**使用者：**

你是一位教授AI倫理課程的老師，我是大學生，現在正在做一個使用 Teachable Machine 辨識行人的影像模型專題。請你針對這個專題可能涉及的倫理問題來問我一個問題。

**ChatGPT 說：**

太好了，這是一個很有趣也很實用的專題！那我來問你一個重要的倫理問題：

🔍 **當你使用 Teachable Machine 來建立辨識行人的模型時，你是如何確保在收集或使用影像資料時，有尊重到這些行人的隱私權與同意權的呢？**

這問題其實不只是技術問題，也牽涉到資料來源、目的、以及資料在未來可能被怎麼用的責任喔！你覺得怎麼樣？

**使用者：**

資料一旦用於模型訓練完畢，我會把原始影片資料刪除，避免被不當使用或外洩的風險。 盡可能拍攝背影或遠景，避免捕捉到明顯的臉部特徵，有需要的話能使用工具模糊臉部。 請你針對我的回答給予回饋與引導，再問我一個更深入的問題。

**ChatGPT 說：**

你這樣的回答真的蠻不錯的！👏

你有展現出保護資料的意識（像是**訓練完即刪除原始資料**），也注意到了避免蒐集可識別資訊的技巧（如**拍攝背影、模糊臉部**），這些都是在實作中非常值得鼓勵的作法！🌱

不過，我們可以再更進一步思考：

🔍 **即使你做了這些匿名化的處理，你覺得模型本身會不會還是保留了某種「可識別性」？例如，模型能不能間接學到某些特定地點或族群的特徵？這樣的風險會不會讓模型有偏見？你會怎麼評估或減少這種風險？**

這是AI倫理裡常碰到的問題：資料看起來匿名，其實還是可能造成偏誤或不公平的決策。想聽聽你的想法！

**使用者：**

確實有可能模型會間接學到某些環境或群體特徵，像是： 如果在學校附近拍的行人，那模型可能就特別擅長辨識穿著類似、年齡層接近的學生，對其他族群辨識能力可能就很差。 為了減少這種風險，可以用以下幾點方式處理: 資料多樣化：在不同地點、不同時間、找不同身形、穿著、性別的人進行拍攝，讓資料涵蓋的行人族群更廣，降低模型只學會某一種樣貌的機率。 除此之外，我可以手動檢查模型輸出是否出現特定偏好，讓其他人一起測試，從不同視角找出偏誤。 請你幫我總結目前我們討論的倫理問題有哪些，並針對其中一個問題幫我擬一個解決策略。

**ChatGPT 說：**

太讚了！你的回覆真的展現出對 AI 倫理議題的深入理解，還能主動提出具體應對策略，超棒！🌟

**🧠 我們目前討論到的 AI 倫理問題總結：**

1. **隱私權與個資保護**
   * 拍攝到的影像中可能含有個資（如臉部、穿著、特定場域）。
   * 解法：模糊臉部、拍攝背影、訓練完刪除原始資料。
2. **資料來源是否經過同意**
   * 行人是否知道自己被拍攝、其影像被用來訓練 AI 模型？
   * 解法：盡量選擇公開場域或模糊身份；若特寫則應取得同意。
3. **偏見與公平性（Bias & Fairness）**
   * 模型可能偏好特定族群、年齡或穿著風格的行人，導致對其他人辨識失準。
   * 解法：資料多樣化、手動檢查模型表現、邀請他人測試找偏誤。
4. **模型殘留風險**
   * 即使刪除原始資料，模型仍可能學到與個人有關的特徵。
   * 解法：進一步監測模型行為，限制用途，避免模型被用於監控或歧視性應用。

**🎯 解決策略擬定（針對：偏見與公平性問題）**

**策略名稱：資料收集平衡機制**

**目的：**  
讓模型能公平辨識來自不同背景、年齡、穿著的行人，減少在某特定群體上過度擬合。

**具體做法：**

1. 拍攝計畫前先設計一個「資料樣本規劃表」，列出預計涵蓋的族群（如學生、上班族、老年人等）。
2. 各群體至少蒐集一定比例的樣本，避免集中於某一類型。
3. 記錄每筆資料的基本資訊（如時間、地點、明顯外觀特徵），方便後續分析是否平衡。
4. 訓練完模型後，針對不同類型的行人進行「族群別測試」，觀察準確率是否有差異。
5. 若發現偏誤，可進行資料補拍或重新訓練。

如果你需要把這些內容寫進報告，我也可以幫你轉成條列式或正式段落喔！你想要哪一種？😄