明新學報 45 卷 第 1 期 PP.26-42
Vol. 45, No. 1, Minghsin Journal, Oct. 2022

淺談人工智慧的倫理議題*

張東文**

明新科技大學通識教育中心

摘要

人工智慧（AI）的迅速發展已為我們的生活帶來便利，但同時也引發各種疑慮與風險，我們不得不認真思考人工智慧所引發的倫理議題。

對於 AI 科技帶來的倫理問題，本文將它分為三大類來討論，第一類可歸為 AI 設計不當導致的倫理問題，它使 AI 使用者在無意間形成偏見，迫害了特定族群；第二類屬於 AI 科技在應用上的倫理問題，這部分包括 AI 對隱私權的威脅、責任歸屬問題以及真假難辨的困境；第三類則是 AI 科技繼續發展對人類生存產生威脅的疑慮。

人工智慧的各種運用，既充滿希望，也潛藏危機，在人工智慧科技不斷發展的同時，也必須制定好相應的制度規範，這將是未來的重要課題。

關鍵字：人工智慧、倫理、偏見、隱私權

The Ethical Issues of Artificial Intelligence

Chang, Tung-Wen

Center of General Education, MinghsinUniversity of Science and Technology

Abstraction

The rapid development of artificial intelligence (AI) has brought convenience to our lives, but it has also caused various doubts and risks. We have to seriously think about the ethical issues caused by artificial intelligence.

*本文已於 2021.11.17 於明新科技大學共同教育學院舉辦之「通識教育創新與實踐教學學術研討會」中發表，感謝主持人蘭陽技術學院通識暨語文中心陳武強教授之建言，經修訂後重新投稿。另經明新學報編輯委員會之審稿後建議，讓論文的架構及論述更為完善，謹申謝忱。

**通訊作者：張東文（明新科技大學通識教育中心講師）地址：新竹縣新豐鄉新興路 1 號
Tel：03-5593142 轉 3453e-mail：tungwen@must.edu.tw

We can divide the ethical issues of AI into three categories. The first category is due to improper AI design, which creates prejudice against specific ethnic groups; The second category is privacy violations, the issue of responsibility,and the dilemma that is difficult to distinguish between true and false; The third category, it is an ethical issue that AI poses a threat to human survival.

The various applications of artificial intelligence are bot 經 h full of hope and hidden crisis. While artificial intelligence technology continues to develop, there should also be relevant regulations. This will be a important topic in the future.

Key words: Artificial Intelligence（AI）、ethics、prejudice、privacy

一、前言

人工智慧（AI）的研究、發展與應用，已是目前最受矚目的新興科技，我國政府和產業界也積極投入相關的研發與應用，以國發會提出的「亞洲・矽谷」計畫為例，將 AI 納入國家未來聚焦推動的關鍵議題；教育部更在各級學校推動人工智慧的基礎教育，各大學校院亦紛紛於計算機概論、電腦應用……等相關課程中融入人工智慧的介紹與應用，甚至特別開設「人工智慧」的專門課程，然這些課程內容多聚焦於對 AI 的基本認識、演算法、操作技術、或實際運用等，缺乏透過人文發展與倫理思維的角度，反思 AI 技術帶來的潛在疑慮與風險、以及對社會的變動與影響。

當人工智慧（AI）由研究室走向產業界，逐漸進入你我的生活，它的迅速發展已經為世界帶來重大變革，舉凡自駕車、人臉辨識、智慧助理……，人工智慧已是無處不在的生活科技。只是，一個新興科技的發展與普及，往往充滿希望也潛藏危機，雖為我們的生活帶來便利，但也會衝擊我們的文化、挑戰我們的倫理原則、改變我們的社會制度，甚至可能影響我們的生存。

因此，教育部在近年提出「人文社會與科技前瞻人才培育計畫」、「數位人文創新人才培育計畫」……，就是希望大學校院能由多元視角，藉著多學科、跨域、跨科際整合，培養能瞻遠、融整人文與跨科技的跨時代人才。因此，在科技突飛猛進的發展趨勢下，除了學習運用新的科技產品與工具外，更應該嘗試融入人文思考，才能讓我們在深刻的思辨中前瞻未來。

二、研究目的

身處知識爆炸的現代社會，新的知識與新的技術帶來新的職業需求，大學教育愈發著重專業知能的灌輸，這種完全的職業取向，所重視的是人力的培養，而不是人性的彰顯；研究者多年來任教於科技大學的通識中心，教授人文類通識課程及專業倫理等學科，深知相對於專業教育，通識教育希望培養學生能具備理性洞察、分析和融貫的能力，在廣博的知識基礎下，由知識的統一進而完成人格的統一，成為兼具人文素養，並能器識恢宏、見識不凡的全面發展之「全人」。

智慧科技及智慧服務是未來發展的主軸，「人工智慧（AI）」不僅是當代的科技趨勢，也已成為大學教育的重要內容，教育部及各大學均紛紛開設相關課程，亦積極鼓勵教師開設創新與跨域等多面向課程，除了教導學生能學習、運用人工智慧科技之知識與技能外，更要具備以人為本的思維、跨域的思維、創新的思維，以面對未來社會變遷與新科技的挑戰。

