March, 2026

THY
Taiwan International
Law Offices

NEWSLETTER

美國人工智慧風險管理框架之介紹

近來人工智慧（Artificial Intelligence, AI）發展迅速，憑藉大數據分析與演算法之深度學習，AI 已能模擬人類智慧，發揮整合資訊、精確分析、持續學習及自主行動之功能，AI 所展現之自動化與高效特質，推動各學科發展取得重大突破，自統計學、軟體工程，延伸至語言學、神經科學，甚至是哲學與人文領域，AI 之運用可謂無遠弗屆。惟正是因 AI 所具備之大數據分析以及表達能力，其運用過程可能涉及個資之蒐集與保密、倫理議題以及歧視性言論等風險，故各國在推動 AI 發展之同時，亦紛紛考量如何避免潛在危害[1]。本次所欲介紹者，為美國於 2023 年所發布之人工智慧風險管理框架（Artificial Intelligence Risk Management Framework, 下稱「AI RMF 1.0」或「本框架」）。

一、AI RMF 1.0 之簡介

為因應美國 2020 年所發布之國家人工智慧倡議法案（National Artificial Intelligence Initiative Act），圍繞 AI 議題展開一系列之應對政策，其中，考慮到 AI 所涉及之各種風險以及其本身之複雜性，現有之風險管理系統難以完整應對 AI 所面臨之隱憂，為此，美國國家標準暨技術研究院（National Institute of Standards and Technology，NIST）在與民間單位及政府機關多方討論後，發布了 AI RMF 1.0，其並非法規範，不具強制力，旨在針對 AI 可能涉及之風險建立一個可資運用之風險管理系統，提供 AI 之開發、設計以及使用者參考並依循之指引，以求打造出安全且值得信賴之 AI 系統或服務，且為因應科技之瞬息萬變，NIST 將會隨時檢視並調整本框架之內容[2]。

二、AI RMF 1.0 之運用

AI 之強大，亦加劇其對世界產生之影響，為最大化正面效益並最小化負面衝擊，建構更可信賴的 AI 系統，風險管理系統之完善是不可或缺的，為達此目的，本框架設計兩大部分，第一部分是基礎資訊，說明風險管理之基本概念與描繪良好風險管理系統之雛形；第二部分是核心理念與文件，闡述風險管理之實施指引，以下簡要介紹之[3]：

[1]本所 2024 年 8 月 Newsletter 曾介紹過我國及歐盟因應 AI 所頒佈之法規範，詳情可參本所之網頁。
[2] 截至 2025 年 05 月為止尚未有修正，詳情可參 https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development (last visited:2025/05/21) .
[3] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1 (last visited:2025/05/21) .

（一）. 第一部分：基礎資訊（Foundational Information）

本部分區分為以下幾點：確認風險範圍（Framing Risk）與目標受眾（Audience）、建構人工智慧風險與可信任度（AI Risks and Trustworthiness）以及檢視本框架之有效性（Effectiveness of the AI RMF）。其核心意涵在於，建立完善之 AI 管理系統，首先需先認知到 AI 所面臨之風險以及目標受眾，畢竟 AI 帶來的負面影響，小至個人之隱私權，擴及對少數群體之歧視，甚至是對整體社會之結構性影響，可能涉及相關規範之違反。在釐清潛在風險後，即須做風險評估與優先排序，同時整合 AI 與其他網路安全、個資保護等風險，以利更有效地配置資源，並提升組織運作與管理效率。至於所謂人工智慧風險與可信任度，則係闡述可信任之 AI 系統所應具備之特性，例如安全、負責且無偏見的，且為呼應 NIST 對此框架之高度適應性與持續演進之期待，NIST 鼓勵使用者定期評估本框架對於其管理 AI 風險之能力是否確實有效提升。

（二）. 第二部分：本框架核心理念與相關文件（Core and Profiles）

本框架於第二部分進一步指出為實現該理念，風險管理系統之核心須涵蓋下列四種功能：治理（Govern）、映射（Map）、測量（Measure）和管理（Manage），此四種功能相互呼應，涵蓋整個風險管理流程的所有階段[4]。其中，關於治理，係指針對風險管理系統擬定政策、建立組織並設立流程，以結合技術與組織價值觀；關於映射，係指評估 AI 之潛在風險，並與跨領域之 AI 參與者協作；關於測量，指評估系統之可信度和風險，並確保決策之準確性；關於管理，指針對既有風險擬定應對計畫，以有效處理風險事件，並透過回饋機制進行系統改善與優化。

三、結論

人工智慧之發展與精進，有賴於國家之推動以及企業之有效運用，我國為在世界站穩腳步，亦頒布不少政策推動 AI 之發展[5]。在發展之同時，其背後之隱憂以及所涉之法規範皆應受到重視，只有當良好的風險管理系統被建構，AI 才是可資信賴之工具，未來我國政府或產業在運用 AI 之際，亦應注意風險管理與系統之適法性。

[4] 除本框架外，NIST 更發布指引手冊作為配套，可參 National Institute of Standards and Technology, *AI RMF PLAYBOOK*, https://airc.nist.gov/docs/AI_RMF_Playbook.pdf (last visited:2025/05/21) .

[5] 例如我國在 2018 年展開臺灣人工智慧行動計畫；於 2023 年公告人工智慧基本法草案；於 2025 年修正產業創新條例，鼓勵各產業導入人工智慧產品與服務，優化產業結構。