圖靈沒有料到的 AI 歷史的今日：

林守德[1]教授「當人類智慧碰到人工智慧」講座紀實

紀錄整理：蘇上雅[2]

人工智慧是一個有趣的學門，和其他領域往往發展後逐漸普及、停滯不同，AI 的發展史有所謂「三起二落」，起了之後掉下、掉下後又再起。每一次興起，都是當一個新技術被開發、給大家期待，但後來人們發現，新技術仍未達成預期目標，便逐漸掉了下來；接著又一個新技術來了，又會再起來。只要人工智慧的謎題沒有被全然解開，起落就會一直持續下去。林守德教授「當人類智慧碰到人工智慧」講座分為三部：首先，回到基礎，談談什麼是人工智慧、我們至今走了多遠？第二，談人工智慧現正在哪個點上、距離最終目標還有多遠？最後，談人工智慧可以帶我們走多遠、多久，帶我們走向什麼未來？

一、 AI 至今走了多遠？人工智慧發展的三起二落

人類探求人工智慧的歷史，可以回溯至 17、18 世紀歐洲世界。當時一些哲學家、數學家，包括當今大家熟知的笛卡兒、霍布斯、羅素等人，已經開始發想：困難的數學推理，是否可能機械化？是否能夠透過一套形式、制式推理過程，來解決令數學家、哲學家困惱的難題。形式推理（formal reasoning）能力，是人類最早對於人工智慧的期待[3]。

時至 1950 年代，被譽為「人工智慧之父」的圖靈（Alan M. Turing）與同期研究者提出，人工智慧的真諦在於「能與人類對話」。著名的「圖靈測試（Turing Test）」，將一機器具有「智能」的標準定性為：當一台機器與人類展開對話，卻不被辨識出其機器身分。「對話」聽起來簡單，實際上，對於文字和語義都需要有深刻的理解，圖靈並沒有預見對話的困難性，樂觀地預言此難題會在 2000 年以前被突破；但時至 2018 年，此難題尚未被攻克。1960 年代，以 Simon, Shaw, Newell 等為首的研究者，認為人工智慧旨在「目標搜尋」，亦即不斷嘗試找到答案的過程。當時 AI 曾被應用於解決困難的代數問題。問題在於：當答案所在搜尋空間太大，又或訊息來源不全盤揭露時，搜尋要耗費太多時間，效果不切實際。同一時期，有人則認為，AI 是擅

[1] 國立臺灣大學資訊工程學系暨研究所教授，建置臺灣法實證研究資料庫第五期實施計畫共同主持人。
[2] 建置臺灣法實證研究資料庫第五期實施計畫專任研究助理，臺灣大學科際整合法律研究所碩士。
[3] 例如：霍布斯曾指出：「推理就是計算（reason is nothing but reckoning）」。

長玩遊戲的電腦。事實上，人工智慧能夠進展，很大一部分確實是在遊戲[4]中產生的：遊戲的規則容易系統化、變數有限，最重要的是，遊戲可以重玩。透過遊戲，電腦可以不斷嘗試錯誤、學習，然後越來越厲害。不過，此說也面臨質疑：用遊戲來評價 AI 的好壞或能力，是否合適？遊戲總是人生的簡化版本：圍棋的盤面有限，人類世界則複雜得多；下棋輸了可以再來一盤，人生許多事情卻無法重新來過。以圖靈測試揭開序幕的人工智慧，在 1970-80 年代面臨第一個寒冬。

經過了第一個起落，人工智慧邁入第二波發展。以沃倫．麥卡洛克（Warren McCulloch），唐納德．赫布（Donald Hebb）等人為首的研究者認為，知識被 encode 在網絡裡面。這些研究者在思考的是：如何將知識從網絡中萃取出來、記憶進去，進而解決人類生活中碰到的問題？另一群人則認為，人工智慧是「專家系統」機制，適用於處理知識密集型的任務，如醫療、司法裁判等等。透過「知識庫」機制存放大量專業知識[5]，加上能從事推論的「推論機」機制，AI 就能夠回答原先專家才能夠回答的問題。以實作專家系統 Prolog 的運作為例：

