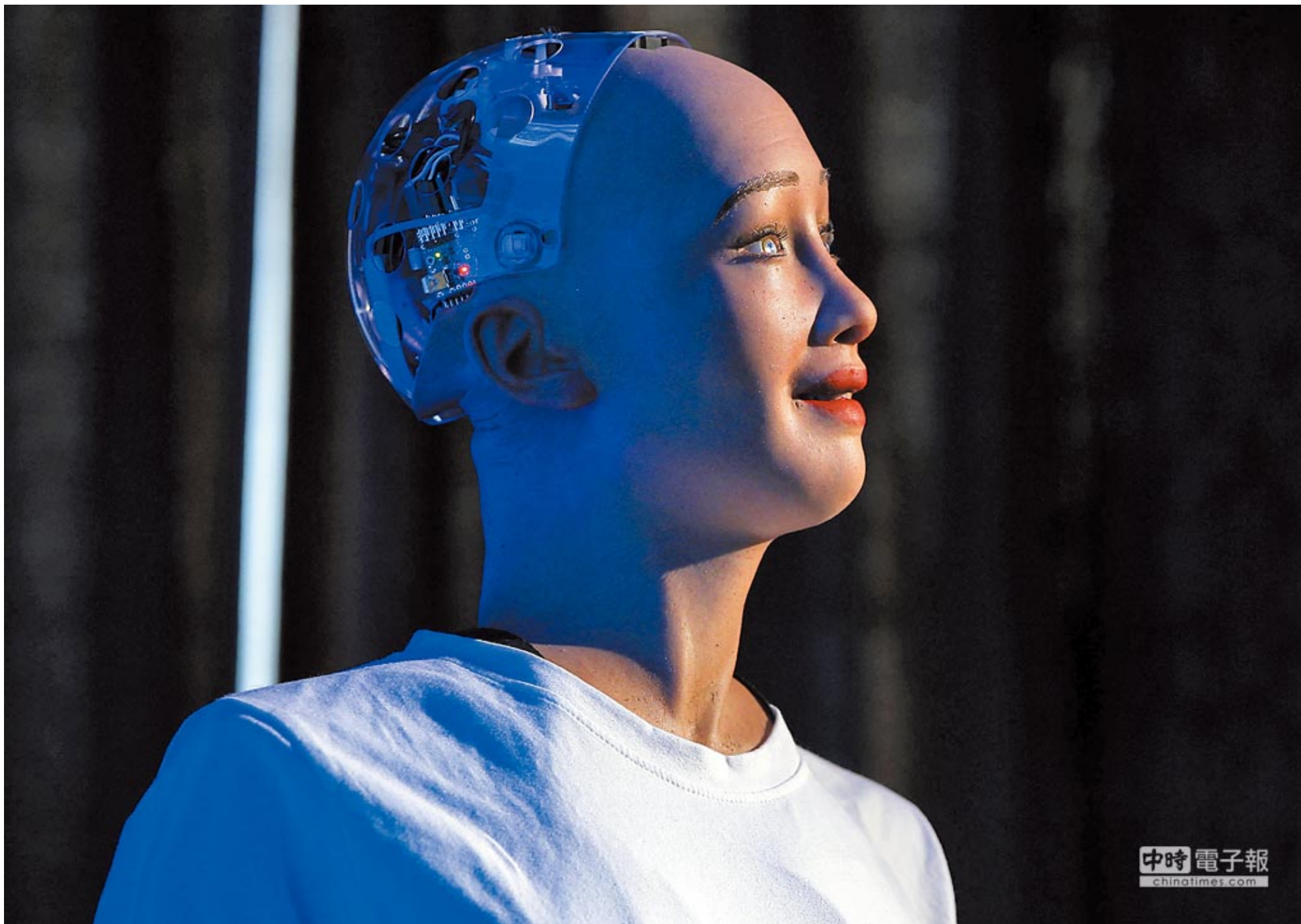


[首頁](#) / [工商時報](#)