因為，人工智慧的應用已進入交通、醫療、銷售、新聞……等生活各層面，我們享受因科技工具的開發帶來的便利和舒適，但也不應忽略伴隨而來的爭議。翻開近幾年的新聞，Google Photos 和 Facebook 都曾經充滿歧視地把黑人標註為「大猩猩」、「靈長類動物」、有自動駕駛功能的特斯拉（Tesla）亦發生過多起源於電腦誤判的交通事故、還有 2020 年台灣、美國總統大選時到處傳播的假新聞……，這林林總總的爭議事件，讓我們無法漠視。因此，本研究試圖採質性研究法，透過文獻的蒐集與分析，以及社會現象的觀察，重新檢視這些不當現象所涉及的倫理問題，反思由於人工智慧設計的思維偏誤、不當使用、或者超越人類控制的過度發展……等，對既有道德規範的衝擊、並探討以往的倫理原則對人工智慧的適用性，試圖對人工智慧相關的倫理議題做一討論。

未來希望能將研究成果融入相關課程的教學中，除了增進個人的教學知能，更希望能幫助學生在認識人工智慧科技的同時，也能有更深入的人文思考，不僅能看見人工智慧的優勢，還能在遇到爭議、困境時能做出正確的判斷，培養學生在專業的學識外兼具宏觀的視野，成為兼具人文精神與科技專業的跨時代人才，以落實「全人教育」之理念。

三、人工智慧的概念與發展

打造一個具有人類智慧的機器人，這是以往科幻小說和電影中的情節，但是當 2017 年 DeepMind 公

司所開發的機器人 AlphaGo 打敗棋王柯潔後，在機器上再造人類智能已是夢想成真，人工智慧勢將成為改變世界的革命性科技。

「人工智慧」（artificial intelligence, AI）一詞的首次提出，要追溯到 1956 年在美國達特茅斯學院（Dartmouth College）所召開的「達特茅斯會議」。但事實上，被稱為「人工智慧之父」的馬文・明斯基（Marvin Minsky），早在電腦剛誕生不久的 1950 年，就造出了史上第一台的神經網絡計算機；「計算機之父」亞倫・圖靈（Alan Mathison Turing）也在同年發表一篇名為《計算機和智慧》的論文，為了解答「機器會思考嗎？」的問題，提出了「圖靈測試」（tuning test）。

從 1956 年出現「人工智慧」一詞到 1974 年，被視為為人工智慧的萌芽期，但直到 1980 年代各式演算法的出現，開展出一些 AI 的核心技術，造就 30 年後 AI 成為改變人類社會的最重要科技之契機，尤其 2012 年在「深度學習」上的突破，更將 AI 推上科技浪潮的尖端。

美國白宮科技政策辦公室(The White House Office of Science and Technology Policy, OSTP)，曾將人工智慧發展分為三個發展階段：第一階段係指 1980 年代的手工知識(handcrafted knowledge)階段，運用規則式專家系統(rule-based expert systems)，是以專家建立好的知識庫，來模擬人類的思維方式；第二階段是大約 2000 年發展出的機器學習(machine learning)，運用大量資料、大規模計算能力、更先進的學習技術，在圖像辨識、語音辨識……等方面有突破性的發展，也開啟了現今第三波的人工智慧發展，因為「深度學習」（deep learning）的演算法，在使用特定領域的大量資料後，不僅能做出最優化的決策，還會自我訓練與精進，運用找到的龐大資料，做出比人類更好的決定，目前的發展成果已相當驚人，但仍屬於「限制領域人工智慧」（Narrow AI）；第三波則聚焦在「強人工智慧」（Strong AI）的發展，企圖達成能勝任人類所有工作、擁有近乎人類智慧表現的人工智慧（李開復，2019）。就在短短的 60 多年間，原本存在科幻小說中的智慧機器人，如今已成為打敗世界棋王的存在。

人工智慧的發展，透過運用大量的資料庫、和愈來愈進步的演算法，其運算能力早已超越人類思考的限制，只是所謂的「智慧」究竟何指？似乎沒有一個確切的答案。從人類的角度來看，雖然常以「智商」（IQ）作為表示，但顯然無法真正的反應出何謂「智慧」，因為它包含了思考、理解、邏輯推理、學習、感知、意識……等各項能力的多層次的總和。

在「達特茅斯會議」中，是以「讓機器有模擬人類智慧的能力」，亦即「使機器能夠使用語言、形成

抽象概念和觀念，並期望能解決各種目前只有人類能解決的問題，並能自我改進」這樣的想法，作為人工智慧的發展方向（三宅陽一郎、森川幸人，2018）。事實上，不同的專家之間，對於人工智慧的定義有不同的定義，但從腦的結構來建構人工智慧的嘗試，是腦科學家和人工智慧研究者相信終有一天將能實現的構想，深度學習即是受到人類大腦運作的方式所啟發，此技術可說是受到人類思考時所仰賴的生理結構啟發而來，以人工的方式創造出像人腦一樣能察覺事情、在資料中形成特徵量、將現象予以模式化的電腦（松尾豐，2016；范雪萊，2020）。

目前世界各國對人工智慧的研究、投資都在加速、加鉅，人工智慧的進化看來似乎沒有終點，只是它能夠進化到什麼層次呢？林守德（2021）指出，這個問題確實存在著樂觀與悲觀的兩種取向，樂觀者認為人工智慧的終極發展，能成為造福世界的無私奉獻者，為人類社會帶來更多可能性；而悲觀者則擔心 AI 會不會對人類帶來始料未及的災難？亦即，人類和人工智慧的臨界點－「奇點」（singularity）會不會真的到來？人類與人工智慧的戰爭會不會真的展開？還是「人工智慧或許不是只單純超越人類，而是以與人類融合的形式持續進化」（三宅陽一郎、森川幸人，2018），隨著時代的演進，也許這個問題的答案將會愈來愈清晰。

