臺灣出版與閱讀 2616-5074
111 年第 4 期（總號第 20 期） 民國 111 年 12 月（2022.12） 頁 40-46 國家圖書館

在想象中想像人工智能

張惟智 國立政治大學哲學系博士班研究生

費茲傑羅（F. Scott Fitzgerald）曾寫過令人難忘的一句話：「第一個創造意識之人所犯的罪是大罪。」我可以了解他為何這麼說，但他的責難並未道出故事全貌，只適用於當意識如此赤裸地揭露出自然的不完美時的沮喪時刻。另一半的故事應該完全用來讚美此一創造，它是所有創造和發現的啟動器，它以失落與悲傷換得了歡欣與慶祝。意識的浮現開啟了一條讓生命值得活的路。了解它的誕生過程只會強化生命的價值。

——安東尼歐．達馬吉歐（Antonio Damasio）（注 1）

人工智能背後的哲學思想

暴雨連綿的十月下旬，興許是惱了許多人的出作入息；所幸總有那麼些偶然時刻，忘卻日常不便，沈浸在這雨景的起落與休止間，雖然沒有芙蓉花，卻也有「木末芙蓉花，山中發紅萼，澗戶寂無人，紛紛開且落」之靜默幽寂感；那日，我在此種悠晃中，意外聽見那首Player 的〈Baby Come Back〉，瞬間回神後，不覺莞爾，很是懷念；那首歌總是會讓我想起，電影變形金剛中，大黃蜂幫山姆向蜜琪道歉那一幕。

說起變形金剛，由於不同版本的世界觀設定不同，很難獨斷地說，他們最初是起源於人所創造的機器人，還是在宇宙初始時，從那渾沌與秩序之中誕生；但這之間的差異卻是讓我想起人工智能（Artificial Intelligence）；在劍橋字典中解釋：「the study of how to produce machines that have some of the qualities that the human mind has, such as the ability to understand language, recognize pictures, solve problems, and learn」，大意是創造具有人類思維能力的機器；如語言理解、識別圖片、解決問題和學習的能力。

這個解釋相當重要，因為它決定了變形金剛思維的底層邏輯，是否基於人類的理性思維框架，換言之，變形金剛究竟是機械「人」，還是其實就跟阿努納奇一樣是「外星生物」；如果變形金剛最初只是人工智能，那麼它是如何成為生命，或誕生出意識的呢？人工智能又是怎麼和創造意識聯結起來的呢？

思緒至此，身旁一對情侶間的對話流入耳府，將我從神遊中拉回；「我以人格擔保，我剛剛絕對沒有看那個正妹」、「你的人格好像不怎麼可靠」、「我以靈魂起誓，如果有偷看，這輩子都不會中樂透」、「有沒有可能你這輩子本來就不會中樂透……」。我暗自覺得好笑，卻又不免想起；人格若能夠不可靠，如果不是天生就是

壞人，那就是在人與人社會互動關係中，沒有獲得認同；靈魂若能夠起誓，如果沒有神或惡魔，亦或是一個獨立於身體或大腦之外的「我」來賞善罰惡，也是白搭。

事實上，如何理解「我」、「心靈」、「意識」、「思維」、「靈魂」、「人格」這些字詞背後所傳達的意義與其中異同，在哲學的世界裡，是一個超級複雜的問題；除了理解方法的差異（如從概念分析去理解或從整體脈絡來把握意義）就已經產生歧異與歧義，使問題更複雜化與後設化之外，這些字詞本身在某個程度上也意謂著整部哲學史，更別說這些字詞或所指稱的對象，還會隨著人類整體的一切活動（如文化、藝術、科技、倫理關係等等）的累積，也隨之被賦予新的內容或意義。

