

# 文學、藝術與 AI 跨領域對談

[ 外文學門跨領域對談 ] AI、倫理與科幻創作<sup>#</sup>

時間：107 年 5 月 25 日（五）13:20-15:00

地點：國立臺灣大學博理館 201 會議室

與談人：陳瑞麟（國立中正大學哲學系教授）

林建光（國立中興大學外國語文學系副教授）

楊乃女（國立高雄師範大學英語學系副教授）

主持人：陳國榮（國立中正大學外國語文學系教授兼文學院院長）

記錄：張澄清（科技部人文社會科學研究中心博士後研究員）

李忠達（科技部人文社會科學研究中心博士後研究員）

胡雲薇（科技部人文社會科學研究中心博士後研究員）

## 一、國立中正大學哲學系陳瑞麟教授

陳瑞麟教授長年關注科學哲學與科幻創作中內含的倫理議題。在本場討論中，陳教授指出 AI 技術突飛猛進之際，人們最憂心的問題不外乎原本屬於人類智慧的領域被機器化，這種原屬科幻的想像已逐漸實現在我們的生活中，因此，人類與 AI 之間的倫理關係成為不可避免的課題，而過去的科幻創作與想像為我們提供很多借鏡。

陳教授首先簡短回顧人類創造擬人物的悠久歷史，猶太人傳說中的魔像（Golem）、日本神話中的「式神」、海地巫術的巫毒娃娃，到煉金術創造的人工生命（Homunculus），這些元素至今仍不斷刺激出各種科幻創作。自從 17 世紀機械技術大突破以來，不僅會下棋的「恐怖的土耳其人」、融合解剖學、電學的「科學怪人」，對智能機器的渴望正在逐漸成形；直到現在擊敗世界西洋棋王的深藍（Deep Blue）和大敗中韓圍棋高手的 Alpha Go，AI 的功能更為強大，人們對 AI 的想像也更具體化。無論是堪比老大哥的街道臉部辨識系統、在美國舊金山上路的自動駕駛汽車，過去曾被認為是科幻想像的技術已經化為現實，這時

---

<sup>#</sup> 本文由科技部人文社會科學研究中心博士後研究員記錄整理，經本場次主持人及 3 位與談人審訂。

問題就產生了：能夠靠深度學習自我進化的 AI，能力超越人類的一天是不是即將到來？

要理解 AI 與人類的關係，陳教授認為應該先精確理解 AI 和 robot 的定義。工程學對以「可移動性」和「自主性」來定義「機器人」。但，自主性也是人類的特質，特別涉及到自由意志（free will）的問題，因此更為關鍵。然而，可移動／不可移動、自主／不自主始終沒有辦法劃出明確的界線，技術的進步——特別是理性與思想的可計算與機械化——更不停地使既有的界線變得模糊。對於 AI 為什麼能運作？AI 對社會將造成什麼樣的影響？人類似乎處在無知的狀態，這種「無知」會帶來高風險。面對科技風險的問題，陳教授認為：我們不只要問 AI 是否可以促進「效益」、「慈善」、「自由」等價值；更應該特別重視「人道」、「公正與豁免」和「永續」等價值原則。

在探討這些議題時，科幻創作可以帶來很大的幫助。陳教授將目前科幻作品對 AI 的想像分成五種類型的腳本。第一種是 AI 與人類發生衝突，比如電影《駭客任務》利用「桶中之腦」的想像，設定出電腦統治人類，將人類意識封鎖在虛擬世界，而將肉體製成發電裝置的未來世界；《魔鬼終結者》也是機器人消滅人類，人類試圖抵抗的劇本。第二種是 AI 自發性的產生道德意識，比如《駭客任務》第二、第三部中的角色「先知」，就是願意同情弱者、權衡長遠利害、轉而幫助人類的 AI。第三種是人類為 AI 設定好道德規範。比如 Asimov 著名的小說中，人類事先為具自我意識的機器人訂下「機器人三大定律」，即（1）機器人不得傷害人類，亦不得坐視人類受到傷害；（2）在不違反第一法則下，機器人必須服從人類命令；（3）在不違反第一與第二法則下，機器人必須保護自己。然而



圖一：陳瑞麟教授將科幻作品對AI的想像分成五種類型探討

在實際應用上，也會出現這三項法則無法處理的矛盾。第四種是從 AI 的立場出發，探討 AI 產生自我意識之後的處境。比如《銀翼殺手》中人造人對自身命運的反抗；《變人》中機器人對於擁有肉身的渴望。第五種是樂觀的看待人類與 AI 和諧共存、情感共鳴的腳本。比如《大英雄天團》中的 Baymax 等等。陳教授認為，這幾種腳本是一套公式，提供科幻創作的劇情骨架；同時也是一種倫理觀點的想像，讓科幻作品有血有肉，刺激人們的思考、引起觀眾的共鳴；更能為我們在未來規範 AI 與人類的關係提供參考。