四、人工智慧的應用

人工智慧科技在我們當前的生活中，早已十分普遍。隨著大數據資料庫的快速發展，配合整體通訊網路技術的提升，與其他硬體設備之強化，AI 科技已在各領域皆能加以運用。李開復（2019）認為，革命性的 AI 產品與應用，「大致上可以分成四波浪潮:1、互聯網 AI（Internet AI）；2、商用 AI（Business AI）；3、感知 AI（Perception AI）；4、自主 AI（Autonomous AI），每一波浪潮運用 AI 的不同能力、顛覆了不同的產業，但都讓 AI 深入我們的日常生活。」，其中前兩波浪潮，已經以幾乎令人無法察覺的方式，發生在我們的四周，改變了數位世界和金融世界。第三波感知 AI 正在數位化實體世界，透過語音、影像辨識等，革新我們和世界互動的方式；第四波自主 AI 對日常生活的影響更為深刻，自駕車、無人機、智慧機器人……，早已進入我們生活的各領域。

表一 四種人工智慧的應用模式（李開復，2019）

	四種人工智慧的應用模式
互聯網 AI	由大型網路公司主導，運用網路上蒐集的資料進行運算，AI 演算法做為推薦引擎，系統學會掌握個人的喜好，為我們精心挑選內容。
商用 AI	就是將 AI 技術運用在既有的各種商業決策中。很多產業有大量結構化資料，連結到特定意義的商業結果，非常適合用商用 AI 來做優化工作。
感知 AI	是指演算法結合了感測裝置後，能夠發揮眼睛或耳朵的功能，辨識語音、影像或物體，並將它延伸到我們的日常生活環境。
自主 AI	結合了前三波 AI，當機器能夠「看到」、「聽到」外在世界的變外，就可以開始「動起來」，也就是我們所想像的 AI 機器人。

人工智慧的應用，為我們的生活帶來很大的變革和好處，愈發精良的硬體設備，配合快速發展的網路通訊技術以及大數據資料庫，人工智慧在各領域都增進了產業的競爭力，不僅減少人力成本，也降低人為失誤，甚至還能協助產業突破以往的發展瓶頸。

以自駕車為例，根據世界衛生組織的統計，全世界每年死於車禍事故者高達上百萬人，其中大多來自於人為的疏忽、視線的侷限、以及疲勞駕駛……等，自動駕駛的功能比人更冷靜、更專注、亦無疲勞駕駛的問題，顯然更能降低車禍事故的機率；再者，AI 對於週遭環境資訊的捕捉，可以幫助我們避開危險傷害，進而拯救很多人的生命，尤其對無法自己駕駛的老年族群、殘障人士更是個福音。

另外，AI 在醫療上的運用，也愈來愈被看重，2017 年在權威學術期刊《自然》（Nature）上，有一篇研究指出，AI 對於皮膚癌的識別，已經達到皮膚科專科醫師的程度，甚至在某些測試中，它比人類醫生更敏感、更精準。（范雪萊，2020）事實上，AI 正在翻轉我們的醫療行為，除了前述的 AI 輔助診斷，還有從自主健康管理出發的穿戴式裝置，到透過基因檢測早期診斷預防疾病發生的精準醫療、在疫情蔓延時減少感染風險的線上問診、輔助醫護人力吃緊的智慧病房、在藥品資料庫中挖掘出有潛力的產品……，人工智慧正在醫療專業中以意想不到的方式迅速發展。

在金融業，「智能投顧」（Robo-Advisor）已是理財市場的新利器，以往銀行理專察看基金過去的績效來幫客戶選擇適合的基金標的，現在「機器人理財顧問」將大數據分析運用到財富管理領域，依據客戶的風險屬性實現自動的資產組合配置，還能不斷地追蹤市場動態，針對基金所持有之資產作持續分析與預測，隨時機動調整資產配置方案，甚至結合 AI 語音助理服務，客戶只要在手機就能透過語音對銀行帳

戶進行轉帳、買賣……等各式金融服務，使金融交易更方便、更即時。

舉凡購物、視聽娛樂、醫療、銷售、導覽、醫療、教育、交通、金融產業……等，人工智慧無所不在地被運用於各不同的產業與領域。想像著，從每天一早，我們在舒適的室溫中醒來，輕鬆地開著自駕車，用 Google Maps 預測路況、規劃路徑然後以最短的時程到達公司，按了指紋進到辦公室處理公文、回覆信件時讓電腦自動檢查郵件上的拼字，中午在無人商店用人臉支付買了食物，邊吃邊坐著欣賞由 Netflix 所推薦的節目……，AI 科技已無所不在地進入我們的生活。

AI 科技確實在我們的生活中發揮著愈來愈大的作用，然而各類爭議事件亦不斷浮出。Nick Bostrom 和 Eliezer Yudkowsky（2014）曾設想，透過機器來審核貸款資格，將可能存在種族歧視的問題，這樣的擔憂在之後的研究已被證實；還有各種來自於機器的誤判、隱私權的侵犯……等各式對人類社會倫理規範的挑戰，這都是在應用人工智慧科技時，不能忽視也不容忽視的問題。

