

涉種族性別歧視 竟分析黑人易犯罪

監督測試法 力遏AI歧視

■隨着AI歧視的問題日趨嚴重，科學家正想方設法解決。
網上圖片

人工智能(AI)技術發展一日千里，當中「機器學習程式」(Machine Learning Algorithms, MLA)能自行學習數據規律，繼而作出準確預測，用途廣泛，但AI的運算明顯存在偏見，例如自行辨別犯罪率較高的黑人社區，惹起歧視黑人之嫌。有科學家於是研發出一款名為「監督式學習平等機會」(Equality of Opportunity in Supervised Learning)的測試法，審視MLA在作出決策時，是否涉及種族及性別歧視。

■黑人女性在職場上易受歧視，連AI於決定時亦受此影響。
網上圖片

■由於新測試法成效良好，美國當局有意用作監管銀行業。
資料圖片

互聯網巨擘Google高級研究科學家哈爾特承認，機器學習作出的決策，對人類生活產生巨大影響，不過「AI偏見」也隨處可見，包括網上廣告平台自動為男性提供高薪行政職位的檢索結果。另外，電腦分析數據時，或得出男性拖欠銀行貸款的機會較女性大，借貨人或因此拒絕向所有男性放債，變相歧視男性。

繞運算過程 直接得出結果

芝加哥豐田技術研究所電腦科學家斯雷布羅，參與研發出「監督式學習平等機會」測試法，他形容新方法繞開MLA繁複及難以理解的運算過程，直接分析輸入的數據及運算結果，審視是否出現歧視情況，過程中避免因研究對象的敏感屬性而出現歧視。有關研究成果已在去年12月於西班牙舉行的「神經信息處理系統」(NIPS)會議中發表。由於成效良好，據報美國消費者金融保護局亦有興趣

引入此方法，評估當地銀行。

學者：決策過程仍欠透明

雖然新方法有助消除AI偏見，但仍然存在不少棘手問題。研究AI的西英格蘭大學電子工程學士溫菲爾德認為，即使運用新方法，外界仍難以得知MLA決策過程，透明度顯然不足，一旦出現爭議，仍須邀請人類專家作出解釋。設菲爾德大學榮休教授夏基稱，在決定輸油管鋪設位置方面，AI可以大派用場，但至今仍未清楚它存在的偏見，利用它作出影響人類生活的決定，着實令人憂慮。

■《衛報》/ Digital Trends網站

面容識別系統偏見 黑人易無端惹禍

美國各地警方要確認嫌疑人身份，通常會把閉路電視拍得的面容影像，與資料庫內的疑犯檔案照作比對。美國喬治城大學法律中心去年4月的研究顯示，當地警方的面容識別軟件資料庫內有較多非裔美國人照片，而且無法準確辨認黑人面孔，令黑人容易無端惹禍。研究人工智能技術的專家認為，面容識別系統依靠人類輸入的資料運作，研發團隊應改善效能，避免黑人社群受針對。

為了解面容識別軟件是否存在種族偏見，喬治城大學法律中心技術員弗蘭克爾

及專家加維，向全美逾100個警察部門索取近1萬頁公共記錄作研究。他們發現警方識別系統的資料庫中，黑人照片比例遠高於其他族裔，而系統本身亦不擅長辨認不同黑人的面孔，造成種族偏見。事實上，此前微軟的動作感應裝置Kinect亦曾無法辨別黑人面孔，Google的相片應用程式甚至誤認黑人為猩猩。

專家：要擴大資料庫採樣數

研究面容識別系統的專家菲利普斯指出，識別系統出現的偏見，其實視乎系統資料庫的樣本多寡，例如東亞地區製作的識別軟件，便較其他地區更能準確辨認東亞人面孔。因此，如果資料庫的照片樣本數量夠多、採樣廣泛，便能更準確辨認各族裔人士。他建議設計團隊多考慮系統對應的社群，從而使軟件運作更公平。他們又認為，軟件工程師及警方並非故意令資料庫帶有種族偏見，但應提高運作透明度，令公眾釋疑。

■《衛報》



■基於面孔難以準確辨認，黑人往往無辜受牽連。
網上圖片

英推機械人指引 行為由人類負責

隨着人工智能(AI)技術日新月異，愈來愈多人類的工作已由機械人取代。但隨着AI涉及歧視的個案頻生，要求機械人遵守人類道德規範的聲音日益高漲。英國標準協會(BSI)早前發佈指引，規範科研人員需設計合乎道德的機械人，規定機械人不應用作殺戮或傷害人類，而它們的行為必須由人類負責。

