美國國家標準與技術研究院公布人工智慧風險管理框架（AI RMF 1.0）

美國國家標準與技術研究院（National Institute of Standards and Technology, NIST）於2023年1月26日公布「人工智慧風險管理框架1.0」（Artificial Intelligence Risk Management Framework, AI RMF 1.0），該自願性框架提供相關資源，以協助組織與個人管理人工智慧風險，並促進可信賴的人工智慧（Trustworthy AI）之設計、開發與使用。NIST曾於2021年7月29日提出「人工智慧風險管理框架」草案進行公眾徵詢，獲得業界之建議包含框架應有明確之衡量方法以及數值指標、人工智慧系統設計時應先思考整體系統之假設於真實世界中運作時，是否會產生公平性或誤差的問題等。本框架將隨著各界使用後的意見回饋持續更新，期待各產業發展出適合自己的使用方式。

本框架首先說明人工智慧技術的風險與其他科技的差異，定義人工智慧與可信賴的人工智慧，並指出設計該自願性框架的目的。再來，其分析人工智慧風險管理的困難，並用人工智慧的生命週期定義出風險管理相關人員（AI actors）。本框架提供七種評估人工智慧系統之信賴度的特徵，包含有效且可靠（valid and reliable）：有客觀證據證明人工智慧系統的有效性與系統穩定度；安全性（safe）：包含生命、健康、財產、環境安全，且應依照安全風險種類決定管理上的優先次序；資安與韌性（secure and resilient）；可歸責與資訊透明度（accountable and transparent）；可解釋性與可詮譯性（explainable and interpretable）；隱私保護（privacy-enhanced）；公平性—有害偏見管理（fair – with harmful bias managed）。

本框架亦提出人工智慧風險管理框架核心（AI RMF Core）概念，包含四項主要功能：治理、映射（mapping）、量測與管理。其中，治理功能為一切的基礎，負責孕育風險管理文化。各項功能皆有具體項目與子項目，並對應特定行動和結果產出。NIST同時公布「人工智慧風險管理框架教戰手冊」（AI RMF Playbook），提供實際做法之建議，並鼓勵業界分享其具體成果供他人參考。

本文為「經濟部產業技術司科技專案成果」

相關連結

- National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023)
- NIST AI Risk Management Framework Playbook, National Institute of Standards and Technology
- AI RISK MANAGEMENT FRAMEWORK, National Institute of Standards and Technology

你可能會想參加

→ 【2023科技法制變革論壇】AI生成時代所帶動的ChatGPT法制與產業新趨勢
→ 「跨域數位協作與管理」講座活動
→ 新創採購-政府新創應用分享會

陳箴
副法律研究員 編譯整理

上稿時間： 2023年04月

資料來源：

National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (last visited Feb. 16, 2023).
NIST AI Risk Management Framework Playbook, National Institute of Standards and Technology, https://pages.nist.gov/AIRMF/ (last visited Feb. 16, 2023).
AI RISK MANAGEMENT FRAMEWORK, National Institute of Standards and Technology, https://www.nist.gov/itl/ai-risk-management-framework (last visited Feb. 16, 2023).

延伸閱讀：

對於該框架之意見，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?no=66&tp=1&d=8727（最後瀏覽日：2023/02/16）。
美國參議院於2022年4月提出《演算法問責法案》對演算法治理再次進行立法嘗試，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?tp=1&d=8912&no=64（最後瀏覽日：2023/02/16）。
美國情報體系發布「情報體系運用人工智慧倫理架構」，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=8520（最後瀏覽日：2023/02/16）。
加拿大政府提交予國會《人工智慧資料法案》，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?tp=1&d=8888&no=64（最後瀏覽日：2023/02/16）。
英國政府提交予國會「人工智慧監管規範政策報告」，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?tp=1&d=8891&no=64（最後瀏覽日：2023/02/16）。

文章標籤

推薦文章