五、人工智慧帶來的倫理爭議

當科技發展到一定程度，倫理問題必然伴隨而來，人工智慧的發展不僅沒有例外，甚至應該說它對人類倫理造成了前所未有的衝擊。早在 1950 年代，美國數學家維納（Norbert Wiener）就曾經在“The Human Use of Human Beings”一書中，探討人與機器的關係，陸續也有學者關心「機器倫理」的議題。尤其隨著機器越來越智能化，現代的人工智慧已經發展到能打敗人類的棋王，但也引發愈來愈多的擔憂。

對於 AI 科技帶來的倫理問題，研究者將它分為三大類來討論，第一類可歸為 AI 設計所導致的相關倫理問題；第二類屬於 AI 科技在應用上的倫理爭議；第三類則是 AI 科技繼續發展將可能威脅到人類生存的問題。

(一)、AI 設計導致的倫理問題

1、擴大既有的社會偏見

范雪萊（2020）提到，儘管人工智慧的應用已經深入我們的生活，但其實它並不完美，因為受制於演算法的限制，人工智慧系統有時會產生錯誤的結果，甚至擴大既有的社會偏見。

人工智慧系統是透過大量的資料訓練和演算法所架構而來，因此可以想見，當演算法的失誤就會造

成錯誤的判斷，大量的訓練資料如果是過時、不足或偏誤，也會帶來災難性的後果。

就演算法的設計觀之，每一種演算法的演算邏輯設定，都是來自於設計者，當設計者有意或無意地將個人喜好、偏見加諸於演算法或訓練過程中，那麼自然就可能導致系統性的偏見發生。2016 年時，微軟公司（Microsoft）與 Bing 開發的聊天機器人 Tay，才在社群網站上一天，就從接收到的語言中學會了歧視黑人的言論；麻省理工學院（MIT）的研究人員 Joy Buolamwini 發現，微軟（Micorsoft）、IBM 等知名科技公司所開發的人工智慧軟體，對於深色人種和女性的辨識，錯誤率特別高，其中確實存在著種族歧視和性別歧視；美國華盛頓大學和康乃爾大學的聯合研究亦指出，使用美國黑人用語的 Twitter 貼文，被 AI 偵測為仇恨言論的數量，相較於一般用語多了 2.2 倍，亦即看似客觀的演算法其實在不知不覺中強化了偏見（戴敏琪，2019）。

畢竟，目前的 AI 演算法，無法從發言者的語氣、身份、習慣、特定時空……等條件中，分辨出該發言到底是基於歧視？仇恨？還是玩笑、自我解嘲？或者其他的可能性？此種由於演算法設計缺陷所導致的歧視問題，很可能會使 AI 使用者在無意間迫害了特定族群，暗中造成了對種族、年齡、性別……等的歧視，或者選擇某些條件做為不良行為的預測，從而使具備該特徵者遭到拒絕或不公平的對待（加里・史密斯，2019），這種現象很難被察覺，但無疑地，AI 系統產生的偏見風險持續擴大，是極不合乎道德的。

雖然目前尚未有能消除 AI 偏見的的方法，但 IBM 已在實驗如何將人類的價值應用在 AI 的決策過程中，理想狀態是希望能打造出具備道德感、不帶偏見的 AI 系統。

2、演算法的失誤會造成錯誤的判斷

現今多數的人工智慧應用程式，採用類似人腦神經網路結構的深度學習機制，每個神經網路都連結大量的數據，每一次的計算是由資料中提取出計算後更適合的結果，但這樣的提取過程充滿不可預測性，有時連設計者都無法解釋自己是如何做出決定的，AI 演算法的「黑箱系統」（black box system）存在不確定性，就像 Siri 有時能給出正確的答案，有時給出的答案卻是荒謬的令人哭笑不得。

我們可以對 Siri 的答案一笑置之，但如果是人命關天的事呢？IBM 公司開發已被實際應用到醫院中的 AI 醫生 Watson，就曾在 2017 年被美國醫學媒體 STAT 拿到 IBM 內部的機密文件，顯示 Watson 在假設性的病例訓練中，可能會給出錯誤的診療判斷，開出錯誤的藥方，嚴重可能導致病患死亡的問題。人工智慧的不可預測性，令人擔憂一次失敗的演算即可能造成誤判，甚或帶來災難性的後果，也會重挫社

會大眾對人工智慧的信任感（范雪萊，2020）。

若就訓練資料觀之，理想中我們提供給 AI 的資料理當全面而完整，然而透過大數據或資料庫所蒐集而來的大量資料，即使齊全亦難保其中內容是否潛藏了存在於人類社會的偏見與歧視；更何況，當提供的訓練資料內容錯誤、過時、不夠完整而全面、或資料本身即存在著好惡偏見時，那麼這樣的資料勢必影響了 AI 的判斷。

再者，我們常過度相信 AI 龐大的計算能力，相信它可以幫助我們做出最佳或最合適的決策。然而，專家指出，人工智慧對於數據的挖掘演算所產生的推測模型，確實可以針對以往情況得到好的結果，但對於預測未來狀況確可能毫無作用（加里・史密斯，2019），因此當我們過度依賴它來做重要決策，反而可能帶來不可知的危機，甚至引發毀滅性的意外。（岡本裕一朗，2020）

范雪萊（2020）認為，由於 AI 設計的缺失，不僅無法如預期地為社會帶來公平、公正，反而可能比人類還不擅於中立、理性地做出影響人生的重大決定。演算法的穩健、可信、沒有會被利用的缺陷是一個重要的標準（Bostrom、Yudkowsky，2014），亦即歐盟的 HLEG 小組(The European Commission's High-level Expert Group on AI)（2019）提出，透過 AI 技術本身、設計者及組織、參與 AI 生命周期的社會技術系統，開發出值得信賴的人工智慧，是最重要也是最有價值的工作。