專家傳真 - 「罪惡感」將是人工智慧難以跨越的鴻溝

04:10 2019/01/02 | 工商時報 | 梁化鵬中正大學企管系博士生



隨著科技進展，在未來的人類社會裡，人工智慧將不再是冰冷的機器，更聰明的人工智慧甚至能自我學習進化，而擔心人工智慧不斷進化、最終毀滅世界的聲音也益加浮現。圖 / 美聯社資料照片

人工智慧 (AI) 的概念早在1956年就被提出，剛開始的時候，人工智慧代表的是未來的概念，隨著數理演算與工程技術的進步，原本只在想像階段的人工智慧漸漸有了雛型，從簡單的語音辨識到類神經網路再到深度學習，高性能的電腦加上以大數據為基礎建構的知識庫讓人工智慧得以成真，人工智慧成為改善人類生活的幫手。

現今以進步的感測技術與強大運算能力開發出來的人工智慧技術應用在醫療、交通、娛樂、教育、生活等領域，在未來的人類社會裡人工智慧將不再是冰冷的機器，是人類不可或缺的夥伴。更聰明的人工智慧甚至能自我學習進化，因此有人擔心人工智慧不斷進化的結果最終會超越人類甚至毀滅世界。人類之所以成為萬物之靈，憑藉的是人類有高於其他物種的智商，而在人類文明高度發展的同時能保持不失序，依靠的是人類社會裡有一股無形的道德力量在制約著人類行為。在未來的世界，人工智慧終會超越人類智慧，至於人工智慧有無毀滅世界的可能，就看在人工智慧的世界裡是否也能建立維繫秩序的道德力量。

「道德」是一個非常抽象的名詞，為了使人工智慧也具備像人類社會的道德力量，科學家開始著手研究哲學演算法，企圖去教育機器人什麼是道德。可是什麼是道德，這個問題可能連人類都無法清楚定義。簡單講「道德」的具體表現是當人們作一件事時，會不會因為作這件事而產生「罪惡感」，人類的罪惡感是來自成長過程中累積的心靈感受，同樣的感受會在機器人身上發生嗎？要使機器人具備像人類一樣的道德能力幾乎是不可能的事。因為「道德」標準會因為種族、年齡、性別、宗教、教育程度而有所不同，因此科學界所研究的哲學演算法絕無可能一體適用，以東西方差異為例，東方世界強調群體主義，西方世界強調個人主義，東方世界強調階級輩份，西方世界強調人人平等。

舉例「媽媽和女友掉到河裡要先救哪一個？」當這樣一個千古難題交給機器人時，機器人會作出什麼樣的處置，設定為東方哲學演算法的機器人因為受東方孝道觀念影響而決定先救媽媽，但是當機器人設定為西方哲學模式時卻可能先救女友，這是因為西方觀念人命等值，但是卻可能因為女友有較高的殘餘價值而選擇先救女友。

這樣的道德差異在設計自動無人駕駛系統時一樣會遭遇生命價值判斷的抉擇，在無法避免的交通事故即將發生時，人工智慧控制的駕駛系統到底要先保護駕駛人還是選擇犧牲駕駛人去保全行人，或甚至決定取舍的條件是以功利階級的角度進行政經地位比較。在人類社會裡會因為東西方道德觀念不同而有不同的價值判斷結果，人工智慧機器人難道也要分為東方哲學機器人和西方哲學機器人，如果機器人用錯地方該怎麼辦？機器人做了決定後是不是也如同人類一般會產生「罪惡感」？

這種道德價值判斷的問題在傳統人類社會裡是不會有標準答案的，有時候模糊抽象是解決道德紛爭最好的方法。但是在人工智慧機器人的世界裡沒有所謂的模糊地帶，因為機器硬體是透過標準製造出來的，決策演算法是預先載入機器內的，因此機器人的所有作為都是「預謀」的，如果機器依照預載的演算法作出人類道德所無法接受的行為時，機器人與其製造者需要為此擔負法律責任嗎？這恐怕是人工智慧發展所要面對最大的挑戰。

- ▶ 專家傳真 - 全球AI健康照護產業投資布局與臺灣商機
- ▶ BMW集團與Amazon合作聯手加速數據驅動的創新

#道德