

# 如何成爲「人」：缺陷及其經驗作爲對人工智能研究之啓發— 以自動駕駛技術爲例

劉育成

## 摘要

本研究以技術人類學家 Bernard Stiegler 有關人類因 Epimetheus 的「失誤」而得以「成爲人」的論述，對人類智能（human intelligence）與人工智能（artificial intelligence）進行探討。以「失誤」概念爲基礎的「缺陷」及對其之經驗，或許是「人之所以爲人」的關鍵元素。人類不僅藉由技術的外部化，以彌補其「失誤」，更透過技術來區別自身於其他物種，進而肯定「成爲人」的意義。本研究提出，若將 Stiegler 的雙關「去—失誤／預設」（de-fault/default）視爲一組區別，那麼從人類智能到人工智能的發展，即是一個從「de-fault/default」到「default/de-fault」的轉變，該轉變有助於理解人工智能現象的內涵。本研究以自動駕駛技術爲例，探討其困境與未來發展。在代結論中，我將提出以「例示／窮盡」（illustration/exhaustion）之區別，作爲對人工智能現象的反思。

- ◎ 關鍵字：人類智能、人工智能、技術人類學、去失誤／預設、自動駕駛技術
- ◎ 本文作者劉育成為南華大學應用社會學系專任副教授。
- ◎ 聯絡方式：Email：ycliu@mail.nhu.edu.tw；通訊處：62249 嘉義縣大林鎮南華路一段55號。
- ◎ 收稿日期：2018/08/05 接受日期：2019/06/22

---

※ 本文作者感謝多位匿名審查人於審查過程中詳細閱讀與細心建議及指正，讓本研究之論述更加完善，也感謝期刊編委會在編輯出版過程中的協助與修訂。

# “Becoming Human” : the idea of fault and its experiences as an inspiration for artificial intelligence — cases from self-driving cars

Yu-cheng Liu

## Abstract

This paper examines the ideas of fault and de-fault developed by Bernard Stiegler in his discussion of the relationship between human beings and technology. “Becoming human” means how we as human beings encounter and cope with what Stiegler terms “original de-fault”. In his discussion, human beings appear in the externalization of this original de-fault, and in doing so human beings can survive with various technologies which are products of this externalization. It becomes the original de-fault for human beings and later “default”. How to become humans or how machines can think like humans refers to the transformation from “de-fault/default” to “default/de-fault” in terms of the theory of form. With the perspective from the anthropology of technology, mainly extended from the idea of Bernard Stiegler, and with the examination of the technologies of artificial intelligence applied around self-driving cars, the argument of this paper may contribute to the future thinking of artificial intelligence. In addition, as one of concluding remarks, the author also suggests another distinction of illustration and exhaustion as a possible way to think about artificial intelligence in the future.

- ⊙ Keywords: human intelligence, artificial intelligence, the anthropology of technology, de-fault/default, driverless technology
- ⊙ Corresponding author : Liu, Yu-cheng is an Associate Professor in Department of Applied Sociology at Nanhua University, e-mail: ycliu@mail.nhu.edu.tw, address: No.55, Sec. 1, Nanhua Rd., Dalin Township, Chiayi County 62249, Taiwan, R.O.C.
- ⊙ Received: 2018/02/08 Accepted: 2018/11/05

## 壹、「像人一樣思考（think like humans）」？

在討論人工智能之前，我們或許應該從何謂人類智能（human intelligence）出發，並且探究何謂「像人一樣思考」<sup>2</sup>。「像人一樣思考」這句話指涉了一個不同於人的對象物：有個東西不是人，但希望可以具有人的思考能力或思考方式。換句話說，其已經預設了一組區別：人與非人。此外，該論述更期待可以讓非人也像人。人工智能概念的提出與發展，則預設了這組區別，以此為前提而想像了一種可以像人一樣思考的機器或演算法。一般認為，1956年夏季在達特茅斯舉辦為期六週的「關於人工智能的夏季研究計畫（Summer Research Project on Artificial Intelligence）」會議，被視為人工智能研究領域的開端（Bostrom, 2014）。Nils Nilsson 在回顧人工智能發展歷史上，則認為有三場對其影響重大的聚會，除了1956年的達特茅斯會議之外，尚有前一年（1955）在洛杉磯舉辦之 Western Joint Computer Conference 中關於「學習機器」（Learning Machines）的場次，以及1958年在英國的國家物理實驗室（National Physical Laboratory）舉辦的「思考過程的機器化」（Mechanization of Thought Processes）論壇（Nilsson, 2010, p. 73）。根據 Nilsson，1955年關於機器學習場次中所發表的論文主要有兩個方向，其一是對神經系統的模仿，其二對心智層級的模仿。儘管兩者有可能是殊途同歸，亦即都是在了解大腦的運作，但實則是在不同層次上對大腦進行描述（Nilsson, 2010, pp. 76-77）。對人工智能的定義，或許首要出現在 John McCarthy 提給洛克斐勒基金會尋求贊助會議的計畫中。他將當前的人工智能問題定義為「使機器以能夠被稱之為智能的方式運作，就如同人類也是如此運作一般」（Nilsson, 2010, p. 77）。這個定義明確區別於過去關於自動理理論（automata

- 
2. 本研究將 artificial intelligence 翻譯為「人工智能」而非大多數所見之「人工智慧」，主要原因在於 intelligence 與 wisdom 無論是在英文或中文（智能 vs. 智慧）均有不同的指涉內涵，但 wisdom 在中文裡經常譯為智慧而非智能。在中文的語用學中，智慧也與智能不同。在教育部編國語辭典修訂本中，智能指的是「智識與才能」，而智慧則是「分析、判斷、創造、思考的能力；聰明才智」等。因此，智能一般被定義為「具有最適切、有效及有效率地解決問題之知識與才能（技能）」。相對地，智慧則不僅止於此，更包含創造、思考等能力。儘管人工智慧一詞在台灣語境中較常出現，但本研究認為區別兩者仍有其意義。這也涉及到我們如何理解人工智能的內涵與應用。「像人一樣思考」是指要跟人一樣具有智慧，還是只需要有解決問題的技能，這或許不僅值得思考，此一區別似乎也指出的是，「人之所以為人」以及「人有別於非人（包括自然界、動植物、技術物等）」兩者在人與非人之互動上具有的辯證關係。

theory) 的研究，以及避免與模控論 (cybernetics) 連結在一起。換句話說，不只是自動化與控制問題，人工智能關注的是人類智能的形式、內容與運作，而非單純的自動化運作或者是控制與反饋的問題 (Nilsson, 2010, p. 78)。第三場聚會是關於思考過程的機器化問題，包括人工思考、特徵與模式辨認、學習、機器語言翻譯、生物學、自動編程、工業計劃與辦公室的機器化等，主要代表人物有 Minsky、McCarthy、Ashby、Selfridge 以及 McCulloch 等 (Nilsson, 2010, pp. 81-82)。

達特茅斯會議中大多數的參與者，對模仿較高層次的人類思想相當有興趣，也就是關於人類如何解決問題 (how human solve problems) 的過程。因此，從一開始關於人工智能的討論，便定調在一定程度上無可避免地要回答何謂人類智能的問題。這些人工智能研究的先驅們提出的一個重要假設是，我們是否能夠製造出具有跟人一樣思考能力的機器 (machines who think like humans)。無論這個答案是肯定或否定的，至少到目前為止，人類的科學知識與技術仍無法達到這個目標。人工智能的發展從 1950 年代後經歷了幾次的寒冬，直到近年來關於腦神經科學研究有大幅進展，再加上相關技術——辨識技術、演算法等——的支援，人工智能似乎又有了新的可能性。然而，對於人工智能的研究，首先得要了解兩個問題，其一是何謂智能，其二是何謂人類智能。此外，這兩個問題均包括諸如形式 (forms)、生成 (generation)、修正 (modification) 及其運作的內涵等面向。在人工智能研究上主要關注兩個部分，其一是資訊的輸入與處理，包括各種辨識技術的軟硬體開發以及適當的演算法等，其二是輸出，也就是在對資訊進行處理與判讀後所執行的動作。本研究所探討的自動駕駛技術，或許是自圖靈測試以來，最有可能是以人工智能為基礎而實現的科技物，但其所面臨及需要解決的問題也不亞於打造一套能夠通過圖靈測試 (Turing Test) 的演算法。完全自動駕駛技術的實現，同樣也是取決於前述兩個部分的實現，也就是辨識技術與演算法。這兩者可類比於人類的各種感官——在自動駕駛部分目前僅是以視覺辨識為主——與大腦或心智的運作。視覺辨識技術也是由兩個部分構成，其一是硬體，例如攝影機、雷達等，其二則是辨識技術，也就是軟體程式的部分。儘管在硬體部分或許有很好的進展，例如高解析度、動態捕捉、高畫質等，然而更重要的是，是否有能夠對這些由攝影機或雷達所蒐集到之資料進行解讀的演算法，而且這個演算法具有像人類駕駛一般的思考能力。換言之，完全自動駕駛所依賴的人工智能技術，或許需

要的是非常接近人類智能的運作。因此，本研究將從技術人類學的觀點出發，一方面嘗試將人類智能的發展與人類自身的缺陷經驗（去—失誤，de-fault）連結起來，另一方面則將此一缺陷經驗視為人類智能發展的關鍵要素，藉此提出人工智能若要能夠像人一樣思考，其或許也要能夠從缺陷經驗出發且具有處理缺陷的能力。