(二)、AI 科技在應用上的倫理問題

1、AI 對隱私權的威脅

隱私權的侵犯，是人工智慧最常被提及的倫理爭議問題。

隱私權是一項基於人性尊嚴與個人主體性的基本權利，保障個人的生活、個人的資訊、及秘密免受他人非法的侵擾而保有自主性，保護個人隱私已是現代社會的共識。然而，以大數據做為人工智慧資料訓練的基礎，卻使個人的隱私權遭受到前所未有的威脅。

在人工智慧的各種應用上，資訊的需要量比以往都要來得更為龐大，而愈來愈先進的各種數據分析系統，也更能輕易地蒐集到我們每一個人的個人資料，除了基本的姓名、年齡、地址、電話外，甚至連婚姻狀況、人際關係、就醫記錄、宗教信仰……等各種各類的訊息，都被巨細靡遺的記錄下來。只是，這樣的資訊蒐集手段和利用方式，很可能遊走在法律邊緣，更造成對個人隱私權和資訊自主權的侵害。

再者，這些資料的運用和儲存，為了讓使用者能更即時地取得所需的資料和數據，這些資料常儲存於相互連接的雲端系統中，加上無線通訊、智慧型手機等設備的普及，就可以輕易地將我們的資訊、所在位置、通信內容……等，暴露給任何有心的政府、私人公司，甚至是駭客或詐騙集團。

早在 1948 年，英國作家喬治・歐威爾（George Orwell，1903-0950）寫下他充滿政治諷喻的預言小說《1984》，書中的”Big brother is watching you.”，預言了人將毫無隱私地完全暴露在政府的監控之下，以往這被視為是歐威爾的多慮，然而人工智慧的發展，已讓我們陷入《1984》一書所預言的困境之中。人工智慧對隱私權侵犯最主要的疑慮，來自在公共場所的人臉辨識系統的運用，結合到處佈設的監視器，使人的一舉一動無所遁形，中國政府即以這樣的技術布建了「天網」、「社會信用評分系統」作為威權統治的工具，使人民活在毫無隱私的監控下，人性尊嚴蕩然無存。

2、責任歸屬難題

隨著人工智慧應用領域愈來愈廣，產品也逐漸普及，然而相較於過往的其他科技，AI 最令人擔憂的問題就是：當機器出錯時，誰該負責？

當人類使用 AI 智慧的產品，其實正意味著我們相信這個系統具有超越人類的優勢或分析決策能力，就如目前的自駕車，它具備有比人更冷靜、更專注、不會疲勞的優點，更能降低車禍事故的機率。然而，當面對緊急的道德兩難困境時，自駕車會如何抉擇？一般人駕駛，如果遇到突發狀況，會依自己過往的經驗，或當時本能的直覺與判斷做出反應，然而透過演算法預設了參數的自駕車，在面對功利論或義務論的困境時，它會如何取捨？還是從大數據中搜尋是否有類似的案例來進行類推？或者只是隨機選取一種方案？話說回來，那麼我們到底應該基於什麼樣的倫理原則來設定自駕車的操作參數呢？

再者，如果自駕車在行駛中因為系統失控或故障，發生事故造成傷亡，那麼又應該由誰來承擔道德與法律責任呢？是負責編寫程式的軟體工程師？還是製造商？又或者是使用者？

同樣的問題也發生在 AI 的醫療系統運作上，運用 AI 科技的達文西手術系統，它需要醫師的介入與操作，但如果手術時由於 AI 應用之決策，因演算法產生不可解釋性與不可預見性，該系統於手術時發生故障或誤判，使病人的生命受到威脅，那麼負責操作的醫師，很可能無法解釋 AI 為何會有這樣的運作，也無法舉證自己的無辜，那麼手術失敗的責任該由醫師來負責嗎？還是如 Nick Bostrom 和 Eliezer Yudkowsky（2014）所言，當一個人工智慧系統在指定的任務中失敗時，為免個人承受所有的指責，可能

更傾向於將所有的責任歸咎於人工智慧系統吧！

綜上可見，在一些複雜的人機環境系統中，事故的責任歸屬恐怕將難以界定，不管人或機器都是當時系統中的一部分，負責完成系統中一部分的功能，但是整體而言，還是可能造成無法挽回的錯誤，更有不知該由誰承擔的責任問題。

3、真假難辨的隱憂

2018 年時，流傳著一則前美國總統歐巴馬（Barack Obama）的影片，片中歐巴馬侃侃而談：「我們正進入一個新時代，敵人可以讓某人隨時說任何話，即使他們從來沒有這樣說過。」，事實上，歐巴馬就未曾說過這些話，這是美國一家網路媒體公司，利用電腦軟體和人工智慧 APP 所創造出來的影片，由於太過擬真，一般人根本無從分辨真假。

人工智慧應用領域愈來愈廣，不僅可以輕易地找出物件的特徵，也能運用特徵合成新的聲音、影像。以往我們在電影中，看到電腦高手運用程式模擬出人類聲音，如今一個名叫 Lyrebird 的網站，宣稱只要錄音一分鐘，系統就能分析並複製出你的聲音，並念出你從未說過的句子；而 deepfakes 技術（深偽技術），幾乎已是「變臉」的代名詞，它可以將人、環境、物體經由程式處理而改變聲音和影像。這些程式原可以用來製造很多娛樂效果，或者創造諸如虛擬模特兒、客製化商品……等做為行銷上的利器，但如今卻也可能成為以假亂真的社會災難。

