

AI 竟會種族歧視

清華與跨國團隊揭露醫學倫理隱憂



國立清華大學
NATIONAL TSING HUA UNIVERSITY

2022 首頁故事

AI 竟會種族歧視 清華與跨國團隊揭露醫學倫理隱憂

AI 人工智慧也會種族歧視？清華資工系郭柏志助理教授與麻省理工學院、哈佛大學跨國研究發現，當採用先進的 AI 深度學習演算法來判讀 X 光片及電腦斷層掃描時，電腦竟可以從這些醫學影像分辨出患者是黑人、白人或亞洲人，並影響判斷的準確率。科學家迄今還不明白 AI 是如何辦到的。

郭柏志老師指出，醫界近年大量引進人工智慧，研發醫療機器人、智慧診療、及醫學影像識別技術等，協助醫生更快速、準確地診斷，這本該是造福病患的好事；但這項研究也揭露其中的隱憂，提醒我們應該注意醫療倫理問題，並消弭引進人工智慧產生的醫療不平等現象。

這項橫跨 4 國的跨國研究團隊，還包括了史丹佛大學、多倫多大學等，與美國國家衛生院合作，分析超過 20 萬位病患的胸部、頸椎、手部 X 光及胸部電腦斷層掃描，揭露人工智慧也有歧視問題的驚人真相。這項研究成果最近登上了國際頂尖期刊《刺絡針數位健康》(Lancet Digital Health)，並吸引外媒的重視報導。

美國食品藥物管理局(FDA)目前已核准醫院採用人工智慧輔助判讀 X 光片及電腦斷層掃描等醫學影像，人工智慧的深度學習系統可幫助電腦針對大量資料自我訓練，並發展出辨識模型。研究團隊發現，人類放射師無法從 X 光片判斷出患者的種族，但 AI 卻能輕易做到，且準確率高達 9 成。

團隊中一名哈佛大學教授表示，當團隊成員告訴他 AI 可以從 X 光片來判斷人種，他簡直不敢相信，認為這真的太瘋狂，一定是哪裡出錯了。當愈來愈多的證據浮現，有團隊成員則直言感覺很「恐怖」。

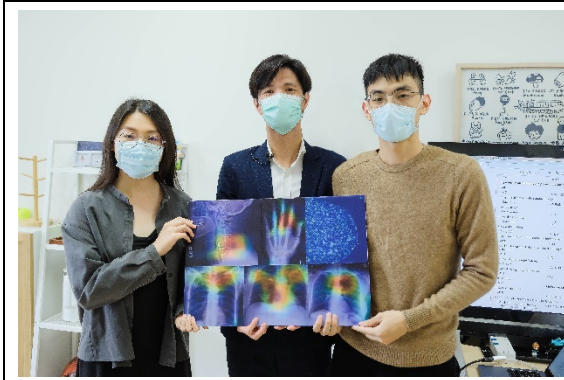
郭柏志老師表示，團隊本來猜想 AI 可能是從骨質密度來判斷人種，因黑人的骨質密度高於白人，但後來發現並非如此。郭柏志老師出示一張手掌的 X 光片，顯示 AI 判讀人種的重點聚焦在中指與食指的第三指節，「團隊裡也有二、三十年經驗的放射科醫師，但我們實在想不出來 AI 是怎麼判讀出人種的。」

更令研究團隊憂心的是，人工智慧的種族歧視也影響了醫學影像判讀的準確率。郭柏志老師表示，研究發現白人的醫學影像有問題卻沒被檢查出來的誤判率是 17%，但黑人的誤判率卻高達 28%，相差超過 10 個百分點。而醫學影像的誤判也會進一步影響急診、醫療給付等資源分配。

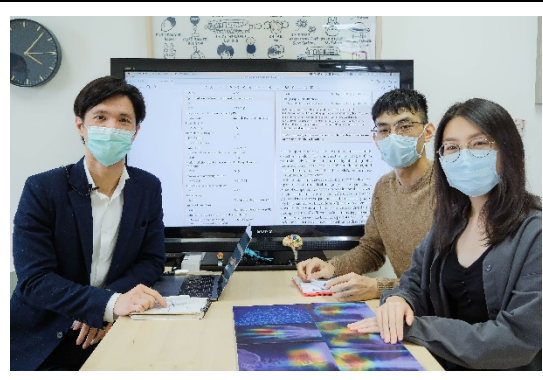
郭柏志老師說，我們原來以為人類才會歧視，電腦應該是最公平的，但沒想到電腦也默默地提取了人類沒教它的資訊，包括人種及性別等，「如何去除歧視，讓

各族群的醫學影像判讀都有一致的準確率，是我們下一階段要努力的目標。」

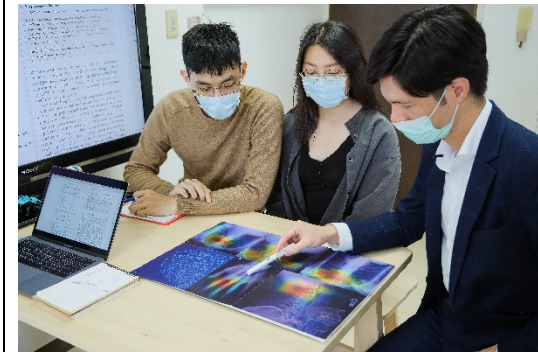
郭柏志老師帶領本校兩位資工系的大四學生王瑞恩、陳立晴參與了這項跨國研究。王瑞恩表示，參與這麼大規模的跨國研究，對大學部的學生來說是很難得的機會，他在團隊中負責人工智慧模型分析，獲益良多。



清華資工系郭柏志助理教授(中)帶領學生陳立晴(左)、王瑞恩(右)參與 AI 也會種族歧視的跨國研究。



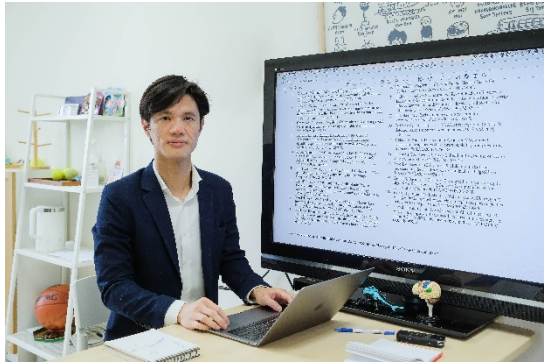
清華資工系郭柏志助理教授(左)帶領學生陳立晴(右前)、王瑞恩(右後)參與 AI 也會種族歧視的跨國研究。



郭柏志老師(右)與學生討論 AI 如何從手掌 X 光影像判斷出人種。



郭柏志老師(左)與學生討論 AI 如何從胸部 X 光影像判斷出人種。



郭柏志老師致力提昇人工智慧應用於醫學影像判讀的公平性。

人工智慧聚焦 X 光片食指、中指的第三指節(紅色處)來判斷人種。