關於人類智能的討論，首先可以從 Howard Gardner 在《心智的框架》一書中對智能的定義出發，其所提出且為教育及認知心理學所沿用的「多重智能」（multiple intelligences）概念，或許可提供探究之基礎。Gardner 指出，所謂的「智能」是：「解決問題的能力、創造產物的能力，而且這些問題或產物在一個或更多文化場景中是有價值的（An intelligence is the ability to solve problems, or to create products, that are valued within one or more cultural settings）」（Gardner, 2011, p. xxviii）。不同於過去從先天的（a priori）觀點來發展關於智能的論述，Gardner 從經驗出發，認為「一般性的智能就是一組解決問題的技巧——使個人能夠解決真正的問題或困難，並且創造有效的產物（結果）——以及發現或創造問題的潛能」（Gardner, 2011, pp. 64-65）。因此，Gardner 將智能連結於文化脈絡，文化在各種不同智能的出現上扮演重要角色，且在不同的文化中有不同的智能需求發展與展現。在其初期研究中，Gardner 區分了六種智能，後增加為八種：語言智能、音樂智能、邏輯——數學智能、空間智能、身體——動覺智能、人際智能、內省智能，以及自然主義智能（Ibid., p. xiv）。我們可以說，這八種智能「共同」構成了「像人一樣思考」的內涵，其中涵蓋了創造力、符號使用與邏輯能力、知覺能力、社交能力以及反身性（reflexivity）能力。如果從 Gardner 所定義的智能以及其所提出的多重智能來看，我們應可假設人工智能研究中所謂的「像人一樣思考」，指涉的也是這些不同面向，以及其混合或綜合體。

如果人類智能的出現與發展無法脫離於文化或社會脈絡，那麼人工智能的研究不僅需要考量的是一般性的智能概念，若要讓其「像人一樣思考」，似乎也得讓人工智能能夠具有在不同文化與社會脈絡中學習的能力，並且根據不同的情境脈絡發展出相對應的問題解決能力。根據 Gardner，「文化」扮演的角色是其動員了這些基本的訊息處理能力——這些核心的智能——以及形塑它們以獲得各自之目標（Gardner, 2011, p. 341）。因此，在不同文化或社會脈絡中，其對智能內容的要求也可能是不同的。換句話說，人類智能是在文化與社會脈絡中成形，也是藉其而得以展現智能的內

涵。這也符應於現象學觀點所主張的，智能乃是具現且鑲嵌在文化與社會脈絡之中（intelligence is embodied and embedded）。此外，在此一對智能內涵及其發展的觀點中，身體扮演了重要角色。例如對空間的認識與身體結構有很大關係，Samuel Todes 在《身體與世界》（*Body and World*）一書中指出，身體的實際結構乃構成了我們用來揭露世界的知覺經驗（Todes, 2001）。例如身體有分前後、上下的傾向，往前走比往後走要來得容易，人類應對在其身體前面的事物要比在後面的事物容易等：我們要在一個垂直場域中保持平衡，才能探索並發現周遭世界的特徵，並且才能適當地回應在該場域中的特定事物（Dreyfus, 2006, p. 48；Todes, 2001）。因此，在 Dreyfus 對 Todes 的討論中指出，知覺的接收性乃是一種體現的（embodied）、有技巧的成就（perceptual receptivity is an embodied, skilled accomplishment）（Ibid.）。如果要了解人類智能是如何解決問題，或許得要先提出的問題是，人類智能是如何認識其周遭環境與社會世界。這也構成了在人工智能發展早期，關於人工智能是否能夠具有像人一樣思考之能力的爭論。Hubert Dreyfus 在 MIT 的研究或許完整地呈現了這一爭辯的內涵，並且在近年來逐漸受到人工智能研究的關注，更讓關於心智與意識運作的哲學與現象學討論得以進入到人工智能研究場域。

早在 1965 年，Dreyfus 便已對達特茅斯會議後有關人工智能的討論有所評論。Dreyfus 並不認為我們能夠製造出像人一樣思考的機器。其對人工智能的評論在當時並未受到關注，且多受到人工智能領域專家的忽視。然而，Dreyfus 在一篇給 RAND 的研究報告，後正式出版名為《煉金術與人工智能》（*Alchemy and Artificial Intelligence*）書中所提出的評斷，不僅正確地指出人工智能的盲點以及錯誤的預設，同時也提出了發展方向上的建議（Dreyfus, 1965）。該書主要回應 Herbert Simon 於 1957 年所發表的一個與人工智能有關的宣言，該宣言主要提到當前已經有機器可以思考、學習與創造。Simon 認為這些能力會快速成長，能夠處理問題的範圍也會與人類並駕齊驅，其也為此做出了四個預言。然而，Dreyfus 並不同意 Simon 的觀點，並透過對其論點的討論以表達對人工智能發展的質疑（Dreyfus, 1965, p. 3）。

Dreyfus 從現象學觀點出發，嘗試探究的即是人類如何認識周遭環境以及世界的問題。對他來說，這個問題應該直指人工智能是否能夠成功的核心關鍵。在該文中，Dreyfus 區別了人類在處理訊息上的三種根本形式，並且主張，透過數位計算機語言

來分析上述這些智能行為的企圖，實際上是排除了這三種形式：Fringe consciousness（邊緣意識）、Essence/accident discrimination（本質/巧合之區分），以及 Ambiguity tolerance（對模糊性的容忍）。因此，對他而言，任何能夠媲美人類表現的系統，至少要具有以下特性：區別某個類型中的某個特別例子之重要與次要特徵、使用那些留存在意識邊緣的提示、解釋脈絡，以及將個體視之為典型（Dreyfus, 1965, pp. 45-46）。換句話說，Dreyfus 事實上是在討論要如何能夠讓人工智能具有人類智能的特性。或者，一個具有人類智能的人工智能，應該會長什麼樣子？在人工智能的發展過程中，主要的假設是認為大腦是在處理訊息，也就是符號處理。因此，無論是早期的下棋程式，或者是本文所探討的自動駕駛技術，都是基於如何處理、辨識與解讀符號的問題。此一觀點將大腦或心智的運作視為訊息處理活動，因此合理的推論是，科學家可以透過電腦等數位設備將這些訊息予以編程，藉此得以實現任何及所有可理解的訊息處理過程。Dreyfus 認為其意指的是，數位計算機器能夠處理任何被完全形式化的訊息，亦即，以窮盡與無模糊性之形式所表現的訊息（Dreyfus, 1965, p. 55）。然而，人類思考或其步驟，乃至於人類智能，是否可以被完全的形式化（completely formalized）？這不僅是科學家至今所相信的，也是 Dreyfus 對人工智能批評的出發點。

Pete Moore 在 *Being Me* 一書中也討論到此一科學家觀點的弔詭。他認為科學主義主張「你（主體）等於你的大腦」（you equals “your brain”），且將大腦視為人類本質的唯一關鍵，這或許是錯誤的開端（Moore, 2003, pp. 8-9）。Moore 引用 Kenan Malik 對此種科學觀之弔詭的討論指出：「人類...既是創造出科學的主體，也是科學（研究或探究）的客體（對象）（...humans...are both the subjects that creates the science and objects of that science）」（Malik, 2000; Moore, 2003, p. 8）。就此而言，科學是否能夠了解人類或人類智能的這個問題，其本身是否具有弔詭，便值得探究。“Being me” 這個問題與本研究標題 “becoming human” 所探究之對象類似，何謂「成為我」或「成為人」？科學是否可以回答這類的問題？科學之所以能夠具有「客觀性」，主要是來自於對「自然」的觀察與解釋。科學的客觀性與其對自然的正確解讀與否有關。之所以如此，有很大部分是因為這個自然在很大程度上對所有人類都有效，或者具有一定程度的普遍性。也因此，由於科學能夠客觀地理解自然，科學家們也認為科

學當然也可以理解人類——包括意識等，因為人類也是自然的一部分。然而，科學的客觀性，或者科學對自然的解釋雖然具有一定程度的普遍性，但是否便代表科學是一種在人類或宇宙萬物運作規則之上的事物？再者，科學的客觀性——包括其對自然的觀察——有很大程度是建立在人類智能的基礎上——就科學也是人類的創造物而言，那麼科學是否也受到人類智能的限制？人類無法理解很多自然界中的現象，是否也是因為此一受到人類智能之限制的科學——人類仍以此來理解世界，而無法跳脫這個由人類或科學所架構出來的世界？

著名的人工智能研究者 Nick Bostrom 對此也有類似觀點。他在對人類智能與人工智能之關係的討論中提到「沙盒」(sandbox)現象，其意思是，相對於人類智能的終極目標或價值若是自我存續，那麼人工智能的終極目標及其價值可能為何的問題(Bostrom, 2014, pp. 146-147)。沙盒現象的弔詭在於，人類智能希望打造出對其自身來說是安全的人工智能，因此，人類智能總是努力地確保其安全無虞。換句話說，人類智能總是期待在一個人工智能——或 Bostrom 的「超智慧」(superintelligence)——受到控制的環境下——亦即「沙盒」——對其進行觀察。在此過程中憑藉的仍是人類智能的「經驗」，因此這個經驗也限制了人工智能得以被觀察或獲得發展的內容：