以往我們相信「眼見為憑」，然而如今假新聞、加工影片的散布已是防不勝防的社會現象，2016 年的美國總統大選和 2017 年的英國大選，均發現有人透過 AI 科技散播不實訊息來影響投票行為（范雪萊，2020）。美國政界也曾擔心這些擬真技術，被有心人利用將可能威脅國家安全或介入選舉，各國領導人也紛紛表態要向假新聞宣戰（曾惠敏，2018），今（2021）年 10 月，台灣也發生使用 deepfakes 技術將情色影片主角的臉部被換上名人的社會事件，只是這種臉部替換不止名人遭殃，其實一般大眾也可能成為受害者；英國資料倫理與創新中心（Centre for Data Ethics and Innovation; CDEI，2019）就擔心，deepfakes 技術造成真假難辨，如被詐騙集團、極端份子……等利用，不僅一般人可能上當受騙或者因照片被誤植而造成心理創傷，對整個社會而言更有極大的風險，譬如它可能影響犯罪調查的影像證據，甚至可能引發嚴重的政治事件，戕害我們的民主體制。

(三)、AI 科技發展可能對人類生存造成衝擊

1、AI 造成人力排擠的效應

隨著生產的智慧化，可預見的未來，智慧機器人將被大規模應用，以提高生產效率，追逐利潤的資本家必然愈來愈傾向以人工智慧取代現有的人力，讓許多工作將由自動化完成。擁有、甚至超越人類某部分智能的機器，正在取代人類從事那些勞累、重複、單調的工作，或者代替人類在有毒、有害的危險環境中工作。

牛津大學 2013 年的研究曾指出，未來十年內有 48%的工作機會將會被機器取代；麥肯錫全球研究院認為 2030 年全球將有 8 億人的工作被機器人和自動化取代;日本學者井上智洋也預測，2045 年勞工將完全被機器人取代，商品全部都是由無人工廠所生產，勞工將失去收入來源（井上智洋，2018）。這將對以往的職場帶來翻天覆地的改變，形成人力被機器排擠的困境，勢將導致失業率升高、勞動者被剝削，最後被社會邊緣化的處境。

即使距離專家預測勞工完全被取代的時間還有 20 多年，但勞工受制於 AI 的困境，卻已是現在進行式，全球最大的電子商務企業亞馬遜（Amazon）公司，以往曾經傳出以 AI 系統追蹤物流部門員工的工作效率，而裁掉數百名員工；就在今年（2021）8 月，一家俄羅斯的支付服務公司 Xsolla，也以 AI 演算法考察員工的數位足跡（Digital Footprint），分析員工的聊天記錄和程式活動，將所算出的「沒有生產力的員工」直接資遣，這個消息在網路上引發爭議，除了窺探員工隱私被認為不道德外，也開始令人擔憂未來在職場工作，是否得看 AI 的臉色來行事？勞工會不會進而成為演算法的奴隸？

2、加劇社會的不公平現象

人們除了擔憂隨著人工智慧的發展，會搶走人類的工作外，伴隨而來的必然是社會不公平現象的加劇。一如有人提到：「當機械能夠自己生產出財富時，人類就失去工作了。不需要勞工、完全自動化的企業，只有股東能獲得財富。這個時候，世界上的人類就分為兩種，一種為股東，一種是非股東。」（井上智洋，2018）。亦即，如果 AI 技術的發展成熟，採用新科技幾乎是有錢人的專屬特權，而最終生活因此變得更富裕的也是少數掌握資本的有錢人，社會的貧富差距必然愈形擴大。

再者，當 AI 的技術應用愈來愈廣泛，未來可能開發出不用人類指導就可以自己學習、而且能包辦各類事務的全方位機器人，即使人人都希望自己能成為利用這種 AI 科技的人，然而新興科技的高昂代價，

恐怕不是人人皆能負擔得起，從而形成社會競爭的另一種不公平現象。

最後，AI 的運算需要大數據，不僅政府、醫療系統或保險公司擁有大量的人民隱私數據，像臉書、谷歌、IBM……等私人大企業也不斷地在蒐集我們的個人信息，他們可以大量的監視消費行為、手機使用習慣、駕駛路線圖、電腦的數位足跡……以便能預測我們的行動，甚至操控我們的行為，一如阿里巴巴創辦人馬雲所說：「得數據者得天下」，然而，渺小的個人只能坐視大企業對個人隱私權的侵犯，猶如小蝦米對抗大鯨魚，在這場不公平的對抗中愈發地無能為力。

3、超智慧 AI 將衝擊人的本質和固有倫常

首先，有研究者認為人機交互、人機結合成為未來趨勢（李開復，2019），當人工智慧科技和生物技術結合的發展，使得我們對自然人本身的認知發生巨大改變，人的自然身體面臨著 AI 的「改造」，人原有的思維、創造力、情感……是否亦被機器人習得？機器人具有思維能力已成為共識，發展具有自主意識的智慧機器人已成為可能，那麼人類的本質、與智慧機器人的關係是否面臨著挑戰？

當智慧機器人的能不斷自我學習，使得智慧愈來愈高甚至超越人類，那麼我們還要讓他們不斷地繼續增長智慧嗎？人類有辦法了解它們嗎？這樣高智慧的「物種」，它們該享有那些權利？如果它們有權利，我們還能像奴隸般地指揮它們做事嗎？如果它們有權利，是否可以自主而不接受我們的指令、不盡義務嗎？那麼，誰該為它們的行為負責呢？