## 二、國立中興大學外國語文學系林建光副教授

林建光副教授以烏托邦 (Utopia) 和反烏托邦 (Dystopia) 的概念來解讀人類對 AI 未來發展的兩種想像。林教授以 Luxgen 於 2009 年推出的廣告為例，說明第一種烏托邦式的想像中，人類和 AI 固然有差異存在，但是雙方共同合作，進而達到人機一體、水乳交融的境界。然而，廣告中以女性形象出現的汽車 AI，強調著機械溫馴、可掌控的性別化意義，並反映出車主凌駕於車子的優越性；而身為人類的男車主與 AI 接吻，模糊化人類與機械的界線，仍然可能帶給人不適感，並擔憂人類喪失其主體性。

在反烏托邦的想像中，人類對 AI 的恐懼感更加強化。比如《機械公敵》中電腦企圖控制人類，以求達到延續此一物種的目的；《2001 太空漫遊》中人工智慧 Hal 不願意被關機，拒絕打開艙門讓太空人回船，並且殺光其他所有冬眠中



圖二：林建光副教授從電影《機械公敵》分析人類對科技恐懼的心理

的人類。這些故事刺激出的恐怖感在於 AI 對人類並沒有抱持恨意，只是透過絕對理性的計算，以違背人類情感和道德法則的手段達成目的。但這恰好是人類珍重視之，不願拋棄的信仰所在。

林副教授也指出理性、思考、情感、靈魂……人類不停的試圖闡釋自己與他者的差異，證明自己的獨一無二。這種觀點可稱為「擁有論」。擁有論者將人類想像為類似容器的主體，主體（身體）內部包含著某些非身體的特質（上述的理性、思考、情感、靈魂等）。弔詭的是，「擁有論」的合理性基礎卻是建立在缺乏內在性的他者（例如機器人、動物、怪物等）之上。也就是說，人類乃是藉由非人、他者的欠缺來建構或想像人類的存有。換言之，擁有論乃是藉由他者的「無」方能自我肯定。這令人反思，我們能不能用一種非擁有論、非本質性的觀點來思考 AI 是什麼這類型的問題？目前大部分科幻作品都傾向維持人類與 AI 之間的固定界線，這又讓我們不禁擔心：我們真的準備好面對何謂人、何謂 AI 等這些問題了嗎？

### 三、國立高雄師範大學英語學系楊乃女副教授

楊乃女副教授從《格列佛遊記》來反省歷史人類對科技進步產生的不信任，以及對藝術、道德、人文精神喪失的憂慮。《格列佛遊記》有強烈的反科技傾向，它描述的飄浮之島（Laputa）上，住著一群熱愛抽象數學，把牛排切成長菱形、把布丁切成圓柱體的怪異學者。這群科學家使用一種寫作機器，可以隨機組成破碎的語言，作者還借格列佛之口諷刺他們「妄想以科技取代藝術」。



圖三：楊乃女副教授從《格列佛遊記》反省科技發展對人類的影響

不過，這些科技在現代已經成真了。Wordsmith 是 Automated Insights 與美聯社合作的自動撰寫軟體，並且已經寫出超過 10 億篇的報導。書寫曾是人類最重要的技術突破，但隨著書寫將記憶資訊化，進而數學化、工業化，人類的身體功能開始大量外包給機械工具，人類意識和大腦神經系統外部化，開始倚賴 AI。現代人大量使用電腦、智慧手機和網路，已經開始改變人類的思維方式和意識狀態。

因此，楊副教授以「奇點」稱呼這個時刻，當人類與科技的關係遭到顛覆，AI 超越人類的時刻即將來到。人類與 AI 將成為離不開彼此的「怨偶」。人類在此時面臨的三個挑戰是：(1) 人類失去「裸眼」，永遠必須透過科技裝置來看世界，這可能導致自我逐漸消失的危機；(2) AI 優於人類的智慧，運算能力遠遠超乎人類大腦所及，但又缺乏人的道德判斷力；(3) 人類的去中心化，過去以地球、以人類、以理性意識為中心的世界觀被哥白尼、達爾文、佛洛伊德打破，現在又將被圖靈打破以為人優於機械的優越感。

最後楊副教授提出，在現代世界中高科技與資本主義的關係越來越密不可分，當我們觀察人與科技的共同演化，思考人類與 AI 的倫理問題時，也必須反省人類、科技與資本主義三角之間的不平衡之關係。