在沙盒中表現得體是友善人工智能與不友善人工智能共通的趨同工具目標。一個智能充足的不友善人工智能會察覺到，如果它一開始表現出友好的態度而被放出沙盒，它的終極目標才有可能實現。等到我們就算發現也沒差的時候——也就是說，當人工智能已經夠強大，而且人類的反抗已經無效時——它才會開始展現它不友善的本質 (Ibid., p. 117)。

相對地，人類智能的發展或其目的，並非在確保人類智能對其自身而言是安全無虞的，也因此發展過程中，我們透過各種失誤、錯誤等來學習。人類智能不需要在其發展之初，即需要經過一個篩選的過程，以讓友善的人類智能勝出並且獲得發展。相反地，人類智能的友善／不友善這組區別是在與其自身環境的互動過程中，關於該區別的內容才獲得形塑。然而，人工智能的發展卻似乎正好相反。Bostrom 指出人工智能在發展過程中，友善／不友善這組區別已經在操作上（先驗上）被用來觀察人工智能自身，在確保其必須具有友善的結果之基礎上，人工智能的學習跟人類智能的學



習過程便有著完全不同的內涵。

Dreyfus 在 60 年代所提出的三種人類認知形式，不僅在後來的四十年間逐漸浮現而成爲人工智能發展領域所承認的困境，這三種認知形式至今也仍未發展出適當的科學知識或科技來解釋或掌握之。在諸多嘗試理解人類認知過程與形式之討論中，現象學觀點或許可以提供另種可能性<sup>3</sup>。現象學觀點——從梅洛龐蒂到海德格等——將人類智能發展或人類認知視爲是體現的（embodied，無法脫離於身體的）也是鑲嵌的（embedded），這兩個元素基本上也類似於前述 Gardner 將人類智能連結於文化脈絡的觀點。心智或意識是否可以獨立於身體以及身體所處的脈絡而存在，這依舊還是個大哉問。就目前關於人類智能的研究來看，或許尚無法肯定前者能夠獨立於後者而存在<sup>4</sup>。身體在意識或心智的形塑與運作上，仍舊扮演重要角色。身體對心智的形塑若有不可忽略之角色，受身體所形塑之心智，也成爲人類認知自身與週遭環境的途徑或方法。倒過來說，如果我們想要了解人類智能如何進行認知，就無法避免地需要了解，身體與心智的交互作用是如何構成或提供人類對其環境的認知。笛卡爾將身體與心智二分的做法，在人工智能發展上面臨了困境<sup>5</sup>。換句話說，心智、身體、環境，三者的交互作用，構成了人類智能在認知上的獨特性。Dreyfus 所謂人類智能之所以能夠處理模糊性、不確定性等，或許也是這三者互動的產物。對人類智能而言，「規則」是用來處理當下的情境，而非作爲窮盡所有可能性的工具。在《電腦不能做什麼》（*What Computer Can't Do*）一書中，Dreyfus 便已主張，「我們所生存在其中的那些有意義的物件，不是一個儲存在我們的心智或大腦中的世界模型；他們就是世界自身」（Dreyfus, 1972, pp. 177-178）<sup>6</sup>。因此，他認爲在電腦內部再現這個世界的

- 
3. 例如，根據 Dreyfus 在 MIT 的經驗，海德格的認知典範影響後來的人工智能研究領域，包括在 MIT 的 Rodney Brooks 的行為取徑、Phil Agre 的實證主義模型，以及 Walter Freeman 的神經動態模型等（Dreyfus, 2007, p. 1139）。這三種方法隱然地接受海德格對笛卡爾的批判，並且支持 John Haugeland 的說法：cognition is embedded and embodied（Haugeland, 1998）。
  4. 例如 Hilary Putnam 有名的「桶中腦（Brain in a vat）」思想實驗（相關討論可參考（Hickey）：<https://www.iep.utm.edu/brainvat/>）。
  5. 根據 Dreyfus 的討論，Agre 在 *Computation and Human Experience* 一書中也指出了同樣困境（Dreyfus, 2007, p. 1142）。
  6. “When we are at home in the world, the meaningful objects embedded in their context of references among which we live are not a model of the world stored in our mind or brain; they are the world itself.”（Italics are original）

豐富性，是非常困難的。人類之所以可以避免這個問題，是因為人類的世界模型就是世界自身<sup>7</sup>。這也描述了在人工智能研究中所面臨以及亟待克服的框架問題（frame problems），也類似前面提及的 Bostrom 所謂的「沙盒」現象。

Andy Clark 在對人類認知強化（cognitive enhancement）的討論中提到，大腦如何透過一個「特定形式的整體問題解決組織（a specific form of overall problem-solving organization）」來運作並處理從外界而來的訊息（Clark, 2013, p. 121）。在其稍早關於人類智能與賽伯格（cyborgs）的討論中，將賽伯格視為是我們自身生物學本質的偽裝想像（a disguised vision of our own biological nature），一如其書名所示：我們是「天生的賽伯格」（natural-born cyborgs）（Clark, 2003, pp. 5-6）。對他而言，所謂的人類大腦以及對人類智能具有的獨特特徵的最好解釋是，「其有能力與非生物學上的建造物、支柱物與輔助物建立深層且複雜的關係（their ability to enter into deep and complex relationships with nonbiological constructs, props, and aids）」（Ibid., p. 5）。這些「非生物學上的建造物、支柱物與輔助物」指涉對象包括技術物與人造物等，或許也包含了文化、社會、政治與經濟制度等。綜上所述，人類智能的發展及其特徵，或許與人類智能與其自身之環境的互動有關，文化也是在其環境之中。據此，反過來說，如果我們期待人工智能要能夠「像人一樣思考」，那麼是否也表示其需要具有與其自身之環境建立「深層且複雜的關係」之能力。如同 Clark 所指出的，人類智能的

- 
7. 這個內部再現（internal representation）的前提或邏輯上的預設是一種「窮盡」。所謂的「內部再現」意指三件事，其一，再現可以在另一個與外界沒有接觸的空間地方中獲得完成，例如電腦。其二，再現是可能的。其三，再現已經被實現過——無論是在現實或想像之中。「儲存」是一個奠基在物質性上的概念，首先，既然可以儲存，就有可以提取的可能性，再者，在存提的過程中，已經假定物件本身不會有所變化。也就是，儲存什麼東西，就會提取出一樣的東西。然而，心智的運作或許並非如此。記憶會出錯，會有缺失，也會有加油添醋。這些不總是那麼精確的運作，構成了世界自身。如果世界自身是由不精確性所構成，那麼期待世界可以被再現的這件事本身，就是一個弔詭。再現就跟提取類似，其也假定了有一個可以被再現的事物，如果這個被認為可以被再現的事物，是處於不斷變動的狀態，也就是身體、心智與環境三者的交互作用，可能每分每秒都在變化的狀態，那麼一開始那個被認為可以被再現的事物，毋寧只是一種對世界的靜態想像，也唯有具有靜態性的事物才有被窮盡的可能性。如果是這樣，秉著窮盡邏輯而對世界進行認知，或許永遠都不會有機會真正認識到世界自身：世界自身就是由不精確性、不確定性與模糊性所構成。這裡的世界指的是人類智能所認知或建構出來的世界，而非指自然界本身。我們也可以說，這個人類所認知到的世界自身，也可以是自然界本身的一部分。

生成不僅是 embodied，其更是在大腦、身體與世界間的流動所展現的產物：

人類從來就不是脫離身體的智能；…在所有受到檢視的例子中，關鍵在於連結了行動——指令、身體動作、環境效果，以及多感官的知覺輸入的複雜反饋迴圈。此一在大腦、身體與世界之間雙向流動的影響是重要的，並且在此基礎上，我們得以建構（以及持續地重構）我們對潛在與當下自我的感知。這個生物學上的皮袋（指身驅）在這裡不具有特別的意義。流動才是關鍵（Clark, 2003, p. 114）。

據此，有關人類智能以及何謂「像人一樣思考」的討論，便可從人類智能與其環境之互動出發，Clark 描述其為一種「流動」（flow）。然而，Clark 並未進一步探究「流動」的可能內涵為何<sup>8</sup>。此外，他認為身體在此流動中似乎並未扮演如同前述現象學觀點所主張的重要角色，這或許反映了依舊是一種偏向人類中心主義的觀點。從技術哲學的角度觀之，就如同前述的沙盒現象，當代對人工智能的期待仍是以人類為中心的（anthropocentric）思考方向，只有在人類智能得以控制的情況下，才有所謂的人工智能。也因此，1960 年代以降的人工智能，乃至於近年來較屬於狹義人工智能的機器學習（machine learning）與深度學習（deep learning）等概念，大多是以最佳化、最適切、有效及有效率等為檢驗人工智能是否成功的原則，而非對 Bostrom 所謂的「人工常態智能」（artificial general intelligence），或者是 Nilsson 稱之為「強人工智能」（strong AI）有更多的關注或進展（Bostrom, 2014, p. 18）。換句話說，當前的人工智能由於各種既有科學知識與技術的限制，大多偏向以支援人類思考的「弱人工智能」（weak AI）方向發展，例如現下流行的聊天機器人（Chatbot）、智慧型手機語音助理（蘋果公司的 Siri、Google 的 Google Assistant，或者是 Amazon 的 Alexa），以及高頻率運算交易程式（algorithmic high-frequency traders programs）等<sup>9</sup>。然而，Bostrom 卻也正確地指出，所謂的「學習能力」對人工常態智能而言，「並非是事後

