

人工智能 與大數據 面對的波折與排斥



- 余創豪 chonghoyu@gmail.com

人工智能與大數據時代已經來臨

不用我多說，相信許多讀者都知道人工智能、機械學習與大數據、數據科學的時代已經來臨，已經影響着我們工作和生活的每一個層面。今年五月，萬國商業機器公司

（IBM）首席執行官阿文德·克里希納（Arvind Krishna）宣稱，近八千個該公司的工作崗位將會被人工智能取代。在接下來的五年中，人工智能可能會取代多達 30% 不需要跟客戶面對面接觸的工作。在不久的將來，機械學習演算法會對金融、會計、人力資源等領域的工作人員構成重大威脅。不過，這種轉變的光明面是，預計到了 2030 年，人工智能將為全球經濟貢獻十六萬億美元。

人工智能、機械學習與大數據、數據科學是兩組不同的東西，我在其他文章已經解釋過，在這裏我只簡單地介紹一下，機械學習是採用數據去不斷增強預測準確性的演算法，它是基於模仿人類大腦神經元網絡的人工智慧技術；數據科學則是嶄新的數據分析方法，它與人工智能有一個共通點，就是基於數據去探索現象或者人類行為的模式

（patterns），而不是純粹倚賴抽象的數學模型；而機械學習則需要龐大的數據，2006 年英國數學家克萊夫漢比（Clive Humby）指出：「數據就是新的燃油。」一部功能強大的機器需要龐大的能量去發動，同樣道理，一套精密的演算法需要很多資料，才能夠得出準確的結果。

由人工智能嚴冬到電腦戰勝人類

上述的趨勢好像是理所當然，其實，任何創新都會受到質疑和排斥，人工智能與數據科學並不是例外。今天兩者大紅大紫，但很多人都不知道，其實它們的發展都經歷過了一



波三折。1958 年，康奈爾大學心理學家弗蘭克·羅森布拉特（Frank Rosenblatt）開發出現代神經元網絡的原始模型，這部能夠學習的機器名為感知器（perceptron），但它只能夠做一個十分簡單的工作，就是區別一張卡片上面是否印上方塊的符號。可惜，一些學者對這發明嗤之以鼻，於是乎這研究項目便無疾而終。雖然還有其他學者繼續研究人工智能，但是成效不大，1970 年代至 1990 年代初期，這領域經歷了兩次「人工智能嚴冬」（AI winters），在這段期間，美國政府和很多研究基金會都不願意撥款予人工智慧的研究，學術期刊亦不願意刊登關於神經元網絡的文章，電腦學者楊立昆

（Yann LeCun）即使在論文中顯示出支持神經

元網絡有效性的證據，但審稿人仍然將他的文章投籃。不過，由 1990 年代中期開始，人工智能突飛猛進，令人刮目相看，例如 1997 年萬國商業機器公司的「深藍」（Deep Blue）電腦擊敗了世界棋王，2011 年，同一公司的「華生」（Watson）電腦在知識問答比賽 Jeopardy 中戰勝了台柱，限於篇幅，我不能一一細說其他里程碑，無論如何，如今天聊天機械人 ChatGPT 和人工智能生成藝術軟件 Midjourney、Stable Diffusion 已經是家傳戶曉的名字。

由古典統計學到大企業的大數據

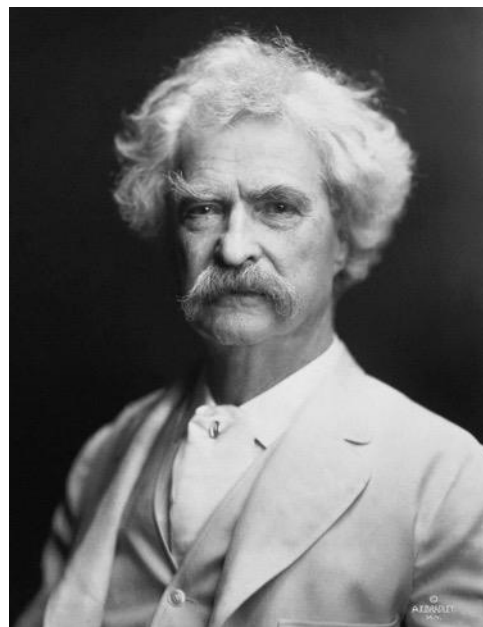
現在談一談數據科學，傳統的統計學基礎奠定於十九世紀末期和二十世紀初期，由 1960 年代開始，已經有不少學者批評這個老古董充滿邏輯上的問題，例如約翰·圖基（John Turkey,）、約翰·錢伯斯（John Chambers）、威廉·克利夫蘭（William Cleveland）利奧·布雷曼（Leo Brieman），他們提出了許多創新的數據分析方法，這些方法能夠彌補古典統計學的缺點，例如圖基認為在開始研究之前不要妄下假設，應該以開放的態度去探索資料所顯示出來的模式；克利夫蘭主張不要過分倚賴數字答案，而是要通過圖像去觀察整體現象和不同事物之間的關係；布雷曼則開發出功能強大的演算法。以上的方法就是現

