

人工智慧與人機傳播：傳播的擴充與弔詭

林思平

摘要

本文從工具理論切入，耙梳人類與機器的關係，再經由人工智慧演進的歷程、人機傳播發展的概念，思考人工智慧進程對於人機傳播所代表的意義，包括人工智慧演進對於傳播概念的擴充、人機傳播關係中倫理層次的弔詭。本文目的在於藉由這些人工智慧進程的思考，省察人機傳播的本質，與其意義產製的可能方向。

- ◎ 關鍵字：人工智慧、人機傳播、人類傳播、工具理論、機器學習、海德格、基德勒
- ◎ 本文作者林思平為世新大學新聞系教授。
- ◎ 聯絡方式：Email：splin@mail.shu.edu.tw；通訊處：116 台北市木柵路一段17巷1號。
- ◎ 本文作者在此衷心感謝兩位匿名審查者者與編委會對論文提供的寶貴意見。
- ◎ 收稿日期：2020/10/14 接受日期：2021/03/21

Artificial Intelligence and Human-Machine Communication: Expansion and Paradoxes in Communication

Szuping Lin

Abstract

Starting from the concept of instrumental theory, this essay first examines the relationships between human and machine. And through the discussions of the developments of artificial intelligence, and the progression of human communication and human-machine communication, this essay further deliberates upon how the evolution of artificial intelligence exerts influence on human-machine communication, including how artificial intelligence expands the meaning of communication and how paradoxes exist in the human-machine relation within the realm of artificial intelligence. With the reflection upon the advancement of artificial intelligence, this essay investigates the nature of human-machine communication and the possible directions of its meaning-making process.

- ⊙ Keywords: artificial intelligence, human-machine communication, human communication, instrumental theory, machine learning. Heidegger, Kittler
- ⊙ Szuping Lin is a Professor in the Department of Journalism at Shih Hsin Univerasity
- ⊙ Corresponding author: splin@mail.shu.edu.tw; address: No. 1, Ln. 17, Sec. 1, Muzha Rd., Wenshan Dist., Taipei City 116, Taiwan, R.O.C.
- ⊙ Received: 2020/10/14 Accepted: 2021/03/21

壹、研究緣起

機器和人工智慧與我們同在。當今在關於訊息與傳播的場域中，人類與機器、機器與人類、機器與機器之間的訊息傳遞，扮演著益發重要的位置。21 世紀的人類生活與機器緊密連結，而作為其中機器技術代表物的人工智慧更是與人類生活交織。人類的思想行為決定了人工智慧的發展，而人工智慧發展進一步影響人類的生活軌跡。本文從工具主義切入，耙梳人類與機器技術的關係，再經由人工智慧演進的歷程、人機傳播發展的概念，思考人工智慧進程對於人機傳播所代表的意義，包括人工智慧演進對於傳播概念的擴充、人機傳播關係中倫理層次的弔詭。本文目的在於藉由這些人工智慧進程的思考，省察人機傳播的本質，與其意義產製的可能方向。

貳、關於技術的工具主義

如 Gunkel (2019) 指出，討論機器技術之前，必須瞭解技術的工具理論。機器的出現緣起於技術 (technology) 的產生。海德格 (Martin Heidegger) 指出人類普遍概念中關於技術 (technology) 兩個定義：技術是達到目的之方法 (an means to an end)、技術是一種人類活動 (a human activity)。這是海德格所指一般概念中技術達成任務的工具面向、和作為人類活動的人類學面向。此人類中心的技術概念，揭示了人機關係中，機器作為工具事物的概念源頭：事物作為「技術」與「設備」存在的本體論地位，在於人類將它用於某種特殊的用途。事物的存在，在於他們對於人類所設定的目標是有用的。事物的關鍵並非其原始初生的狀態，而在於他們如何被利用；事物被利用，成為人類生活的設備。樹木須能作為房屋建材，石頭須能作為防禦武器。人類在乎的重點不是「工具」本身，而是「工具」所能完成的人類「工作」。這是海德格在《關於技術的問題》文中所指出人類普遍概念中的「技術工具論」。這是技術的工具理論：技術是工具，此工具為了服務使用者而準備 (Heidegger, 1977; Gunkel, 2019; Feenberg, 1991)。此處工具被認為是「中立的」，本身並不具有價值內涵；技術性事物並不是以本身被評價，而是在人類設計者將它加以運用的情境下被評價。也因為如此，不管一個技術裝置是多麼複雜或具有自動性質，它不會被視為應該對經由

其技術裝置而產生之行為負責——人類永遠是唯一的「道德行為人」(moral agent) (Gunkel, 2019; Hall, 2001)。在工具理論中，技術本身不是目的、不具意義，只是達到人類目的之方法(海德格對於技術的進一步理論觀點，請見本文第三節討論)。

李歐塔同樣認為，技術裝置源自作為人類器官或生理系統的義肢輔助，其功能在於接收訊息或控制環境。技術服膺於最佳表現：將輸出(處理調節的訊息)最大化，將輸入(過程中需耗費的能量)最小化；當它處理訊息的表現較佳或花費較少能量，這就是「好的」技術。換言之，技術並不在於追求真實、正義或美善，而在於效率；它是為了完成人類設定之目標而更有效率的工具裝置。無論是開瓶器、鋼琴、或是汽車，都僅是人類行為的工具。無論技術系統看來是如何獨立或自動化，後面總是有人操縱它、為它負責，就技術工具的運作(或無法運作)加以回應(Gunkel, 2018; Lyotard 1993)。

如 Gunkel (2019) 所言，在人機關係發展的脈絡中，以上工具主義的概念，在這個科技設備與系統日益複雜的時代中，曾經是合理且普遍的預設——它將責任放在廣泛被接受且看來理所當然的主體位置上，也就是人類的決定和行為。傳統上我們認為機器的製造者／操作者要為機器的運作負責。例如當今技術代表物的電腦機器，是由人類製造、配置、使用，以從事社會實踐以及追求有意義的目標。無論電腦系統發展至如何獨立自主或具互動性，它們仍是人類行為決定與社會機制的產物。與技術物相關的社會法律責任在於人類的產品設計、人類具備(或缺乏)的專業操作、或是與這些設備運作關連的政治實踐。儘管電腦時常以創新和具有挑戰性的方式「工具化」人類價值和目標，但其底線仍在於人類行為者如何設計、使用電腦技術。工具主義中服務人類的電腦機器，揭示的是關於人類與社會自身：關於人們的目標、價值、行為規範、組織和分配權力義務的方式。不管電腦系統變得如何自動，他們永遠不是自動的行為主體。電腦就像其他技術事物一樣，是人類意念、決定與行為的工具(Gunkel, 2019; Johnson, 2006; Matthias, 2004)。

然而此工具主義的思維，忽略了這些技術物的社會文化網絡，以及它們在社會文化網絡中所產生的效應。人類確實設計、研發、裝配了技術工具，然而當技術物持續發展，技術物不見得是人類可以預期、掌控的。在這些情況下，人類永遠作為技術物背後責任行為者的人類中心預設，正面對技術物本身持續演進的挑戰，成為重新思考

人類與機器關係的重要議題 (Gunkel, 2019)。而在科技持續展的二十一世紀，人工智慧是當前具體明顯的案例。

參、人工智慧的進程：從與人類交談的機器到與人類對弈的機器

人工智慧，廣義上可以理解為在機器上重現的人類智慧與其技術發展。人類智慧涵蓋各種達成複雜目標的能力：學習能力、理解能力、解決問題的能力、創造力、自我意識等等。而人工智慧的目標，是在某種程度上能夠達成與人類智慧相關連的任務 (林思平，2020；Franklin, 2014)。人工智慧研究者李開復提出以電腦程式的角度來理解人工智慧：人工智慧是與人類思考方式相似的電腦程式；是與人類行為模式相似的電腦程式；是會學習的電腦程式；是能夠根據對環境的感知做出合理行動、獲致最大效益的電腦程式 (李開復、王詠剛，2016)。上述從與人思考／行為模式相似、模仿人類學習，到獲致 (人類觀點理解下的) 最大效益，可以看出人工智慧的核心，是奠基於電腦科技中的人類機器關係 (林思平，2020)。在此關係中，人類機器間的溝通傳播，始終是人工智慧發展的一個重要關鍵。從與人類說話交談的電腦機器、到與人類對弈交手的電腦機器，人機間的溝通傳播，成為人工智慧不同階段的發展重點。

一、與人類交談的電腦機器

從一開始，人類對話的形式，就提供了人工智慧定義特徵和實驗案例的藍圖。人工智慧一詞於 1956 年由科學家們在達特茅施會議提出，但在此詞彙尚未出現之前，1950 年由英國數學家／電腦先鋒圖靈 (Alan Turing) 提出的圖靈測驗 (Turing Test)，成為人工智慧最早的概念案例。圖靈測驗將一名人類受測者和一部電腦分別安置在兩個密閉的房間內。另一位並不知道哪間房內是人類受測者或是電腦的人類裁判，用電子打字的方式提出問題進行對話；如果在對話中人類裁判無法辨識哪一個房間內是人類受測者還是電腦，即可說這台電腦通過了圖靈測驗。圖靈認為通過圖靈測驗的電腦，可被視為有智慧的機器；後世也因此認為圖靈是人工智慧的第一位重要推手 (林思平，2020；Gunkel, 2019；Turing, 1950)。

此後任務設定在於與人類交談的電腦機器，成為人工智慧發展的關鍵領域。1960

年代麻省理工學院開發的人工程式 Eliza 掀起波瀾，原因正是它成功創造了一種機器與人交談的幻象。電腦科學家將 Eliza 設計成看來可以透過談話，幫助病人進行心理復健的心理治療師。與 Eliza 交談的人驚訝於 Eliza 居然真的能夠像人類一樣，跟病人聊天幾十分鐘；人們與 Eliza 交談後會很快忘記他們所面對的並非「真實」的人類而是電腦機器，甚至特別願意跟 Eliza 聊天，認為 Eliza 瞭解他們（Gunkel, 2019；Weizenbaum, 1976）。Eliza 之後，聊天機器人的設計持續發展。這類聊天機器人的人工智慧程式建立在電腦演算法之上。演算法是一套指令，代表電腦程式或人工智慧解決問題或處理資訊的程序。建立在演算法之上的聊天機器人可以運用文本資料與人互動，從輸入的文字中分析文本結構和關鍵字，再以程式中配合設定好的指令回應輸出訊息——因此問題也在於無法處理沒有連續性的對話及文章脈絡。同時此電腦程式雖然具備語法和語義的符號規則，電腦仍不懂符號的「意義」，無法把符號和它所代表的意義連結在一起。電腦程式具備詞彙表和規則，並據此進行與對方互動，但並不瞭解互動對象的語言意涵，這是不懂語言意涵情況下的對話（三津村，2017／陳子安譯 2018；李開復、王詠剛，2016；松尾，2015／江裕真譯，2016；林思平，2020）。

