實現「負責任的AI」的關鍵在於強化數位資料歷程「證明」

「負責任的AI」為各國關注焦點。為幫助組織負責任地研發、維護、持續改善AI系統，SC 42委員會（ISO/IEC JTC 1/SC 42）於2023年12月18日發布AI管理系統標準「ISO/IEC 42001:2023（以下稱ISO 42001）」。

ISO 42001強調「保存、管理AI系統生命週期的數位資料」，有利於舉證。資策會科法所創意智財中心鄒宗萱副主任也表示：「本所協助司法聯盟鏈所推出的『b-JADE證明標章』，同樣採取管理數位資料歷程，用『完整的證據鏈』說話，讓民眾、產業更能信任司法。產業如果想證明AI可以信賴，可以參考司法機關作法，提出AI完整資料歷程證據。」

ISO 42001的AI系統資料可依流程，分為以下3大類型：

1. 管理AI系統須先有管理依據，組織的最高管理階層應制訂AI政策，滿足組織目的、適用要求、制定AI框架及承諾持續改善AI系統，且留存紀錄。
   組織須建立並維護AI風險標準，規範「可接受與不可接受的風險等級」、風險評估、風險處理、有效性評估以及AI系統影響評估流程，並要求保留前述歷程資料為證。

2. 組織應依規範流程執行AI系統，留存AI研發及使用資料歷程檔案，且數位資料應透過「分散儲存位置」、「設定接觸權限」等方式控管檔案，確保組織人員可以透過檔案名稱、日期、語言、軟體版本或圖片等方式來區分內外部檔案，並取得其所需的檔案。

3. 組織應持續監督、審查業務執行的有效性，如果不合格，應採取矯正措施，並留存審查紀錄為證。

總結而言，產業如想發展「負責任的AI」，應留意數位資料的生成、保護以及維護作法，強化管理AI生命週期所涉及的數位資料。

上稿時間： 2024年03月08日

新聞來源： https://www.cna.com.tw/postwrite/chi/364983