附件 6 去識別化處理程序

計畫名稱：可信賴的生活場景文字識別技術 ____（下稱本計畫）

計畫主持人：陳祝嵩 ____

填表人：__ 李偉華

填表日期：111 年 8 月 10 日

備註：下表一項資料集請填寫一份，表格不敷使用請自行增列。

※若技術團隊從既有資料庫取得資料，則不須繼續回答本附件。

一、本機構對於「涉及個資之原始資料集」進行「**去識別化**」之**意願**

事項內容	是否同意
對於涉及個資之原始資料集，本機構（本技術團隊）須先進行「去識別化」處理再供共享使用，以符合資料最少化原則、維護資訊安全、降低洩漏個資風險。 備註：所謂「去識別化」，係指透過一定程序的處理，使個人資料**不再具有直接或間接識別性**。（依據：個資法第 2 條第 1 款有關個人資料定義之反面解釋、法務部涵釋《法律字第 10703512280 號》）	■ 同意 □ 不同意，理由：_____ _____ _____ _____ _____ _____
關於去識別化，技術團隊採用哪些措施（方法）？（例如：將人名、性別、年齡、職業……去除）	措施 / 方法： _____替換影像中車牌之文字，以及對人臉進行模糊化 _____ _____ _____

二、去識別化**處理程序**

事項內容	回應
當計畫需要去識別化時，在符合計畫及組織需求下，須遵循下列階段，進行識別化處理： 1. **定義需求：**若資料有特定接收對象，須與資料需求者溝通，確定**須納入哪些資料**。有些資料對於計畫目的雖有幫助，但不是關鍵性資料，此種資料是否需要保留**必須有所取捨**。 （考量因素：對資料使用者信任程度、計畫團隊對隱私 保護政策要求等。確定資料範圍後，須分析出**直接識別符**，並進行處理） 2. **「重新識別」風險規劃：**設計並記錄**重新識別風險分析**的方法及過程。 （內容包含：採用去識別化之方法、**威脅模型建立**、處理步驟及**相關風險參數之確立**等） 3. **去識別化實做及驗證：**使用選擇的去識別化方法**重複**對資料進行**去識別化處理**，計算實施結果直到符合相關風險參數。 （若資料量很大，考量效率，實做上可採用「抽樣技術」輔助**事先預演操作**，在確定符合風險參數後再套入原本的資料集進行處理） 4. **解決方案驗證：**完成去識別化處理之後，須對特殊資料進行處理，**防止**具**特殊疏離群值的的 PII（**個人可識別資訊）**被重新識別**。此外，為了防止攻擊者導入外部資料比對獲得 PII（個人可識別資訊），建議須**抽樣相當數量**的資訊到常被使用的網站搜尋，**分析是否可查詢到個人資訊**，以**驗證**去識別化的**完整性**。最後須記錄整個作業程序，作為善盡個資保護盡義務之佐證，並可供後續類似工作參考。 5. **週期性持續審視：**在資料的整個生命週期過程中**持續、** **週期性重新評估**「現行的去識別化機制**是否仍符合風險需求**」。計畫團隊須因應威脅發展及技術改變，**適度調整**去識別化處理的程序。 （相關的審視須包含適當矯正措施及預防動作程序）	■ 同意 □ 不同意，不同意之處及其理由： _____ _____ _____ _____ _____ _____

資料集內容若含有個人資料，則須說明資料集已經**由何種方式進行去識別化**，如：遵循本專案（人工智慧主題研究專案）共同制訂方式或計畫自行定義方式（需逐欄位說明各欄位之處理作法）。	■ 同意 □ 不同意
在資料集內容皆須去識別化的情況下，目前技術上主要實現去識別化技術有下列幾種： 1. 壓抑(Suppression, Redaction)：將資料移除，或是使用 其他不洩漏個資的值取代。 2. 模糊(Fuzzing)：在資料中加入”雜訊”。 3. 概化(Generalization)：將資料的精確度降低，轉換為較 不容易被識別或是高階表述的值或是型態。 4. 縱向資料一致性(Longitudinal Consistency)：將縱向資 料依一致性相符的原則進行修改。 5. 開放式本文處理(Text Processing)：手動(Manual)方式 處理開放式格式(Free-format)文件。 6. 保留原狀(Pass-through)：不處理，保留資料原貌。 以上去識別化處理技術的主要目的為處理夠大的資料集，讓攻擊者不容易自釋放出的資料集中辨識出其中的PII 當事人。	■ 同意 □ 不同意