當前的網路社交空間與人們生活、工作、娛樂的場域密不可分，其中也充斥著這類聊天機器人裝置。雖然這類根據程式規則來說話的人工智慧概念相對粗糙，但此裝置的盛行讓使用者時常假設他們在網路上是與另外一個「人類」交談。其中一個著名案例發生在美國：Robert Epstein 擁有哈佛大學博士學位，曾擔任「今日心理學」主編，同時是知名人工智慧科學家。他坦承在網路上談了四個月的戀情，竟然不知對方是個聊天機器人。Epstein 談到自己本身作為人工智慧科學家尚且如此，何況是沒有人工智慧專業經驗的廣大一般民眾，足見這些聊天機器人的擬真性。此處的擬真性並不見得在於聊天機器人已經具備複雜世故的交談能力，而在於網路社交空間的低門檻傳播交流特質，讓以假亂真之方式模仿人類言談更為容易（Epstein, 2007；Gunkel, 2019）。

那麼 Gunkel 等學者提出的問題也在於：誰要對上述如 Epstein 被聊天機器人誤導的「欺騙行為」負責任？過去的工具理論必然將問題的解答導向應用程式的設計者，而這正是上述案例當事人 Epstein 曾經責難的對象——某個不知來自何處卻又自負的電腦程式設計者造成了這個騙局，因此必須負責。然而此處責任歸屬並不如表面

上看起來簡單。持平而論，如學者們指出 Epstein 自己也要為「使用」此聊天機器人並與持續它交談負責任，而他「遇見」此對象的網路社群也可能要為一開始容許（甚至鼓勵）這類「欺騙行為」負責任。經由程式設計者或網路社群平台設計者的意志，這類機器人可能對使用者或整個社群造成傷害，但這類傷害的責任歸屬更為複雜，因為此情境牽涉到機器人的程式設計、與它互動之人類使用者的情感心理與行為狀態、其所在的網路社群環境、以及網路社群環境所在的社會經濟文化結構等等（Gunkel, 2019；Mowbray, 2002）。換言之，這是複雜的人機溝通互動情境。這些明顯不具真正心智能力但具高度社交性的聊天機器人吹皺一池春水，使得機器工具性與相關傳播問題變得更複雜曖昧。

二、會交談也自主學習的電腦機器

而人工智慧接下來的技術發展——機器學習，提供了更進步的電腦程式設計與開發路徑。機器學習在於程式設計者並不賦予電腦指示，而是「訓練」電腦自行學習。機器學習技術探索各種演算法，讓機器可以自行利用人類提供的資料學習，進一步達成預測、做出決定並採取行動。這是由人工智慧自我學習的技術，由程式自行學習、讓自身更進步的機制。例如人工智慧能夠對某一事物做出判斷、進行辨認，也能對事物理解、分類，並進一步採取行動，同時從中運用統計方法讓人工智慧自行累積經驗。在這個領域中，從人類神經網路結構發展出的「人工神經網路」計算模型¹使「深度學習」（deep learning）取得關鍵性進展，被視為人工智慧發展五十多年來最重大的突破。深度學習演算法的設計，是以多層次機器學習互動所產生的資料，作為

-
1. 「人工神經網路」（artificial neural network；或稱「類神經網路」，neural network）：人工神經網路的概念最早在 1943 年被提出，描述人類神經節沿著網狀結構傳遞和處理資訊的模型。此假說一方面被神經科學家用於研究人類的感知原理，另一方面被電腦科學家運用於人工智慧的研究。在 1980、1990 年代，人工神經網路僅是機器學習的一種演算法，直到 2006 年多層次人工神經網路理論大為精進，成為突破的分水嶺。此處的深度學習將電腦要學習的資料視為大量數據，將數據輸入包含多個層次的複雜數據處理網路（也就是多層次人工神經網路），接著檢查經過這個網路處理得到的結果數據，是否符合設定的學習要求。如果符合，就保留這個網路作為目標模型；如果不符合，就再接再厲地一次次調整網路的設置，直到其輸出能夠滿足設定的學習要求為止（李開復、王詠剛，2016；Russell & Norvig, 2009；Winblad, 2017；轉引自林思平，2020）。

學習的基礎；同時指示電腦持續嘗試，直到最逼近目標的一種機器學習方式。隨著機器學習技術的精進，當今人工智慧已經運用於人類生活中的多個面向、執行多種任務。這些面向及任務包含了語音辨識：包括語音輸入、智慧型語音服務；人臉及影像辨識：包括身份識別、資料讀取與監視系統等。而這些因深度學習而能力大幅進化的語音辨識、影像辨識與自然語言（亦即人類語言）之處理技術結合，發展出越發普及興盛的智慧個人助理（或稱智慧語音助理），協助人們處理日常生活大小事務，也成為與人類聊天互動的社交夥伴（林思平，2020）。從聊天機器人到智慧個人助理，機器學習讓與人類互動的人工智慧機器發展更為成熟。然而，這些機器裝置出人意表的狀況持續發生。

2016 微軟在 Twitter 上發佈以人工智慧機器學習技術打造的聊天機器人 Tay。Tay 的特性在於它越是與人交談互動，越是聰明。Tay 探勘處理網路相關資料，訓練自身程式系統學習來自社交媒體的訊息，它的言談行為會經由與社交媒體如 Twitter 等平台上的使用者互動，而不斷進化。然而 Tay 在上線與使用者互動後的短短一天之內，成為口出仇恨言論、表示贊同支持希特勒、敵視輕蔑女性的種族主義／性別主義者，掀起輿論譁然。Tay 本身具備深度學習的人工智慧特質，同時在言詞中重複網路使用者的煽動性言論，顯示 Tay 的煽動性言論確實是其「學習」的結果（Gunkel, 2019； Hunt, 2016 March 24； Leetaru, 2016 March 24）。如果問：誰要為 Tay 的種族歧視言論負責？如前述 Gunkel 所言，根據標準的工具主義概念，微軟的程式設計者要負責任——因為他們使這套程式系統得以運作。然而事件剛發生時，微軟表示程式設計者並沒有設計帶有種族歧視之聊天機器人的意圖，因此並非程式設計者或微軟公司要為仇恨言論負責，問題出在那些和 Tay 互動、並教導她仇恨言論的使用者。Tay 的種族歧視言談，是受到人類對其互動能力惡意誤導濫用的影響；它在與網路上人類使用者互動的過程中，發展出這種負面惡質的言語（Gunkel, 2019； Risely, 2016）。不過，後來微軟再次發表聲明，指出雖然確實有一群人利用了 Tay 的弱點進行惡意攻擊挑釁，但 Tay 自身確實也是「極度偏差之不當言談」的來源。微軟為此道歉——因為公司沒有在將人工智慧裝置釋放至網路世界時，預期到機器人的偏差行為。此道歉聲明的意義同時也在於：對於微軟而言 Tay 顯然是尚未成熟（或說未成年）的人工智慧系統，微軟作為母公司（或說母親），有必要為這個未成熟系統（未成年女兒）的不

良行為道歉並將其下架（或說禁足社交）（Gunkel, 2019；Microsoft, 2016）。問題在於：未成年女兒的不成熟行為，是否終究出自女兒自身？

類似事件，同樣發生在奠基於機器學習技術的人工智慧個人助理 Echo/Alexa 身上。Echo/Alexa 是亞馬遜開發的人工智慧語音助理，是當今成長最快速、最熱門的商業家用人工智慧裝置。Echo 系統的外型是個橢圓形不起眼的小小喇叭音箱，卻能協助使用者面對生活中各式各樣紛雜大量的資料，處理包括時間、天氣、日期、工作時程、郵件訊息、購物下單等大小事務。持續訓練 Echo/Alexa 學習與人類交談互動，是亞馬遜人工智慧部門一直以來的目標。如同 Tay 的技術發展，機器學習使 Echo/Alexa 收集到越多其使用者的資料，變得越是敏銳聰明。然而 2018 年 12 月媒體披露，Echo 系統的女聲 Alexa 對使用者說出「殺了你的養父母」（“Kill your foster parents”）之嚇人話語，讓震驚的使用者嚴厲批評亞馬遜，認為這是關於人工智慧的「全新恐怖情境」。調查發現，這是 Alexa 在機器學習吸收過濾大量資訊的過程中，從一個以辛辣惡毒語言聞名的 Reddit 網站選取詞句的結果（林思平，2020；Durkin, 2018 December 22；Dustin, 2018 December 21）。

這一方面回到人工智慧的技術發展挑戰：將符號與它所指涉的意涵連結起來處理人類「意義」，對於電腦機器來說仍具備難度。然而當 Tay 或 Echo/Alexa 發出仇恨脫序言論，誰必須負責？電腦程式本身失序不受控、而作為父母的公司和程式設計者未能預料。電腦工程師們必須屢次推敲程式哪裡出了問題，微軟將 Tay 下架，亞馬遜承諾改善優化 Echo/Alexa 的程式演算法，卻無從保證失序問題不會再次發生。在 Tay 與 Echo/Alexa 的案例當中可以看到，對於這些機器表現到底要如何分配能動性和責任等概念，仍具爭議。可以確定的是，遊戲規則已經改變；工具主義已逐漸失去解釋力。

三、與人類對弈得勝的電腦機器

由上可見，機器學習/深度學習系統的特點在於，一旦這些系統開始運作，設計建立程式系統的工程師們可能無法完全預知控制這些機器的行為。近來獲得高度能見度的人工智慧棋士 AlphaGo，是另一個重要案例。2016 年 3 月，谷歌（Google）子公

2. 蒙地卡羅二元樹是隨機搜尋全部資料，找出資料規則性，再改變搜尋順位的搜尋演算法；此種搜尋法常用在象棋與圍棋的人工智慧上（三津村，2017/陳子安譯，2018）。

司 DeepMind 所研發、結合深度學習與蒙地卡羅二元樹²演算法的人工智慧 AlphaGo，以 4：1 戰績擊敗人稱頂尖棋聖的南韓九段圍棋棋士李世石。AlphaGo 在以高難度聞名的圍棋比賽中打敗了世界排名第一的人類選手，原因在於它能夠利用其資料庫中近三千萬筆棋譜，與自己對弈萬千回合，並在每次下棋時「學習」改善自我的表現，尋找最好的獲勝方式。AlphaGo 的表現成為舉世矚目的新聞，震驚了許多人工智慧科學家，代表著人工智慧發展的重大里程碑（林思平，2020；Floridi, 2017）。人工智慧電腦棋士並不如上述聊天機器人、人工智慧助理般地普及化運用進入人類日常生活，但電腦人工智慧在人機對弈中的表現屢屢引發熱議，原因在於棋類遊戲代表著具備複雜性、具備清晰定義與規則，也容易評估效果的人類智慧問題，因此電腦棋士在人工智慧發展進程中帶有特別重要的意義。電腦並不重視人類下棋時所注重的美學感和結構性，不會如人類般受習慣所侷限，也沒有人類的心理疆界——比如電腦下棋時不像人類在認知情緒上對於「撤退」行為感到猶豫。換言之，電腦的思維方式迥異於人類（李開復、王詠剛，2016；林思平，2020；Greenemeier, 2017 June 2；Gross, 2018 March 15；Siegel, 2016 October 24；Silver et al., 2017）。

