

AI 治理的實踐方向

透過 ISO 42001 打造 AIMS AI 治理需要穩扎穩打

如同聯合國宣佈永續發展目標 (SDGs)，AI 治理 (Governance) 正由聯合國高級別 AI 諮詢機構 (AI Advisory Body, AIAB) 倡議，於各國政府間與 AI 業界形成實務與法遵共識，而人工智慧管理系統 (AIMS) ISO 42001 正是 AI 治理得以落地的基石，支撐著 AI 治理所需的各項元素與數據。

文／梁日誠

有鑑於 AI 治理日漸重要，如何下手進行也成了需要研習的課題，下面便以遵循 ISO 42001 作為施行步驟，提供發展 AI 的團隊作為參考。

AI 實作管理的實踐

基於 ISO 42001 的 AIMS 扮演 AI 治理模型

中「AI 實作管理」的基石的角色，主責包含了各項 AIMS 流程、程序的建立、部署、維護、持續改善作業與績效數據提供的工作，並提供可稽核 (Auditable) 與可驗證 (Certifiable) 的機制 (經由 ISO&IEC 認證體制)，AIMS 的各項作業亦支持 UN AI 治理功能的 7 項機構功能。AIMS 包含了

| ISO 42001 章節 | 章節名稱 | 重要內容 |
|--------------|---------------------|-------------------------------------------------------------------------------------------------|
| 第4節 | 組織全景 | 全景 (包含法規、文化、道德、合規)、關注方、範圍 |
| 第5節 | 領導 | 領導與承諾、政策、角色、責任、權限 |
| 第6節 | 規劃 | 風險與機會、AI 風險評鑑、AI 風險處理、AI 系統衝擊評鑑、目標、變更 |
| 第7節 | 支援 | 資源、能力、認知、溝通、文件化資訊 |
| 第8節 | 運作 | 作業規劃與控制、AI 風險評鑑、AI 風險處理、AI 系統衝擊評鑑 |
| 第9節 | 績效評估 | 監督、量測、分析、評估、內部稽核、管理審查 |
| 第10節 | 改善 | 持續改善、不符合項目與矯正措施 |
| 附錄 A | 控制目標與控制措施 | 政策、內部組織、AI 系統的資源、評鑑 A 系統衝擊 (包含公平、偏見、社會、人權)、AI 系統生命週期、AI 系統資料、AI 系統的相關利害團體的資訊、AI 系統的使用、第三方關係等要求 |
| 附錄 B | AI 控制措施實作指引 | 政策、內部組織、AI 系統的資源、評鑑 AI 系統衝擊 (包含公平、偏見、社會、人權)、AI 系統生命週期、AI 系統資料、AI 系統的相關利害團體的資訊、AI 系統的使用、第三方關係等指引 |
| 附錄 C | 潛在的 AI 相關的組織目標與風險來源 | 目標: 公平、資安、安全、隱私、穩健、透通性與可解釋性、當責、可用性、可維護性、可用性與訓練資料品質 |
| | | 風險來源: 自動化等級、缺乏透通性與可解釋性、環境複雜度、系統生命週期議題、系統硬體議題、科技準備度、機器學習相關的風險來源 |
| 附錄 D | 跨領域或行業的 AI 管理系統使用 | 領域、ISO 27001、ISO 27701、ISO 9001 |

◀附表一▶ ISO 42001 結構

| AI 相關項目 | 歐盟 | 美國 | 國際社會 |
|---------|---------------------------------------------------------|--------------------------------------------------------------------------------|-------------------------------------------------------------|
| 法規 | EU AI Act 於 2024 年 2 月於歐洲議會獲得委員會層級的通過 | 總統令 (EO 14110) 於 2023 年 10 月公布 | 各國 AI 法規制定進度不一 |
| 標準 | 歐盟 CEN/CENELEC/JTC 21 與 ISO/IEC 合作制定國際/歐盟等同 AI 標準 | 聯邦政府：NIST 標準並將 NIST 標準對應其他國際 AI 標準。 非政府機構：由 ANSI 代表美國參與 ISO/IEC 制定國際 AI 標準。 | 參與 ISO/IEC 制定國際 AI 標準 |
| 符合性評鑑 | 評鑑機制： EU AI Act Article 43 "Conformity Assessment" | 評鑑機制： NTIA 建議 AI 系統稽核、稽核師登錄與獨立評估 | 評鑑機制： ISO/IEC 認證體制 (ISO 17021-1、ISO 17065、ISO 42006-發展中) |
| | 評鑑標準： CEN/CENELEC/JTC 21 與 ISO/IEC 合作制定國際/歐盟等同 AI 標準 | 評鑑標準： 發展中 | 評鑑標準： ISO 42001 |
| 社會與永續 | 聯合國 SDGs 與 Climate Change | 聯合國 SDGs 與 Climate Change | 聯合國 SDGs 與 Climate Change |

◀附表二> AI治理環境與發展現狀

AI 風險管理 (可參考 ISO 23894 標準) 與 AI 系統衝擊評鑑 (可參考 ISO 42005 標準-正發展中) 等議題, ISO 42001 的結構如<附表一>, 並可與其他領域的管理系統 (如: ISMS 資訊安全管理系統、PIMS 隱私資訊管理系統、QMS 品質管理系統) 整合, 以支持組織的整體治理工作, 包含 ESG 相關的各 SDGs。

AI 治理環境與發展

AI 治理 (含 AI 管理) 在全球都是新興且持續發展中的議題, 國際間與各經濟體的發展路徑與進度或有不同, 然藉由 UN 的 AI 治理倡議, 正逐漸形成共識及未來可預見的合意, <附表二>例舉經濟體有關 AI 治理環境與發展的 AI 法規、標準、符合性評鑑、社會與永續等現狀供參。

AI 治理可於組織內結合 AIMS 進行 (如: 依據 ISO 38507 與 ISO 42001), 也可運行於社群 (Society) 或行業 (如: 各中央目的事業主管機關所管的行業與機構)、國家 (如: 由中央主管機關主責) 與國際組織 (如: 由 UN AIAB 主責) 間, 宏

觀的涵蓋 AI 國際標準 (ISO) 的治理原則與 UN 的指導原則, 並與國際社會接軌, ISO/IEC 所制定或發展中的國際標準, 正提供 AI 治理、AI 管理、AI 風險管理、AI 系統衝擊評鑑、資料品質與管理等領域的指引與要求。考量其他法規 (如: 資通安全管理法) 採用或建議 CNS/ISO 27001、CNS/ISO 27701, 以及於專業證照領域培養了為數不少的 ISO 資安/隱私保護/營運持續相關主導稽核師課程證照持有者, 由國人熟悉的 ISO/IEC 搭配 CNS 體制來推動 AI 治理與管理、相關 AI 認證制度及 AI 治理與管理能力建置, 不失為中肯且可行的考量。



梁日誠 (GPM-bi CISSPI CCISMI CCISAI PII CCPI CCAI AIMPI FIAAIS I FHCA-EU AI ActI CAIEI CDEI CDAI DSFMI CBAEI FHCA-GDPRI) 現為加拿大 SCC/MC ISO/IEC JTC1/SC42、SC27、ISO/TC22/SC32、IEC/TC65 技術組成員, ISO 42001/ISO 27001/ISO 27701/ISO 22301/ISO 20000-1/IEC 62443-2-1 稽核師及講師, TCIC 環奧國際驗證公司全球營運總經理。