有一道問題：「若蔣中是蔣經的爸爸、蔣經是蔣孝的爸爸，則誰是蔣中的孫子？」

Prolog 系統必須事先被給定一個邏輯運算規則「孫子（A，B） :-爸爸（X，A） AND 爸爸（B，X)」[6]，並且被置入大量相關資料，例如：爸爸（蔣經，蔣孝），爸爸（康熙，雍正），爸爸（雍正，乾隆）……等等。在上二前提滿足下，面臨上述提問，Prolog 會先搜尋資料庫中與「孫子」有關的邏輯規則，找出邏輯式「孫子（A，B） :-爸爸（X，A） AND 爸爸（B，X)」，進而從資料庫裡存放的父子關係 data 中，抓取出「爸爸（蔣中，蔣經)」、「爸爸（蔣經，蔣孝)」，進而推導出「孫子（X，蔣中)」之 X 為「蔣孝」。

但是，專家系統存在難以克服的問題：一來，知識不易產生，而且知識需要持續維持（maintain）、更新（input），才能夠確保最新。最麻煩的難題在於：沒有碰過的東西，AI 即無法推論[7]。1984 年 Douglas Lenat 發展的 Cyc 計畫，提出人工智慧即是「知識工程」，目標是蒐集全世界所有知識，並用電腦可理解的語言表達。然，究竟如何蒐集所有知識？又何謂適合電腦理解的語言？皆是有待釐清的難題。而手動添加知識永遠比知識增加的速度慢、缺乏對知識的驗證機制、應用目標不夠明確等等，也都是「知識工程」面臨的挑戰。因為上述種種難關無法克服，第二波 AI 在 90 年代面臨寒冬。

在介紹 1990 年代以降第三波人工智慧的發展以前，林守德教授先帶大家釐清三種不同的

[4] 所謂遊戲，依訊息態樣可細分為「完全訊息」遊戲、「部分訊息」遊戲兩種，前者如西洋棋、圍棋，AI 可以看到整個盤面、沒有任何不確定性；後者則如「德州撲克」、「星海爭霸」等等，AI 看得見自己的牌、附近局面，但部分訊息需要自己去猜，因此搜尋空間更大、比前者困難。

[5] 專家系統通常使用"if…then…"規則來表示。例如，一醫療專業知識系統，可以設定「if 吃多、喝多、尿多 then 有較大機會得到糖尿病」。

[6] 此邏輯模型意指：若 A 是 B 的孫子，則存在 X 使「X 是 A 的爸爸」且「B 是 X 的爸爸」。

[7] 例如上述舉例的實作專家系統 prolog，必須以事先給定的邏輯推論模型、知識庫來做判斷。

AI 概念：

1. 強人工智慧（Strong AI）：可思考且有心靈/自我意識的 AI。
2. 泛人工智慧（Artificial General Intelligence，簡稱 AGI）：外在所有行為就像一個有廣泛智慧的人類，在所有領域都做得一點點好。
3. 弱人工智慧（Weak AI）：又稱「狹隘人工智慧」，指在特定領域上，表現得很有智慧。

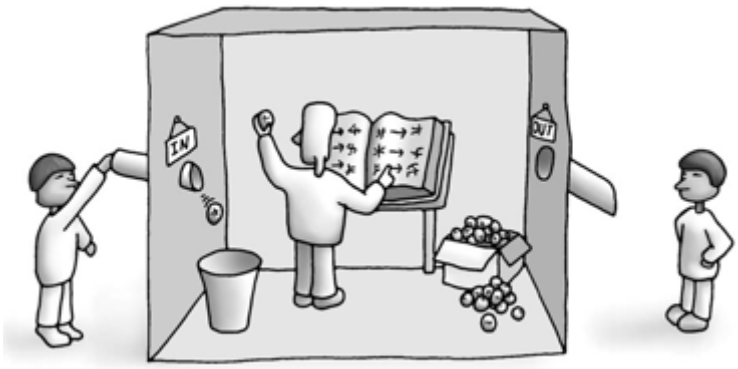

圖 1：哲學家 Searle 提出的 "Chinese Room Argument" 示意圖。圖片引用來源：https://goo.gl/images/Tp54Tm

