產品經營</td></tr>
<tr><td>6</td><td>本 AI 開發過程中，是否需用到與人相關之資料（無論資料是否先經去識別化處理）</td><td colspan="3">□是
□否</td></tr>
<tr><td rowspan="7">7</td><td rowspan="7">所有參與本 AI 評估之人員
(表格不敷使用請自行增列)</td><td>姓名</td><td>職稱</td><td>在可信賴 AI 評估過程中所負責之事項</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

1.透明與可解釋性(Transparency & Explainability)

為增進社會公眾對 AI 研發及應用之信賴，AI 的發展應符合透明性及可解釋性。透明性與可解釋性彼此間具有目的與手段之關係。透明性是指讓利害關係人可從外部檢驗 AI 系統的資料、特徵(features)、模型、演算法、訓練方法與品質保證等過程，以幫助利害關係人衡量 AI 系統的開發及運作是否合乎所期待的價值。而可解釋性是達成透明性的手段之一，是指能說明 AI 之所以能達成預期部署或應用目的之理由，例如但不限於解釋 AI 產出結果或決策的因果關係、AI 產出結果或決策背後的科學知識、AI 產出結果或決策有效的正當理由。此外，為確保達到透明性與可解釋性，往往需要可追溯性(Traceability)原則作為輔助，以保存 AI 系統設計、開發、部署到應用等生命過程中的相關紀錄（例如但不限於訓練及測試 AI 所使用之資料、演算法或規則；所採行之訓練方法；驗證過程與歷次產出結果或決策等），以便受 AI 產出結果或決策影響之人提出挑戰、申訴或救濟時，能有跡可循。

AI 生命週期中各階段應關注事項：

1、 建模前：瞭解開發 AI 所用的資料

2、 建模中：解釋 AI 模型所產出之結果

3、 建模後（含系統部署至場域端經營及監控）：解釋 AI 系統所作之結果或決策

建模中	1-2-1（目的達成）	您是否確認**所開發的 AI 模型確實能達成預期部署或應用的功能**？採取什麼方法驗證或確認[1]？	• 低度：未確認 • 中度：單次確認 • 高度：建立程序化確認流程	□是（成熟度為□中度/□高度），驗證或確認方式說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.6.2.4

[1] 驗證或確認的方法，不限於技術方法，也可能包括利害關係人的參與或意見徵詢。

2.隱私與資料治理(Privacy & Data Governance)

AI 的研發與資料之間具有密切關連，無論是訓練、驗證或測試 AI，皆須使用資料。研究人員及 AI 系統生命週期中的相關參與者，應採取良好的隱私保護及資料治理，以負責任的方式蒐集、處理、利用及共享資料，提升 AI 研發成果的準確度、公平性及可靠性。AI 訓練、驗證或測試資料的蒐集、處理、利用手段應正當合法，並避免在研發與應用過程中侵害他人隱私、智慧財產權或影響國家安全。蒐集個人資料及資料再利用之前，應顧及個人資料當事人之自主權，並在符合資料蒐集目的特定、資料最少化等原則下，處理、利用資料。研究人員及相關參與者，應建立適當的資料管理與安全維護措施，以達成良好的隱私及資料治理。

AI 生命週期中各階段應關注事項：

1、 建模前：AI 開發所需資料之蒐集、處理、利用（尤其是與人相關之資料）
2、 建模中：AI 模型建立過程與產出結果對隱私、營業秘密、核心科技發展與國家安全的影響
3、 建模後（含系統部署至場域端經營及監控）：AI 系統運作或應用對隱私、營業秘密、核心科技發展及國家安全的影響

建模中	2-2-1（資料最少化原則）	您是否確保 AI 建模過程中，僅處理、利用 AI 預期部署或應用目的所必要的資料？ （例如：為訓練自駕車蒐集道路街景，偶然拍攝到人臉或車牌	• 低度：未確保 • 中度：單次確保 • 高度：建立程序化確保機制	□是（成熟度為□中度/□高度），確保方式說明如下： □否（成熟度為低度），理由說明如下： □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 3# 隱私和資料治理（資料治理）、

		號碼，預先於 AI 訓練前，將人臉或車牌號碼移除或模糊化）			ISO/IEC 42001:2023 控制措施 A.7.5
	2-2-2（預防風險）	當發現 AI 模型產出結果可能有下列風險時，您是否採取行動，以降低或避免風險[2]？ 1. 影響或侵害個人隱私 2. 影響或侵害營業秘密或影響 3. 影響國家安全及核心科技發展	• 低度：未處理 • 中度：單次處理 • 高度：已建立程序化處理機制	□是（成熟度為□中度/□高度），行動說明如下： 涉及之風險項目如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5

3.人類自主(Human Autonomy)

AI 的應用是為輔助人類決策，不應導致脅迫、欺騙、操縱人類甚至取代人類。因此，AI 的開發及應用應循以人為本的原則，作為強化或補充人類認知、社會或文化技能，確保人類與 AI 系統交流的過程中，仍保有作出有意義的選擇權，並能保持充分而有效的自主性與控制權。AI 參與者應採取符合具體情況並與現有技術相符之機制和保障措施，透過建立相關監督機制以確保系統不會侵害人類自主性或是引發其他負面效果。

AI 生命週期中各階段應關注事項：

[2] 三項風險只要有其中一項即須採取風險處理行動。

1、 建模前：瞭解 AI 開發目的 2、 建模中~建模後（含系統部署至場域端經營及監控）： （1）瞭解 AI 操控人類之風險 （2）人類挑戰 AI 系統產出結果之可能性