Nick Bostrom 和 Eliezer Yudkowsky（2014）和 Ryan（2020）都提到，人們很容易將人工智慧擬人化，而試圖將人類的道德活動加諸其上，而這將會衍生出更多的倫理問題，因為當人們受到嚴格的道德約束，意謂著人們也該享有相對的合法利益與福祉；然而，人們普遍地都認為超智慧 AI 即使未來發展到具有感知力（sentience）或相當的智能（sapience），仍舊不過是一系列可任意由我們改變、刪除、終止的電腦程式，那麼我們如何能要求這樣的 AI 要符合人類的道德規範，是否合理呢？

再者，我們相信人是情感的動物，情緒是重要而獨特的人類生命經驗，然而目前科學界和產業界正努力為人工智慧科技，灌注各式各樣的能力，希望透過「情感運算」（affective computing），能理解、複製、回應人類的情感反應，甚至還可能有親身體驗的情緒（理查．楊克，2017）。因此有研究預估到 2050 年人形機器人將變得和真人一樣，不僅有精緻的五官和外貌、柔潤的膚觸、聽話溫順的性情，一如電影「變人」（Bicentennial Man）的場景，它們能扮演各種角色，例如：褓姆、寵物、情人……等，以各種不

同身份進入家庭。那麼，人與智慧機器人之間會不會產生感情？會不會產生利益糾紛？會不會對人倫關係造成衝擊？會不會對傳統的家庭結構產生改變？

4、人類的生存威脅

人工智慧的不斷發展，究竟會達到什麼程度呢？當處理的數據愈來愈龐雜，系統愈來愈進化，與人類智慧相當 AI 是否會成為現實？代表人工智慧與人類智慧融合時間點，亦即人工智慧與人類相互存在形式出現本質上改變的「奇點」（singularity）會不會真的到來？（三宅陽一郎、森川幸人，2018）

有百分之九十的 AI 領域專家，預測在 2075 年以前將會出現具有與人類同等智力的 AI 機器人，甚至極有可能的在 21 世紀結束前就能見證科技奇點的來臨，人工智慧超越人類，「超智慧 AI」（superintelligent AI）將會崛起（范雪萊，2020），當這一天來臨，人類文明會有什麼天翻地覆的改變？還是就此覆滅？

當超智慧 AI 出現，它們的智慧超越人類，那麼以人類的智商還有辦法了解這些智慧機器人嗎？這些機器人還願意為人類工作嗎？它們的行為該由誰來負責任呢？我們能要求或教育它們應有的道德義務嗎？公平、正義、道德的概念，它們也能演算出來嗎？或者說，它們會不會因為人類的「蠢笨」、懶惰、和殘暴，而決定毀滅人類呢？

著名的物理學家史蒂芬．霍金（Stephen Hawking）2014 年接受 BBC 電視台訪問時，曾說過：「總有一天，自律化人工智慧出現後，說不定將用迅雷不及掩耳的速度開始自我改造。受限於生物演化速度的人類應該毫無能力抵抗，最終將被超越。」（岡本裕一朗，2020）；事實上，特斯拉汽車（Tesla Motors）執行長伊隆．馬斯克（Elon Musk）、現任的 Google 公司總監雷蒙．庫茲韋爾（Raymond Kurzweil）、英國劍橋大學名譽教授馬丁．里斯（Martin Rees）……，都曾發表過類似的悲觀論調，這類人工智慧超越人類，進而奴役人類、毀滅人類的，就像電影《魔鬼終結者》（The Terminator）、《駭客任務》（The Matrix）……等電影的劇情一樣，AI 取代了人類的主宰地位，人類反成為智慧機器人的附庸，甚至被消滅，這可說是人類一直以來的擔憂，也可能是人類最大的生存威脅。

六、結語

AI 科技為人類帶來創新與優勢，但也帶來潛在疑慮與風險，為了考量 AI 可能引發對人文社會的變動與影響，國內外早已有許多針對法律、安全與倫理的相關研究。李崇僖（2020）認為，人工智慧的倫

理議題具有其特殊性，與一般科技的倫理有本質上的差異；因為對於一般科技的倫理議題，主要聚焦於使用者的濫用或誤用，然而人工智慧的有其「黑盒子效應」(black-box effect)，其產生的倫理爭議並非開發者或使用者所能完全控制。

目前積極發展 AI 的國家或地區，都逐漸由政府組織研究相關的基本準則，例如：歐盟在 2018 年的「Ethics Guidelines for Trustworthy AI（值得信賴的人工智慧倫理準則）」提出了 AI 的七大倫理原則、2020 年經由各界回饋後再提出「White Paper on Artificial Intelligence（人工智慧白皮書）」、日本「び人間中心の AI 社会原則(以人為本 AI 社會原則)」、OECD (經濟合作暨發展組織)「Principles on Artificial Intelligence (AI 準則)」及美國 IEEE (Institute of Electrical and Electronics Engineers)「Ethically Aligned Design-Version II (AI 道德設計準則)」等。2019 年 9 月，我國科技部亦邀集各領域的專家學者，共同發布「人工智慧科研發展指引」，亦有立委呼籲政府應儘速三讀通過「人工智慧基本法」，希望能夠樹立我國人工智慧發展的基本倫理原則。聯合國教科文組織（UNESCO）也在 2020 年 3 月組成了專家小組，目的就是為了為全球起草一份人工智慧的倫理建議書。(UNESCO，2020)

