美國商務部國家電信和資訊管理局呼籲透過第三方評測提高AI系統透明度

2024年3月27日，美國商務部國家電信和資訊管理局（National Telecommunications and Information Administration, NTIA）發布「人工智慧問責政策報告」（AI Accountability Policy Report），該報告呼籲對人工智慧系統進行獨立評估（Independent Evaluations）或是第三方評測，期待藉此提高人工智慧系統的透明度。

人工智慧問責政策報告就如何對人工智慧系統進行第三方評測提出八項建議作法，分別如下：

1.人工智慧稽核指引：聯邦政府應為稽核人員制定適合的人工智慧稽核指引，該指引須包含評估標準與合適的稽核員證書。

2.改善資訊揭露：人工智慧系統雖然已經應用在許多領域，但其運作模式尚缺乏透明度。NTIA認為未來可以透過類似營養標籤（Nutrition Label）的方式，使人工智慧模型的架構、訓練資料、限制與偏差等重要資訊更加透明。

3.責任標準（Liability Standards）：聯邦政府應盡快訂定相關責任歸屬標準，以解決現行制度下，人工智慧系統造成損害的法律責任問題。

4.增加第三方評測所需資源：聯邦政府應投入必要的資源，以滿足國家對人工智慧系統獨立評估的需求。相關必要資源如：

(1)資助美國人工智慧安全研究所（U.S. Artificial Intelligence Safety Institute）；

(2)嚴格評估所需的運算資源與雲端基礎設施（Cloud Infrastructure）；

(3)提供獎金和研究資源，以鼓勵參與紅隊測試的個人或團隊；

(4)培養第三方評測機構的專家人才。

5.開發及使用驗證工具：NTIA呼籲聯邦機關開發及使用可靠的評測工具，以評估人工智慧系統之使用情況，例如透明度工具（Transparency Tools）、認驗證工具（Verification and Validation Tools）等。

6.獨立評估：NTIA建議聯邦機關應針對高風險的人工智慧類別進行第三方評測與監管，特別是可能侵害權利或安全的模型，應在其發布或應用前進行評測。

7.提升聯邦機關風險管控能力：NTIA建議各機關應記錄人工智慧的不良事件、建立人工智慧系統稽核的登記冊，並根據需求提供評測、認證與文件紀錄。

8.契約：透過採購契約要求政府之供應商、承包商採用符合標準的人工智慧治理方式與實踐。

NTIA將持續與利害關係各方合作，以建立人工智慧風險的問責機制，並確保該問責報告之建議得以落實。

相關連結

NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION, NTIA calls for audits and investments in trustworthy AI systems (Mar. 27, 2024)

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

→ 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

陳郁潔
副法律研究員 編譯整理

上稿時間：2024年07月

資料來源：
NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION, NTIA calls for audits and investments in trustworthy AI systems (Mar. 27, 2024), https://www.ntia.gov/press-release/2024/ntia-calls-audits-and-investments-trustworthy-ai-systems (last visited April 26, 2024).

延伸閱讀：
National Telecommunications and Information Administration, FACT SHEET: NTIA Urges Policy Changes to Boost Accountability and Trustworthiness in Artificial Intelligence Systems (Mar. 27, 2024), https://www.ntia.gov/other-publication/2024/fact-sheet-ntia-artificial-intelligency-policy-accountability (last visited April 26, 2024).

文章標籤

人工智慧 數位政府 人工智慧監管 人工智慧風險管理

推薦文章