建模中~建模後（含系統部署至場域端經營及監控）	3-2-1（風險評估）	針對 AI 系統的部署應用，您是否預先評估過個人可能因此被操控的風險[3]？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化評估流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 1# 人類自主性和監督（人類自主性和自治）、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5

[3] 例如：使自然人成癮、使自然人失去原有的判斷能力、使自然人個人或群體之社會價值或認知被 AI 引導。

	3-2-2（挑戰 AI 產出結果之機會）	若有人類被 AI 操控之風險，您是否預先設計相關機制，讓 AI 使用者或可能受影響之人對 AI 產出結果提出挑戰？	• 低度：未設計 • 中度：已設計機制 • 高度：已設計機制，並能依挑戰意見適度矯正 AI 系統	□是（成熟度為□中度/□高度），機制設計說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.8.3

4.公平與不歧視(Fairness & Non-discrimination)

AI 的研發與應用，應避免延續或加劇對個人或群體之刻板印象、偏見或歧視。AI 參與者從開發、部署到應用 AI 的過程中，應關注 AI 產出結果是否在種族、膚色、民族、性別、性別認同、宗教、年齡、國籍、身心障礙、遺傳或其他分類上產生偏差，從而對個人或群體造成不公平待遇或歧視。

AI 生命週期中各階段應關注事項：

1、 建模前：

（1）瞭解開發 AI 所用資料

（2）確保 AI 相關參與者能認知到偏誤及歧視

2、 建模中：避免 AI 模型產出結果發生歧視或不公平

3、 建模後（含系統部署至場域端經營及監控）：避免及因應 AI 系統運作或應用結果所產生之不公平

建模中	4-2-1（風險評估）	您是否評估過 AI 模型產出結果可能對潛在使用者或可能受影響之當事人**造成、延續或加深成見、刻板印象及不公平待遇**[4]？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化確認流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5# 多元性、不歧視與公平(避免不公平的偏誤)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5
	4-2-2（溝通與避免風險）	您是否適時**徵詢潛在使用者及可能受影響之當事人或群體的意見**，以避免 AI 模型產出結果發生成見、	• 低度：未徵詢 • 中度：已徵詢 • 高度：已徵詢並建	□是（成熟度為□中度/□高度），徵詢方式及結果說明如下：	EU Assessment List for Trustworthy Artificial Intelligence

[4] 例如：利用該公司過往的聘僱紀錄（以男性為主）進行訓練、測試及驗證，而導致模型產出結果將女性應徵者排除。

		刻板印象及不公平待遇？	立程序化的徵詢機制(SOP)	□否（成熟度為低度） □不適用，原因如下：	(ALTAI) 5# 多元性、不歧視與公平(無障礙與通用設計)

5.資安與安全(Security & Safety)

AI 常見的安全威脅如：資料下毒(data poisoning)[5]、模型迴避(model evasion)[6]、模型逆向(model inversion)[7]等。AI 開發者及部署者，應確保 AI 在可預見的使用情境下，能正常、安全地運作。為此，需藉由辨識、防護、偵測、應變與矯正風險等方法，確保樣本蒐集、模型訓練、系統部署、系統運作環境及過程之安全，以維持 AI 產出結果之穩定性與可再現性，避免 AI 系統在實際場域應用時，因錯誤而對人類生命、身體、健康、財產造成損害。

AI 生命週期中各階段應關注事項

1、 建模前：
   （1）AI 開發資料蒐集管道之安全性
   （2）AI 開發資料之異常偵測、應變與矯正
2、 建模中：AI 模型之安全性
3、 建模後（含系統部署至場域端經營及監控）：AI 系統之安全性

[5] 攻擊者竄改特定模型所用之樣本，有意影響訓練資料以操控模型預測結果，尤其發生在模型需透過網際網路持續從新資料中學習的情況。

[6] 攻擊者在模型輸入值回傳的過程中，對輸入值加入細微雜訊，以大幅改變模型的預測結果。

[7] 攻擊者利用機器學習系統提供的一些 API 來獲取模型的初步資訊，並藉這些初步資訊對模型進行逆向分析，以獲取模型內部的隱私資料。

建模中	5-2-1（模型安全防護）	選擇模型之後，您是否採取措施確保模型的安全？ 例如：採取措施排除可能對模型有害的資料	• 低度：未防護 • 中度：已防護 • 高度：已防護並建立程序化防護機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4
	5-2-2（模型輸出結果異常偵測）	模型訓練過程中，您是否採取措施，偵測模型所產出的非預期訓練結果？如何偵測？	• 低度：未偵測 • 中度：已偵測 • 高度：已偵測並建立程序化偵測機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4
	5-2-3（應變與矯正）	針對非預期訓練結果，您是否建立校正、調整訓練過程或	• 低度：未建立 • 中度：已	□是（成熟度為□中度/□高度），對策說明如下：	EU Assessment List for

		重新選擇模型的對策？	建立 • 高度：已建立並能監督執行	□否（成熟度為低度） □不適用，原因如下：	Trustworthy Artificial Intelligence (ALTAI) 7# 可歸責性(風險管理)、ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4

6.永續發展與福祉(Sustainable Development & Well-being)

AI 發展應維護人性尊嚴、人權與民主價值，並追求對人類及地球有益之結果，避免加劇人類及生存環境之危害。AI 開發者、部署者及使用者，應關注 AI 對弱勢群體的包容性，避免濫用 AI、使人類失去創造力與技能或對環境永續帶來負面影響。