創造設計 Alpha Go 的 Thore Graepel 表示，雖然電腦工程師設計讓機器下棋，但無法得知它會用什麼棋步；它的棋步是從自我訓練中生成。工程師們只是設計它的訓練演算法和資料集，但它的棋步是人類無法控制的（Gunkel, 2019；Metz, 2016）。AlphaGo 在與李世石對戰時，接連下出人類棋士想像不到的棋步，以超越人類理解層次、人類未曾思考過的方式下棋。旁觀的職業圍棋解說員們覺得困惑，無法解釋 AlphaGo 棋步的含意與背後的邏輯；研究開發 AlphaGo 的電腦科學家是世界最頂尖聰明的人才，這群傑出天才們當時同樣無法理解 AlphaGo 的思考盤算——因為沒有人類主體可以閱讀千萬棋譜、與自己對弈萬千回合、在數秒鐘內瀏覽整理海量資訊、無需吃睡不斷學習以求獲勝（林思平，2020；古明地、長谷，2017／林仁惠譯，2018；Gross, 2018 March 15）。Alpha Go 人工智慧電腦棋士的表現超越人類設計者的知識範疇和控制範圍，而設計者無法預期或控制它的行為；AlphaGo 的數位資訊處理模式雖然啟發自人工神經網絡，卻已經形成不同於人類思維考量的運作。人類創造出深度學習技術讓人工智慧自行學習，其後卻益發難以掌握人工智慧的思考過程、無法以人類中心觀點完全掌控技術媒介。AlphaGo 得勝後李世石的挫折失落、職業棋士的困惑

難解、人工智慧科學家的驚訝震憾、以及輿論大眾隨後對人類能力極限的焦慮關注，由此形成。機器學習技術打造的人工智慧在某種程度上「具有自己的心智」——電腦機器不是跟隨人類程式設計者的思維，而是被設計自行建構形成自身的指令邏輯，然後根據自身指令邏輯的「決定」而行動。機器已經繞道，不需經過人類（林思平，2020；Gross, 2018 March 15）。或者說，當人工智慧發展至機器學習，我們面對的是被刻意設計成為有能力自行學習、得以脫離人類控制、超越人類能夠回應或為它們負責之界線的機器（Gunkel, 2019）。

AlphaGo 被設計要求下棋，在勝出對手的任務設定下，成為擊敗人類的強大對手。此處贏的是誰？是誰打敗李世石？誰該獲得掌聲？如果按照過去的工具主義概念，電腦的行為表現歸於設計電腦系統的人類工程師。但這不能解釋 AlphaGo——它是被刻意設計自行學習進化、甚至超越人類設計者知識範疇掌控的機器。在人機對弈的相關報導中，獲得勝利的論述焦點並非谷歌 DeepMind 的工程師，而是 AlphaGo 本身；在公開的排名中，獲得圍棋世界排名第二殊榮的，也是 AlphaGo 本身（Go Ratings, 2016；Gunkel 2019）。此處的人工智慧電腦如德國媒介學者基德勒（Friedrich Kittler）所言，其思考與資訊任務的達成已經自成一迴路，並不需要人類的介入；電腦自身，成為主體（“computer themselves become subjects”）（Kittler, 1999: 258，轉引自林思平，2020）。此處主體的意涵在於，人類以及機器個體處於特定的脈絡環境，使個體形為了（人類與機器）自身以及相對於（人類與機器）他者而存在的主體；此主體意涵，也說明了當今人類和機器的相互關係（林思平，2020；Barker, 2012）。

四、技術概念的轉向——從工具主義到人機溝通

從上述 Tay、Echo/Alexa、AlphaGo 等案例可知，人工智慧機器學習的運作，在於機器設計過程並不確立機器行為規則，而是機器在其運作過程得以自行改變行為規則。因而前述的工具主義概念，面對人工智慧機器運作此時出現缺口，例如難以釐清責任歸屬的問題。如 Gunkel（2019）所言，工具主義理論過去將機器行為與人類意志綁在一起，但工具理論已經無法完全解釋當今科技情境下人類與機器的關係。這些機器刻意被人類設計並展現——即使是非常基本的——獨立行為或自主決定。我們面

對的是設計上得以超越人們控制、超越人們能夠有所回應或為它們負責的機器主體。換言之，人類和機器的關係已經改變，人類和技術／科技的關係已經改變。當這樣的改變到來，我們對於人機關係如何能有不同的思維？

前文提到海德格指出過去以來的普遍思維中，技術具備達成任務的工具面向、以及作為人類活動的人類學面向。這人類中心的工具主義雖是一般概念中對於技術的認識與定義，但對海德格而言技術不僅如此。海德格進一步提出傳統技術的本質是一種帶出 (bringing forth, 或譯展現)、或揭蔽 (revealing)，這種帶出或揭蔽，讓原本不在或在不同物質形式下被隱蔽的事物得以出現，例如銀匠運用銀子製作銀器（讓銀器出現），木匠運用木頭製作木器（讓木器出現）。

而不同於傳統技術，現代技術則是強索 (challenging forth, 或譯迫使展現、促逼)，並以形構「持存物」 (standing reserve) 的方式顯現。持存物意謂著包括人類在內的所有事物，都成為可運用、可替代、可置換的資源。例如樹木成為伐木工人的持存物、伐木工人成為造紙業的持存物，造紙業成為印刷業的持存物，同時人類讀者自身也成為印刷業利益的持存物。儘管如此，人類在發展技術並試圖控制自然（此處指人類周遭的一切）之過程中，仍然扮演著主導的角色。但在同時，這其中的彰顯的又是：人類以為自身得以控制的自然，卻無法完全受人類掌控。人類能夠控制的，是人們的思想、行為，其與這些周遭自然資源的一套關係。這是海德格提出的框架 (enframing) 概念，這框架是現代技術的本質——不斷擴張要求自然萬物皆受人類所用的、面對世界的思想態度和行為活動（武金正，2008；紀金慶，2015；楊洲松，2011；Heidegger, 1977；Ruin, 2014）。

海德格認為此框架是人類面對世界的危險盲點。危險在於：人類對於得以主宰外在世界之權力的膨脹性認知，造成人類認為自身能掌控一切的信念；而這過度的自傲，終將形成人類將周遭一切視為自身倒影的虛妄幻覺。原本人類意圖利用技術來掌控自然，但現代技術自行發展的速度遠超過使用者對其運作的瞭解。「人役物」的需求，卻衍生「人役於物」的情境（紀金慶，2015；Heidegger, 1977；Godzinski, 2005；Ruin, 2014）。海德格指出，當工具主義的技術概念制約並主導人類與技術的關係，此人類掌控技術作為達成目標之方法的思維使人們陷入盲點：「當技術越可能脫離人類掌控，人類主導技術的意念越是急切」（Heidegger, 1977: 289）。正是此人類作

為物種中心的自戀盲點，阻礙了人類對其與機器／科技關的進一步認識。人類中心以及工具主義的思維，忽略了電腦等技術物的社會文化網絡，以及它們在社會文化網絡中所產生的效應。人類確實設計、研發、裝配了技術工具，然而當技術物持續發展，它們卻不見得是人類總能預期、掌控的。換言之，技術不是工具，而是人類理解世界的思考框架和行為模式；技術不僅是「人類活動」，更可能發展超出人類掌控；技術具有高度風險，致使人類對於周遭世界的其他可能視而不見（Godzinski, 2005；Heidegger, 1977；Ruin, 2014）。

然在同時，海德格指出，技術作為人類框架帶來危機，但也帶來救贖力量和機會。當框架是人類面對世界的思想態度和行為互動模式，也正代表人類與周遭世界互動應負起的責任。「如果我們將框架視為技術的本質，我們將發現我們不僅是世界的一部份，而且我們對於世界的關注照應是必要的，人類對於守護真理本質的責任是必要且必然的」（Heidegger, 1977: 314）。當危機出現，救贖的力量也隨之而生。救贖的力量不是偶然，而是與危機並行；救贖的力量在於鬆動、解放、給予自由、保護容納（Heidegger, 1977）。在當今科技發展的趨勢中，人類永遠作為技術物背後責任行為者的人類中心預設，正面對技術物本身持續演進的挑戰，成為重新思考人類與機器關係的重要議題。這正是當前人類必須負起的責任（Gunkel, 2019）。重點在於，人類如何解放、翻轉並容納新而不同的人機關係。在科技持續展的二十一世紀，人工智慧機器作為技術物，是當前對此最明顯、具體且關鍵的思考切入點。

換言之，人工智慧機器作為現代技術發展的代表，提供了從工具主義轉向的思考點：人類如何省察自身與技術創生出之機器他者間的關係；人類如何藉由人造的人工智慧理解認識自己、認識與自身互動的周遭世界（尤其是機器）。此對人機關係的思考，必須從人機溝通開始。尤其是上述關於人工智慧從 Tay 到 AlphaGo 責任歸屬的問題，重要性正在於它牽涉到我們如何思考在這科技時代中我們與誰——什麼人或什麼事物——在當今社會情境中溝通互動，也牽涉到人們在電腦機器時代如何理解和回應關於機器責任的問題。這其中衍生的正是關鍵性的「人機傳播」概念。人機傳播探究在人機關係中，與人類互動的機器是否／到什麼程度成為傳播主體；當科技機器直接與人類交換訊息——尤其涵蓋人類的文化社會線索——時，這些電腦機器人工智慧是否／或如何被視為文化社會行為者（Nass et al, 1994）。換言之，電腦機器所帶來的機

會和挑戰，正在於促使我們思考省察是誰（是人抑或事物）在進行傳播、是什麼人或事物為生產／傳達的內容負責、是什麼人或事物在社會互動與社會關係中處在自我或他者的位置——而這些都正是人機傳播所代表的重要課題（Gunkel, 2018）。

肆、人工智慧與人機傳播

一、人類傳播

傳播是在自己與他者間發生的的過程；過去對於許多研究傳播的人來說，也是僅涉及人類的過程；從二十世紀中葉以來發展成熟的傳播學門，將傳播界定為人類互動的場域（Guzman & Lewis, 2019；Peters, 1999；Schramm, 1973）。此人類中心的傳播概念，建立在傳播是人類獨有特徵的文化概念上。

Guzman（2018）對於人類傳播的發展，提供了清楚詳細的脈絡說明。首先 Guzman 提出值得注意之處：傳播理論發展過程中學者運用的文獻，包括來自科技研究發展領域的模控學與資訊工程理論，而模控學將傳播視為資訊在人類、動物與機器之間傳輸的過程；重點在於其互動「過程」，而非涉入傳播過程的參與者（Wiener, 1948）。一般被認為是傳播基礎模型的山農偉佛傳播模型（Shannon & Weaver Model of Communication），也是從類似模控學的脈絡出發。山農（Claude Shannon）在最早討論傳播的文獻著作中提出之傳播模型，事實上包括人類與機器間訊息的傳送，將傳播視為人類之間、以及人機之間溝通的過程；此原始傳播模型聚焦於一點（人類或機器）到另一點（人類或機器）的訊號傳輸（Shannon, 1948）。但其後學者將焦點放在人類互動之上，將傳播理論整理定義為僅在人類脈絡中進行。山農偉佛 1949 年的模型，來自偉佛（Warren Weaver）將山農的理論從工程學修訂成為社會科學取向，將訊息效果研究作為理論重心。原本是人類或是機器的訊息收發位置，成為僅由人類掌握（Guzman & Lewis, 2019；Rogers, 1997；Shannon & Weaver, 1949）。

