落實全球AI治理 美國發布可信賴AI行政命令

人工智慧（Artificial Intelligence, AI）為現今國家科技發展不可或缺之關鍵，然如何安全、透明、合法的使用AI，是近年來國際上最關注的議題。財團法人資訊工業策進會科技法律研究所（資策會科法所）長期關注國際AI法制趨勢，致力於規劃促進AI發展的資料流通機制，期望兼顧AI倫理及資料治理並與國際接軌。

資策會科法所許嘉芳副研究員觀察近期國際發展趨勢表示，多數國際組織與先進國家皆肯定落實AI倫理價值與原則之重要性，嘗試提出輔助落實的文件或工具。美國2023年發布2項重要規則文件，期以完備可信賴或負責任AI目標之達成。

其一，美國國家標準暨技術研究院1月26日發布第一版人工智慧風險管理框架（Artificial Intelligence Risk Management Framework, AI RMF），提供各種規模且不限特定領域的社會組織靈活運用AI之自願性標準。期望將可信賴之考量融入AI系統設計、開發、部署及使用過程中，以幫助AI應用之決策與預期目標及價值保持一致，並降低AI應用風險。

圖一：AI RMF 四大功能。

其二，美國總統拜登於10月31日簽署美國安全與值得信賴AI行政命令（Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI），透過全面性的行政命令，引領各領域因應AI風險，包括：制定AI安全新標準、隱私保護、促進公平與公民權利、維護消費者、病患與學生權利、支援勞工、促進創新與競爭、提升美國海外領導地位、確保政府負責任且有效地使用AI等8項具體行動目標，藉以推動AI技術之研發與應用，進而確保AI技術及應用之安全性及促進人類福祉。

圖二：美國安全與值得信賴 AI 行政命令重點摘要。

針對國際對於AI倫理相關倡議，已逐漸形成共通性原則，並開始探討實際落實執行之相關機制。許嘉芳認為，未來落實AI治理應定期點AI研發過程中蒐集、處理、利用之資料、資料蒐集管道及利用目的，以進行隱私風險管理，且明確標示或揭露使用生成式AI所產出之結果。另建議應從AI設計、研發、部署及應用各階段，評估AI應用或輸出結果對個人或特定族群造成歧視之風險，預先採取風險管理措施。國安全與值得信賴AI行政命令提出後，後續亦將按時程規劃提出一系列具體落實措施或輔助工具，值得持續關注。

上稿時間： 2023年12月20日

新聞來源： https://www.cna.com.tw/postwrite/chi/360136

文章標籤