AI 生命週期中各階段應關注事項：

1、 建模前：AI 開發所需資料的多元與代表性

2、 建模中：AI 模型產出結果對整體社會與環境的影響

3、 建模後（含系統部署至場域端經營及監控）：

（1）AI 系統包容性

（2）AI 系統運作及應用之整體社會及環境影響

建模中	6-2-1（負面影響評估）	您是否評估 AI 模型產出結果對民主、人權、環境永續價值可能帶來的負面衝擊？採取什麼方式評估？發現哪些潛在負面衝擊？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化評估機制(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： 潛在負面衝擊如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 6# 社會與環境福祉(環境福祉、對社會全體或民主之影響)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5
				□否（成熟度為低度）	
				□不適用，原因如下：	

7.問責(Accountability)

問責原則橫向貫串前述六項原則，以確保AI從開發、部署至應用生命過程可追溯與可檢驗。AI 生命過程中的參與者，應針對前述六項原則之落實進行權責分工，負責任地管理開發 AI 所用之資料、模型與演算法，保存 AI 系統設計、開發、部署到應用等生命過程中的相關紀錄，瞭解並確保 AI 開發、部署及應用遵循相關法律、規範或契約要求。而 AI 實際應用時所產出之結果或決策，將影響到利害關係人之權益，應給予受 AI 產出結果或決策影響之

人提出挑戰、申訴或救濟之機會。 **AI 生命週期中各階段應關注事項：** 1、 建模前：可信賴 AI 生命週期管理與法遵責任 2、 建模中：AI 建模的管理責任 3、 建模後（含系統部署至場域端經營及監控）：AI 系統管理責任

生命週期階段	**對應問項**	**成熟度判斷準則**	**評估結果與說明** （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）	**問項參考來源**
建模中	7-2-1（建模過程可檢驗）	您是否採取相關措施，以促進未來 AI 系統的部署、運作或應用可被檢驗[8]？	• 低度：無措施 • 中度：已建立程序化措施 • 高度：已建立程序化措施，並確實執行稽核	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 7# 可歸責性(風險管理)、ISO/IEC 42001:2023 控制措施

[8] 例如：系統整合溯源、系統運作紀錄、系統應用風險管理紀錄等。

					A.5.3、 A.6.2.2~A.6.2.8

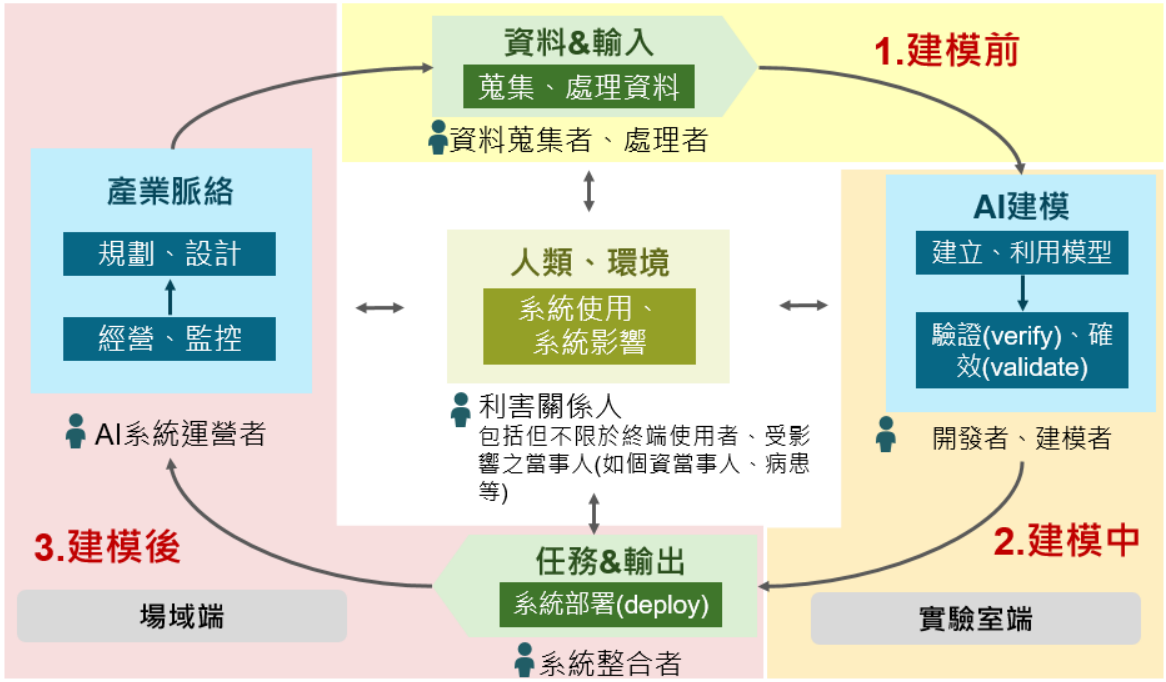
國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
可信賴 AI 評估表-建模後（徵案階段僅供計畫團隊參考，不需 填寫）

前言

為避免 AI 發展對民主、自由、人權等基本價值造成難以逆轉的負面衝擊，國科會臺灣 AI 卓越中心（Taiwan AICoE）集結 AI 專案各計畫團隊之力，共同建立「可信賴 AI 評估表」，將民主、自由、人權等基本價值融入整體 AI 系統生命週期各階段（如：訓練及測試資料之蒐集、模型建立、系統部署、系統經營及應用），增進研究人員及參與者對 AI 研發及應用相關倫理議題之敏感度，並促進研究人員、潛在 AI 使用者及可能受影響之人互動對話，及早消弭 AI 偏見與歧視，從而強化社會大眾對 AI 發展之信任。

本評估表先以 OECD「AI 建議」及國科會「AI 科研發展指引」作為基礎框架，列出七項可信賴 AI 基本原則，包括：透明與可解釋性、隱私與資料治理、公平與不歧視、人類自主、資安與安全、永續發展與福祉、問責，再依 AI 系統生命週期區分建模前、中、後（詳圖 1），預擬前述各項原則在不同生命週期階段，分別對應的評估問項及成熟度判斷準則。

