

# 人工智慧科研發展指引

## AI Technology R&D Guidelines

科技部 108 年 9 月版

人工智慧 (AI) 作為引領未來的關鍵浪潮，對人類社會、經濟、政治生活造成重大改變，並於醫療、教育、自駕車等各領域均引起破壞式創新。為持續創造 AI 正面效益，應連結政府、產業、研發人員、民眾等利害關係人之能量，全體齊心協力善盡職守，共同打造值得信賴之 AI 環境。

研發人員因其角色與性質特殊性，若能於 AI 研發過程善盡能力，適時注意與利害關係人互動對話，符合普世期望協助創建「以人為本」、「永續發展」及「多元與包容」三大核心價值之 AI 社會，其研究成果也將更為廣泛接納。(一)**以人為本 Human-centred Values**：AI 科研應遵循以人為本之價值，以提升人類生活、增進人類福祉為宗旨，構築尊重人性尊嚴、自由與基本人權的人工智慧社會。(二)**永續發展 Sustainable Development**：AI 科研應追求經濟成長、社會進步與環境保護間之利益平衡，以人類、社會、環境間的共存共榮為目標，創造永續願景。(三)**多元包容 Diversity and Inclusion**：AI 科研應以創建及包容多元價值觀與背景之 AI 社會為發展目標，並且積極啟動跨領域對話機制，普惠全民對 AI 的理解與認知。

惟人工智慧牽涉之複雜度極高，若不謹慎利用恐造成重大歧視、濫用之社會疑慮現象，使研發人員心血浪費，AI 益處無法充分發揮。為抑制潛在疑慮與風險，消弭 AI 偏見、歧視與排除等情形，先進各國已陸續藉由研擬人工智慧相關基本準則，對人工智慧核心技術、人才教育、標準等行為進行導引。為持續完善我國科研發展，科技部亦訂定科研人員依循之「人工智慧科研發展指引」(以下稱本指引)，希冀透過倫理思維及應對之遵循參考達到加乘效果，促進人機合作，解

決人類過往於科技上所遭遇瓶頸，增進 AI 研究、發展及應用各科研階段能量，同造人類與 AI 共榮共利的美好未來。

為達成使用 AI 改善人民生活之願景，科技部由三大價值衍生構築後續 8 大指引，提供我國 AI 科研人員於學術自由及研究創新發展階段下之遵循，藉以完善 AI 科研環境與社會，接軌我國科研與國際 AI 脈動，引領臺灣科研新局。

### **1. 共榮共利 Common Good and Well-being**

AI 科研人員應追求人類、社會、環境間的利益平衡與共同福祉，並致力於多元文化、社會包容、環境永續等，達成保障人類身心健康、創建全體人民利益、總體環境躍升之 AI 社會。

### **2. 公平性與非歧視性 Fairness and Non-discrimination**

AI 科研人員應致力於 AI 系統、軟體、演算法等技術及進行決策時，落實以人為本，平等尊重所有人之基本人權與人性尊嚴，避免產生偏差與歧視等風險，並建立外部回饋機制。

### **3. 自主權與控制權 Autonomy and Control**

AI 應用係以輔助人類決策，AI 科研人員對於 AI 系統、軟體、演算法等技術開發，應致力讓人類能擁有完整且有效的自主權與控制權。

### **4. 安全性 Safety**

AI 科研人員應致力於 AI 系統、軟體、演算法等技術運作環境之安全性，包括但不限於穩健性、網路與資訊安全、風險控管與監測等，並追求 AI 系統合理且善意的使用，構築安全可靠之 AI 環境。

### **5. 個人隱私與數據治理 Privacy and Data Governance**

個人資料隱私侵害的預防，必須建立有效的數據治理，在 AI 研發與應用上，AI 科研人員應致力注意個人資料蒐集、處理及利用符合相關法令規範，以保障人性尊嚴與基本人權，並針對 AI 系統內

部之個人資料架構有適當的管理措施，以維護資料當事人權益。

## **6. 透明性與可追溯性 Transparency and Traceability**

AI 所生成之決策對於利害關係人有重大影響，為保障決策過程之公正性，在 AI 系統、軟體及演算法等技術發展與應用上，包括但不限於對於模組、機制、組成、參數及計算等進行最低限度的資訊提供與揭露，以確保一般人得以知悉人工智慧系統生成決策之要素。

AI 技術之發展與應用須遵循可追溯性要求，對於決策過程中包括但不限於資料收集、資料標籤以及所使用的演算法進行適當記錄，並建立相關紀錄保存制度，以利受 AI 技術決策影響之利害關係人得為事後救濟及釐清。

## **7. 可解釋性 Explainability**

AI 發展與應用階段，應致力權衡決策生成之準確性與可解釋性，兼顧使用者及受影響者權益，故 AI 技術所生成之決策，應盡力以文字、視覺、範例等關係人可理解之方式與內容，對於 AI 系統、軟體、演算法等技術之使用者與受影響者，進行事後的說明、展現與解釋。

## **8. 問責與溝通 Accountability and Communication**

基於社會公益與關係人利益之維護，AI 的發展與應用應致力於建立 AI 系統、軟體、演算法等技術之問責與溝通機制，包括但不限於決策程序與結果的說明、使用者與受影響者之回饋管道等。