8. 俗民方法學（ethnomethodology）或許可以為此「流動」概念之內涵，提供很好的例證與說明，然而因篇幅有限，期待在來文中能夠對此部分進行深入探究。

9. 高頻率運算交易程式被廣泛地用在全球金融市場，其主要用途是為大型投資公司自動化股票交易系統，包括將投資管理人的買賣執行指令予以自動化，或者是用來適應多變的市場狀況，進行複雜的交易策略。高頻率運算交易程式得以在非常短地時間內回應，其基於理性邏輯的運算也終究導致了 2010 年的美國股票市場大崩盤（Bostrom, 2014, pp. 16-17）。

的添加延伸，而應該是核心設計的必要特徵」（Bostrom, 2014, p. 23）。此種學習能力包括對不確定性與機率資訊的有效處理。這也讓後來的人工智能研究轉向腦神經研究與生物遺傳學研究。此外，Bostrom 也提到，人工智能不見得要與人類的心智相類似，相反地，人工智能可能會「有著與生物智能徹底相異的認知架構，且在發展初期就會有截然不同的認知強弱項」，且「人工智能的目標系統可能會和人類的大相徑庭」（Bostrom, 2014, p. 29）。

若是綜合 Gardner 將人類智能發展與文化相連結、Dreyfus 從現象學觀點將人類智能視為 embodied 與 embedded，以及 Clark 將人類智能的發展視為一種由大腦、身體與其環境（包括文化）之間各種流動的討論，關於人工智能是否能夠「像人一樣思考」的可能性，或許便得將這些觀點納入考量。每一種文化與社會脈絡的展現及其內涵，相對於人類智能整體而言，總是局部的（part and partial）。透過個別文化或特定脈絡，我們僅看得到人類智能的某個或某些面向，而非其全貌。從技術人類學的觀點來看，文化不僅是技術的，同時也是外在於人的。人在文化之中且受其影響，但人卻不屬於文化自身。據此，本研究將從技術人類學的觀點出發，亦即認為，文化的角色、發展及其內容，乃是人類透過技術將其自身之缺乏予以外部化（exteriorization）的產物之一。這部分也將延伸 Stiegler 對技術與人類之關係的討論，亦即將人類自身原初之缺乏（original absence）的外部化過程，視為人之所以能夠「成為人」（becoming human）的重要元素。接下來的第二節繼續延伸 Bernard Stiegler 的技術人類學觀點來探討缺陷及其經驗，以及在第三節中討論由此所衍生出來的「去一失誤／預設」（de-fault/default）這組區別是如何重構人類智能與人工智能在起源與發展上的不同路徑，最後以自動駕駛技術為例說明這樣的區別可能具有的啟發。最後在代結論的部分，我將初步提出以「例示／窮盡」（illustration/exhaustion）觀點作為對人工智能發展的反思與建議。

## 貳、缺陷及對其之經驗作為「成為人」的基礎

一般來說，所有的生物體都會面臨死亡的這個事實，可以說是最大也是最無可避免的「缺陷」之一。然而，為什麼死亡會被視為一種缺陷？站在人類智能的立場，我

們將死亡視為是生命體的終極命運，在科學研究上，我們也將其視為一種「缺陷」，而非欣然接受自然的安排。因此，在人類歷史上，對於死亡的延續不僅在各種文化或神話中可窺見一二，或者有各種對於長命的祝福，以及對長生不老的期盼等。甚至我們透過科學與科技來改善可能影響死亡的因素，或者強化延長生命的可能性。

技術人類學者 Bernard Stiegler 在《技術與時間（第一冊）》（*Technics and Time, I*, 1998）一書中深刻分析了一則著名的希臘神話——關於艾比米修斯的失誤，以及普羅米修斯如何彌補這個失誤的故事。Stiegler 指出，人類的缺陷是從一個失誤（*fault*）而來——一個眾神在創造人類與其他生物時所犯下的錯誤。這個故事的通常版本如下：在創世之初，艾比米修斯（*Epimetheus*）負責分配給每一個生物體一種本能，但他最後卻忘了要分配給人類。也因此，人類便不像其他生物體，一出生便具有某種得以生存的能力。另一位希臘神普羅米修斯（*Prometheus*），在知道艾比米修斯的失誤後，由於其憐憫並且想保護人類（根據神話，人類是普羅米修斯用土與水捏製出來的），因此從眾神處盜了火與各種技術給予人類，以使人類得以同其他生物競逐生存。儘管艾比米修斯忘記給予人類任何本能，但其失誤本身卻正好變成了人類的本能，這個本能便表現為一種「缺乏」（*absence*）狀態。普羅米修斯的行為則為人類填補了這個缺乏狀態，以彌補艾比米修斯因為遺忘所造成的失誤。然而，此一缺乏狀態以及稍後的「填補」，卻讓人類具有了將失誤予以外部化的能力——這也是技術意義上的人類之起源。該神話的啟發通常是描述了人類用火與技術的開端，然而更重要的意義是，人類因為這個缺乏狀態且為了填補這個狀態而開啓的人類本質從 *de-fault*（自然意義上）到 *default*（技術意義上）的轉變過程。這個轉變也構成了現代性中對科學與科技發展的想望，以及各種版本的科技烏托邦或者是科技樂觀主義。

根據神話故事，普羅米修斯的行為在眾神眼中也是一種失誤。因此，宙斯將其綁在高加索山上，每天派一隻老鷹吃掉他的肝，隔天肝會重新長回來，如此日覆一日地承受被啄食肝臟的痛苦。作為對艾比米修斯之失誤的彌補，普羅米修斯的失誤讓人類得以有生存的能力。艾比米修斯的失誤，則因為普羅米修斯的失誤，卻成為人類最重要也是根本的缺陷，且普羅米修斯的失誤更是讓艾比米修斯之失誤獲得外部化（*externalization*）的契機。此一失誤的外部化，乃成為人類透過技術重新發現自身——成為人——的開端。這裏的「重新發現」指的是，人類只有在使用技術之後，

才得以認識到自身作為人類——而非其他物種——而存在。因此，Stiegler 指出，「人類的存有就是外在本身（外在於自身）」，而普羅米修斯給予了人類「將其放在自身之外」的禮物（Stiegler, 1998, p. 193）。這個禮物指的就是「技術」。就此而言，技術就是讓人類能夠「將自身放到自身之外」的東西。換句話說，也就是讓人類具有可以與自身（原本作為自然之一部分）建立距離，並且透過技術的協助而能夠將此距離予以客體化（objectification）的能力，甚至讓其成為是一種必須性。這也是「輔具」（prosthesis）這個字的意思，pros-thesis 在字義上指的是那些被放在前面的東西，也就是被放在外面的東西。假如那個「被放在外面的東西」其實就是那個將其放在外面之東西外部化的產物，那麼此存有就是外在本身，同時也是外在於自身。此也正回應了馬克思對資本主義作為一種技術的觀察，資本主義不僅是外加於人類，更是將人類異化於自身（Marx, 1887[1867]）。

艾比米修斯的「失誤」，也因為普羅米修斯的「失誤」而獲得彌補，此一雙重失誤構作了人類在本能上的預設（default），亦即「原初的去—失誤」（original de-fault）。就此而言，這個「原初的去—失誤」，不僅象徵了一個不斷去—失誤的過程，其也成為「成為人」的預設（default）<sup>10</sup>。Stiegler 指出，包括語言在內，諸如發明、發現、洞見、想像等，都是此一 original de-fault 的特徵。對他而言，de-fault 指

10. 僅管本文將 default 翻譯為「預設」，對該字的理解還是得要回到 Stiegler 對 fault 之討論的脈絡中。Fault 有缺點、缺陷、錯誤、故障、誤差等之意。就英文的使用而言，在某字前面加上 de-，通常表示「離開、否定、減少、失去」其所連接之字的原意。據此，de-fault 也可以理解為否定失誤、減少失誤之意，在中文語境中通常將 de- 譯為「去—」或「解—」，例如 de-colonization（去殖民化或解殖民化）、de-composition（解組）、de-mystify（除去…中的神祕因素）等。Stiegler 也是在此意義上使用 de-fault 這個字，對他而言，艾比米修斯的「失誤」透過普羅米修斯的「失誤」——盜火與技術給與人類——而得以去除，而其結果就是讓人類不斷地透過透過技術而得以「成為人」。普羅米修斯的失誤，對人類而言就是那個原初的去—失誤（original de-fault），自此之後，人類只能在不斷地 de-fault 過程中，找尋那個肯定自身為人類而非其他物種的可能性。相對地，default 一字的 de 並非是「去—」之意，根據字典的說明，其來自於舊法語，default，或者 defaillir，其意為「to fail」，亦即「無法完成」、「失敗於…」之意。在英文中有兩種意義，其一為「無法履行（債務、款項）」，其二為「默認、預設，當沒有其他選項或使用者沒有特別指定時」。本研究由於是以人工智能為主題，因此在使用 default 這個字上採用的是第二個意義。這也與本文指出 Stiegler 使用的是一種雙關語有關。De-fault 跟 default 並非是相同的字，其字根與使用上都不盡相同，但由於其字的構成是相同的，在意義上反而提供了一種對比性，de-fault 是在描述人之所以為人的可能性，這個可能性卻成為人之所以為人的預設（default）。相反地，人工智能從一開始在實踐的就只是