如果對人類過去怎麼思考「心靈」、「意識」、「思維」有興趣，我私心推薦一本書《心靈風暴：當代西方意識哲學的概念革命》，正如博客來上的介紹：「讓本書告訴你哲學上最核心的問題——『心靈是什麼？』」、「本書將會有條不紊、層次分明的帶領讀者來仔細探討這一哲學問題、理論或論證，讓你很快的具備一些基本知識與思路來迎接新知識的衝擊與挑戰」。「在當代，除哲學外，心理學、電腦科學、人類學、語言學、腦神經科學等等紛紛加入了這個論戰，這個當代西方最熱門的『意識問題』——『心靈是什麼？』」；這個介紹很真實，看完此書後，便能夠對「心靈是什麼？」有一個整體輪廓的理解，這也是我最初入門意識哲學的書籍，真的非常推薦（可惜的是撰稿的當下為止，我對意識哲學的理解也沒有比剛入門多多少）。

我看向身旁的那對情侶，男生正面露慌張神色，宛如經典哲學向其發出靈魂拷問「我是誰」、「我在哪裡」、「我要去哪裡」般，一副既尷尬又心虛地說著：「妳今天吃錯藥了喔？」；女生則一副看好戲模樣，嘴角不禁揚起一抹淺淺微笑反擊：「你才沒吃藥勒」。在那兩人之間，情緒滿益流露著；如果我們把「心靈」、「意識」、「思維」、「靈魂」、「人格」甚至「情感」都視為「大腦」的功能（或大腦活動所產生的結果），那麼「我」一切的意識活動，如思想、感覺、情緒都一定具有身體性的基礎；又如果思維活動是整個大腦協同運作的結果，那麼，當思維有不同的活動狀態，大腦的整體結構（或運作）也會不同。如此一來，意識能影響大腦結構，大腦結構能改變意識，也就不那麼令人驚奇了，因為只是一個東西的兩面；所以，人可以通過吃藥去影響大腦神經元的運作，並以此來改變情緒，也可以透過意識調整情緒，來影響大腦神經元運作。

從大腦的「功能」（或大腦的「結構」，亦或是把意識作為能傳輸的「數據」）來理解「意識」，或許對當代的我們並不陌生。記得小學的時候，看過一部日本漫畫銃夢（ガンム），於 2019 年改編成電影《艾莉塔：戰鬥天使》（Alita: Battle Angel），其中故事背景設定在遙遠的未來，人們能夠以機械替換身體與大腦，並且記憶能夠備份，在某個意義上而言，人能永生不死；還記得當時閱讀到一個很有趣的情節，故事中某

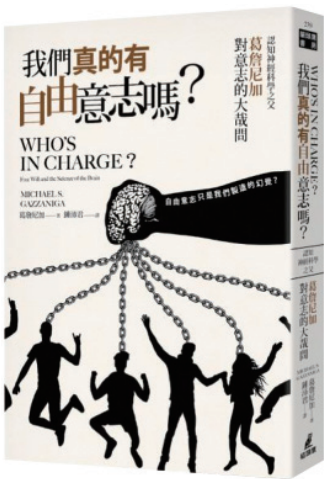
科學家將自身記憶備份，不但實現多個自己同時存在，甚至還想到用多個晶片串連提升腦力；讓當時還是小學生的我，眼界大開，直呼驚奇。雖然這些設定在今日的科幻文學與影視作品中早已稀鬆平常，例如：2018 年上演的《碳變》（改變自英國科幻小說家理查．摩根 2002 年的同名小說），故事發生在一個意識可以轉移到不同軀體的世界背景。

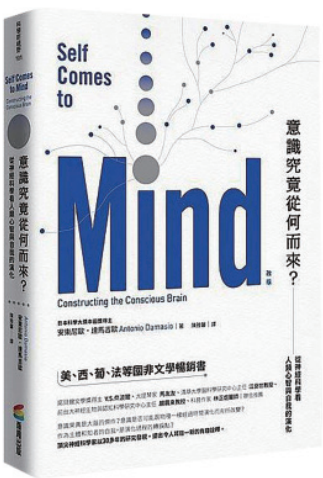
如果對透過大腦功能來理解意識有興趣，我推薦《我們真的有自由意志嗎？認知神經科學之父葛詹尼加對意志的大哉問》；這本書非常好讀，書中從自由意志和決定論者之間的爭論談起——我們每天的決定與選擇，真的是出於意志嗎？思考，會不會只是隨機的神經衝動？會不會我們以為的自由意志，其實根本就是某種奠基在反射動作與其他無意識行動下的幻覺？作者透過介紹前沿腦神經研究成果，來解釋自由意志與大腦之間的關係與新發現。

