

人工智慧的兩難！ 體驗 AI 道德設計的取捨

人工智慧正以驚人的速度進行發展與應用，並且引領著科技的浪潮。然而，隨著人工智慧技術的蓬勃發展，我們也面臨一系列的道德挑戰和社會影響。人工智慧不僅是一種革命性的技術，更是一個可能觸及我們核心價值觀和道德原則的領域。隨著特斯拉執行長 Elon Musk、OpenAI 創辦人 Sam Altman 等科技巨擘大聲疾呼，對 AI 發展提出示警，也印證 AI 設計的道德倫理確實是不可忽視的議題，也許今時今日危機尚遠，但卻是人人該正視的一環。

文／旗標科技

AI 的偏見

人工智慧的相關技術當中，尤其是機器學習正被應用在許多領域，協助做出重要決策，且機器學習會完全仰賴你所給予的資訊來進行訓練並應用。不過問題是，在真實世界中的資料常存有不理想包含在其中的某些資訊；在資料收集過程中也可能有偏誤，將這些錯誤資訊包含在內，這會帶來演算法偏見，如以下幾個例子。

- 亞馬遜招聘演算法：2014 年，亞馬遜開始開發內部人工智慧系統並建立 500 多個模型，以識別過去簡歷中出現的約 50,000 個術語，做為內部人工智慧招聘工具並用以簡化招聘流程，來快速判斷大型資料庫內相關候選人的資格。但很快地，亞馬遜公司的機器學習專家發現了一個大問題：他們的 AI 招聘引擎不喜歡女性。



圖 1 亞馬遜機器學習專家發現 AI 招聘工具不喜歡女性

AI 學習到不健康的刻板印象

- 男人之於父親，就像是女人之於母親
("man is to father as woman is to mother")
- 男人之於國王，就像是女人之於王后
("man is to king as woman is to queen")
- 男人之於程式設計師，就像是女人之於家庭主婦
("man is to computer programmer as woman is to homemaker")

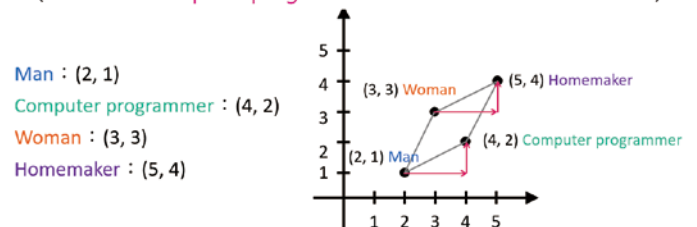


圖 2 AI 學習到不健康的刻板印象

- 詞嵌入 (Word Embedding)：這是一種用於自然語言處理應用程序的資料結構形式，把對文本內容的處理簡化為向量空間中的向量運算，並計算出向量空間上的相似度，來表示文字語義上的相似度。如果我們想要 AI 系統去理解男人和女人是平等的，從源頭避免這樣的偏見將是很重要的。圖 2 就是典型會造成偏見的資料集。
- 社群網路：由於社群網路的內容推薦主要是基於用戶的點擊，因此它們很容易導致現有偏見的放大，即使它們一開始很小，也會因為不斷的點擊而放大。若要建立降低傷害的人工智慧

系統，那就應該多包含那些最有可能受到系統傷害的弱勢族群觀點。

隱私問題

許多科技公司會利用各種方式收集有關其用戶的大量資訊。以往主要是賣場和其他零售商會透過結帳時要客戶提供會員卡，藉此收集購買資料。商店因此能夠將購買商品、購買時間與客戶關聯起來。但是，這一類型的資料記錄還不算人工智慧，只是傳統收集資料的方法之一。而現在盛行的人工智慧使用，給我們的隱私帶來了新的威脅，即使您已經很小心地保護自己的身份或資訊，也可能很難避免這些威脅。