部分地區的警方已開始採用AI技術，辨別構成潛在威脅的可疑

人物，但機械人學習時使用的數據可能存在偏頗，導致AI出現歧視，例如AI系統會偏好中年白人男士，少數族裔特別容易受針對。

過分依賴 人類可能被騙

科學家指，深度學習系統透過互聯網學習，問題在於這些數據本身存在偏頗，而機械人又照單全收，結果出現性別及種族歧視的傾向。BSI指引提出，當機械人

的應用愈來愈廣，人類遲早會過分依賴，引發嚴峻社會問題。設菲爾德大學教授夏基稱，如果長時間與機械人合作，漸漸習慣由它們提供正確答案，人類最終會忽略機械人出錯的可能。

根據BSI的指引，設計員應注重透明度，讓外界了解機械人的運算情況。不過，現時的AI太複雜，以深度學習系統為例，有別於由設計員編寫運作程序，機械人會自行嘗試數百萬次，直至發

展出一套有效的策略，設計員根本無從得知它的「思考」過程。指引亦提出一些較少人探討的問題，例如應否鼓勵機械人與人類產生情感聯繫。研究發現，在託兒所引入小型機械人陪伴兒童，會令孩子們「愛」上機械人，甚至認為它們比寵物更具靈性。有學者指出，當人類與機械人的感情日漸親密，不排除出現人類被機械人無意中欺騙的事件。

■《衛報》



■當機械人的應用愈來愈廣，人類遲早會過分依賴。
資料圖片

Google誤標用戶為「大猩猩」人當狗

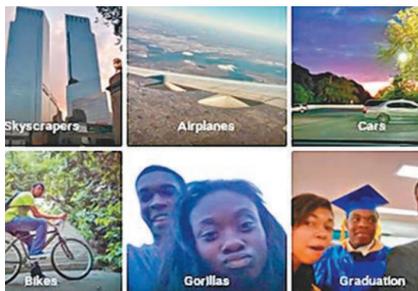
Google旗下的相片應用程式Google Photos，早前以人工智能為相片加上標籤，卻誤將兩名黑人標籤為「大猩猩」。事主在twitter投訴她和其朋友的相片被誤認，Google社交平台首席設計師祖格回應時稱，會在文字應用及影像識別上花更多工夫改善。

祖格指出，面容模糊不清以及膚色和光暗程度不同，均會影響影像識別系統，以往曾有不同種族的人被誤認作狗隻，團隊正在加上標籤方面改善用字及影像識別，包括有效辨認深色皮膚面容。

網搜「黑人屋」竟顯示奧巴馬居所

Google地圖程式也曾發生涉嫌種族歧視，在搜尋華盛頓地區的「黑人屋」或「黑人皇帝」時，搜尋結果便會顯示白宮，亦即黑人前總統奧巴馬的居所。Google發言人事後道歉，承諾會着手解決問題，以防同類事件再次發生。

除Google外，相片分享網站Flickr也曾將黑人影像



■Google曾誤將黑人標籤為「大猩猩」(下圖中)。
資料圖片
標籤為「猩猩」或「動物」，又將集中營的相片標籤為「運動」和「攀爬架」，發言人事後承認自動標籤功能仍有缺陷，將不斷作出改善。
■《衛報》

AI首當選美裁判 白人大獲全勝

雖然AI已經可以在很多範疇上代替人類做決定，但一旦涉及辨別美醜這種判斷，便會發現電腦的局限性。由微軟資助研發的「Beauty.AI」系統，去年舉行全球首次由電腦擔任評判的國際選美比賽，研究員本來預期系統會根據面部對稱性、皺紋數量等客觀因素評判，結果卻令人大跌眼鏡：44名勝出者中，絕大部分是白人，只有一名黑人脫穎而出。研究員指，結果反映AI的演算法在建立期間，已經會因樣本質素和數量產生偏見，從而影響判斷力。

Beauty.AI由微軟資助的「深度學習」研究組織Youth Laboratories研發，系統去年收集全球100個國家共6,000名佳麗的照片，從中選出最

能代表「人類美貌」的美女。參加者中白人佔多數，但也有不少有色人種參與，包括大批印度和非洲佳麗。

缺有色美女參考 產生偏見

Beauty.AI首席科學家扎沃龍科夫表示，雖然系統正式運作前，已分析大量美女的照片作為演算參考，但由於資料庫中有有色人種相片不足，導致系統誤以為美女以白人為主，繼而產生「白人就是美」的偏見。

同樣問題都出現在其他AI系統中，例如用來預測未來罪犯的軟件，往往會對黑人存有偏見。哈佛福德學院電腦科學教授弗里德勒指，由於少數族裔所佔樣本不多，



■即使在選美方面，AI亦會傾向選擇白人勝出。圖為2016年環球小姐選舉。
資料圖片

容易被演算法忽略，例如AI即使對佔美國人口2%的原住民有偏見，準確度仍可達98%。弗里德勒認為，研發者可藉改善輸入資料庫的數據，以及加入篩選，以確保不同種族人士可獲平等對待。
■《衛報》