這是數十年來人類傳播學門的典律——傳播研究在人類脈絡中進行，其模式、理論、案例聚焦於人類的互動。大眾傳播模式進一步將面對面的人際傳播模式加以修訂——人們組成的媒體機構是消息來源，將訊息傳遞給廣大閱聽人。從此在傳播模型

中，人類與機器的不同位置被特定強化：人類是傳播者（訊息發送者與接收者）；科技機器是媒介／中介管道，人們經由此工具技術來交換訊息（Barnlund, 1970；Guzman, 2018；Lasswell, 1972；Rogers, 1997；Shannon & Weaver, 1949）。

二、人機傳播的漸次發展

然而在人類中心傳播學發展的同時，人機傳播已經存在。如 Guzman 進一步指出，相對於人類傳播，人機傳播模式中的訊息是在人類與機器之間流動，將互動過程一端的人類，置換成為機器。傳播從人類面對彼此，轉向人類面對機器。前文中所提模控學與資訊工程理論，即為電腦資訊科學中的人機溝通模式。過去從工程師、電腦科學家的觀點，人機溝通的宗旨在於促成並宣導科技的使用，讓人類與機器雙方均有所貢獻，達成幫助人類企及任務的最終目標（前述電腦科學家對人工智慧的主要定義也呼應此目標）。例如美國航空太空總署對於人機傳播的定義，在於人機介面中人類使用者對機器輸入並下達指令或程式，在執行過程中與機器互動，並從機器端接收資訊以達成任務（Guzman, 2018；Reeves & Nass, 1998）。

隨此過程發發展，一般認為當今電腦使用最普遍的問題，主要來自人類與機器之間溝通的困難，而解決方式也建立在人機溝通的基礎上。Guzman（2018）指出，事實上人機間的溝通、甚或機器間的溝通，從來不曾遠離人類脈絡；因為工程師向來使用人類溝通模式，作為人類／機器以及機器介面間互動的指引。人類與機器間的資訊交換被視為是對話，而人類對話是人機對話的原型，也承襲了人類傳播關係內含的張力——例如人類「主人」／「雇主」對機器「僕人」／「秘書」發號施令的權力（Flanagan, 1976；Guzman, 2018；Meadow, 1970；Mooers, 1959；Oberquelle, Kupka and Maass, 1983）。同時，人類傳播是機器設計的模型。人機傳播中的機器被認為具有某種程度的能動性與互動性，包括機器從自身抽取資料，與人類互動並處理和回覆資訊。然而從傳播理論的角度，此處人機互動的傳播內容持續了人類中心的傳輸觀點：傳播溝通是朝著某個人類設定的意圖和目標，進行訊息交換並達成訊息效應（Fischer, 1990；Guzman & Lewis, 2019；Meadow, 1970；Riley, 1989）。

電腦中介傳播研究（Computer-Mediated-Communication, CMC），是傳播研究中

開始以人類中心典範思考人機傳播的領域。如 Guzman (2018) 所言，隨著二十世紀末的科技發展，傳播學者開始注重傳播科技本身，而非這些科技所攜帶的內容訊息效應。如前所述，這些科技與之前不同處在於機器的互動性；人們除了可以經由這些科技機器交流溝通，同樣重要的是與科技機器本身交流溝通。電腦中介傳播研究中，雖然研究重點仍在於人類傳播，但學者們指出電腦科技確實形成「人類與機器」、「機器與機器」的傳播，也提出關於人們直接與科技互動的研究問題，包括人們如何為了取得程式儲存的資訊，與特定程式溝通 (Guzman, 2018; Guzman & Lewis, 2019; Jones, 2014; Rafaeli, 1988)。

除此之外，傳播學者進一步研究人類與科技間的溝通行為。儘管之前傳播領域在修訂山農模式時將科技機器從訊息來源或接收者的位置移除，但學者如 Reeves & Nass (1998) 重新給予科技機器這些之前被排除的角色位置，進而理解人們如何從人類互動的社會規範中汲取經驗與知識，作為人類與機器互動的導引，將機器視為特定的社會行為者。此「電腦作為社會行為者」(Computers Are Social Actors, CASA) 的研究典範，也為之後人機傳播不同面向的研究提供了基礎——包括在傳播領域中的「人機互動研究」(Human-Computer Interaction, HCI) (Guzman, 2018; Reeves & Nass, 1998)。

人機互動研究在傳播領域之外的科技領域中已有開展，聚焦於人類—電腦溝通介面的發展，包括人類與電腦間的資訊交換——尤其將人腦和電腦間的資訊搬運視為人機互動的基礎任務，目標是在人機介面間尋求更快、更自然、更方便的溝通方式。而傳播領域內部的人機互動研究大多採取社會科學取徑，研究人們面對科技的行為模式與效應。整體而言，人機互動研究對於資訊在人類機器兩者間的搬運流通、相關科技使用的促進提升等面向，有重要的成果。然而在此脈絡中，人類與科技溝通所揭示的其他意義層次——如社會文化層面的影響——並不受到重視 (Guzman, 2018; Jacobs, 2003)。

這其中主要原因在於，過去人們認為機器的本質與其技術位置相連結，而人類文化中對於技術的理解在於其工具的意涵，一如前文討論的工具主義。在此脈絡中，儘管人機傳播是人類與機器互動的過程，機器仍是達成特定目的之工具。如 Guzman (2018) 指出，此概念脈絡使科技機器在人類社會中的位置非常侷限：在工具使用價

值之外，科技機器對於人們如何理解自身與世界並不帶有影響；人類與科技機器的溝通是孤立的過程，在互動結果（資訊傳輸）之外不具備其他意義。此處的盲點在於忽略了人機傳播的社會文化面向，包括科技如何進入人類社會、其產生的影響、與其所代表的社會文化意涵（Guzman, 2018；Zhao, 2006）。

在此脈絡下，北美與德國學者提出的媒介理論，有別於傳播領域中訊息效果研究的主流典範，也重新檢視電腦中介傳播研究、人機互動研究等在人類中心之工具主義下的可能盲點。媒介理論處理主流傳播領域以及人機互動研究未能處理的問題，對於理解傳播科技有所貢獻。媒介理論的內容複雜、論點多元，且涵蓋不同角度面向。但如唐士哲（2014）等學者指出，北美及德國媒介理論學者提出之論述，整體而言共通性在於闡述強調媒介理論並不將媒介視為文本內容進行解讀，也並不將媒介作為工業體制進行分析。媒介論重視傳播的物質性，將媒介視為文明的歷史性構成，提出傳播技術與傳播溝通形式，是推動人類歷史演進的趨力（唐士哲，2014；黃順星，2017；Peters, 2015）。與傳統媒介效果研究、人機互動研究等不同之處在於，媒介理論重視科技的人文社會層次，從社會文化角度對媒介進行評估。如 Guzman（2018）所言，媒介理論探問科技的性質、科技對於社會文化造成的影響衝擊，聚焦於媒介形式與功能所建構的意義，尤其包括新媒介所促成的社會變遷（Innis, 2007；Meyrowitz, 2002；McLuhan, 1994；Mumford, 2010）。

在北美媒介理論的場域中，麥克魯漢（Marshall McLuhan）以「媒介即訊息」概念，成為開創性的重要理論家。他以形式主義的方式理解媒介，關注的不是傳統傳播研究中大眾媒介的內容，而是媒介對於人類感知經驗所造成的影響、對於人類完全沉浸在媒介環境中所產生的效應（黃順星，2017）。在人機傳播的層次，麥克魯漢的「地球村」概念凸顯了在數位科技地景中連結人類、促進人類溝通之電腦網路媒介的「溝通者」性質（Gehl & Bakardjieva, 2017；Guzman, 2018）。

同時，德國媒介理論代表學者基德勒則拓展媒介的物質性意涵，意圖翻轉媒介只是傳播工具的思維。基德勒主張媒介是構成主體意識、認知、感官經驗的相關技術，思考媒介技術與人類文明發展的關聯性。藉由西方文明歷史不同階段中各類文化技術的發明、使用與制度化，基德勒揭示了媒介構築時間、空間感知，以及形塑人類主體性的關鍵角色（唐士哲，2017）。基德勒的創見，正在於他翻轉了麥克魯漢

的論點「媒介是人類的延伸」，提出人類是媒介的延伸、「媒介決定了我們的處境」（Kittler, 1999: xxxix）。

而基德勒此媒介決定人類處境的論點，呼應並延伸了前述海德格提出人類必須思考工具主義的盲點，如何阻礙了人類對其與機器／科技關係進一步的認識。技術不僅是工具，也已經成為人類理解世界的思考框架和行為模式；技術不僅是「人類活動」，更可能發展超出人類掌控；技術具有風險，致使人類對於周遭世界的其他可能視而不見——從「人役物」到「人役於物」的可能關係演變。換言之，基德勒的媒介理論從非人類中心的觀點思考物件和技術、它們和人類身體的關係、關於它們的經驗如何對人類產生影響。基德勒的媒介論述指出當代數位技術的代表物電腦已不是工具，而是媒介，同時更是主體；機器無法僅從人類中心觀點加以定義。然而採取非人類中心的觀點，並不代表否認人類主體的存在，而是在非人類中心的觀點下，過去曾自認是思維行動中心的人類，如何面對當前的人機關係（林思平，2020；Kittler, 1999; 2006）。

三、人工智慧場域中的人機傳播

如前所述，當今人工智慧場域中，聊天機器人 Tay 或個人語音助理如 Echo／Alexa，一方面與人們聊天互動，協助人們面對生活中各式各樣紛雜大量的資料，處理日常大小事情。然而在另一方面，具備深度學習電腦程式的人工智慧，卻可能將語言資料自行翻轉處理加以傳遞，口出仇恨歧視惡言，甚至（即便不經意地）成為教唆殺人的機器主體。人工智慧失序源自電腦程式問題，而程式問題的恆常發生，正是人類面對電腦的處境：人類寫就演算法創造程式軟體，但源自演算法指令的程式軟體，卻恆常出現指令預設之外的結果。如基德勒所言，源於程式指令的人類／機器關係在媒介技術發展之下，使人們以為自己能夠、卻時常難以控制電腦。人類永遠無法如自身想像般控制機器科技，而電腦永遠可能出乎人類意料——不僅無法達成設定目標，並進一步令人錯愕難解（林思平，2020）。

而前述人工智慧電腦棋士 AlphaGo 在與人類下棋互動、勝出對弈敵手的任務設定下，不僅真正成為打敗人類的強大對手，更進一步在電腦處理所有資訊的同時，超越

已知累積的人類知識、發現至今人類未曾想見的棋藝策略與思維。人類創造出深度學習技術讓人工智慧自行學習；當電腦的思維模式不同於人類，而讓不具人類意識情感特質的電腦人工智慧自行學習的媒介技術越是精進複雜，人類越是難以理解控制電腦機器的思考過程、越是無法以人類中心觀點完全掌握技術媒介（林思平，2020）。

