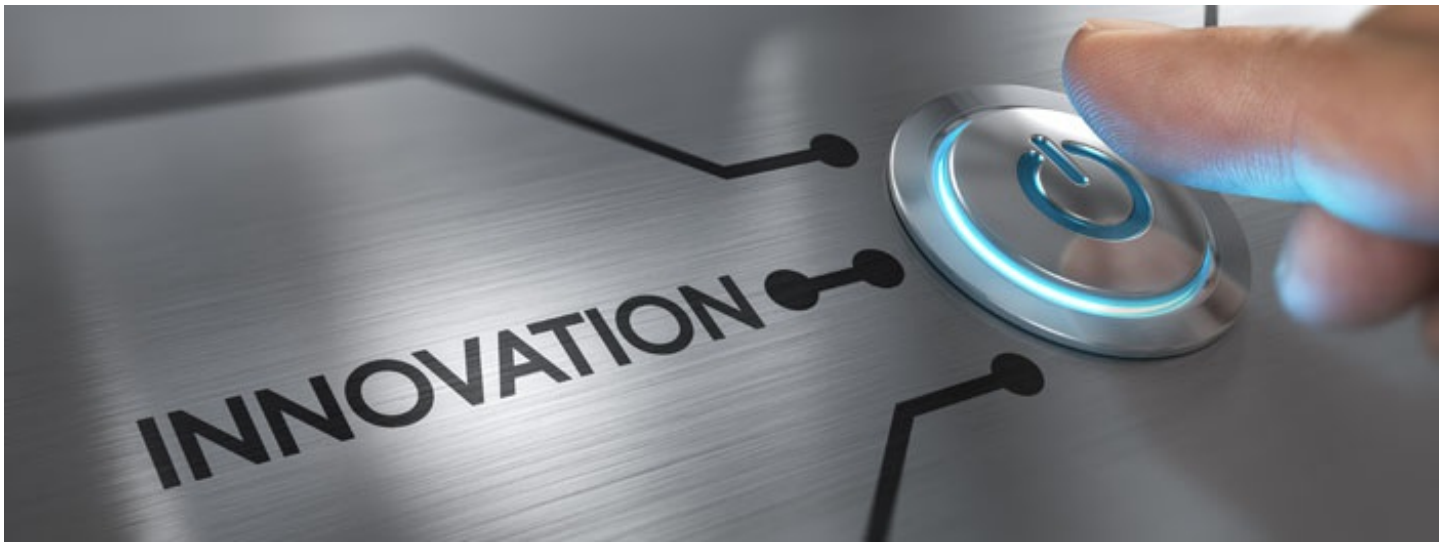
美國國家標準暨技術研究院規劃建立「人工智慧風險管理框架」，並徵詢公眾對於該框架之意見



美國國家標準暨技術研究院（National Institute of Standards and Technology, NIST）為管理人工智慧對於個人、組織以及社會所帶來之風險，於2021年7月29日提出將建立「人工智慧風險管理框架」（Artificial Intelligence Risk Management Framework, AI RMF）之規畫並徵詢公眾意見，截止日為9月15日，並預計於10月發布正式報告。

依照NIST說明，公眾所建議之人工智慧風險管理框架，可促進人工智慧之可信賴性，其中包含如何應對並解決人工智慧於設計、發展及使用過程中所遭遇之「精確度」（accuracy）、「可解釋性」（explainability）、「偏見」（bias）等議題。此外，上開管理框架預計為非強制性、供企業自願性使用於人工智慧設計、發展、使用、衡量及評估之人工智慧標準。

依現有公眾意見徵詢結果，其中DeepMind公司建議於人工智慧設計初期，必須預先構思整體系統之假設是否符合真正社會因果關係。舉例言之，當設計一套可預測民眾健保需求程度之系統時，如輸入參數僅考量民眾於醫療上的花費，將使僅有可負擔較高醫療費用之民眾被歸類為健保需求程度較高者，從而導致健保制度排擠經濟負擔程度較差之公民，故在設計系統時，應從預先設定之假設事實反面（counter-factual）思考並驗證是否會產生誤差或公平性之問題（例如預先思考並驗證「醫療費用支出較低之民眾是否即可被正確歸類為健保需求度低之民眾」）。惟進行上述驗證需要大量社會資料，因此DeepMind也建議NIST應建立相關機制，使這些社會資料可以被蒐集、使用。

此外，亦有民眾建議管理框架應有明確之衡量方法以及數值指標，以供工程界遵循。同時鑒於人工智慧發展極為快速，未來可能有不同於以往之人工智慧類型出現，故亦建議NIST應思考如何在「建構一套完整且詳細之人工智慧治理框架」與「保持人工智慧治理框架之彈性與靈活性」之間取得平衡。

最後，目前也有許多徵詢意見指出，許多人工智慧治理之目標會相互衝突。舉例言之，當NIST要求人工智慧系統應符合可解釋性，則人工智慧公司勢必需要經常抽取人工智慧系統中之「數據軌跡」（audit logs），惟數據軌跡可能被認為是使用者之個人資料，因此如何平衡或完善不同治理框架下之目標，為未來應持續關注之議題。

本文為「經濟部產業技術司科技專案成果」

相關連結

- Comments Received for RFI on Artificial Intelligence Risk Management Framework
- AI RMF Development & Request for Information
- 歐盟提出人工智慧法律調和規則草案

相關附件

- Final Report [pdf]
- Request for Information: Artificial Intelligence Risk Management Framework [pdf]

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
→ 供應鏈資安國際法制與政策趨勢分享會
→ 112年度「領航臺灣數位轉型」國際研討會-實體場
→ 112年度「領航臺灣數位轉型」國際研討會-直播場
→ 「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

→ 【線上場】113年「新創採購機制及鼓勵照護機構參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 113年新創採購-照護機構獎勵說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【中部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【臺北場】113年度新創採購-招標作業廠商說明會
→ 【臺中場】113年度新創採購-招標作業廠商說明會
→ 【高雄場】113年度新創採購-招標作業廠商說明會

王柏霆
法律研究員 編譯整理

上稿時間： 2021年10月

資料來源：

National Security Commission on Artificial Intelligence, *Final Report*, https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf (last visited, Sep. 28, 2021).
DeepMind, Request for Information: Artificial Intelligence Risk Management Framework, https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0105.pdf (last visited, Sep. 28, 2021).
Comments Received for RFI on Artificial Intelligence Risk Management Framework, NIST, https://www.nist.gov/itl/ai-risk-management-framework/comments-received-rfi-artificial-intelligence-risk-management (last visited, Sep. 28, 2021).
AI RMF Development & Request for Information, NIST, https://www.nist.gov/itl/ai-risk-management-framework (last visited, Sep. 28, 2021).

延伸閱讀：

〈歐盟提出人工智慧法律調和規則草案〉，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?no=55&tp=1&d=8674（最後瀏覽日期：2021/09/28）。

文章標籤

推薦文章