歐盟執委會提出「具可信度之人工智慧倫理指引」

歐盟執委會人工智慧高級專家小組（High-Level Expert Group on Artificial Intelligence）於2019年4月8日公布「具可信度之人工智慧倫理指引」（Ethics Guidelines For Trustworthy AI）。該指引首先指出，具可信度之人工智慧需具備三個關鍵特徵：（1）合法（Lawful）：應遵守所有適用於人工智慧之法規；（2）合乎倫理（Ethical）：確保人工智慧符合倫理原則與價值；（3）健全（Robust）：自技術與社會層面觀之，避免人工智慧於無意間造成傷害。

該指引並進一步指出人工智慧應遵守以下四項倫理原則：

(1) 尊重人類之自主權（Respect for Human Autonomy）：歐盟之核心價值在於尊重人類之自由與自主，與人工智慧系統互動之個人，仍應享有充分且有效之自我決定空間。因此，人工智慧之運用，不應脅迫、欺騙或操縱人類，人工智慧應被設計為輔助與增強人類之社會文化技能與認知。

(2) 避免傷害（Prevention of Harm）：人工智慧不應對人類造成不利之影響，亦不應加劇既有的衝突或傷害。人工智慧之系統運行環境應具備安全性，技術上則應健全，且確保不會被惡意濫用。此外，弱勢族群應於人工智慧運用中受到更多關注，並被視為服務對象。

(3) 公平（Fairness）：人工智慧系統之開發、布建與利用，必須具備公平性。除了透過實質承諾與規範，進行平等與公正之利益與成本分配外，亦須透過救濟程序確保個人或特定族群不受到歧視與偏見之侵害，並可對人工智慧之自動化決策結果提出質疑，且獲得有效之補救。

(4) 可解釋性（Explicability）：人工智慧應盡量避免黑箱（Black Box）決策，其系統處理程序須公開透明，並盡可能使相關決策結果具備可解釋性，分析特定訊息可能導致之決策結果，此外亦需具備可溯性且可接受審核。

本文為「經濟部產業技術司科技專案成果」

余和謙
法律研究員 編譯整理

上稿時間：2019年05月

文章標籤

人工智慧 智慧機械 機器倫理 人工智慧倫理準則

推薦文章