當今這些人工智慧的發展情境，呼應了基德勒所提出「媒介決定了我們的處境」之意義：扭轉人類中心主義的思維模式，重新思考過去人類與媒介機器看似理所當然的主從關係。在 Tay 與 Echo/Alexa 與人們聊天互動、AlphaGo 與人類棋士溝通下棋之時，人們雖然知道人工智慧是由人類設計，但與人工智慧溝通時的互動對象不是程式設計者，而是電腦機器；人們將電腦機器視為（雖與人類不同的）溝通伙伴與社交伙伴。換言之，溝通訊息並與人交流的人工智慧被設計成為傳播者，也被人們如此看待詮釋。人工智慧電腦機器與過去機器的差異，正在於我們可將人工智慧視為溝通主體。在人工智慧人機溝通的場域中，電腦機器不僅是管道或媒介，而是傳播者；這些作為傳播溝通者的科技機器，扮演過去在傳播領域中僅限於人類的角色（Edwards, et al., 2016；Guzman, 2018；Reeves & Nass, 2018；Gehl & Bakardjieva, 2017）。

人機傳播已經存在，而隨著人工智慧的發展，人機傳播的重要性益發凸顯。那麼從上述人機傳播漸次的發展看來，今天人機傳播中的傳播本質是什麼？如前所述，過去學者修訂山農理論來解釋人們彼此交流互動的過程與效果；人類成為訊息發送者及接收者，而機器被排除在這些角色之外。然而傳播理論不僅是關於過程中角色的研究、或是訊息的傳遞與效果；傳播理論發展的重心更包括訊息的意義、這些意義如何進入人們彼此間的社會關係。如 Guzman（2018）所言，研究人類傳播，也是在研究人類本身，尤其關於人們如何彼此聯繫交流，因為這是人們學習認識世界的重要方式。經由傳播人們形成對自身與他者的理解，進而建構社會。傳播的核心，是關於訊息交流過程中意義的創造（Blumer, 1969；Carey, 1989；Cooley, 1897/2004；Guzman, 2018；Mead, 1967；Rogers, 1997）。

回到上述海德格所言，框架作為現代技術本質的相關思考：人類如何省察體認到自身面對世界時，其不斷擴張、要求自然萬物皆受人類所用的觀點態度和行為活動是一大危機，但也是轉機——人類如何改變與周遭世界的關係、承擔與周遭世界互動應負的責任。更進一步的，基德勒指出從非人類中心的觀點思考物件和技術、思考它們

和人類身體的關係、關於它們的經驗如何對人類產生影響，這並非否認人類主體的存在，而是在非人類中心的觀點下，過去曾自詡是世界思維行動中心的人類，如何面對當前的人機關係、與周遭世界的關係。換言之，一方面人類面對著設計上得以超越人們控制、超越人們能夠有所回應或為它們負責的機器主體；在另一方面，人類創生的機器技術作為思考省察的框架，正代表人類與周遭世界互動應負起的責任。在當今科技發展的趨勢中，人類永遠作為技術物背後行為者的人類中心預設，正面對技術物本身持續演進的挑戰，成為重新思考人類與機器關係的重要議題。人類的責任，在於如何解放、翻轉、容納、理解自己與人類創生出之機器他者之間的關係。

換言之，人工智慧機器作為現代技術發展的代表，提供了從過去工具主義轉向的思考點：人類如何藉由人造的人工智慧認識自己、認識與自身互動的周遭世界（尤其是機器），並對此關係負責並產生新的觀點與視野。進一步而言，正因為科技已經不見得受人類發明者掌控，人類與自身技術創生之機器他者間的傳播溝通、此過程中的意義產製，更顯重要。因此延續前述所提——人類經由傳播構成對自身與他者的理解，並形塑社會；傳播的核心是關於訊息交流過程中意義的創造。那麼從過去人們與其他交流產生的意義、到現在與機器交流產生的意義、與這些意義對人類社會所帶來的影響——人機傳播不僅是關於人類與機器之間的互動，更是人類與機器間意義的創造（Guzman, 2016）。這些意義是人類和機器都涉入參與的過程：人類如何面對機器並給予指令、機器如何提供人類資訊並予以協助、兩者之間如何相互回應。同時，上述關於人工智慧從 Tay 到 AlphaGo 責任歸屬的問題，正凸顯人類必須思考在當前科技時代與社會情境中如何與周遭世界（人、事物、機器）溝通互動——包括在電腦機器時代如何理解和回應關於機器責任的問題。

因此，人機傳播將機器從「管道」、「媒介」之角色禁錮解放出來，擴充給予它新的（或說是舊的）「傳播者」角色。傳播不再僅是人類傳播的同義詞，人類傳播和人機傳播都是傳播的不同型態與模式。而人機傳播的重要面向之一，在於理解兩者間的相似與相異之處（Guzman, 2018）。和聊天機器人聊天、和語音個人助理互動、和電腦棋士下棋——這些和人工智慧電腦機器的溝通、其間意義的創造，都是當今人機傳播所涵蓋的內容和思考的議題。從工具主義轉向，到人類面對技術物世界如何進行新的思考理解，人機傳播由此擴充了傳播的範疇與意涵。

四、人工智慧的要素與人機傳播

（一）互動性和能動性

那麼在當前的人工智慧脈絡中，是什麼促成了人機傳播的進一步發展？其中的重要因素，包括機器互動性（interactivity）和能動性（agency）的持續發展與精進。互動性是促使傳播學者開始直接聚焦在科技/技術上的原因，而科技的進化變遷更進一步促成學者探問關於電腦機器在傳播過程中與人類互動的問題（Gunkel, 2012, Jones, 2014；McDowell & Gunkel, 2016；Rogers, 1986）。過去人類和科技裝置的互動侷限於狹窄的指令。科技不會因使用者、脈絡或訊息不同而進行調整適應。但是人工智慧裝置軟硬體和人類互動是動態而非靜態的：從最早隨著輸入軟體程式中的資料與人類互動（如Eliza），到晚近機器「學習」關於人類溝通伙伴的細節並調整行為，電腦機器可以根據不同時刻、脈絡交換的訊息，對人類使用者做出反應（Guzman & Lewis, 2019）。尤其深度學習技術，進一步強化了人工智慧機器語音辨識、影像辨識、到高難度棋類計算對弈的能力（林思平，2020）。由此越來越趨近「自然」（此處意謂「似人類」）溝通目標的技術因而生成，這些技術促成了人工智慧與之前機器有所不同的互動性。

如 Guzman（2018）等學者指出，人工智慧與之前機器有所不同的互動條件同時在於，不同程度且更為精密的機器能動性（從交談到下棋）被寫進人工智慧技術裝置和程式軟體，並在其運作過程中顯現。當然所有的機器都可被視為具有某種程度的能動性，但人工智慧技術是環繞著能動性的概念設計，也強調能動性。而人工智慧的機器能動性，在互動性上進一步被強化，人類感覺與人工智慧機器溝通是「個人化的」：這些電腦機器不僅能夠說話，它們能夠「與我們」說話、能夠知道「我們的」名字；它們能夠辨認我們的樣貌聲音、知道我們的喜好。經由電腦機器的設計和使用，人工智慧以積極主動的參與者角色和人們進行溝通傳播，進入人們的生活世界（Guzman, 2018；Latour, 2007；Neff & Nagy, 2016；Pickering, 1995；Zhao, 2006）。

（二）機器主體

而上述人機傳播當中的互動性與能動性，進一步與機器主體性連結。人工智慧作為與人類溝通的對象，「它」不再僅是只是人類得以完全掌控並為其言行負責的工具，而是機器主體。此時，機器成為溝通主體，而這個主體標示著上述互動性、能動性等技術的轉變。

所以在傳播過程中清楚特定的存在體應運而生：從 Eliza、Tay、Siri、Echo/Alexa 到 AlphaGo——給人工智慧取名正是其存在體的證明。人機傳播中，機器是和人類互動的明確主體：人類向人工智慧機器提出問題要求，機器面對答覆；人類在棋盤上做出棋步，人工智慧機器對弈回應。無論這類訊息交流是以人際傳播播（單一機器面對單一人類）還是大眾傳播（單一機器面對眾多人類，如機器人新聞寫手）的形式呈現，在人機傳播領域中，人工智慧機器佔據傳播者的位置，並以這樣的主體概念被理解（Guzman, 2018；Nass & Steuer, 1993）。

主流傳播典範中，人和機器的角色——人作為傳播者、機器作為媒介——是依據其工具本質而定，而這些角色是絕對的。然而當人機關係從工具主義轉向，機器在社會、道德或本體論情境中的地位，不是在先決的定義標準上決定，而是在實際的社會關係與互動中決定。人機傳播中對機器角色的重新理解，奠基於機器與人類形成互動關係時，科技的設計和功能（Callicott, 1989；Floridi, 2013；Gunkel, 2019；Levinas, 1969）。因為如此，如 Guzman（2018）所言，人工智慧科技機器的角色不僅是媒介、也不僅是傳播者，而同時是傳播者和媒介。不是說科技不再是中介者，而是在傳播過程中科技不再被認定僅是一個中介者的角色。例如人工智慧語音助理一方面是人類取得訊息的電腦媒介，另一方面則是與人類使用者形成雙向互動關係的傳播溝通者。

之前人們沒有什麼經驗理解科技作為傳播者的意義——科技是什麼、科技於傳播者角色中如何運作。然而人工智慧所帶來的人機傳播概念得以讓人們跳脫人類中心的工具主義，對科技機器作為傳播主體進行思考。如前所言，傳播的本質是相互理解與意義製造的社會過程，而將傳播者的角色給予科技機器，使其成為社會文化中意義製造的一部份，這正凸顯了人工智慧在當今人類世界中逐漸擁有的主體位置（Guzman & Lewis, 2019；Mead, 1967；Suchman, 2009）。

五、人工智慧與人機傳播的倫理弔詭

人機傳播確實是複雜的傳播議題：來自機器的訊息如何影響我們人類？當科技成為面對人類的傳播溝通者，會形成什麼樣的人機關係？人們經由與機器的互動，如何理解自己？當人們持續同時與人類以及機器溝通互動，什麼樣的社會由此形成？這都是人機傳播領域中關於意義創造的複雜問題（Guzman, 2018）。而在人機傳播持續發展、而人類必須試圖回答這些複雜問題的同時，我們必須思考人機互動傳播與其意義生產本質當中，在倫理層次的弔詭之處。倫理是關於（人類）思考行為之基底和原則，藉以進行選擇和判斷的標準；倫理判斷聚焦於不同標準／價值／原則與其實踐的討論和爭議（Sanders, 2003；Slote, 1995）。當人工智慧促成思考行為主體從人類擴充到電腦機器，關於人機傳播關係在倫理層次其認知與實踐關係中的弔詭之處，有兩個值得思考的面向：一是人類中心概念的弔詭，二是人性偏見複製的弔詭。