出的是「存有的缺陷」（a flaw of being）。所有可以用來 de-fault 的事物都是外在於人的，因此是技術的（非自然的），也就是具有輔助性、是人造的。「成為人」意謂著，人具有某種「人的性質」。然而，在 Stiegler 對人與其他物種、人與其環境之互動的討論中指出，「人性沒有性質，沒有預定，其必須發明、實現、製造出性質，並且也沒有任何暗示說，被製造出來的這些性質將會實現人性，或者這些性質將會變成是人類的性質，因為這些性質或許更會變成的是技術的性質」（Stiegler, 1998, pp. 193-194）。然而，何謂人性（humanity）？如何理解技術與人性的關係？如果人性沒有固定的內容（預設），如同艾比米修斯的失誤所指出的，那麼我們如何理解人性？或者，因為遺忘所給予人類的預設就是，「人性的預設（default）就是不斷地去失誤（de-fault）」，我們或可說，人性本身就是技術的。人性的內容取決於技術的內容或性質。然而，反過來說卻不一定成立，這些被製造出來的性質並無法用來完全地實現或展現人性，或告訴我們人性的真正內容，或其預設。相對地，我們只能從這些不斷透過技術所製造出來的性質中，窺見或接近人性的內容，永無止息——就像普羅米修斯的懲罰一樣，日復一日地在痛苦與恢復痛苦的循環中，證明自身依舊是那個已經與原初人類（being-in-fault）分離開來的「人」（being-in-de-fault）。

「成為人」乃是人類將其自身的那個原初失誤透過技術予以外部化的後果。這些技術包括傳統、信仰、文化，甚至是科學知識等。原初失誤並不會讓人類意識到自身與其他物種的不同，原初失誤作為一個遺忘的產物，其仍舊可以是一種屬於自然界中的本能。「成為人」的關鍵在於原初失誤透過技術予以外部化的過程。原初失誤的外部化，意謂著人類獲得或發展出去除這個原初失誤的可能性，也就是填補這些失誤所造成的缺乏狀態。然而，要能夠看到這個原初失誤並且更重要的是將其「當作是」失誤而需要去除，其首要條件即是創造出與其之距離，也就是將其放到人類自身之外，不再將其視為自身及自然的一部分。這個外部化過程也是人類將自身予以去自然化（denaturalization）的過程。人類本質（原初失誤）的去自然化，得以讓人類作

---

default，其預設並非來自於自身，而是來自於程式設計師、軟硬體工程師等外部環境。就此而言，人工智能與人類智能的本質各有不同，前者是 default，後者則是 de-fault。「本質」一詞僅是用來描述可能為人工智能與人類智能之區別的內容，在本文脈絡中，有時與「預設」一詞之意是相通的。此外，Stiegler 也對該詞之使用有其說明(Stiegler, 1998, p. 280)。

為人類而出現，而與動物有所區別。但也正是去自然化，亦即透過技術或技術物來彌補原初之失誤，讓人類成為非人類而消失在自然之中。這樣的弔詭或許已然是預設在其本質之中，也就是那個缺乏狀態。「讓人類成為非人類而消失」意思是，人類在去自然化的過程中，將其自身予以客體化或外部化，在客體化後便可在客觀上予以消除。這個經由將人類予以客體化而出現的非人類，乃是一種非自然意義下的人類，這種非自然意義下的人類是技術的，或者是與技術相結合的（例如賽伯格，cyborgs）。這樣的結合讓屬於自然那部份的人類消失，留下的是人與技術之結合的後人類（posthuman）。

從這個角度來看後人類，或許可以提問的是，後人類是不是非人類？如果是，那麼後人類事實上意味著人類的消失，消失在與技術的結合過程之中。去自然化讓人類得以作為與其他物種或自然之區別而「現身」在歷史或世界發展之中，但這樣的現身卻也標示了屬於自然的人類的消失，人類在歷史上的現身是一種技術意義上的現身（technologically present），而非自然意義上的現身。「一出現即消逝」或者，「以消逝為前提的現身」，不僅作為「成為人」的重要元素，此一元素也符應於技術在人類歷史與社會中的發展傾向，「轉瞬即逝」——無論是時間、訊息、事件——或許便回應了人類從一開始就具備的傾向或本質（Tomlinson, 2007）。因此，Stiegler 指出，人類的現身，必須依賴於其自身的消逝，消逝於技術將其外部化的過程中：「人類在其本質的去自然化中，乃是其自身之消失（Man is his disappearance in the denaturalization of his essence.）」（Stiegler, 1998, p. 121）。

在自然狀態，人類不會感受或有意識地理解到何謂死亡，因為死亡是作為自然的一部份而存在於有機體。在許多神話中，人死後並未消失，而是進入到自然或者以各種「存有」形式繼續存在。隨著科學、科技與各種技術物的介入，將人的死亡與自然切割開來，如今死後也不再是自然的一部分：死了就是死了，那些讓人「成為人」的事物都不會留下，也不會回到自然。技術介入帶來了將死亡從自然切割出去，變成是與人類自身不同的東西之可能性，也就是前述Stiegler所謂的外部化，或者客體化、對象化等過程。人類開始可以將死亡當作是外在於自身的東西來經驗、觀看或理解——藉由技術。這也是 Stiegler 所探討的可完美性（perfectibility）所具有的虛擬性（virtuality），一旦這個虛擬性得以透過技術或技術物而獲得實現（actualize），人也



就得以跟自然與其他物種區別開來，自然狀態下的人也就因為技術或技術物的介入而逐漸消失。這也是 Stiegler 在《技術與時間》一書中所要呈現的主要概念。然而，他大概沒有想到的是，現代科技發展讓這一天更快速的到來。他說：

只要野蠻人不在自由的不安定中，只要其可完美性依舊是虛構的且不會擾亂自然與其自身之本質之間的原初計畫，只要其虛構性不會變成真實的——也就是技術的，他就不會對死亡有所感並且不會預做準備：他不在時間之中（Stiegler, 1998, p. 122）。

據此，原初的人類並不會對死亡產生畏懼。當人類開始害怕死亡時，他便受到了前述那種「自身的缺席」的損害（影響），因為其已經與那個原初之預設的（自然）自身分離了。Stiegler 說，原初人類害怕的不是死亡這件事，而是苦痛（pain）（Ibid., p. 123）。因此，「原初的人類並非是會死的」（Stiegler, 1998, p. 126）。這句話的意思是，原初人類並未意識到死亡本身是一種需要被獨立出來、被關注、被害怕的事物：死亡是自然的一部分。換句話說：死亡尚未被外部化，死亡仍然是作為自然的人類的一部分。對死亡的意識、恐懼，亦即一種改變對自然時間的看法，取決於「形勢」（*circumstances*）（Ibid.）。人類在面對各種形勢時，事實上是在面對人類自身的缺陷（或缺乏），也就是艾比米修斯的失誤。當人類被野獸包圍，或者是需要可以填飽肚子的東西時，人類便需要與這些形勢產生互動。形勢這個字的英文是 *circumstance*，其字源指的是 *encircle*、*encompass*，也就是包圍、圍繞之意。人類被某個情境所包圍或圍繞，因而使得人類必須回應或採取應變行為。沒有形勢，人類也將不需要回應，形勢開了一條路，或者作為以下之前提：讓沒有任何本能的人類，得以或必須依靠技術或技術物的介入來回應這些圍繞在其周圍的各種情況。也因為形勢的多變、存在及對其之回應，人類得以或必須脫離其原初之預設，因而建立與原初之預設的距離。這個距離經由技術（物）來彌補之，也就是 Stiegler 所謂的「縫隙」。這個縫隙（距離）因為科技發展而越來越大或越遠，這也是人類所面臨之形勢，橋接或彌補的方式也只能是透過技術（物）而行之。因此，*circumstance* 提供或建立了兩種距離，或者使兩種距離得以可能，一種是人與其原初預設的距離，一種是人與其所面對之事物的距離。這個「圍繞」所造成或提供的「與原初預設之分離（距離）」之可

能性，也或許是人類——作為人類而非自然中的動物——開始作為觀察者的起點。人類區別自身與其環境的可能性，乃是建立在人類與其原初預設之分離的基礎上。這也是 Stiegler 所謂的，內環境（原初預設作為在人類自身內的環境）變成了外環境（技術作為人類自身之外的環境）。