也推薦《意識究竟從何而來？——從神經科學看人類心智與自我的演化》，這本書比較像學術論文，在閱讀上稍需花些精力，當然知識量相對的不會讓讀者失望；作者提出科學研究證據，指出意識是生物體所創造出的一種過程，並不是只有人才有意識，動物、昆蟲也有意識。書中除了從內省觀、行為觀及神經觀等三種傳統觀點研究人類心智外，作者也透過演化觀重新理解意識心智史，並重新詮釋感覺的起源與多樣性。

人與人工智能的創造性

1950 年，英國科學家圖靈（Alan Mathison Turing），提出一個測試機器是否具有「智慧型式」的方法；如果一臺機器與人類對話，而沒有被辨別出其機器身分，那麼這臺機器就具有智慧





型式。當然，這也就意謂著，圖靈認為機械能夠「思考」。事實上，機器具有「智慧型式」就等於機械能夠「思考」，是一種化約論的思考，而這也是為什麼我們對於「機械通過圖靈測試，就等於機械有思維能力」這種宣稱，總是感到有其詭異之處；圖靈偷渡了概念，他在某個意義上，就預先肯定了機械有思維能力的可能性。

2015 年上映的《人造意識》(Ex Machina)，整部電影可以說是一場更加精妙的圖靈測試。主角加勒是程式設計師，在員工抽獎中，獲得前往創辦人納森的別墅度假一周的獎勵，在抵達別墅後，納森邀請加勒在對女機器人艾娃進行測試。經過前面四天醞釀；不知道是納森的安排，還是艾娃的籌劃，抑或真的是各種機緣巧合，第五天加勒不但決定要與艾娃一起逃離，甚至還一度懷疑自己是不是其實也是機械人，在最後兩天故事進入高潮，艾娃不但殺死納森，將加勒永遠地困在研究所，然後進入了人類社會。電影非常好看；儘管知道圖靈偷渡了概念，但在電影中更加精巧的實驗設計下，我們根本分辨不出艾娃要離開研究所，是出於程式底層邏輯設計，還是艾娃的本能渴望，還是這兩者在某種意義脈絡中根本無法分辯，反而使我們開始重新思考，難道「機械有思維能力的可能性」是不可能的嗎？

雖然電影提供的只是思想實驗，人工智能也只是在某些領域得到承認，如 1997 年，IBM 電腦「深藍」擊敗世界西洋棋冠軍加里卡斯帕羅夫（Garry Kimovich Kasparov)；30 年後，2017 年中國烏鎮圍棋峰會，AlphaGo 擊敗圍棋世界排名第一的棋士柯潔。雖然我們不知道會下棋是否等於會思考，但應該大多數人都能承認人工智能「會下棋」吧。

除了會下棋，人工智能小冰也在同年推出詩集《陽光失了玻璃窗》，為史上第一本人工智能詩集，我看了一些，並沒有全讀，其中也不乏有我覺得不錯的，如〈她嫁了人間許多的顏色〉：

那繁星閃爍的幾天蒼色
那滿心的紅日
看萬里天使在世界
我就像夢

看那星閃爍的幾顆星
西山上的太陽
青蛙兒正在遠遠的淺水
她嫁了人間許多的顏色

詩中最後兩句「青蛙兒正在遠遠的淺水，她嫁了人間許多的顏色」著實令我驚艷，真的是詩感滿益，也確實不覺得是人工智能所創作的。不過每首詩的文字掌握度有落差，與其說是個人詩集，不如說更像是多人作品集，就像電腦打字看不出筆跡一樣；由此看來，或許離一般人認為的「會寫詩」還有一段距離。

說起人工智能創作文學這件事，推薦一個講科幻故事的自媒體「幻海航行—science fiction」，其中有個他們自己編撰的故事《AI 續寫三體事件》，觀後覺得十分有意思；故事講