林守德教授提到，在 90 年代以前，AI 研究者還懷抱著「如何使 AI 變得和人類智慧一樣」的理想，這種思維，也連結到哲學家 John Searle 提出的著名反 AI 論述：「中文房間論證（Chinese Room Argument）」。Searle 假設有一不懂中文的人，被關到一個「中文房間」裡，看到一篇中文，查詢裏頭的各式字典、百科全書翻譯後，再把查到的東西抄上丟出去，外頭的人還以為裏頭的人懂中文，實際上卻不然。「中文房間論證」用來強調：AI 不可能和人類智慧一樣（下圖）。但從弱人工智慧的發展角度來看，只要學習的外顯成果達到目標就好，是不是以和人類一樣的方式產生智慧、是不是「真有智慧」，不是重點。90 年代以降，有目的性的弱人工智慧就成為主流。

弱人工智慧為何會崛起？主要有三個原因：首先是 Internet 崛起，使大數據資料（Big Data）蒐集更便利。第二，電腦計算能力變快、儲存空間變大，使過去無法實現的演算法（如深度學習）變得可行。林守德教授提到過去少人注意、實際上非常重要的：第三，有明確的驗證機制、具體目標、各種競賽，來評估 AI 的能力，進而促進研究發展。舉例來說，世界知名 AI 西洋棋程式 Deep Blue，評估具有智慧的目標即「下棋要贏過人類」，在 1996 年就達成目標。2004 年由美國國防部資助的 DARPA Grand Challenge 提供高額獎金給「第一台橫越內華達沙漠的自動車」，2005 年即有 5 台車成功橫越。「有明確的 goal，就能夠去 measure 機器人有沒有達成。」林守德教授提到資料探勘領域最重要的年度比賽 ACM KDD Cup，每年比賽的題目都不同[8]，但目標一致：參賽者建構預測模型，競逐預測準確率，台大團隊組隊參與賽事，多次榮獲優異成績。The Netflix Challenge 亦是著名 AI 賽事之一，Netflix 公司為增進營收舉辦競賽，目標為

[8] 例如：乳癌預測、使用者行為預測、學習成果預測、樂迷推薦歌單系統等等。

「公司推薦片單準確率提升 10%」，使 AI 推薦系統演算法突飛猛進。另外，IBM Watson 益智問答、google 的 AlphaGo 圍棋，都是藉由蒐羅大量資訊，在特定領域上贏過人類的例子。只是，棋局可以模擬、下輸可以重開一局，現實生活中人工智慧沒辦法無窮「模擬」現象，因此還是需要資料。而近日 AI 紅起來的原因即是：能夠進行「深度學習（deep learning）」，多層神經網絡不同層次各司其職，抓出不同層次的資訊。深度學習是機器學習的一支，目前已主導了現在的 AI。

為什麼我們要瞭解這麼長久的 AI 發展/進展史？林守德教授指出：因為技術可能逆襲、歷史可能重來。時下很「夯」的神經網絡 AI 模型，早先發展成果不如其他較簡單的 AI、被認為是 dead end，但客觀環境、主觀技術成熟後，現在反而展露頭角。換言之，瞭解不同的 AI 分支做了什麼、為什麼會走到現在的樣子，是非常重要的，因為他們仍可能在未來扮演重要角色。下一波 AI 革命，有可能就是來自這些以往被拋棄的知識分支。