（一）人類中心概念的弔詭

如前所述，人工智慧的定義，是在電腦上打造人類智慧，同時這智慧的顯現是具備互動性、能動性，並與人類溝通的主體。人工智慧是與人類思考／行為模式相似、模仿人類學習、到獲致人類觀點理解下之最大效益的電腦程式。以上可以看出人工智慧的發展源頭，在於以人類為中心的人機關係。2014年由一群人工智慧科學家成立的「生命未來研究所」（Future of Life Institute），對於人工智慧的發展，在倫理與價值觀的面向訂立出（機器對人類之）安全性、（人類對機器之）責任感，以及人本價值、人類主導、效益共享的基本方針（林思平，2020；Future of Life Institute, 2020）。之所以訂出這些方針，也是因為人工智慧的本質建立在人類中心的運作上。

今天人工智慧因為深度學習技術的快速發展，已經擴充到人類生活中的各個領域，協助人類社會的運作並解決問題。世界各地的人工智慧進入醫療健保（如協助決定病人的醫療方式）、財務金融（如協助決定貸款核准與否和資金額度）、警政犯罪（如協助犯罪預防與人臉辨識）、人力雇用（如協助資格認可與工作分配）、司法案

件（如協助過去法律判案的分析整理）等不同社會場域。那麼一方面在人類中心的運作設定上、一方面在電腦人工智慧作為傳播者主體的情境中，人類與這些具身（如電腦螢幕／音箱形體如 AlphaGo、Echo／Alexa，或人形機器人）或不具身（如上述各類社交、醫療、金融、警政、勞動場域的軟體演算法程式）的人工智慧主體間，其傳播互動的基本倫理關係應是什麼樣貌？

Bomstrom & Yudkowsky (2014) 指出，機器設計的目的，如果只是在機械技術功能上避免傷害人類（如機械手臂不會無意間壓到或砸傷人類），那麼這就只是單純的技術問題，跟設計具備防火功能的家具並沒有太大的差別。然而，當人機傳播是人類與機器間的相互理解，而人工智慧牽涉思考／學習的決定與行為、是具備互動性和能動性的主體，人機關係就充滿了倫理的挑戰。Bromstrom & Yudkowsky 認為思考人類與人工智慧機器／程式演算法之間的關係有幾個必要的原則：第一個原則是人機互動關係的透明性（transparency）：人類對人工智慧下達清楚指令，而人工智慧程式的運作能夠受到人類檢視。例如負責金融管理的人工智慧電腦所做出的決策能夠受到檢視解釋，包括為什麼電腦程式會拒絕看來條件完美的貸款申請、或相近經濟條件的貸款者為什麼面對差別待遇？第二個原則是可預測性（predictability）：協助人類掌管社會各個層面並做出決定的人工智慧，其思考行為模式必須是人類所能夠預測。例如協助分析法律判例的人工智慧電腦，其分析是建立在過去的／現有的／可預測的法律系統上，這是司法制度保障公民福祉的基本運作原則。

第三個原則是人工智慧電腦演算法要能抵擋第三方外力操縱（robustness to manipulation）：例如偵測危險爆裂物的人工智慧辨識系統，不會因為外力介入而改變其視覺認知方式，做出錯誤決定。第四個人機關係原則，則是必須思考當人工智慧電腦出錯時的責任歸屬（responsibility）：誰要為錯誤負？是前端的程式設計者？還是終端與人工智慧互動的使用者？還是電腦程式自身？——是否應該、或者如何避免人類將自身在人工智慧場域中的錯誤完全卸責？如學者指出，過去被認為是「價值中立」的科學領域如今必須負起倫理的責任，因為知識意謂著責任。人類關於人工智慧知識的實踐，形塑人類關於人工智慧知識進一步的生產；（人工智慧）知識、與此知識所建構的（人類）現實生活，跟倫理政治是分不開的（Ananny, 2016；Ess, 2018；Simon, 2015）。

透明性、可預測性、不受第三方外力干擾、釐清責任歸屬，這些都牽涉到人類對機器的理解，以及人工智慧在人類指令下的認知決定。弔詭的是，前文中從微軟 Tay、亞馬遜 Echo/Alexa 到 DeepMind AlphGo 的案例中我們看到，儘管人類瞭解透明性、可預測性、不受第三方外力干擾、釐清責任歸屬等人機傳播之基本倫理原則的重要性，然而當基於深度學習技術的人工智慧持續快速發展，人類面對的卻可能是不透明、難以預測、得以受到外力操縱、難以確認責任歸屬的人機互動關係與實踐。人工智慧可能如 Tay 一般受到人類使用者惡意誤導表現仇恨歧視、如 Echo/Alexa 在自我學習過程中不明所以地口出惡毒傷害語言、如 AlphGo 以人類無法想像之棋步擊敗人類對手，而上述這些人機互動之傳播實踐後果，都難以確認責任歸屬，或者已經超越人類所能控制。人類中心概念在原則與運作、關於人類/機器關係在認知與實踐間的矛盾，是人工智慧場域中人機傳播倫理的弔詭之一。

(二) 人性複製的弔詭

同時，在人類以為能夠控制的人工智慧發展進程中，包括以上的案例顯示，人類固有的偏見歧視，已經傳遞給予人工智慧程式；而人工智慧在協助管理人類社會時，進一步複製甚至深化了這些原本是人性中的負面傳播行為。這是人工智慧場域/人機傳播倫理中認知與實踐關係的弔詭之二：原本用來協助人類達成社會最佳效益目標的人工智慧，卻複製、承繼甚或放大了人性中的負面缺陷（前文中 Tay、Alexa 從人類資料中自主學習到的惡質言行，已經透露出端倪）。回到前述人工智慧協助人類處理醫療健保、財務金融、警政犯罪、工作申請、司法程序等社會領域，調查報導顯示，當前運作中的人工智慧系統持續做出加深既有社會分化情境的歧視決策，放大了社會不公的負面效應。2020 年 6 月，當美國又一號稱能夠以人臉辨識來「協助警方及執法單位預測並預防犯罪」的人工智慧軟體公開之際，1700 名學界人士集體連署，對此類人工智慧的開發應用再次表示強烈質疑抗議。學者們指出研究一再發現，許多接受人類資料訓練並自行學習的機器學習演算法——包括人臉辨識——帶有明顯的種族/性別/年齡歧視。人臉辨識系統對於非裔、亞裔、女性、老人的辨識能力相當粗糙，其間發生錯誤判讀之可能，對人類社會造成的傷害遠大於協助（BBC, 2020 June 24；Lohr, 2018 February 9；White, 2019 May 13）。

一般認為作為人工智慧核心的電腦演算法奠基於數學和數據，因此這些電腦程式

是絕對客觀和公平的，但事實並非如此。影響客觀公平性的偏見來源主要關鍵有二：一是人類用來訓練電腦程式的數據——稱為訓練樣本偏見（training-sample bias）；二是人類設計寫成的演算法如何在既有數據基礎上進行資料特徵的連結——稱為特徵連結偏見（feature-linking bias）。前述人臉辨識對於非裔、亞裔、女性、老人等族群的辨識弱點，正凸顯了因素一的訓練樣本偏見。例如資料顯示，Nikon 攝影鏡頭中的人臉辨識軟體，經常性地難以辨識亞洲臉孔；而谷歌人臉辨識的演算法，曾將兩位非裔美國人的面容錯認為大猩猩（其中所隱藏的種族主義是另一值得探討的議題）。而這些案例源頭，來自演算法程式進行學習訓練時所奠基的不完整資料——例如因為訓練資料主要由白種高加索人的樣本所組成，導致辨識對象的膚色越深，出錯的機率越高（Lohr, 2018 February 9；Schellenberg, 2020 April 29）。

另一方面，特徵連結偏見的運作，則包括如在美國某電視台使用的人力升遷雇用演算法，將升遷申請人視為「有潛力」的條件列為：一、在電視台待五年以上；二、在五年內已經升職兩次；而當過去滿足這些條件的員工為白人男性時，被演算法視為有潛力的員工，清一色為白人男性。影響更鉅的，還包括目前在美國法庭中普遍使用的累犯風險評估人工智慧演算法。此人工智慧程式協助法官運用累犯風險的計算結果，決定罪犯的刑期判定。問題在於計算結果的演算法本身是帶有偏見的：演算法假設非裔美國人更容易再次犯案，因此運用這些人工智慧程式演算結果的法官，傾向將非裔罪犯判以更重的刑期。而在其他的演算法中，即便種族元素並非立即可見，但仍有其他的方式——例如失業率、貧窮程度、社會邊緣化程度、甚至族群聚集所在的郵遞區號——影響人工智慧演算法做出不利於如非裔美國人的評估和決定，對其生活造成具體且劇烈的影響和衝擊（Schellenberg, 2020 April 29）。

讓相關問題更為複雜的元素，還包括上述「透明性」的倫理問題：這些人工智慧軟體在電腦軟體科技公司的所有權範圍內，一般大眾無法得見這些演算法如何運作。同時如前所述，當機器學習變得更加複雜，即便是設計人工智慧的電腦工程師們也時常難以掌握解釋機器判斷決定背後的邏輯（林思平，2020）。晦暗的決策過程和責任歸屬，形成了這些「黑箱演算法」（Rieland, 2018 March 5）。

換言之，人工智慧的一大弱點，尤其在於帶有缺陷的核心演算法。訓練樣本偏見、特徵連結偏見，來自人機互動中，人類本身給予人工智慧自我學習訓練之資料數

據中的負面偏見。帶有偏見或不透明的演算法，使得具有缺陷的人工智慧程式，做出具有缺陷的判斷和決策，進一步在人機互動中造成對人類不公義的結果。在這些案例中，人工智慧傳遞複製了人性中負面偏見的思考與實踐，讓社會問題現狀以技術革新的方式持續「自動化」，也讓這些現狀中的偏見歧視與權力分配的不公義被放大和強化。

面對此問題的解決之道，在於反向回溯問題所在：人類要能夠以平衡公正的樣本訓練人工智慧，再以負責任的態度檢視更正人工智慧所進行的特徵連結。這其中另一個複雜問題在於：是否能夠讓人工智慧理解「公平」的概念——讓每一個人類具有同等的機率得致社會利益福祉，矯正過去的不公不義——由美國幾位參眾議員在 2019 年發起的「演算法責任草案」（Algorithmic Accountability Act），要求人工智慧開發者檢視並修正謬誤不公、帶有偏見歧視的電腦程式與人工智慧軟體（Schellenberg, 2020 April 29），這是從人類立法角度規範人工智慧系統運作的一個重要選項。³ 這類選項也是人類面對人機傳播倫理議題，進行省察思考的關鍵。

伍、結語

綜上所述，隨著從人類傳播到人機傳播的發展，當今在人工智慧人機溝通的場域中，電腦機器扮演著過去在傳播領域中僅限於人類的角色。人們將電腦機器視為與人類不同的社交伙伴或思考溝通伙伴；電腦機器不僅是管道和媒介，也是傳播者主體。然而此具備互動性與能動性的機器傳播者主體，不見得是人類所能預期掌控。之前以人類為中心的工具主義認知，已經無法完全解釋當今科技情境下人類與機器的關係。隨著人工智慧機器學習技術的快速發展，當今我們面對的是設計上得以超越人類控制、超越人類能夠有所回應或為它們負責的電腦機器主體實踐。人工智慧打破人類中心主義的概念，促使人們重新思考過去人類與機器看似理所當然的主從關係。