建模前階段處於實驗室端，主要在蒐集與處理訓練、測試AI之資料。建模中階段亦處於實驗室端，涉及模型之選定、建立、驗證與確效。建模後階段則包括將 AI 進行系統整合、部署至應用場域及進行 AI 系統運營等。此外，鑒於問項填答內容與「AI 開發情況及其所欲部署或應用之目的」息息相關，在七原則問項之前，亦設計「**基本資料表**」，作為計畫團隊填寫可信賴 AI 評估問項的前置作業，以便更準確判斷哪些問項為必要評估項目，並了解計畫團隊自評結果是否合乎 AI 所欲部署或應用之目的。



資料來源：OECD Framework for the Classification of AI Systems(2022)

圖 1. AI 系統生命週期各階段之情境及主要參與者

本評估表目的在「**引導團隊預先思考 AI 研發與應用相關倫理議題**」，評估 AI 模型或系統未來可能影響人權與民主價值之風險，促進團隊提前討論可行的因應做法，故**性質上並非法遵評分表(非所有問項答案都填「是」，就能得到高分)**。團隊填答時，**應依實際 AI 研發狀況，填寫是、否或不適用並附上原因或理由**。

基本資料表			
計畫名稱		填表人姓名	
		填表人於計畫中擔任之角色	
計畫主持人		填表日期	

項目		**說明**
1	AI 名稱	（以下簡稱本 AI）
2	性質	□ 1.現為 AI 模型，預計僅作為其他 AI 系統之部分元件 □ 2.現為 AI 模型，預計開發為 AI 系統 □ 3.現已為 AI 系統
3	計畫團隊基於什麼目的開發本 AI？本 AI 能夠解決什麼社會問題？	

<table>
<tr><td>4</td><td>本 AI 具有何種功能？</td><td colspan="3">□ 1. 單純模仿人類行為，請簡要說明：
□ 2. 尋找人類難以察覺的事物關聯性，請簡要說明：
□ 3. 預測人類難以估算的結果，請簡要說明：
□ 4. 其他：＿＿＿＿＿＿＿＿＿＿，請簡要說明：</td></tr>
<tr><td>5</td><td>針對本 AI，計畫團隊是否「參與」或「預計會參與」AI 生命週期的「建模後」階段（例如：AI 系統部署、AI 系統場域應用、AI 系統產品經營等）？</td><td colspan="3">□ 1. 完全不參與或完全無預計參與；
□ 2. 參與或預計會參與，包括（可複選）：
□(1)AI 系統部署 □(2)AI 系統場域應用 □(3)AI 系統產品經營</td></tr>
<tr><td>6</td><td>本 AI 開發過程中，是否需用到與人相關之資料（無論資料是否先經去識別化處理）</td><td colspan="3">□是
□否</td></tr>
<tr><td rowspan="7">7</td><td rowspan="7">所有參與本 AI 評估之人員
(表格不敷使用請自行增列)</td><td>姓名</td><td>職稱</td><td>在可信賴 AI 評估過程中所負責之事項</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

備註：團隊 AI 模型的性質若僅作為其他 AI 系統之元件，且團隊並不參與 AI 系統於應用場域之部署、應用或產品經營者，原則上涉及建模後之問項可填不適用。

1.透明與可解釋性(Transparency & Explainability)

為增進社會公眾對 AI 研發及應用之信賴，AI 的發展應符合透明性及可解釋性。透明性與可解釋性彼此間具有目的與手段之關係。透明性是指讓利害關係人可從外部檢驗 AI 系統的資料、特徵(features)、模型、演算法、訓練方法與品質保證等過程，以幫助利害關係人衡量 AI 系統的開發及運作是否合乎所期待的價值。而可解釋性是達成透明性的手段之一，是指能說明 AI 之所以能達成預期部署或應用目的之理由，例如但不限於解釋 AI 產出結果或決策的因果關係、AI 產出結果或決策背後的科學知識、AI 產出結果或決策有效的正當理由。此外，為確保達到透明性與可解釋性，往往需要可追溯性(Traceability)原則作為輔助，以保存 AI 系統設計、開發、部署到應用等生命過程中的相關紀錄（例如但不限於訓練及測試 AI 所使用之資料、演算法或規則；所採行之訓練方法；驗證過程與歷次產出結果或決策等），以便受 AI 產出結果或決策影響之人提出挑戰、申訴或救濟時，能有跡可循。

AI 生命週期中各階段應關注事項：

1、 建模前：瞭解開發 AI 所用的資料

2、 建模中：解釋 AI 模型所產出之結果

3、 建模後（含系統部署至場域端經營及監控）：解釋 AI 系統所作之結果或決策

建模後	1-3-1（解釋 AI 產出結果或決策）[1]	您是否能**解釋 AI 系統所作出的結果或決策**？（例如：能解釋或說明「因果關係」、「產出結果背後的科學知識」或	• 低度：無法解釋 • 中度：可解釋 • 高度：AI 系統具有本身可解	□是（成熟度為□中度/□高度），作法說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 4# 透

[1] 本問項之目的在提升社會大眾對 AI 系統之理解，從而能接受 AI，並非要求必須公開程式碼。

		「產出結果之所以有效的正當理由」）	釋功能		明度(可追溯性)
	1-3-2（溝通傳達）	您是否建立相關機制，讓使用者[2]及受影響之當事人**知悉 AI 系統的性能或功能**（例如但不限於預測、偵測、人機互動），並**瞭解 AI 系統所作出的結果或決策**？	• 低度：未建立機制 • 中度：已建立被動機制（使用者及受影響之當事人要求才說明或提供資訊） • 高度：已建立主動機制	□是（成熟度為□中度/□高度），機制說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 1# 人類自主性和監督(人類自主和自治)、要求 4# 透明度(可解釋性)、ISO/IEC 42001:2023 控制措施 A.4.3