二、 AI 距離終點有多遠：人工智慧 vs.人類智慧

於演講第二部分，林守德教授說明現階段主流 AI 所從事的「機器學習（Machine Learning）」之原理、特徵、與人類產生智慧過程的差異。機器學習，簡單來說就是：從大量資料中學習函數 f（x）。f（x）可能有各種用途，例如分類、機率等等。不過，AI 擅長處理的是二選一的「是非題」、多選一的「選擇題」，對於需要深入論述的「問答題」[9]，AI 並不擅長。

人工智慧與人類智慧的不同，可以「作詩」為例加以說明。微軟公司研發了一款取名為「小冰」的 AI，不僅出了詩集，也曾匿名投稿被出版社接受刊登。林守德教授說明，「小冰」生出一首詩的過程，概略可分為「辨識圖片」、「預處理圖片關鍵字」、「前後生成」等三個步驟：於第一步，透過圖片辨識技術[10]，「小冰」從輸入的圖片中，抽取與圖片相關的字詞。於第二步，小冰將圖片關鍵字比對原詩訓練集[11]，從中萃取出頻率較高的名詞與形容詞作為關鍵字，同時從訓練集中找出經常與關鍵字搭配使用的名詞/形容詞/副詞，生成詩的關鍵詞組。第三步，「小冰」採用「前後遞迴生成 （recursive generation）」方式，從關鍵字詞往前、往後各生出關聯字，以此類推，直到向前生到句首、向後生到句尾為止，一句含有關鍵字的詩句便生成完成。上述遞迴生成過程，是仰賴「語言模型」來生字，而所謂語言模型，是電腦從 50000 行詩句裡，針對給定搜尋的關鍵字，記錄該關鍵字下一字出現的機率，統整出的字詞搭配機率模型；「小冰」便依憑此機率規則前後生字。又，為避免以既存詩篇為模型來生字，出現「抄襲」問題，「小冰」尚有「自動評價」系統，能根據生成候選詩句的流暢度、詞性、原創度加以評分，進而篩選出最好的詩句選擇。此外，為使詩前、後句內容連貫，在上句詩生成後，會將該詩句所含資

[9] 例如：「為什麼」支持/反對年金改革？
[10] 透過卷積神經網絡達成 (GoogleNet, Szegedy et al, 2015.)，在 AI 圖片辨識訓練過程中，要餵入數以萬計的檔案。
[11] 餵入五萬多首 1920-80 年代的現代詩。

訊以編碼形式傳到下句，使後續詩句在生成時都會考慮前句詩意，確保句與句之間內容的連貫性。

不過，「小冰」寫作的詩仍有其侷限，包括使用特定慣用字詞、不能控制較多變化（例如主題、情緒生成）、對特定關鍵字生成困難（只能透過關鍵字擴張來修正）、圖片辨識錯誤等等，另外，透過辨識圖片來生詩，小冰接收到的訊息是靜態的，難以辨識出「動作」，生詩過程受限於所看到的東西、辨識能力，創意有限。總歸來說，比較人類與 AI「小冰」作詩的過程，可以發現下述差異（下表 1）：

表 1：人類寫詩與 AI 作詩的差異（表格提供：林守德教授）

	人類寫詩	AI「小冰」作詩
動機	抒發情感	人類按下執行
意識到在寫詩？	是	否
了解什麼是詩？	是	否
機率計算？	否	是
深度？	可深可淺	類似

林守德教授指出，上述差異可進一步說明，人類智慧與人工智慧非常不同。對人類而言，學齡前兒童就具備的「自我意識」，AI 做不到；一般運動、理解、推理等基礎或中階智慧，對 AI 來說也是很大的挑戰；但對人類來說屬於高階智慧的「決策」，AI 能力則遠勝於人。這些差異也映證了「莫拉維克悖論」：人覺得困難的東西，電腦覺得簡單；電腦覺得難的東西，人覺得簡單。其原因，其實也很合理：因為人類和 AI 的知識構造，是不一樣的。林守德教授認為，在此前提之下，應該重新思考的是：如何使人類做得很好的地方，AI 也能做得好？如何讓 AI 已經做得很好的地方，可以做得更好，以幫助人類生活應用？談到這裡，林守德教授以日本研發、原先以「考取東大」為目標的人工智慧程式「東 Robo 君」（圖 2）近年宣布放棄考東大為例，說明人工智慧與人類智慧的不同：入學考試測驗的是在給予少量資訊前提下的理解力，人工智慧的長處在於大量資訊下「學習」，無法在給予少量資訊下自己「思考」。