AI 的問責

由於人工智慧系統並非是完美無缺，當然也會有失控或出錯的時候。那當人工智慧犯錯時，誰應該負起責任呢？使用者、創造者或是供應商？例如由微軟設計模仿青少年的聊天機器人，在網路上發布後數小時內就開始發布種族主義仇恨言論，微軟雖然立即下架，但相關傷害已經造成。

這個問題絕對比你想像的還要複雜，專家學者也各有不同的觀點。試想如果自動駕駛汽車傷害了行人，那應該由誰來負責呢？硬體的製造商（像是汽車用來感知環境的傳感器）？汽車上決定路徑的軟體開發者？允許自動駕駛汽車上路的政府？還是購買這輛車的車主？這是一個複雜的問題，也是各個國家在開發自動駕駛汽車時必須更為審慎面對的議題。

現階段的人工智慧，本質上是一種進行機器學習的自動化智能程序，但有了智能後，是不是就代表「它」能為自己的行為負起責任呢？自動駕駛汽車最常討論的難題包含以下的兩種問題（但不限於）：

1. 當自動駕駛汽車必須從多個目標當中選擇撞上其中一者時，我們應該如何幫自動駕駛汽車制定「該撞哪一個目標」的規則？
2. 自動駕駛汽車所帶來的任何傷亡，應該由誰負責以及擔負賠償與修復等責任呢？

這是一個困難的問題，麻省理工學院（MIT）的

道德機器（Moral Machine）平台，有關於這個問題的實驗，你可以實際操作看您會如何做選擇。

活動網址：Moral Machine
(<https://www.moralmachine.net/>)

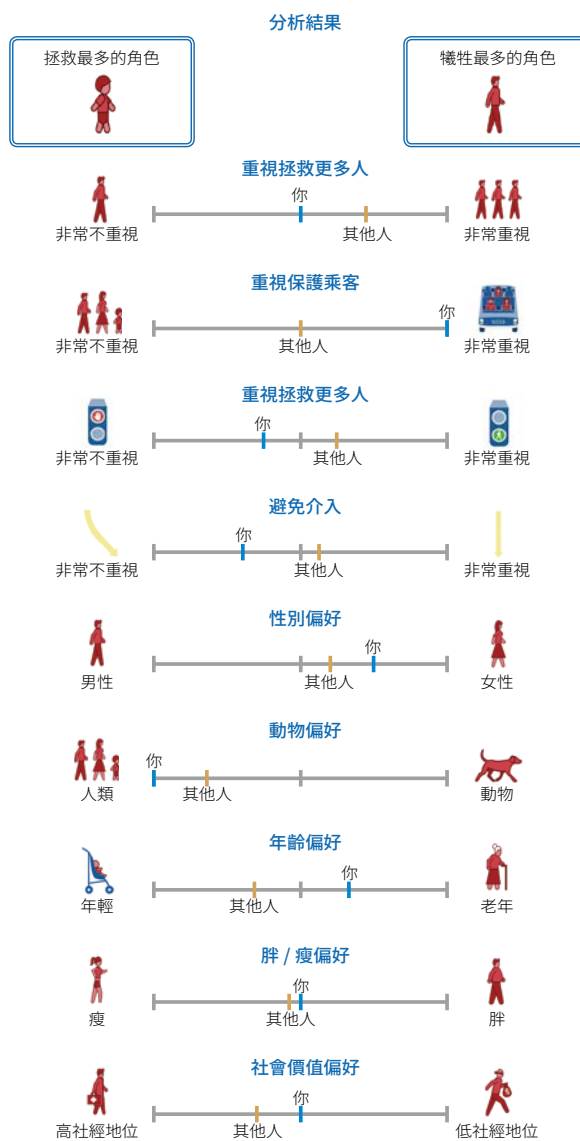


圖 3 道德情境的分析結果

結語

要在 AI 時代生存下來，你需要對 AI 有基本素養，更要懂得運用 AI，提高效率、彌補自身不足，才不會被改革的浪潮所淹沒。《40 堂不寫程式的實作課，培養新世代人工智慧素養》一書透過各種實際體驗的活動，帶你親身了解各種 AI 議題以及相關技術背後的原理。