

英國與美國為人工智慧安全共同開發簽署合作備忘錄



英國技術大臣（U.K. Secretary of State for Science）蜜雪兒·多尼蘭（Michelle Donelan）和美國商務部長（U.S. Secretary of Commerce）吉娜·雷蒙多（Gina Raimondo）於2024年4月1日在華盛頓特區簽署一份合作備忘錄（MOU），雙方將共同開發先進人工智慧（frontier AI）模型及測試，成為首批就測試和評估人工智慧模型風險等進行正式合作之國家。

此備忘錄之簽署，是為履行2023年11月在英國的布萊切利公園（Bletchley Park）所舉行的首屆人工智慧安全峰會（AI Safety Summit）上之承諾，諸如先進AI的急速進步及濫用風險、開發者應負責任地測試和評估應採取之適當措施、重視國際合作和資訊共享之必要性等等，以此為基礎羅列出兩國政府將如何在人工智慧安全方面匯集技術知識、資訊和人才，並開展以下幾項聯合活動：

1. 制定模型評估的共用框架（model evaluations），包括基礎方法（underpinning methodologies）、基礎設施（infrastructures）和流程（processes）。
2. 對可公開近用模型執行至少一次聯合測試演習（joint testing exercise）。
3. 在人工智慧安全技術研究方面進行合作，以推進先進人工智慧模型之國際科學知識，並促進人工智慧安全和技術政策的一致性。
4. 讓英、美兩國安全研究所（AI Safety Institute）間的人員互相交流利用其團體知識。
5. 在其活動範圍內，依據國家法律、法規和契約規定來相互共享資訊。

換言之，兩國的機構將共同制定人工智慧安全測試之國際標準，以及適用於先進人工智慧模型設計、開發、部署、使用之其他標準。確立一套通用人工智慧安全測試方法，並向其他合作夥伴分享該能力，以確保能夠有效應對這些風險。就如英國技術大臣蜜雪兒·多尼蘭強調的，確保人工智慧的安全發展是全球性問題，只有通過共同努力，我們才能面對技術所帶來的風險，並利用這項技術幫助人類過上更好的生活。

相關連結

- [The Bletchley Declaration\(2023, November 1\), The Bletchley Declaration by Countries Attending the AI Safety Summit 1-2 November 2023, GOV.UK](#)
- [U.S. and UK Announce Partnership.\(2024, April 1\), U.S. and UK Announce Partnership on Science of AI Safety, U.S. Department of Commerce](#)
- [Michelle Donelan, UK-US Partnership on AI safety research, UK Parliament, April 16, 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**
- 「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
- 電商零售業法制宣導說明會暨產學研座談會

林允中 編譯整理

上稿時間：2024年07月

資料來源：

The Bletchley Declaration(2023, November 1), *The Bletchley Declaration by Countries Attending the AI Safety Summit 1-2 November 2023*, GOV.UK, <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> (last visit June 30, 2024).

U.S. and UK Announce Partnership.(2024, April 1), *U.S. and UK Announce Partnership on Science of AI Safety*, U.S. Department of Commerce, <https://www.commerce.gov/news/press-releases/2024/04/us-and-uk-announce-partnership-science-ai-safety> (last visit June 30, 2024).

Michelle Donelan, *UK-US Partnership on AI safety research*, UK Parliament, April 16, 2024, <https://questions-statements.parliament.uk/written-statements/detail/2024-04-16/hcws409> (last visit June 30, 2024).

文章標籤

人工智慧

巨量資料

資料運用

人工智慧監管

人工智慧風險管理

 推薦文章