從 Stiegler 的討論來看，*circumstance* 是人類之所以要發展出各種應對或外部化自身——概念的使用也是其一——的重要契機。在面對每一個 *circumstance* 時，其就像是將人類包圍起來，人類只能在其之中進行回應。當人類自身沒有——內環境中的——本能來應對時（一種缺乏狀態），他只能透過將其自身予以外部化的方式，也就是以非其本能的方式——也就是各種技術——來應對（填補這個缺乏狀態）。對 *circumstance* 的知覺，一開始並非是建立在理性的基礎上，但對其之回應或處理，則是以技術或外部化自身為基礎。透過這個知覺，人類意識到自身處於缺乏狀態，而為了填補這個缺乏狀態，人類外部化自身，將其轉變為所謂的去失誤——*original default*。換句話說，前者也就是意識到自身處於缺乏狀態且無法回應其 *circumstance*，並非是以理性為基礎，也因此沒有概念的問題。在人類得以因為意識到並且將「原初失誤」轉變成「原初去失誤」後，才開啓了所謂理性化、外部化、技術化的進程。目前的人工智能針對後者有一定的成果，也就是「再現」觀點的人工智能發展：人工智能抓住了去失誤（*de-fault*）到預設（*default*）的這部分，透過以理性為基礎所發展出來的邏輯作為其預設。然而，針對前者以及該轉變本身，或許仍未在人工智能中獲得實現。也就是說，人類如何在缺乏狀態下應對環境，或者說，如何「在缺乏狀態下」回應各種 *circumstances*，並非是人工智能發展關注或嘗試要解決的面向。

*De-fault* 是一個雙重失誤的產物：艾比米修斯的失誤（忘記給予人類一種性質或本能），再加上普羅米修斯的失誤（從眾神處盜火與技術給人類），因而產生了「去—失誤」（Stiegler, 1998, p. 194）。這個「去失誤」也成為人類的預設狀態（*default*）。這個雙重失誤或許作為一種辯證關係而出現，構成了「人之所以為人」的內涵。這邊的「人」意指的已經是那個與自身之缺乏性質建立了距離的「技術人」。與自身的缺乏性質建立距離，才能夠讓自己變成為是技術人，也才能夠知道「為什麼自己是人」，而不是其他物種。換句話說，如果是處在一種沒有與自身之缺乏性質建立距離的狀態（無距離狀態），那麼人依舊是動物或自然的一部分。然而，

因為普羅米修斯的失誤，讓人類與自身之缺乏性質得以建立了距離，也正是因為這個距離，讓人類得以發展技術，用以彌補自身之缺乏性質，從而脫離動物或自然，成為凡人——知道何謂死亡、會對死亡產生畏懼與害怕的凡人。因此，技術將原本作為在場的失誤（*fault as presence*），轉變為作為缺乏的失誤（*fault as absence*），去失誤（*de-fault*）也就轉變成了預設（*default*）。原本艾比米修斯的失誤，並非會造成一種缺乏的狀態，如果仍然是自然的一部分，與其他動物一般跟自然和諧共存，那麼這個失誤也不會變成是一種缺乏，失誤就只是失誤。但普羅米修斯將技術給了人類，讓這個原本只是在自然上的失誤，具有了非自然的意義。人類與自然分開了，原本的失誤也轉變為一種缺乏，在人類與自然拉開距離的過程中。這樣的缺乏，便只能透過非屬於自然的技術來填補，而非是回到自然本身，在與自然的互動中找到出口。據此，普羅米修斯的失誤為人類帶來了原初的去失誤，也讓人類自此必須要不斷地透過外部化（或技術）來與那個原初預設劃定界線。

缺陷——艾比米修斯的失誤——在不斷透過技術予以彌補的過程中，出現了 Stiegler 所謂的從虛擬的可完美性（*virtual perfectibility*）到實際的可完美性（*actual perfectibility*）之轉化。人類的可完美性具有的本只是一種虛擬性，因為遺忘所帶來的失誤，最終透過外部化過程，將其放到人類自身——自然——之外來處理。後者即是 Stiegler 所謂實際的可完美性概念。實際的可完美性，乃是在意識到自身之缺乏時才得以出現，並且會嘗試透過技術來達成。原初的可完美性之所以是虛擬的，是因為原初的失誤並未脫離人類自身，人類也沒有與其建立距離（脫離自身或外部化）。實際的可完美性則發生在外部化過程之後，在人類與其自身之原初失誤建立了距離，也因此，人類逐漸認為似乎可以透過技術達至一種完美狀態<sup>11</sup>。缺陷及其經驗的外部化，是人類用來不斷辨認出自身作為人類之基礎或證明。缺陷經驗的外部化，也帶給人類原初失誤本身便具有之潛力，亦即，提供或建立與動物或自然之距離，然後再透過技術來彌補該距離的可能性。從「去—失誤」出發，我們更可能接近何謂「成為人」的內涵：去失誤或許是使人類與自身之原初失誤建立距離，而讓「（技術意義下的）人

---

11. 近年來發展出一項新的科技——仿生手，也就是讓失去手的人能藉此恢復功能，其功能甚至包括各種感知，也就是人類身體所具有的能力。試用者指出，仿生手也讓他感覺到自己完整的。

類」——而非動物或原初的人類、自然的人類——得以出現的關鍵元素。若是回到人工智能本身的發展及運作，其並未有這樣的一種從內環境到外環境、從原初失誤到去失誤等過程，換句話說，人工智能與其自身是沒有距離的，其無法成為觀察者而觀看自身，也無法具有反身性地對自身的觀察進行觀察，也就無法出現所謂「像人一樣思考」的可能性。然而，對「人工智能是否能夠思考？」這個問題而言，或許較適當的描述應該是，其可能更接近的是「像人工智能一樣思考」，而不一定是「像人一樣思考」。這也回應了 Bostrom 對人工智能的發展所提出的觀點：我們不應將其擬人化，「陷入人類的框架，都會萌生毫無根據的期待」（Bostrom, 2014, p. 116）。下節中便以自動駕駛技術為例，延伸「去失誤／預設」這組區別的內涵，作為重新理解人工智能的另種可能性。

### 參、「去一失誤／預設」作為重新理解人工智能發展的可能性——以自動駕駛技術為例

近年來的自動駕駛技術，不僅已經成為一個熱門的議題，同時也是一個對於未來移動的想像。全世界有許多企業與政府均投入相關技術與產業的研發與發展，例如 Google、Apple、Tesla、BMW、Uber、Lyft 等（Davies, 2018）。自動駕駛技術是以人工智能科技為基礎，企圖實現無人駕駛的各種運輸載具，包括汽車、公車、飛機、船隻等。2016 年，美國國家公路交通安全管理局（National Highway Traffic Safety Administration, NHTSA）依據美國汽車工程師協會（Society of Automotive Engineers, SAE）所定義的不同自動化程度，將自動駕駛分為五個等級：

等級 0：無自動駕駛。完全手動操作。

等級 1：弱駕駛輔助。人類負責主要駕駛任務，但機器可單一輔助方向或保持速度，例如定速巡航。

等級 2：部分自動駕駛。人類負責主要駕駛任務，但機器可根據環境因素進行自動操作，例如自適應巡航。

---

12. 資料來源：<https://technews.tw/2017/01/18/what-is-autopilot>

等級 3：有條件全自動。系統完成所有操作，人類做部分選擇回應。

等級 4：高度自動化。系統完成所有操作，人類不一定要做選擇回應。

等級 5：完全自動化。系統完成所有操作，人類不用做選擇回應，在所有情況下皆可正常行駛。

自動駕駛技術的原理基本上就是將汽車視為一種機器人，透過各種感測器對路況與周邊情況進行偵測，將蒐集到的數據傳輸到處理器，處理器透過人工智能演算法對數據進行辨識與判讀後，透過電機系統根據判讀之數據來操控機械裝置，進而控制車輛執行各種動作<sup>13</sup>。換句話說，自動駕駛技術基本上就是感測技術、定位技術與決策技術（人工智能）三者結合的產物<sup>14</sup>。其中，感測技術以影像偵測辨識技術等視覺科技為最重要的一環，也是目前較廣泛的應用範疇。目前的自動駕駛技術均配備有攝影機或雷射感測器，其目的就是要讓車子可以「看見」並且可以據此進行判斷與動作<sup>15</sup>。這也是人工智能技術非常仰賴 Nvidia 這類的影像處理晶片技術大廠的主要原因之一。Nvidia 自動駕駛部門總監 Danny Shapiro 便指出，人工智能的精準程度，在很大程度上取決於「極端的」計算能力與大量代碼，以及強大的電腦及影像處理元件，「人工智能需要決定前面的物體是一個人、車輛、消防栓還是什麼別的東西」<sup>16</sup>。然而，何謂「看見」？心理學家 Peter Hancock 從另一位心理學家 James Gibson 有關知覺與行動之關聯性研究中指出，「對人類駕駛而言，視線（vision）是最重要的。然而，駕駛所看到的事物取決於其如何駕駛車輛：煞車、加速及行進等，都會改變車輛位置以及駕駛的視野」（Hancock, 2018）。此一知覺與行動的相互依賴，意味著人類駕駛在面對真實環境時，也會對其他人類駕駛的行動產生特定的預期，而這些期待很

13. 資料來源：<https://www.stockfeel.com.tw/%E7%84%A1%E4%BA%BA%E9%A7%95%E9%A7%9B%E5%88%86%E5%9B%9B%E5%80%8B%E9%9A%8E%E6%AE%B5%EF%BC%8C%E6%88%91%E5%80%91%E7%8F%BE%E5%9C%A8%E5%88%B0%E5%93%AA%E5%80%8B%E9%9A%8E%E6%AE%B5%E4%BA%86%EF%BC%9F/>