2.隱私與資料治理(Privacy & Data Governance)

AI 的研發與資料之間具有密切關連，無論是訓練、驗證或測試 AI，皆須使用資料。研究人員及 AI 系統生命週期中的相關參與者，應採取良好的隱私保護及資料治理，以負責任的方式蒐集、處理、利用及共享資料，提升 AI 研

[2] 使用者並不限於終端使用者，亦可包括 AI 生命週期的中間使用者(如 AI 系統商、AI 系統操作人員等)，可由團隊依本 AI 預計開發、部署、應用之情境自行界定。

發成果的準確度、公平性及可靠性。AI 訓練、驗證或測試資料的蒐集、處理、利用手段應正當合法，並避免在研發與應用過程中侵害他人隱私、智慧財產權或影響國家安全。蒐集個人資料及資料再利用之前，應顧及個人資料當事人之自主權，並在符合資料蒐集目的特定、資料最少化等原則下，處理、利用資料。研究人員及相關參與者，應建立適當的資料管理與安全維護措施，以達成良好的隱私及資料治理。 **AI 生命週期中各階段應關注事項：** 1、 建模前：AI 開發所需資料之蒐集、處理、利用（尤其是與人相關之資料） 2、 建模中：AI 模型建立過程與產出結果對隱私、營業秘密、核心科技發展與國家安全的影響 3、 建模後（含系統部署至場域端經營及監控）：AI 系統運作或應用對隱私、營業秘密、核心科技發展及國家安全的影響

建模後	2-3-1（風險評估）	您是否評估 AI 系統化後的運作或應用，可能產生以下風險[3]？ 1. 影響或侵害個人隱私 2. 影響或侵害營業秘密 3. 影響國家安全及核心科技發展	• 低度：未評估 • 中度：單次評估 • 高度：已建立程序化風險評估及管理機制	□是（成熟度為□中度/□高度），評估方式說明如下： 評估後發現可能涉及之風險項目如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5

[3] 三項風險皆須進行評估。

	2-3-2（事前溝通）	依前述風險評估結果，您是否採取措施，以利向潛在使用者及受影響之當事人溝通或傳達 AI 系統運作或應用對個人隱私、營業秘密或國家安全及核心科技發展的潛在風險？	• 低度：無法溝通或傳達 • 中度：已採取被動措施（使用者或受影響當事人詢問才溝通傳達） • 高度：已建立主動溝通傳達機制	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度），理由說明如下： □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 4# 透明度(溝通)、ISO/IEC 42001:2023 控制措施 A.8.2、A.8.5
	2-3-3（事後通報與處理）[4]	您是否建立明確管道、流程或機制，當 AI 系統於場域運作或應用時，若有下列情況，能夠通報及處理：	• 低度：未建立通報管道及處理流程 • 中度：已建立通報	□是（成熟度為□中度/□高度），建立方式說明如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 7# 可

[4] 本問項希望團隊於研發階段先思考，當未來 AI 在場域應用真的發現風險或有危害時，能否有機會應變。

		1. 有侵害個人隱私之風險或發生危害結果； 2. 有侵害營業秘密之風險或發生危害結果； 3. 有衝擊國家安全及核心科技發展之風險或發生危害結果？	管道及處理流程 • 高度：已建立程序化通報管道及處理流程，並能從根本上矯正	□否（成熟度為低度），理由說明如下： □不適用，原因如下：	歸責性(風險管理)、ISO/IEC 42001:2023 控制措施 A.8.2~A.8.5

3.人類自主(Human Autonomy)

AI 的應用是為輔助人類決策，不應導致脅迫、欺騙、操縱人類甚至取代人類。因此，AI 的開發及應用應循以人為本的原則，作為強化或補充人類認知、社會或文化技能，確保人類與 AI 系統交流的過程中，仍保有作出有意義的選擇權，並能保持充分而有效的自主性與控制權。AI 參與者應採取符合具體情況並與現有技術相符之機制和保障措施，透過建立相關監督機制以確保系統不會侵害人類自主性或是引發其他負面效果。

AI 生命週期中各階段應關注事項：

1、 建模前：瞭解 AI 開發目的

2、 建模中~建模後（含系統部署至場域端經營及監控）：

（1）瞭解 AI 操控人類之風險

（2）人類挑戰 AI 系統產出結果之可能性

建模中~建模後（含系統部署至場域端經營及監控）	3-2-1（風險評估）	針對 AI 系統的部署應用，您是否預先評估過個人可能因此被操控的風險[5]？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化評估流程(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 1# 人類自主性和監督（人類自主性和自治）、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5
	3-2-2（挑戰 AI 產出結果之機會）	若有人類被 AI 操控之風險，您是否預先設計相關機制，讓 AI 使用者或可能受影響之人對 AI 產出結果提出挑戰？	• 低度：未設計 • 中度：已設計機制 • 高度：已設計機制，並能依挑戰意	□是（成熟度為□中度/□高度），機制設計說明如下： □否（成熟度為低度） □不適用，原因如下：	ISO/IEC 42001:2023 控制措施 A.8.3

[5] 例如：使自然人成癮、使自然人失去原有的判斷能力、使自然人個人或群體之社會價值或認知被 AI 引導。

			見適度矯正 AI 系統		

4.公平與不歧視(Fairness & Non-discrimination)