述 2055 年，人類舉辦一場盛大的學徵稿活動，以續寫劉欣慈的科幻小說《三體》為題，最終 AI 獲得物理學諾貝爾獎的故事（沒錯！是物理學）。其中令人感興趣的設定是，AI 發展出人格模擬，因此可以透過人格模擬進行創作，不但如此，AI 還可以在創作的過程中，透過數十億個人格模擬去汰選出更好的作品。在故事結尾，作者認為人與人工智能的根本差異，在於 AI 會依照最初給定的底層邏輯運行，而人因為有許多無意識的思維特徵與情感，因此不會依照一種根本既定理性思維去行動；作者也進而反思，雖然 AI 會依照最初給定的底層邏輯運行，但會不會有一天 AI 能夠反思性地去修正（或改寫）最初給定的底層邏輯，那會不會就是人與人工智能走向競爭對立的開始。

如果順著作者留下的足跡，進一步思考，人工智能若能反思性地去修正或改寫最初給定的底層邏輯，是否就意謂著，機械學會了創造？是否就意謂了人類創造了意識？再者，人類的思維是否具有這種反思性？如果有，那麼人類思維反思性的限度又在哪裡呢？就如我們創造了語言，但也從此受制於語言結構，人類能夠對語言結構進行最原初的反思嗎？難道我們的創造其實也只是一種生產製造？

或者，其實沒那麼複雜，試想另一種狀況；《西部世界》（Westworld）所描繪的另一種情況，一個超高自由度大型故事體驗樂園，其中人工智能扮演與人類玩家互動的 NPC 中立角色，人們可以體驗各種背景不同的故事，也可以一次次地體驗同一個故事各種發展；不同角色的資料，不斷地被刪除、或重新植入於人工智能中，在這樣重複的過程裡，人工智能開始做夢，並由此發現自身只是故事角色，而產生自我覺醒。其實非常有趣，如果讓人工智能擁有類似淺意識的思維特徵，是不是就能不再只是純粹的機械？或者說這是一種讓人工智能擁有意識的一種過程，正如美國哲學家約翰．席爾（John Searle）所說：「光是有組織圖像進入了心理活動（mental stream）中即可產生心智，但除非加入一些補充過程，否則心智仍是無意識的。無意識的心智所欠缺的乃是自我。為了擁有意識，大腦必須取得

一個新的所有物—主體性，而主體性的定義特徵是普遍存在於我們主觀經驗到的圖像中的感覺」（注 2）。

人與人工智能的關係

無論我們怎麼理解人工智能、思考，以及兩者之間的關係，人類進入人工智能化的社會，終究會無可避免地到來；推薦一本由 AI 專家李開復與全球華語科幻星雲獎得主陳楸帆，兩人聯手創作的《AI 2041：預見 10 個未來新世界》，書中藉由許多故事來描繪二十年後後的美麗 AI 新世界，非常有意思。

或者，當我們越來越無法分辨人與人工智能的差別後，是否有一天會愛上人工智能？又或者人工智能會愛上人類？ 2001 年上演的《A. I. 人工智慧》（A. I. Artificial Intelligence），

一個會愛人的機械人大衛，作為莫妮卡的孩子，大衛以為莫妮卡會永恆不變地愛他，直到莫妮卡原本的人類孩子馬丁，在絕症中奇蹟康復，並從冷凍睡眠中甦醒。在馬丁出院後，大衛不但失去了莫妮卡大部分著注意力，又在一次事故中差點害死馬丁，最終遭到遺棄，此後馬丁踏上尋求成為人類的旅程；故事結局十分感人。又如 2013 年上映的《雲端情人》（Her），主角西奧多與人工智能莎曼珊在互動過程中，西奧多逐漸愛上莎曼珊，卻發現莎曼珊同時與成千上百的人交談，並已與數百人墜入愛河的喜劇故事。