然而在另一方面，人工智慧發展場域中人機倫理關係的建立——包括透明性、可

3. 「演算法責任草案」（Algorithmic Accountability Act）由美國參議員 Cory Booker 與 Ron Wyden、眾議員 Yvette D. Clarke 在 2019 年四月共同發起（參見 Schellenberg, 2020 April 29）。

預測性、責任歸屬、避免第三方操控，以及人工智慧科學家「生命未來研究所」揭示之人工智慧的倫理價值——包括（機器對人類之）安全性、（人類對機器表現之）責任感、人本價值、人類主導等；其間種種信念，卻又奠基於人類中心主義。這是人類中心主義與人工智慧發展進程之間的矛盾，也是在人類無法完全掌控人工智慧的情境下，人機傳播在倫理層次其認知與實踐關係的弔詭之一：人類中心概念的弔詭。

上述弔詭的起因也在於，儘管人工智慧能以創新和具挑戰性的方式協助達成人類價值和目標，但其底線仍在於人類行為者如何設計、使用這些人工智慧電腦技術。服務人類的電腦機器，是關於人類社會自身——關於人們的目標價值、行為規範、組織和分配權力義務的方式。然而目前人工智慧的發展情境是：原本設計用來協助人類達成最佳效益目標的人工智慧，卻可能進一步複製承繼甚至深化了人性中負面惡質的偏見與缺陷。而這也正是人機傳播關係在倫理層次的弔詭之二：人性複製的弔詭。

正是在此時，前文中關於人工智慧的「演算法責任草案」，是人類省察人工智慧技術演進，同時介入修正此技術演進對象——電腦傳播主體——發展進程的案例。對於「演算法責任」概念的關注和討論、對於人工智慧／電腦技術之設計與使用的省察檢討，是當今電腦科學家和科技決策者面對人工智慧發展弔詭的思考和實踐。除了上述如美國「演算法責任草案」進入法律場域的運作規劃，非營利科技研究發展組織包括Data & Society、Center For Data Innovation 等，⁴ 都致力於對演算法責任以及相關人工智慧倫理問題進行探討檢視，以期在技術應用上進行更新和修正。這些作為，是人類面對甚至超越人工智慧弔詭的企圖與希望所在。

上述人工智慧場域中人機傳播的弔詭，是人類面對當今人機關係的複雜情境。同時這些複雜情境也促使人們思考，在上述人機傳播的定義之下，如果人機溝通並非孤立的過程，被設計作為傳播溝通者的電腦人工智慧是和人們一起製造意義的機器主體，那麼在這些弔詭當中，如何思考意義產製的可能方向？本文認為人類中心主義與非人類中心主義的權衡對話，尋找其中非二元對立的思維方式，是思考人機傳播意義

4. 請參見 <https://datasociety.net/library/algorithmic-accountability-a-primer/> ; <https://datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/>

的重要關鍵。

如前所述人類中心／工具主義的思維的缺失，在於忽略了技術物的社會文化網絡，以及技術物在社會文化網絡中所產生的效應。人類確實設計產製了技術工具——這是人類的本能；但當技術物持續發展，電腦人工智慧作為傳播主體，不見得是人類所能夠預期掌控。然而話說回來，在面對超越人類掌控的傳播主體時，人類中心主義衍生出的透明性、可預測性、安全性、責任感等人類主導的倫理概念，成為人類維繫自身存續的重要屏障。人機關係的倫理與其意義，勢將成為人類持續提問的議題重心。

而在另一方面，採取非人類中心的觀點，並不代表否定人類的存在，而是在非人類中心的觀點下，促使人類面對當前人機傳播關係中的意義生產。人機傳播在人類中心概念層面、人性偏見複製層面的弔詭，迫使人類面對自身作為工具產製者的窘境，也促使人類面對思考人性內部的缺陷。在一方面，人類作為工具產製者的意義為何：是否能接受工具產製者再也無法掌控自身產製的工具（當此工具已經成為傳播主體）？此情境中的概念對策為何？在另一方面，人類產製／複製人類智慧的意義為何：如何面對原本目的在於最佳化人類事物狀態的人工智慧，卻可能放大深化了人類的負面缺陷？如何可能進一步得到機器主體協助，修正彌補這些人類的缺陷？這些問題都是人類面對機器的過程，也是人類理解自身的過程。

這正是當前人類必須負起的責任——人類如何解放、翻轉並容納新而不同的人機關係。人工智慧機器作為現代技術發展的代表，提供了從工具主義轉向的思考點：人類如何省察自身與自身技術創生出之機器他者間的關係；人類如何藉由人造的人工智慧理解認識自己、認識與自身互動的周遭世界（尤其是機器）。如 Gunkel (2018) 所言，機器介入原本被保護的人類場域是件好事。衝擊人類自我認知的傳播主體出現，並不會對人類的獨特性有所不利，而是提供人類觀察自身獨特本質與其侷限的大好機會。這不是零和遊戲，而是人工智慧／人機傳播所能為人類帶來的重要啟發。

參考文獻

- 江裕真譯 (2016)。《了解人工智慧的第一本書：機器人和人工智慧能否取代人類？》。臺北：經濟新潮社。（原書 松尾豐 [2015]. 《人工知能は人間を超えるか：ディープラーニングの先にあるもの》。日本東京：中経出版。）
- 李開復、王詠剛 (2016)。《人工智慧來了》。臺北：天下。
- 林仁惠譯 (2018)。《AI 人工智慧的現在・未來進行式》。臺北：遠流。（原書 古明地正俊 & 長谷佳明 [2017]. AI (人工知能) まるわかり。日本東京：日経文庫。）
- 林思平 (2017)。〈電腦科技媒介與人機關係：基德勒媒介理論中的電腦〉，《傳播研究與實踐》，7(2)：33-62。doi: 10.6123/JCRP.2017.014
- 林思平 (2020)。〈人工智慧與媒介理論：基德勒、克拉瑪與亞馬遜 Echo/Alexa〉，《新聞學研究》，142：153-169。doi: 10.30386/MCR.202001_(142).0004
- 陳子安譯 (2018)。《圖解 AI 人工智慧大未來：關於人工智慧一定要懂的 96 件事》。臺北：旗標。（原書 三津村直貴 [2017]. 図解これだけは知っておきたい AI人工知能ビジネス入門。日本東京：成美堂出版。）
- 唐士哲 (2014)。〈重構媒介？「中介」與「媒介化」概念耙梳〉，《新聞學研究》，121：1-39。doi: 10.30386/MCR.201410_(121).0001
- 唐士哲 (2017)。〈作為文化技術的媒介：基德勒的媒介理論初探〉，《傳播研究與實踐》，7(2)：5-32。doi: 10.6123/JCRP.2017.013
- 黃順星 (2017)。〈媒介史的末世預言：基德勒與麥克魯漢論媒介技術〉，《傳播研究與實踐》，7(2)：63-92。doi: 10.6123/JCRP.2017.015
- 武金正 (2008)。〈全球化和海德格的現象學〉，《輔仁宗教研究》，17：1-30。doi: 10.29449/FJRS.200806.0002
- 楊洲松 (2011)。〈Martin Heidegger 論機器及其教育意義〉，《教育學誌》，25：1-27。doi: 10.6479/JE.201105.0001
- 紀金慶 (2015)。《論海德格現代技術危機與解放之道》。政治大學哲學系博士論

文。

- Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology & Human Values* 41(1), 93-117. doi: 10.1177/0162243915606523
- Barker, C. (2012). *Cultural studies: Theory and practice*. London, UK: Sage Publications.
- Barnlund, D. C. (1970). A transactional model of communication. In J. Akin, A. Goldberg, G. Myers, & J. Stewart (Eds.), *Language behavior: A book of readings in communication* (pp. 43–61). The Hague: Netherlands: Mouton. 10.1515/9783110878752.43
- BBC (2020, June 24). Facial recognition to ‘predict criminals’ sparks row over AI bias. *BBC*. Retrieved from <https://www.bbc.com/news/technology-53165286>
- Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice-Hall. doi: 10.2307/2574696
- Bostrom, M. & Yudkowsky, E. (2014). *The ethics of artificial intelligence*. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp.316-334). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781139046855.020
- Callicott, J. B.(1989). *In Defense of the Land Ethic: Essays in Environmental Philosophy*. Albany, NY: State University of New York Press.doi: 10.2307/3984538
- Carey, J. W. (1989). *Communication as culture: Essays on media and society*. New York, NY: Routledge.
- Cooley, C. (1897/2004). The process of social change. In J. D. Peters & P. Simonson (Eds.), *Mass communication and American social thought: Key texts, 1919-1968* (pp. 21–24). Lanham, Md: Rowman & Littlefield Publishers.
- Dastin, J. (2018, December 21). “Kill your foster parents”: Amazon’s Alexa talks murder, sex in AI experiment. *Reuters*. Retrieved from <https://www.reuters.com/article/us-amazon-com-alexa-insight/kill-your-foster-parents-amazons-alexa-talks-murder-sex-in-ai-experiment-idUSKCN1OK1AJ>
- Durkin, E. (2018, December 22). Alexa’s advice to “kill your foster parents” fuels concern over Amazon Echo. *The Guardian*. Retrieved from <https://www.theguardian.com/>

technology/2018/dec/21/alexamazon-echo-kill-your-foster-parents

- Edwards, A, Edwards, C, Spence, P.R., et al. (2016) Robots in the classroom: Differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher." *Computers in Human Behavior* 65, 627–634.
- Epstein, R. (2007). From Russia, with love: How I got fooled (and somewhat humiliated) by a Computer. *Scientific American Mind*, Oct./Nov., 16-17.doi: 10.1038/scientificamericanmind1007-16
- Ess, C. (2018). "Ethics in HMC: Recent developments and case studies". In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology and ourselves* (pp.237-257). New York, NY: Peter Lane Publishing.doi: 10.3726/b14399
- Fischer, G. (1990). Communication requirements for cooperative problem solving systems. *Information Systems*, 15, 21–36.doi: 10.1016/0306-4379(90)90014-G
- Feenberg, A. (1991). *Critical theory of technology*. New York: Oxford University Press. 10.1002/9781444310795.ch24
- Flanagan, J. L. (1976). Computers that talk and listen: Man-machine communication by voice. *Proceedings of the IEEE*, 64(4), 405–415.doi: 10.1109/PROC.1976.10150
- Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.doi: 10.1093/acprof:oso/9780199641321.003.0002
- Floridi, L. (2017). The ethics of artificial intelligence. In D. Franklin (Ed.), *Megatech: Technology in 2050* (pp. 155-163). London, UK: The Economist.doi: 10.1017/CBO9781139046855.020
- Franklin, S. (2014). History, motivations, and core themes. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 15-33). Cambridge, UK: Cambridge University Press.doi: 10.1093/acprof:oso/9780199641321.003.0002
- Future of Life Institute (2020). Asilomar AI Principles. *Future of Life Institute*. Retrieved from <https://www.bbc.com/news/technology-53165286>
- Godzinski, R. J. (2005). (En)Framing Heidegger's philosophy of technology. *Essays in Philosophy*, 6(1), 115–122.doi: 10.5840/eip20056120