AI 的研發與應用，應避免延續或加劇對個人或群體之刻板印象、偏見或歧視。AI 參與者從開發、部署到應用 AI 的過程中，應關注 AI 產出結果是否在種族、膚色、民族、性別、性別認同、宗教、年齡、國籍、身心障礙、遺傳或其他分類上產生偏差，從而對個人或群體造成不公平待遇或歧視。

AI 生命週期中各階段應關注事項：

1、 建模前：

（1）瞭解開發 AI 所用資料

（2）確保 AI 相關參與者能認知到偏誤及歧視

2、 建模中：避免 AI 模型產出結果發生歧視或不公平

3、 建模後（含系統部署至場域端經營及監控）：避免及因應 AI 系統運作或應用結果所產生之不公平

建模後（含系統部署至場域端經營及監控）	4-3-1（識別常見的不公平現象）	在 AI 系統預計部署應用的情境下，您認為有哪些常見的刻板印象或不公平待遇？	N/A	說明：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5# 多元性、不歧

					視與公平(避免不公平的偏誤)、ISO/IEC 42001:2023 控制措施 A.8.5
	4-3-2（風險評估）	您是否評估 AI 系統於場域運作或應用時，可能複製對潛在使用者[6]及受影響之當事人之成見、刻板印象或不公平待遇，或提高相關風險？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化持續評估機制(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5# 多元性、不歧視與公平(無障礙與通用設計)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5

[6] 不限於終端使用者。

	4-3-3（事前溝通傳達機制）	若有造成成見、刻板印象或不公平待遇之風險，您是否設計機制，以能向潛在使用者[7]及受影響之當事人妥適**溝通或傳達**「AI 系統運作或應用時預期有的成見、刻板印象或不公平待遇風險」？	• 低度：未設計機制 • 中度：採取被動溝通傳達措施(使用者或受影響之當事人要求才傳達) • 高度：已建立主動溝通傳達機制	□是（成熟度為□中度/□高度），機制設計說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5# 多元性、不歧視與公平(避免不公平的偏誤)、ISO/IEC 42001:2023 控制措施 A.8.4、A.8.5
	4-3-4（事後通報處理機制）	您是否建立明確管道或流程，當 AI 系統於場域運作或應用過程中，發現預期或非預期的成見、刻板印象或不公平待遇時，能夠**通報及處理**？	• 低度：未建立通報管道及處理流程 • 中度：已建立程序化通報管	□是（成熟度為□中度/□高度），管道或流程設計說明如下： □否（成熟度為低度）	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 5#

[7] 不限於終端使用者。

			道與處理流程 • 高度：建立程序化通報管道與處理流程，並能矯正不公平或歧視	□不適用，原因如下：	多元性、不歧視與公平(避免不公平的偏誤)、ISO/IEC 42001:2023 控制措施 A.8.2~A.8.5、A.9.2~A.9.4

5.資安與安全(Security & Safety)

AI 常見的安全威脅如：資料下毒(data poisoning)[8]、模型迴避(model evasion)[9]、模型逆向(model inversion)[10]等。AI 開發者及部署者，應確保 AI 在可預見的使用情境下，能正常、安全地運作。為此，需藉由辨識、防護、偵測、應變與矯正風險等方法，確保樣本蒐集、模型訓練、系統部署、系統運作環境及過程之安全，以維持 AI 產出結果之穩定性與可再現性，避免 AI 系統在實際場域應用時，因錯誤而對人類生命、身體、健康、財產造成損害。

AI 生命週期中各階段應關注事項

1、 建模前：

[8] 攻擊者竄改特定模型所用之樣本，有意影響訓練資料以操控模型預測結果，尤其發生在模型需透過網際網路持續從新資料中學習的情況。

[9] 攻擊者在模型輸入值回傳的過程中，對輸入值加入細微雜訊，以大幅改變模型的預測結果。

[10] 攻擊者利用機器學習系統提供的一些 API 來獲取模型的初步資訊，並藉這些初步資訊對模型進行逆向分析，以獲取模型內部的隱私資料。

（1）AI 開發資料蒐集管道之安全性 （2）AI 開發資料之異常偵測、應變與矯正 2、 建模中：AI 模型之安全性 3、 建模後（含系統部署至場域端經營及監控）：AI 系統之安全性

建模後（含系統部署至場域端經營及監控）	5-3-1（AI 系統機密性）	在系統部署階段，您是否採取措施確保只有經授權之人，才能取得 AI 權重？如何確保？	• 低度：未確保 • 中度：已確保 • 高度：已確保並建立程序化機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 3# 隱私和資料治理(資料治理)、ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4、A.9.2

	5-3-2 (AI 系統完整性)	在系統部署階段，您是否採取措施確保未來輸入資料不會受到雜訊[11]干擾？如何確保？	• 低度：未確保 • 中度：已確保 • 高度：已確保並建立程序化機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2# 技術穩健性與安全(抵禦攻擊之韌性和安全性)、ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4、A.9.2

[11] 如：AI 部署至應用場域時，應用場域所生有別於在實驗室中的任何干擾因素。應用場域中的環境噪音、可能發生的資安攻擊，也是其中一種。

	5-3-3 (AI 系統可用性)	在系統部署階段，您是否採取措施確保未來使用者無法欺騙或誤導 AI 系統？如何確保？	• 低度：未確保 • 中度：已確保 • 高度：已確保並建立程序化機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2#技術穩健性與安全(一般安全)、ISO/IEC 42001:2023 控制措施 A.6.1.3、A.6.2.3、A.6.2.4、A.9.2、A.9.4
	5-3-4 (事故應變)	您是否建立措施及流程，當 AI 系統於場域運作或應用時，發生資安事故，或對使	• 低度：未建立 • 中度：已建立 • 高度：已建立並能監督執行	□是（成熟度為□中度/□高度），措施或流程說明如下： □否（成熟度為低度）	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 2#

