

臺灣主權 AI 訓練語料庫申請須知及使用規範

114.12.09

一、申請資格

從事人工智慧(AI)訓練及學習之開發、應用或相關研究之國內外政府機關(構)、公私法人、研究機構、學校、非法人團體或自然人。

二、申請及審核

1. 帳號申請

於臺灣主權 AI 訓練語料庫（以下簡稱語料庫）線上申請帳號及填寫基本資料。

2. 應備文件

2.1 請具體說明從事 AI 訓練及學習之開發、應用或相關研究內容，包含專案應用情境、使用模型、預計使用算力資源、所需訓練語料資料集類型（如客服諮詢問答、法規資料）及預期效益等。

2.2 上傳 2.1 所提說明內容之佐證文件（如研究計畫書、專案計畫書、模型訓練說明文件等）。

3. 評估審查

3.1 申請資料不齊全或資格不符者，語料庫維運管理單位得通知限期補正，逾期未補正則視同放棄。

3.2 語料庫維運管理單位得就申請資格及應備文件進行綜合評估審查，符合下列情形之一者得優先審查：

(1) 曾提供語料資料予語料庫。

(2) 數位發展部數位產業署「數位產業跨域軟體基盤暨數位服務躍升計畫」算力平臺之入選名單。

- 3.3 語料庫維運管理單位原則於收受申請案後 7 個工作日內完成審查，並以電子郵件通知帳號啟用；為確保帳號安全性及有效性，帳號啟用後每 3 個月，須重新進行憑證驗證。

三、使用規範

1. 使用者所查詢下載之語料資料，僅得作為 AI 模型訓練與學習之用途，並應遵守各該語料資料所採用之授權條款及語料庫相關使用規範，並同意語料庫得揭露其使用語料資料所產出或輸出之相關成果。
2. 使用者應妥善保管並安全運用所下載之語料資料於 AI 訓練及學習，且不得進行任何意圖重新識別個人身分之行為；倘若發現語料資料中涉及個人資訊，應通報語料庫維運管理單位，並主動刪除已下載之相關資料。
3. 使用者應就語料資料之使用成果，包括人工智慧訓練之產出與輸出，於合理範圍內設置適當防護機制，避免與原語料資料產生實質近似，或肇生著作權侵害，以降低對原語料資料市場或價值產生潛在負面影響之風險。
4. 使用者使用語料庫提供之語料資料，涉及下列法律及倫理爭議行為時，應自負其責，授權人不因提供語料資料，而被認為對該等行為有表達同意、許可或核准之意思表示：
 - (1)違反國際或當地法規；
 - (2)傷害未成年人或依法應受特別保護之人；
 - (3)製造虛假資訊；
 - (4)用於個人資訊追蹤；
 - (5)用於污蔑、誹謗或騷擾；
 - (6)偽裝他人身份；
 - (7)侵犯他人合法權利；
 - (8)對特定族群進行歧視或傷害；

(9)對特定群體的弱點進行分析並加以傷害；

(10)用於法律禁止的專業用途（如醫療或其他法律保留之相關行為等）。

5. 若使用者之使用目的、方式等有違合法性或不當用途，語料庫維運管理單位得暫停或終止該使用者帳號下載語料資料權限；對於已下載之語料資料，使用者應配合停止使用。

6. 語料庫維運管理單位辦理推廣活動或進行成果調查時，請配合提供相關執行進度或成果報告等資料，以提升語料庫整體服務效益。

四、聯絡方式

若有申請相關問題，請洽語料庫維運管理單位客服信箱：

tsaitc@moda.gov.tw