14. 資料來源：<https://www.techbang.com/posts/49294-is-your-exclusive-driving-the-car-the-car-a-get-operation-principle-and-development-of-unmanned-vehicles>

15. 例如：<https://www.bnext.com.tw/article/42885/how-self-driving-work>

16. 資料來源：<http://www.knowledgeatwharton.com.cn/zh-hant/article/9327/>。該文中提及之「極端」運算能力，依照慣常中文翻譯，推測應指的是「邊緣運算能力」（Edge computing）。

多卻是以約定俗成或者是非概念性——也就是非形式化——的行為或認知為基礎。

Uber 高級工程經理、美國卡耐基梅隆大學教授 Jeff Schneider 也指出，汽車交通事故大多肇因於兩類錯誤，其一是認知錯誤，其二是決策錯誤。他認為，認知錯誤可以透過感測器、雷達、攝影機、Lidar（遠端感應系統）等技術來避免，這些資料再透過人工智能系統進行計算，最後形成決策。此外，他也提到，提高精確度的方式之一就是確保自動駕駛車輛不會搞錯各種不確定、模糊不清的情況。自動駕駛技術涉及到的同時是對外部環境與車輛內部資訊的搜集、判讀、動作與監控，並且兩者的結合與互動或許是自動駕駛技術是否能夠成功的關鍵（Davies, 2018）<sup>17</sup>。然而，這其中仰賴的人工智能技術，仍有許多問題待解決，最重要的是在面對道路上複雜情況時，能否做出精確判斷與動作的問題<sup>18</sup>。這部分也因為人工智能並未能夠像人類智能一樣，透過經驗、情感、個人認知等對真實路況進行判斷，因此對外部資訊的搜集與判讀之需求便更加重要。

自動駕駛技術在高速道路上或者是相對封閉之區域的試驗通常較為成功，原因是這些場景中的物件與可能發生的情況都較為單純許多。一般道路上的情況更為複雜，不只是交通號誌、行人、自行車等各種車輛，還有許多突發性的情況，例如自行車或行人經常有穿越道路的行為舉動，或者道路邊可能有障礙物等。這些也都需要仰賴更多外部資訊的搜集以及更精確地對其之判讀的人工智能技術。Abdesalam Soudi 的文章標題或許更精確地指出了自動駕駛技術關乎的不只是道路規則的問題，更是有關駕駛語言（the language of driving）的問題（Soudi, 2018）。其觀點與前述 Gibson 有關知覺與行動的研究的討論是殊途同歸，也就是關於「溝通」的問題。人工智能科技如何與其環境進行「溝通」，以及這樣的溝通與人類智能的溝通是否有異同之處？Soudi 指出，駕駛的語言是相當複雜的。如果以前一節中有關智能與脈絡之關係的討論為基礎，駕駛對其環境的認知可能還得包括諸如時間點、道路類型、氣候等。這些脈絡對人類駕駛所具有的意義，以及對駕駛行為的可能影響，不僅單純地是一種輸入／輸出的運作，而是一種詮釋關係（interpretation），這些詮釋也都構成了駕駛語言的一部

---

17. 資料來源同上注。

18. 例如：<http://www.sciencemag.org/news/2017/12/are-we-going-too-fast-driverless-cars>

19. 資料來源：<https://tw.appledaily.com/new/realtime/20180322/1319658/>

分。

儘管近年來全世界對自動駕駛技術的未來都抱持樂觀看法，但近日在美國亞利桑那州所發生的一起 Uber 自駕車撞死橫越馬路之行人的事件，也凸顯了一些過去未受到關注的問題<sup>19</sup>。事件發生後，Uber 也立即停止其在北美的測試計畫，待問題釐清後再議。在稍後公布的現場影片中，我們則可以看到行人突然出現在車輛行經道路上，且當下駕駛並未接管自動駕駛，因此造成遺憾。在事後的報導中，除了探討自動駕駛車輛似乎未偵測到異狀而進行減速，以及駕駛當下的分心行為之外，也有報導指稱可能是行人的錯，而非自駕車的問題<sup>20</sup>。無論行人是否突然出現，至少可以肯定的是，自動駕駛技術在面對如此突如其來的情況時，依舊難以快速回應。自動駕駛車輛在測試期間，便已經有不少問題出現，這些問題大多被歸因於與人類駕駛的魯莽行為有關<sup>21</sup>。道路情況的複雜性，不僅考驗人工智能技術，同時也在形塑人類智能與人工智能如何共存的問題<sup>22</sup>。隨著人工智能的發展，人類智能似乎越來越被視為是對其之阻礙，而非是與之合作的對象。這在理解人工智能與人類智能之差異上，似乎不僅是道德上的議題，更是一個科技哲學上的問題。

就前述自動駕駛的相關技術為例，這些技術的首要目標均在於廣泛且儘可能詳細的資料搜集，以及對這些資料進行精準辨識與判讀的能力。換句話說，技術的目的在窮盡（exhaust）關於自身與其環境中可能存在或即將發生的現象，以便發展出更準確的預測與行動。其邏輯假定，假如搜集的資料越齊全、完整，就越有可能進行更精確的解讀，以便有更正確的回應與動作。換句話說，儘可能地且完整地羅列（counting

20. 資料來源：<https://buzzorange.com/techorange/2018/03/22/temp-police-release-uber-car-crash/>

21. 資料來源：<https://chinese.engadget.com/2016/09/26/googles-self-driving-car-is-the-victim-in-a-serious-crash/>

22. BMW 在中國負責自動駕駛的 Robert Bruckmeier 即指出，自動駕駛在中國的試驗遇到的困難可能要比想像中多。其中的問題在於人工智能並非是在學習，而仍只是依照由外部輸入的資料庫作為其運行之框架，其學習並非如同人類智能一般。該文章提到，「如果要按照西方國家的資料訓練自動駕駛汽車拿到中國用，在理論階段，那都不太能實行。所以在中國做自動駕駛，就得中國研究中心來做，得用中國的數據，學習中國的風格」。這也呼應了在本研究第二節中提到的關於智能是否能夠依著不同文化與脈絡而發展出相對應的學習內容，似乎對人工智能發展而言是關鍵的元素。資料來源：<https://www.techbang.com/posts/53627-bmw-tried-autonomous-ai-in-germany-to-china-found-that-simply-not-possible>

out) 各種可能性與選項，乃是當前的人工智能科技所依賴最重要的原理原則之一。人工智能科技的運行要有更精準的解讀，仍需要搜集儘可能詳細的資料以供處理。在交通號誌的辨識、行人與車輛的行徑與動作的解讀等，都還是在一個「窮盡所有可能性」的選項下來發展。此一透過窮盡以作為理性與客觀之基礎，或許也構成了人工智能乃至於現代科技的道德角色內涵。其背後的語言似乎是，如果窮盡一切後仍有失誤，那麼就是人類智能的問題，而非技術的問題。因此，對自動駕駛技術而言，窮盡其預設的內涵，而非去失誤 (de-fault)，去失誤反而成為其環境。這是說，窮盡是作為解決去失誤的基礎，在操作上則是其主要方式。

我們可以先用 Spencer-Brown 的形式理論觀點來說明。Spencer-Brown 說：「做出一個區別 (Draw a distinction)」 (Spencer-Brown, 1969, p. 3)。對任何觀察者而言，除非先做出一個區別，否則我們無法「指出」任何事物。例如，只有在「合法／不合法」這組區別下，我們才能夠認識到哪些事物或行為是合法的，哪些又是不合法的，但合法或不合法的內涵為何，則會隨著時間、空間、脈絡而有所變化，其並非是固定不變的。就此而言，「合法／不合法」這組區別就是一個「形式 (form)」。以這組區別為例，Spencer-Brown 指出，一個區別總是用來標示出該區別的其中一面，而非另一面。也就是說，「合法／不合法」這組區別所要標示出來的是合法的這一面，另一面 (不合法) 只是作為其環境而存在。假若這組區別要具有持續的有效性，該組區別則需要不斷地能夠標示出「合法」的這一面。這樣的運作如何可能？從形式理論觀點而言，這組區別需要不斷地在前者 (該區別所要標示出來的一面，也就是「合法」這一面) 中不斷地引入 (introduce)、重複 (repeat) 或複製 (copy) 該組區別。換

- 
23. Spencer-Brown 指出，當一個區別的兩面之間的界線被標示出來之後，這個界線便可稱之為「形式」 (form)。因此，一個「形式」有兩面。Luhmann 則進一步認為指出，「形式」就是關於兩面的事物，亦即系統與環境。如果我們用「秩序」 (order) 與「擾素」 (noise) 來說明，那麼系統便是某種具有「秩序」或具有組織性的實在或現象，而其環境則是充滿了各種「擾素」。此外，秩序與「擾素」之間的關係也不是直接的，例如系統與從環境中突現或茁生出來的。Luhmann 以 Saussure 所使用的 signifiant (signifier) 與 signifié (signified) 來說明「符號」 (sign) 乃是一個具有兩個面向的形式 (a form with two sides)。當作為一個符號而使用時，我們必須總是移入該形式的內部，並且從其內部進行運作，亦即 signifier 的這一邊。換句話說，一個形式區別出來的兩邊，標示出來的乃是系統，也就是 signifier，沒有被標示出來的為其環境，亦即 signified。語言也是使用相同的假定，亦即文字 (words) 指出了某種我們沒有非常清楚知道的東西 (Luhmann, 1995[1984], p. 45)。