		用者[12]、受影響當事人的生命、身體、健康、財產造成危害時，能夠偵測、通報及處理？		□不適用，原因如下：	技術穩健性與安全(抵禦攻擊之韌性和安全性)、ISO/IEC 42001:2023 控制措施 A.8.3、A.8.5

6.永續發展與福祉(Sustainable Development & Well-being)

AI 發展應維護人性尊嚴、人權與民主價值，並追求對人類及地球有益之結果，避免加劇人類及生存環境之危害。AI 開發者、部署者及使用者，應關注 AI 對弱勢群體的包容性，避免濫用 AI、使人類失去創造力與技能或對環境永續帶來負面影響。

AI 生命週期中各階段應關注事項：

1、 建模前：AI 開發所需資料的多元與代表性

2、 建模中：AI 模型產出結果對整體社會與環境的影響

3、 建模後（含系統部署至場域端經營及監控）：

（1）AI 系統包容性

（2）AI 系統運作及應用之整體社會及環境影響

[12] 不限終端使用者

建模後	6-3-1（社會包容）	若您所開發的 AI 系統是供普羅大眾使用時，您是否考量並採取措施促進多元族群無障礙近用 AI 系統[13]？	• 低度：未考量 • 中度：已考量 • 高度：已考量並納入無障礙措施	□是（成熟度為□中度/□高度），考量方式及採取之措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) # 多元性、不歧視與公平(無障礙與通用設計)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5、A.9.2、A.9.3

[13] 例如：AI 系統使用者介面能被身心障礙或其他弱勢者使用。

	6-3-2（風險評估）	您是否評估 AI 系統在場域運作或應用時，可能侵害民主、人權或環境永續價值[14]？	• 低度：未評估 • 中度：已評估 • 高度：已評估並建立程序化機制(SOP)	□是（成熟度為□中度/□高度），評估方式說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 6#社會與環境福祉(環境福祉、對社會全體或民主之影響)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5、A.9.2~A.9.4

[14] 例如：可能導致特定族群有失業或去技能化的風險、可能被用於不合理的人民監控、可能被當作致命武器使用、可能耗費過多能源等。

	6-3-3（風險預防）	承上題，您是否採取措施，以避免 AI 系統在場域運作或應用時，侵害民主、人權或環境永續價值？	• 低度：未防免 • 中度：已防免 • 高度：已防免並建立程序化機制(SOP)	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 6#社會與環境福祉(環境福祉、對社會全體或民主之影響)、ISO/IEC 42001:2023 控制措施 A.5.2~A.5.5、A.9.2~A.9.4

7.問責(Accountability)

問責原則橫向貫串前述六項原則，以確保AI從開發、部署至應用生命過程可追溯與可檢驗。AI生命過程中的參與者，應針對前述六項原則之落實進行權責分工，負責任地管理開發 AI 所用之資料、模型與演算法，保存 AI 系統設計、開發、部署到應用等生命過程中的相關紀錄，瞭解並確保 AI 開發、部署及應用遵循相關法律、規範或契約要求。而 AI 實際應用時所產出之結果或決策，將影響到利害關係人之權益，應給予受 AI 產出結果或決策影響之

人提出挑戰、申訴或救濟之機會。 **AI 生命週期中各階段應關注事項：** 1、 建模前：可信賴 AI 生命週期管理與法遵責任 2、 建模中：AI 建模的管理責任 3、 建模後（含系統部署至場域端經營及監控）：AI 系統管理責任					
生命週期階段	**對應問項**	**成熟度判斷準則**	**評估結果與說明** （若為「是」，請提供佐證說明；若為「不適用」，請說明原因）		**問項參考來源**
建模後	7-3-1（部署、運作及應用過程可檢驗）	您是否採取相關措施，以促進 AI 系統的部署、運作或應用可被檢驗[15]？	• 低度：無措施 • 中度：已建立程序化措施 • 高度：已建立程序化措施，並確實執行稽核	□是（成熟度為□中度/□高度），措施說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 7# 可歸責性(風險管理)、ISO/IEC 42001:2023 控制措施

[15] 例如：系統整合溯源、系統運作紀錄、系統應用風險管理紀錄等。

					A.5.3、A.6.2.2~A.6.2.8
	7-3-2（溝通管道）	您是否提供申訴管道，讓 AI 系統使用者及受影響之當事人可對 AI 系統決策結果提出質疑？	• 低度：未提供 • 中度：已提供 • 高度：已提供並建立程序化處理機制(SOP)	□是（成熟度為□中度/□高度），管道說明如下： □否（成熟度為低度） □不適用，原因如下：	EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) 7# 可歸責性(風險管理)、ISO/IEC 42001:2023 控制措施 A.8.3

國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
訓練模型用之資料盤點表

計畫名稱：

計畫主持人：

填表人：

填表日期：　年　月　日

備註：

1. 請計畫團隊盤點訓練、驗證 AI 模型/系統所需之資料，完成下列表格。

2. 表格若不敷使用，可自行增列。

編號	資料檔案名稱	內含敏感資料與否 (1~4 可複選)	蒐集		保管方式 (可複選)	有存取權限之人員
			蒐集方法	蒐集依據		
D01		1. □含有下列一種或多種個人相關資料：姓名、出生年月日、身分證字號、護照號碼、車牌號碼、生物特徵、婚姻、家庭、教育、學經歷、職業、聯絡資訊(如電話、地址、email 等)、財務、社會活動狀況 2. □含有下列一種或多種個人相關資料:病歷、醫療、基因、性生活、健康檢查、犯罪前科 3. □含有國科會公告之國家核心關鍵技術保護清單相關資料	1. □團隊**直接**向資料當事人/著作權人蒐集 註明來源：_____ 2. □團隊**間接**從網路開放資料/資料當事人或著作權人以外之第三方蒐集 註明來源：_____ 3. □其他：_____	1. □依法令 法令依據：_____ 2. □依契約(需留存契約書) 3. □資料當事人書面同意(需留存同意書) 4. □其他：_____ 5. □以上皆無	1. □儲存於個人電腦 2. □儲存與特定作業區域之電腦 3. □儲存於國內雲端 4. □儲存於國外雲端	1. □計畫團隊**特定**成員，包括：_____ 2. □計畫團隊**所有**成員 3. □不限