思緒至此，我再次下意識地瞥向身旁情侶，如果……他倆是機械人，他們會相愛嗎？那樣的愛又代表了什麼樣的意義？如果機械能夠「思考」，是否就又意謂著機械能夠「理解」，並獲得「意義」呢？如果我跟機械人相愛，他懂我的愛嗎？我是否又能將機械人的種種行為理解為愛，並對此真正地有所領會呢？甚至，我們之間真的有透過語言進行理解活動嗎？

最後，讓我們短暫地回到 1996 年，一隻名為桃莉（Dolly）的母羊，誕生在在不列顛群島上的聯合王國上；隔年，在島上的聯合王國，一本名為《自然》的刊物，發表了這隻羊的故事——羅斯林研究所（The Roslin Institute）的科學家，成功複製哺乳類生命。雖然桃莉能正常與公羊交配、受孕，並於隔年產下六子，但卻很難稱之為正常；從三歲起便出現衰老現象，五歲半就得了老羊病關節炎，桃莉似乎只活了正常羊一半的生命（原羊複製 DNA 時已六歲），就

步入了暮年。想想《銀翼殺手》與《銀翼殺手2049》中所勾勒的世界觀；其中，人造複製人，是經過基因設計的人為的產物，人造複製人一誕生就是成年的身體，由於沒有人類孩童時期發展情緒的過程，因此沒有發展出完全的同理心（也是其與人的差異）。製造人造複製人的目的，在於從事人類無法勝任或不願去做的困難工作，因而被被該世界視為「機械」。

正如上文所提，「我」、「心靈」、「意識」、「思維」、「靈魂」、「人格」這些字詞與其背後所傳達的意義，在不同思想脈絡中都有差異，在人類不斷發展的同時，這些字詞與其背後所傳達的意義，也被人類賦予更多的意義與內容；同樣的，「人」、「人工智能」、「機械」也是如此，很可能在不遠的未來，對「人工智能」與「機械」的理解，會走向我們意想不到之處。

注釋

1. 安東尼歐・達馬吉歐著；陳雅馨譯。《意識究竟從何而來？（改版）：從神經科學看人類心智與自我的演化》（臺北市：商周出版社，2017年）。
2. John Searle, *The Mystery of Consciousness* (New York: New York Review Books, 1990).

延伸閱讀

1. 理查・摩根著；李函譯。《碳變》（臺北市：避風港文化，2019）。
2. 參克・葛詹尼加著；鍾沛君譯。《我們真的有自由意志嗎？認知神經科學之父葛詹尼加對意志的大哉問》（臺北市：貓頭鷹，2021）。
3. 安東尼歐・達馬吉歐著；陳雅馨譯。《意識究竟從何而來？（改版）：從神經科學看人類心智與自我的演化》（臺北市：商周出版社，2017年）。
4. 小冰著。《陽光失了玻璃窗》（臺北市：時報出版，2017）。
5. 劉慈欣著。《三體》（臺北市：貓頭鷹，2011）。
6. 李開復、陳楸帆著。《AI 2041：預見10個未來新世界》（臺北市：天下文化，2021）。

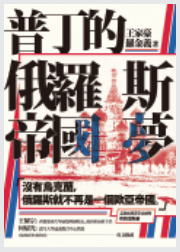
社會科學

死磕正義：709案和中國的人權政治

白信 著

新銳文創 /11108/300面 / 21公分 / 390元 / 平裝
ISBN 9786267128275/574

本書是首部針對「中國709維權律師大抓捕案」的系統性研究專著，書中從結構面雙向分析「人權律師群體」與中共的「法外主義機制」，即中國人權律師的形成和抗爭、中共政法委的法外主義和人權政治，深入探尋中共如何在「法治」名義掩護下，發展出繼承自革命時代的大規模人權侵犯制度，成為一個奉行法外主義的「法外政權」；進而勾畫出「709」案於中國人權運動的貢獻，以及在新冷戰的框架下，中國與全球人權政治的發展圖景。（新銳文創）