人工智慧將往什麼方向發展？會取代人類嗎？它將引導人類走向何方？許我們一個什麼樣的未來？這些答案將取決於我們怎麼使用這種科技。有句廣告詞說：「科技始終來自於人性」，日新月異的科技發展不單單只是技術面的變革，也在持續性地改變人類各方面的生活，帶來便利的同時也可能帶來災難，我們希望未來人工智慧的發展，能像歐盟的倫理準則中所昭示的一樣，要能安全可靠、重視隱私、透明、多元公平、尊重、負責，亦即如聯合國教科文組織（UNESCO）所言，我們需要的是以人為本的人工智慧，才能在促進人權與維護人性尊嚴的基礎下，幫助人類不斷成長、超越，真正為全人類帶來福祉。

參考文獻：

三宅陽一郎、森川幸人（2018）。**從人到人工智慧，破解 AI 革命的 68 個核心概念**。台北市：臉譜出版。

井上智洋（2018）。**2030 僱用大崩壞**。新北市：大牌出版。

加里・史密斯（Gary Smith）(2019)。**錯覺**。中國北京市：中信出版集團。

尼可拉斯・卡爾（Nicholas Carr）(2016)。**被科技綁架的世界**。台北市：行人文化實驗室。

托比・沃爾許（2019）。**2062 人工智慧創造的世界**。台北市：經濟新潮社。

自由時報（2019.4.5）。AI 竟然也會種族歧視科學家呼籲科技公司改善。自由時報，檢自：https://www.ltn.com.tw/

李崇僖（2020）。**人工智慧競爭與法則**。台北市：翰蘆圖書。

李開復（2019）。**AI 新世界**。台北市：天下文化。

林守德（2021）。當人類智慧碰到人工智慧。載於林守德、高涌泉（主編），**智慧新世界**（頁 1-32）。台北市：三民。

岡本裕一朗（2020）。**當人工智慧懂哲學**。新北市：楓葉社文化事業。

松尾豐（2016）。**了解人工智慧的第一本書**。台北市：經濟新潮社。

紀品志（2016.3.25）。聊天機器人 Tay 一天學會「種族歧視」，微軟緊急消音，**數位時代**，檢自：https://www.bnext.com.tw/

范雪萊（Shelly Fan）（2020）。**AI 可不可以當總統或法官？**。台北市：臉譜出版。

班・格林（Ben Green）（2020）。**被科技綁架的智慧城市**。台北市：行人文化實驗室。

理查・楊克（Richard Yonck）（2017）。**情感運算革命**。台北市：商周出版。

張麗卿（2021.3.3）。AI 倫理準則及其對臺灣法制的影響，**臺灣人工智慧行動網**，檢自：https://ai.iias.sinica.edu.tw/ai-ethics-guidelines-in-taiwan/

陳耕彥（2021.8.9）。金流公司 Xsolla 通過 AI 分析裁員 150 人 CEO：不認真工作就滾蛋！**ETtoday 新聞雲**，檢自：https://www.ettoday.net/

郭一璞、閆耕（2018.7.30）。IBM Watson 爆出致命 Bug：開錯藥給病人，醫死人算誰的責任？**科技橘報**，檢自：https://buzzorange.com/techorange/

富蘭克林（Daniel Franklin）（主編）（2018）。**巨科技：解碼未來三十年的科技社會大趨勢**。台北市：遠見天下文化。

曾惠敏（2018.7.18）。AI 擬真 影片人物真假難辨。**公視新聞網**，檢自：https://news.pts.org.tw/

漢娜・弗萊（2019）。**打開演算法黑箱**。台北市：臉譜出版。

戴敏琪（2019.8.16）。AI 演算法帶有偏見!新研究：黑人用語較易被 AI 視為仇恨言論。**新頭殼**，檢自：https://newtalk.tw/

Boddington,P., Millican, P. & Wooldridge, M. (2017) Minds and Machines Special Issue: Ethics and Artificial Intelligence. *Minds & Machines*, 檢自： https://doi.org/10.1007/s11023-017-9449-y

Bostrom, N. & Yudkowsky, E. (2014）.The Ethics of Artificial Intelligence. *The Cambridge Handbook of Artificial Intelligence*, 1, 316-334

Centre for Data Ethics and Innovation（CDEI）（2019）. Snapshop Paper – Deepfakes and Audiovisual Disinformation. 檢自 https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. 檢自 https://doi.org/10.1007/s10676-018-9450-z

HLEG AI.（2019）. Ethics guidelines for trustworthy AI. 檢自 https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

Pavaloiu, A. & Kose, U. (2017) . Ethical Artificial Intelligence - An Open Question. *Journal of Multidisciplinary Developments*. 2(2), 15-27.

Ryan, M.（2020）. In AI We Trust : Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, Vol.26(5), 2749-2767

Thomsen, K.（2019）. Ethics for Artificial Intelligence, Ethics for All. *Paladyn*,Vol.10(1), 359-363

Vieweg, S.H.（2021）. *AI for the Good : Artificial Intelligence and Ethics*. New York, NY: Springer.

Zou, J. & Schiebinger, L.（2018.7.18）. AI can be sexist and racist – it's time to make it fair. *Nature*, 檢自： https://www.nature.com/

UNESCO（2020）. UNESCE appoints international expert group to draft global recommendation on the ethics of AI. 檢自 https://en.unesco.org/news/unesco-appoints-international-expert-group-draft-global-recommendation-ethics-ai