美國國家標準暨技術研究院發布「全球AI安全機構合作策略願景目標」，期能推動全球AI安全合作

美國國家標準暨技術研究院（National Institute of Standards and Technology, NIST）於2024年5月21日提出「全球AI安全機構合作策略願景目標」（The United States Artificial Intelligence Safety Institute: Vision, Mission, and Strategic Goals，下稱本策略願景），美國商務部（Department of Commerce）亦於2024年參與AI首爾峰會（AI Seoul Summit）期間對外揭示本策略願景，期能與其他國家攜手打造安全、可靠且可信賴之AI生態系。

由於AI可信賴與否往往取決於安全性，NIST指出當前AI安全所面臨的挑戰包含：一、欠缺對先進AI之標準化衡量指標；二、風險測試、評估、驗證及確效（Test, Evaluation, Validation, and Verification, TEVV）方法不健全；三、欠缺對AI建模後模型架構與模型表現間因果關係的了解；四、產業、公民社會、國內外參與者等在實踐AI安全一事上合作程度極為有限。

為因應上述挑戰並促進AI創新，NIST在本策略願景中擬定以下三大戰略目標：(1)推動AI安全科學發展：為建立安全準則與工具進行技術合作研究，並預先部署TEVV方法，以利評估先進AI模型之潛在風險與應對措施；(2)推展AI安全實務作法：制定並發布不同領域AI風險管理之相關準則與指標，以達到負責任設計、開發、部署與應用AI模型與系統之目的；(3)支持AI安全合作：促進各界採用前述安全準則、工具或指標，並推動全球合作，以發展國際通用的AI安全風險應對機制。

相關連結

- U.S. Secretary of Commerce Gina Raimondo Releases Strategic Vision on AI Safety, Announces Plan for Global Cooperation Among AI Safety Institutes, U.S. Department of Commerce (May 21, 2024)
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY [NIST], The United States Artificial Intelligence Safety Institute: Vision, Mission, and Strategic Goals (2024)

你可能會想參加

→ 【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

楊承昕 編譯整理

上稿時間：2024年08月

資料來源：
U.S. Secretary of Commerce Gina Raimondo Releases Strategic Vision on AI Safety, Announces Plan for Global Cooperation Among AI Safety Institutes, U.S. Department of Commerce (May 21, 2024), https://www.commerce.gov/news/press-releases/2024/05/us-secretary-commerce-gina-raimondo-releases-strategic-vision-ai-safety (last visited Aug. 1, 2024).
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY [NIST], *The United States Artificial Intelligence Safety Institute: Vision, Mission, and Strategic Goals* (2024), https://www.nist.gov/system/files/documents/2024/05/21/AISI-vision-21May2024.pdf (last visited Aug. 1, 2024).

延伸閱讀：
劉心妍，〈美國國土安全部發布「2024人工智慧路線圖」，確保AI安全開發與部署〉，資策會科技法律研究所，2024年6月，https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=9184（最後瀏覽日：2024/08/01）。

文章標籤

人工智慧 人工智慧風險管理

推薦文章