代數據科學的前身，但如你所料，他們都是講人自講，絕大多數教科書和大學課程都不願意引進這些新觀念，學術期刊亦排斥這些「非標準化」的方法。不過，隨着大數據逐漸普遍，商業機構開始意識到傳統方法的局限，舉例說，亞馬遜每一天有過百萬宗交易、谷歌每一天有超過八十億人次搜尋、臉書每一天有超過十億貼文，分析師不得不轉向適用於大數據的數據科學。

不自覺地找尋藉口

可以說，人工智能與數據科學都是鹹魚翻生！我們不應該因著一時一地的成敗而對新事物蓋棺定論。有時候你會輸掉很多場戰役，但最後仍然贏了戰爭（lose the battles but win the war）；有時候一些做法在戰術上可能是愚蠢的舉動，但在整體戰略上卻是神來之筆。

人類都是充滿惰性的，都傾向於蕭規曹隨、照本宣科，這包括筆者在內。由於嶄新的觀念對原有秩序構成了威脅，故此人們會不自覺地提出反對的理由，但往往這只不過是原地踏步的藉口。例如曾經有人對我說：「並不是每一個新點子都是可行的，在美國大約 90% 的創業公司都以失敗告終。」我的回應是：「那麼應該找出這個新主意是屬於成功的 10% 還是失敗的 90%，若果完全抹煞任何新嘗試，那麼成功率就是 0%！大文豪馬克吐溫曾經說：『我們對沒有做的事情感到遺憾，多於我們對做了的事情而感到遺憾。』我情願嘗試過而失敗，而不是什麼也不做。」



亦有人對我如此說：「並不是最好的東西一定會勝出，大約 50 年前，日本索尼公司推出的錄影帶格式 beta 比 JVC 的 VHS 優勝，但最後市場選擇了 VHS；蘋果麥金塔的功能和使用者介面都勝過微軟視窗和英特爾晶片的電腦，但後者雄霸了個人電腦市場。」正如我在上面提過，我們不應該只執着一時一地的成敗，日本索尼陸續有其他創新的產品，在無反攝影機市場上，索尼已經超越了尼康和錦囊，擁有最大的市場份額，去年索尼是日本第九大企業；2011 年 JVC 和 Kenwood 合併之後，在電子業中已經無足輕重。蘋果電腦現在是全世界市值最高的企業，微軟緊隨其後，而蘋果採用台積電五納米技術製造的 M2 Ultra 晶片，已經將英特爾遠遠拋離。

一些反對採用人工智能和大數據的理由如下：「大多數教科書仍然涵蓋傳統的內
容。」「大多數學術期刊仍然沿用傳統的標準。」「大數據無關重要，我可以採用現有的軟
件去完成所有工作。」然而，學生畢業後所任職的公司，真的會要求他們施展教科書教
導的東西嗎？有幾多個畢業生的工作性質是投稿予學術期刊？六月初，根據一個一千人
的樣本，美國〈發現學生債務〉（**Discovery Student Loans**）一項調查研究顯示：**68%** 的
父母擔心自己的孩子在大學畢業後將會背負大量債務，**46%** 的父母質疑大學是否值得投
資。我完全理解他們的關注，沒有人會願意投資十幾萬美元去接受訓練，但四年之後卻發
現學習到的技能已經過時，但事實上，人工智能與數據科學已經是不可逆轉的趨勢，在文
章的開首提過，未來很多工作都會被人工智能取代。無論如何，危機也是契機，到了
2030 年，預計人工智能將為全球經濟貢獻十六萬億美元，如果學生能夠掌握人工智能與
數據科學的技術，便可以在這龐大的經濟效益中分一杯羹。

2023 年 6 月 10 日

〈人工智能與大數據面對的波折與排斥〉

原載於澳洲《同路人》雜誌

[更多資訊](#)