圖 2：以考取東大為目標的 AI「東 Robo 君」。

圖片來源：https://goo.gl/images/kYzQXA

三、 人工智慧能帶人類走多久？

最後，林守德教授談到人工智慧發展的未來。對於 AI 未來發展、對人類社會的影響，分為樂觀論、悲觀論兩派。樂觀論者認為「泛人工智慧」終將發展成功，世界將會增加許多「有能力但不自私」、「願意犧牲奉獻」的代理人[12]。樂觀論者認為，AI 是下一次資訊革命的工具，能夠為人類所利用，創造更多機會、可能性。

另一群人則採悲觀論，認為 AI 的發展將對人類社會產生不良影響。常見的論述是認為 AI 可能做出危害人類的事。不過，林守德教授提到，必須澄清的是：AI 並不像許多科幻片的設定會產生自我意識、希望統治人類；需要擔心的應當是：AI 作為一技術，因為發明者、使用者動機不良，或者因為訓練不夠、不夠聰明，而在遂行命令時，做出危害人類的事情。此外，林守德教授分享先前參與臺大法學院「人工智慧、心靈與演算法社會」論文討論會[13]，法律系顏厥安教授提出一有趣的思維：「演算法社會」，其概念意即，當 AI 越來越強、融入社會，人們為了和 AI 競爭、與 AI 溝通，而必須要用 AI 的模式來思考。比如 AI 慣用知識生產方式是機器學習，人們的思考模式也會漸漸傾向以 AI 演算法的方式來思考，而屆時社會的運作方式，將與現在大不相同。

另外， Yuval Noah Harari 提出，AI 發展將抵消民主政治優勢、侵蝕自由與平等，使權力集中在少數菁英身上。Harari 的看法是：民主政治之所以存在、被偏好，是因此制度乃「代議政治」、集眾人之智，避免獨裁者決策不夠周全。一旦 AI 智慧決策發展起來，「眾人之智」很可能不再被需要，以至於未來 AI 說什麼、人照做就好了，結果可能使少數權力擁有者，不需經過實質的諮詢，就做出決策。最後林守德教授也提到，AI 目前安全性仍堪慮，例如深度學習的人臉辨識技術其實有可能被破解而騙過 AI。

總結以上討論，林守德教授認為，純粹機器（深度）學習，是很難達到泛人工智慧的，短時間內 AI 不大可能具備意識與心靈能力；結合其他 AI 技術（例如知識導向），或許是未來可行的方向。不過，林守德教授亦呼應「演算法社會」觀點指出，若 AI 做出決策、預測能更準確，自然可能使人類決策模式轉變、更信任（依賴）AI 的判斷。最後，林守德教授總結：由 AI 發展史可見，人工智慧領域是技術帶領應用：新的技術出現後，才帶動下一波人工智慧應用的高峰。現在有越來越多團隊投入以人工智慧做基礎科學研究；然而，距離一個安全、透明、有倫理觀念、能夠與人類協作的「強人工智慧」，在研發上仍有很長的路要走。不論從研究或應用的角度，這些都是尚待鑽研討論的議題，需要大家未來一起努力、共同參與。

---

[12] 因為電腦不知道「自我」是什麼，當然不會有「自私」行為。

[13] 2018 年 6 月 8 日論文討論會紀實，亦收錄於《基礎法學與人權研究通訊》，第 21 期，頁 22-25，篇名為〈機器-人-明日社會：「人工智慧、心靈與演算法社會」討論會紀實〉（電子報請參：http://tadels.law.ntu.edu.tw/upload/edm_file/Issue21.pdf#page=22 ）。