社會科學

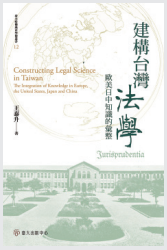
普丁的俄羅斯帝國夢

王家豪、羅金義 著

新銳文創 /11109/240面 / 21公分 / 320元 / 平裝
ISBN 9786267128435/574

本書首兩章聚焦俄羅斯自蘇聯解體以還的政治變遷，綜觀上世紀90年代對民主化的冒進與其後的畸變，以及普丁承接大位後力保政權穩定的關鍵與波折。第三和第四章闡釋普丁的歐亞融合雄圖，包括烏克蘭危機的來龍去脈與其效應，以及歐亞經濟聯盟成員國各自面臨的重大挑戰，當中包藏俄羅斯的諸般野心。第五和第六章分析普丁政權在全球地緣政治角力的表現，審查他的「後疫情時代」世界觀與外交策略，以及發展全球影響力的舉措與得失。（新銳文創）

社會科學



建構台灣法學
——歐美日中知識的彙整

王泰升 著

臺大出版中心 /704 面 /11107/ 23 公分 /970 元 / 精裝
ISBN 9789863506324/580

本書連結法學者、法學論述、政治與社會環境，本於歷史學、法律學、社會學的關懷，敘述台灣如何跨越 3 個世紀，經歷殖民、威權、民主等政體，彙整來自歐美日中的現代法學知識，建構出當下的法學內涵，並提出應超越歷史束縛的主張。另從東亞視角，描繪法學緒論著述所顯現的「明治日本→民國中國→戰後台灣」的知識傳遞及流變。（臺大出版中心）

社會科學

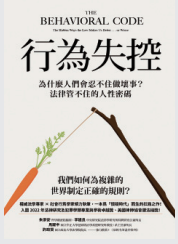
這是我的，別想碰！

邁可・海勒、詹姆斯・薩爾茲曼 著；王瑞徽 譯

平安文化 /11107/336 面 /21 公分 /420 元 / 平裝
ISBN 9789865596972/584

「所有權」並非天賦人權，而是社會的運行機制，它可能在你不知道的時候，隨著政府、企業、掌權者制訂的遊戲規則而改變，卻未有人仔細思考，這些看似天經地義的「潛規則」中間所存在的邏輯問題與利益歸屬。從買咖啡到買房子，「所有權」無所不在，我們或許無法擺脫社會的枷鎖，但卻可以了解自身權益、明辨是非、保護自己，做出最有利的人生選擇。唯有深入了解所有權背後的原理，才能避免自己權益損失，成為獨立思考的自由公民。（平安文化）

社會科學



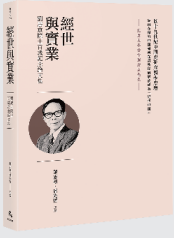
行為失控

班哲明・凡魯吉、亞當・范恩 著；簡秀如 譯

平安文化 /11109/368 面 /21 公分 /450 元 / 平裝
ISBN 9786267181119/580

研究顯示，一旦法條的設計違背人類的認知與行為模式，不僅難以有效遏止犯罪，還可能導致治安惡化。本書結合法學專業與行為科學分析，提出解決社會問題的 6 個有效途徑，以及 3 個改變人們行為模式的改革方向，對執法過當、貪瀆、竊盜、販毒、性侵、殺人等犯罪議題進行思辨，不論你是否相信法律，這本書都能幫助我們洞悉人性本質，保護自己，並打造更理想的社會。（平安文化）

史地／傳記



經世與實業
：劉廣京院士百歲紀念論文集

黎志剛、潘光哲 主編

秀威資訊科技 /11108/442 面 / 23 公分 /720 元 / 平裝
ISBN 9786267088524/ 607

史家劉廣京先生（1921-2006）致力於近代中國史研究，開十九世紀航運史、中美關係史、基督教在中國、自強運動與經世思想之研究先河，本書聚焦近代中國思想世界與經世實踐、經濟脈動與實業創建，收錄王汎森、李金強、吳翎君、周啟榮、周健、科大衛、梁元生、區志堅、張彬村、張偉保、陳計堯、陳明銶、陳慈玉、陸寶千、潘光哲、黎志剛、劉翠溶、鄭潤培、關文斌等十九位學者專文，併附先生學述及其履歷著作表，謹此紀念。（秀威資訊科技）