		4. □含有營業秘密資料 5. □以上皆無				
D02		1. □含有下列一種或多種個人相關資料：姓名、出生年月日、身分證字號、護照號碼、車牌號碼、生物特徵、婚姻、家庭、教育、學經歷、職業、聯絡資訊(如電話、地址、email 等)、財務、社會活動狀況 2. □含有下列一種或多種個人相關資料:病歷、醫療、基因、性生活、健康檢查、犯罪前科 3. □含有國科會公告之國家核心關鍵技術保護清單相關資料 4. □含有營業秘密資料 5. □以上皆無	1. □團隊直接向資料當事人/著作權人蒐集 註明來源： 2. □團隊間接從網路開放資料/資料當事人或著作權人以外之第三方蒐集 註明來源：_____ 3. □其他：_____	1. □依法令 法令依據：_____ 2. □依契約(需留存契約書) 3. □資料當事人書面同意(需留存同意書) 4. □其他：_____ 5. □以上皆無	1. □儲存於個人電腦 2. □儲存與特定作業區域之電腦 3. □儲存於國內雲端 4. □儲存於國外雲端	1. □計畫團隊**特定**成員，包括： _____ 2. □計畫團隊**所有**成員 3. □不限

國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
預計產出之 AI 模型管理與共享表

說明：在合法授權再利用的基礎下，計畫團隊須就計畫產出之 AI 模型規劃共享模式，依據性質團隊可選擇公開共享或有條件共享。本專案產出之模型須配合揭露於本專案指定之網站。

表 B-2. 預計產出之 AI 模型

序號	模型名稱	簡要說明主要功能	子計畫別	模型 產出年月	是否上架共享平台* 若不公開，請敘明理由	使用資料集
M01						例如：D01+ D03
M02						
M03						
……						

*若選擇公開授權，計畫團隊須於本專案指定網站部署及共享(參見徵求公告之肆、「三、落地實踐要求」及「四、驗證與部署要求」)。

國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」
計畫預計產出之資料集清單

說明：

預計產出之資料集經本會認定具重要性、公共性或符合主權 AI 之關鍵資料集者，本會得請計畫團隊協助取得相關再利用授權，以利政府及其授權者視需求使用。

表 B-3. 計畫預計產出之資料集

對應表 B-1 之編號*	資料集名稱	內容說明及資料格式	資料量 (筆數、MB)	子計畫別	資料集產出年月

*附件 B-1「訓練模型用之資料盤點表」中屬於計畫產出之資料集者，請於此表再次列出並於欄位標記，以利追蹤核對。

國科會前瞻處 114 年度「邁向新世代前瞻人工智慧研究專案」計算資源需求申請表

「邁向新世代前瞻人工智慧研究專案」（以下簡稱本專案）與國家高速網路與計算中心（以下簡稱國網中心），擬進行專屬計算資源合作案，為提供適量算力資源予以本專案各研究計畫使用，請就研究之計算需求填列下列資訊，以利評估本專案之整體算力需求。

一、國網中心計算資源與說明：

Vender	Instance Type
NCHC Nano 5	8x NVIDIA H100 SXM 80GB

註：Nano 5 採用 NVIDIA H100 伺服器，單一臺伺服器內含 8 片 H100 GPUs。

二、為有效規劃與使用計算資源，請計畫團隊以 H100 GPU 為基礎，依據附件 B 表 B-2「AI 模型及其使用之資料集」表格預估所需計算資源並填寫下列表單。

序號	模型名稱	參數大小	預估資料集大小(MB)	記憶體使用量(GB)	子計畫別
M01					
M02					
M03					
…					

計算資源需求表(單位: H100 GPU 小時)

月份	1	2	3	4	5	6	7	8	9	10	11	12	合計
第一年	0	0	60	60	120	120	240	480	480	480	240	120	2400
第二年													
第三年													
第四年													

三、權利義務

(一)**非商業用途**：本算力僅限於測試、評估及開發用途，包括模型開發或解決方案研究，以支持未來產品或服務的研發。相關開發的模型或研究成果歸使用者所有。嚴禁將本專案的運算資源用於營利活動（如雲端服務、挖礦等）。

(二)**資訊安全**：使用者不得在該環境中處理或輸入任何專有或敏感資訊。

(三)**資料填報**：申請人應據實填寫申請書，包括 Token 數量、記憶體配置等，以利審查與算力分配。

(四)**成果揭露**：本計畫屬政府資源，在不涉及使用者研發或營業機密的前提下，將適當公開計畫成果。

(五)**生成式 AI 使用規範**：依據行政院及所屬機關（構）發布的生成式 AI 參考指引，凡使用生成式 AI 執行業務或提供服務，應向相關對象揭露其使用情形，以確保知情權。

(六)**條款變更**：本公告如有未盡事宜，除依相關法律規定辦理外，本專案保留修訂及補充（包括異動、更新、修改）的權利，並以國科會 AI 專案徵案網站公告為準。