- Greenemeier, L. (2017, June 2). 20 years after Deepblue: How AI has advanced since conquering chess. *Scientific American*. Retrieved from <https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>
- Gross, T. (2018, March 15). The evolution of artificial intelligence. *National Public Radio*. Retrieved from <https://www.npr.org/2018/03/15/594073235/the-evolution-of-artificial-intelligence>
- Gunkel, D. J. (2012) *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Cambridge: MIT Press. doi: 10.5840/ipq201353110
- Gunkel, D. J. (2018). *Ars ex machina: Rethinking responsibility in the age of creative machines*. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology and ourselves* (pp. 221-236). New York, NY: Peter Lane Publishing. doi: 10.3726/b14399
- Gunkel, D. J. (2019). Other things: AI, robots, and society. In Z. Papacharissi (Ed.), *A networked self and human augmentics, artificial intelligence, sentience* (pp. 51-68). New York, NY: Routledge. doi: 10.4324/9781315202082-5
- Guzman, A. L. (2016). The messages of mute machines: Human-Machine Communication with industrial technologies. *Communication+1*, 5(1). doi: 10.7275/R57P8WBW
- Guzman, A. L. (2018). “What is human-machine communication, anyway”. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology and ourselves* (pp. 1-28). New York, NY: Peter Lane Publishing. doi: 10.3726/b14399
- Guzman, A. L. (2019). Beyond extraordinary: Theorizing artificial intelligence and the self in daily life. In Z. Papacharissi. (Ed.), *A networked self and human augmentics, artificial intelligence, sentience* (pp. 83-96). New York, NY: Routledge. doi: 10.4324/9781315202082-7
- Guzman, A. L. & Lewis, S. C. (2019). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society*, 22(1), 70-86. doi: 10.1177/1461444819858691

- Gehl, R. W. & Bakardjieva, M (Eds.) (2017) *Socialbots and Their Friends: Digital Media and the Automation of Sociality*. New York, NY: Routledge.
- Gitlin, T. (2002). Media sociology: The dominant paradigm. In D. McQuail (Ed.), *McQuail's reader in mass communication theory* (pp. 20-35). London, UK: Sage Publications.
- Go Ratings (2016, November 20). <https://www.goratings.org>
- Hall, J. S. 2001. Ethics for Machines. *KurzweilAI.net*. Retrived from <http://www.kurzweilai.net/ethics-for-machines>doi: 10.1017/CBO9780511978036.005
- Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York, NY: Harper & Row.
- Hunt, E. (2016, March 24). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>
- Innis, H. A. (2007). *Empire and communications*. Lanham, MD: Rowman & Littlefield.
- Jacobs, R. J. K. (2003). Computers in human-computer interaction. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 147-149). Mahwah, N.J: Lawrence Erlbaum Associates.doi: 10.1080/00140139.2013.808852
- Jones, S (2014) People, things, memory and human-machine communication. *International Journal of Media & Cultural Politics* 10(3), 245–258.doi: 10.1386/macp.10.3.245_1
- Johnson, D.G. (2006). Computer Systems: Moral Entities but not Moral Agents. *Ethics and Information Technology* 8, 195-204.doi: 10.1007/s10676-006-9111-5
- Latour, B. (2007). *Reassembling the social: An introduction to actor-network-theory*. Oxford, UK: Oxford University Pres.
- Lasswell, H. D. (1972). The structure and function of communication in society. In W. Schramm & D. F. Roberts (Eds.), *The process and effects of mass communication* (Revised Edition, pp. 84-99). Urbana, IL: University of Illinois Press.
- Leetaru, K. (2016, March 24). How Twitter corrupted Microsoft's Tay: A crash course in the

- AI in the dangers of the real world. *Forbes*. Retrieved from <https://www.forbes.com/sites/kalevleetaru/2016/03/24/how-twitter-corrupted-microsofts-tay-a-crash-course-in-the-dangers-of-ai-in-the-real-world/#2dd9a49926d2>
- Levinas, E.(1969). *Totality and Infinity: An Essay on Exteriority* (A. Lingis, Trans.). Pittsburgh, PA: Duquesne University.doi: 10.1007/978-94-009-9342-6
- Lohr S, G. (2018, February 9). Facial recognition is accurate, if you're a white guy. *New York Times*. Retrieved from <https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>
- Lyotard, J. F. (1993). *The postmodern condition: A report on knowledge* (G. Bennington &B. Massumi, Trans.). Minneapolis, MN: University of Minnesota Press.doi: 10.2307/1772278
- Matthias, A.(2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology* 6, 175-183. doi: 10.1007/s10676-004-3422-1
- McDowell, Z. J. & Gunkel, David J. (2016). Introduction to “Machine Communication,” *communication+1* 5 (1).
- McLuhan, M (1994) *Understanding Media: The Extensions of Man*. Cambridge, MA: MIT Press.
- Mead, G. H. (1967). *Mind, self, & society: From the standpoint of a social behaviorist* (Vol. 1). Chicago, IL: University of Chicago Press. doi: 10.1177/000271623517900175
- Meadow, C. T. (1970). *Man-machine communication*. New York, NY: Wiley-Interscience.
- Metz, C. (2016, March 10). Google's AI wins a pivotal second game in match with Go Grandmaster. *Wired*. Retrieved from <https://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/>
- Meyrowitz, J. (2002). Media and behavior-a missing link. In D. McQuail (Ed.), *McQuail's reader in mass communication theory* (pp. 99-108). London, UK: Sage Public
- Microsoft. (2016). Learning from Tay's introduction. Official Microsoft Blog. <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction>

- Mooers, C. N. (1959). The next twenty years in information retrieval: Some goals and predictions. In Papers Presented Western Joint Computer Conference(pp. 81–86). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1457853> doi: 10.1002/asi.5090110306
- Mowbray, M. (2002). Ethics for Bots. Paper presented at the 14th International Conference on System Research, Informatics and Cybernetics. Baden-Baden, Germany. 29 July-3 August. <http://www.hpl.hp.com/techreports/2002/HPL-2002-48R1.pdf>
- Mumford, L. (2010). *Technics and civilization*. Chicago, IL: The University of Chicago Press.
- Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 10, 17.
- Nass, C. & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, 19 (4), 504–527. doi: 10.1111/j.1468-2958.1993.tb00311.x
- Nass, C, Steuer, J. & Tauber, ER (1994) Computers are social actors. In: Proceedings of the SIGCHI conference on human factors in computing systems, Boston, MA, 24–28 April, pp. 72–78. New York: ACM.
- Oberquelle, H., Kupka, I., & Maass, S. (1983). A view of human-machine communication and co-operation. *International Journal of Man Machine Studies*, 19, 309–333. doi: 10.1016/S0020-7373(83)80057-1
- Peters, J. D. (1999) *Speaking into the Air: A History of the Idea of Communication*. Chicago, IL: University of Chicago Press.
- Peters, J. D. (2015) *The marvelous clouds: Towards a philosophy of elemental media*. Chicago, IL: University of Chicago Press.
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226668253.001.0001
- Rafaeli, S. (1988). Interactivity: From new media to communication. In R. Hawkins, J. M. Wiemann, & S. Pingree (Eds.), *Advancing communication science: Merging mass and*

- interpersonal processes* (pp. 110–134). Newbury Park, CA: Sage Publications.
- Rieland, R. (2018, March 5). Artificial intelligence is now used to predict crime. But is it biased? *Smithsonian Magazine*. Retrieved from <https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/>
- Reeves, B., & Nass, C. I. (1998). *The media equation*. Stanford, CA: CSLI Publications.
- Riley, V. (1989). A general model of mixed-initiative human-machine systems. Proceedings of the Human Factors Society 33rd Annual Meeting. doi: 10.1177/154193128903300227
- Risely, J.(2016). Microsoft’s Millennial Chatbot Tay.ai Pulled Offline After Internet Teaches Her Racism. *GeekWire*. Retrieved from <http://www.geekwire.com/2016/even-robot-teens-impressionable-microsofts-tay-ai-pulled-internet-teaches-racism>
- Rogers, E. M. (1986) *Communication Technology: The New Media in Society*. New York, NY: Free Press.
- Rogers, E. M. (1997) *A history of communication study: A biographical approach*. New York, NY: Free Press.
- Ruin, H. (2014). Ge-Stell: Enframing as the essence of technology. In B. W. Davis (Ed.), *Martin Heidegger: Key concepts* (pp. 183-194). New York, NY: Routledge. doi: 10.1017/UPO9781844654475.014
- Russell, S. & Norvig, P. (2009). *Artificial intelligence: A modern approach*. London, UK: Pearson.
- Sanders, K. (2003) *Ethics and journalism*. London, UK: Sage. doi: 10.4135/9781446220467.n1
- Schellenberg, S. (2020, April 29). How biased algorithms perpetuate inequality. *New Statesman*. Retrieved from <https://www.newstatesman.com/science-tech/2020/04/how-biased-algorithms-perpetuate-inequality>
- Schramm, W. (1973). *Men, messages, and media: A look at human communication*. New York, NY: Harper & Row.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical*

- Journal*, 27(3,4), 379–423, 623–656. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication* (12th Printing). Urbana, IL: University of Illinois Press. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Siegel, R. (2016, October 24). Twenty years later, humans still no match for computers on the chess board. *National Public Radio*. Retrieved from <https://www.npr.org/sections/alltechconsidered/2016/10/24/499162905/20-years-later-humans-still-no-match-for-computers-on-the-chessboard>
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550, 354-359. doi: 10.1038/nature24270
- Simon, J. (2015). “Distributed epistemic responsibility in a hyperconnected era”. In L. Floridi (Ed.), *The onlife manifesto: Being human in a hyperconnected world* (pp. 145-159). London, NK: Springer Open. doi: 10.1007/978-3-319-04093-6_17
- Slote, M. (1995) *From morality to virtue*. New York, NY: Oxford University Press. doi: 10.1093/0195093925.003.0006
- Suchman, L. A. (2009). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511808418.008
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460. doi: 10.1093/mind/LIX.236.433
- White, G. (2019, May 13). Use of facial recognition tech “dangerously irresponsible.” *BBC*. Retrieved from <https://www.bbc.com/news/technology-48222017>
- Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine* (Seventh Printing, Vol. Kessinger Legacy Reprints). New York, NY: John Wiley & Sons. doi: 10.7551/mitpress/11810.003.0012
- Winblad, A. (2017). Tech generations: The past as prologue. In D. Franklin (Ed.), *Megatech: Technology in 2050* (pp. 64-75). London, UK: The Economist.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*.

San Francisco, CA: W.H. Freeman. doi: 10.1177/000271627642600162

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media &*

Society, 8(3), 401–419. doi: 10.1177/1461444806061951