言之，前者乃是持續地以「合法／不合法」這組區別——作為一個形式——來觀察其自身之環境（那些無法被標示為「合法」的那些事物）。這便是 Spencer-Brown 或 Luhmann 所謂的區別的「再進入」（re-entry）（Luhmann, 1995[1984]）<sup>23</sup>。因此，合法的這一面所涵括的內容是相對而言確定的，那些無法被標示為合法的事物都會被放到其環境中。此外，這兩個面無法同時被使用：沒有一個事物可以同時是合法與不合法。據此，這個有關區別的形式理論或許為人類智能與人工智能之討論提供了洞見。如果我們將人類智能視為一種形式，而人工智能是另外一種，這兩種形式便也代表使用了兩組區別。就本研究而言，我們視人類智能之形式是使用了「de-fault/default」這組區別，而人工智能則是使用了「default/de-fault」之區別。人工智能在努力邁向「像人一樣思考」的道路上之發展，更可能用來說明這兩個形式本身的內涵。

Default 除了譯為「預設」之外，經常也會看到有「默認」之譯。人工智能系統在形成自身的運作上使用了「default/de-fault」這組區別，並且在「default」這一端複製了這組區別，以此而使其得以持續地維持該區別之運作，也就是透過羅列所有可能性，以對其環境進行溝通（communicate about）。就此而言，去失誤乃為其預設之環境，前者所具有的複雜性便可能遠高於後者。人工智能系統以窮盡作為其預設，而對來自其環境中的 de-fault 進行化約，以使人工智能系統得以持續運作。也就是說，以自動駕駛技術來看，其所使用的人工智能科技必須首要能夠辨識並理解周遭的事物，以及這些事物及其動作代表的可能意義。然而，不只是事物本身，還有這些事物的運動，以及事物彼此之間、事物的運動彼此之間所可能具有的關係，都必須要能夠為人工智能科技所觀察到並且予以解讀，否則其回應將會遇到一定程度的困難或阻礙，甚至造成誤判。對人工智能而言，這個「預設」便表現為「窮盡」，其多半是一組相對其環境而言簡單或簡化的「程式」，這組程式可以在一定程度上對其環境進行窮盡的運作——亦即化約來自於環境中的複雜性，儘管大多數情況並非能夠如其所願，但「去失誤」作為其環境，其複雜性或許便已經遠高於人工智能系統的「預設」本身。

人類智能則不同於人工智能，如前面所述，去失誤乃為其預設，其表現為「去失誤/預設」（de-fault/default），預設反為去失誤的環境。就 default 是一種缺乏狀態而言，de-fault 的運作是期待能夠補充或彌補這個缺乏狀態，而非要窮盡它。這或許也是人類智能至今仍具有諸如創造力、想像力等人工智能尚未能夠具有的能力之原因。

從 Stiegler 的觀點來看，being-in-de-fault 是發明、發現、創造、想像的來源，正因為人類沒有被指定的性質，所以人類反而具有跳脫其僅作為自然之一部分的潛力。人類智能面對的是原初之缺乏、缺陷以及對其之經驗，就人類智能而言，default 作為其環境反而提供了豐富的可能性，亦即具有較 de-fault 高的複雜性。人類智能的學習與認知世界——彌補其缺乏狀態——之過程，更像是 Garfinkel 的索引式表達 (indexical expression) (Garfinkel, 1992[1967], pp. 4-6)。<sup>24</sup>

相較於索引式表達，人工智能技術已經可以非常輕易且快速地完成各種客觀式表達 (objective expression)，也就是理性且客觀的資料搜集與分析。然而，儘管客觀式表達的應用乃持續地嘗試發展出更有效地將索引式表達轉化為客觀式表達的方法——例如將顏色轉變成 0 與 1 的組合，Garfinkel 認為這樣的嘗試並非總是令人滿意的 (Garfinkel, 1992[1967], p. 4)。例如，當我們想要很精確地將性別歸於兩類時 (客觀性表達)，就會發現問題，因為總是會有例外出現，而這些例外要不就是被粗暴地對待而硬是被歸到其中一類，要不就是以原本的分類來指出例外的錯誤，進而將其排除在外，甚至消除之。換句話說，社會成員的意識索引系統或許都不同，在運作上也會受到脈絡、經驗等影響。然而，索引式表達也是構成社會成員對某個場景或脈絡之解明的重要元素，社會成員對場景的解明呈現或多或少的差異性，也正是因為每個人

- 
24. 「Indexical expression」一詞在中文世界中沒有特定的翻譯，就 index 一詞而言，除了有索引、目錄之意外，還有標誌、跡象等意。在英文字典中，indexical 為 deictic 同義字，後者為「relating to or denoting a word or expression whose meaning is dependent on the context in which it is used」。換句話說，某字或某個表達，其意義取決於該字或該表達所使用的脈絡而定。更進一步說，某字或某表達本身沒有固定的意義或內涵，若想要理解該字或該表達的意義，就得要將其放回到其所使用的脈絡中，這些字或表達提供了某些線索或跡象，透過這些線索或跡象，才能夠「找到」或者是獲得其所代表的意義。Garfinkel 也是在此意義上使用該字，他指出社會成員對某個場景的解明 (accountable)，也就是使該場景對社會成員而言是「可觀察的且可說明白的」(observable and reportable)，乃是取決於社會成員所使用到的關於該場景中的線索或跡象 (Garfinkel, 1992[1967], p. 1)。因此，相同的場景對不同的社會成員而言，不見得在一開始就具有相同的意義。此外，社會成員對該場景的解明，也會受到其如何取用該場景中所存在或不存在的線索或跡象之引導。因此，對 Garfinkel 而言，社會成員對場景的解明之所以具有反身性，正是因為「社會成員對場景的解明也會構成該場景的一部分」。所謂的「對場景之解明」即是在此一反身性運作的過程中成為一個「持續地完成」(ongoing accomplishment)。順此，就字詞所放回去的脈絡、字詞所提供的跡象或線索等意義而言，在本文中將 indexical expression 一詞譯為「索引式表達」，似乎更能傳遞 Garfinkel 對該字詞之使用的意義與脈絡。

所使用的索引都有些許差異——不會完全不一樣，但也不會完全一樣。因此，同樣都是過馬路的場景，不同的參與者因為攫取的索引不同，也會導致其對該場景的解明有差異，進而引導其對場景的回應有所差異。

據此，人類智能並非是以清晰、精確、理性、客觀等方式來獲取關於其環境（default）的資訊，而毋寧是一種索引式的，必須要依賴線索的、相對來說是主觀的方式。在 Dreyfus 的討論中也指出，「對模糊性的寬容允許人類使用關於目的與脈絡的資訊，以縮小剩下的可能語法分析與意義的光譜，而不用要求該對結果的詮釋是要絕對沒有模糊性的」（Dreyfus, 1965, pp. 34-3）。如果我們以神經科學哲學研究的粗淺論述來看，大腦神經元的運作對人類智能（部分展現為意識）之生成的最可能貢獻，或許是其生產了「模糊性」這樣的狀態。如果模糊性是神經元運作過程中所突現出來的性質，其可能也是意識生成的基礎，或者也可以用來解釋主觀經驗的內涵。心理學研究早已指出，大腦提供了記憶的功能，不論是長期或短期記憶。無論是哪一種記憶，在某種程度上，這些記憶都不是精確的（Bernecker & Michaelian, 2017; Glannon, 2019）。大腦運用各種方式來記錄，但每一種記錄都只是一種索引式的紀錄，只有記憶的部分或片段被記錄下來，而非所有細節。心理學研究也指出，即便是再精確的記憶都具有一定程度的模糊性，也就是都有著不那麼精確的部分。這些部分或許即是意識運作的基礎。相反地，人工智能以及電子計算機的記憶，都是以精確、無遺漏、鉅細彌遺為基礎，以至於無法從中茁生出模糊性。若再回到自動駕駛技術，以辨識物件而言，假如大腦的辨識運作經常呈現的是一種模糊性操作，綜合了諸如物件的顏色、紋理、形狀、出現的位置等資訊，以及這些特徵與記憶的連結等。然而，這些連結並非都是確定的，而是在一堆可能的連結中進行模糊性地比對，然後產生出可能是正確或僅是對當下情境而言具有適切性的回應。我們不需要精確辨認出物件的紋理、顏色、形狀，使用了以模糊性為基礎而建立的各種連結，在這些連結中找到最可能的答案。所謂在意識層次上的「確定」，諸如「我確定這是一件白色背心，且我知道這是我的衣服」，可能也只是一種大腦運作出來暫時性的結果，這些暫時性的運作產物也是以模糊性為基礎所建立的各種可能連結<sup>25</sup>。

---

25. 例如 Walter Freeman 對兔子的學習模型研究中主張，大腦是一個非線性的動態系統（nonlinear dynamical system）：「大腦不只是單純的抽離出特徵，其結合了感官訊息與

這個預設狀態就如同其另一半「去失誤」一樣，只是此一過程的一部分。如果將「預設」當作是系統運作上的真值，也就是被標示出來的那一端，那麼無論是強人工智能或者是深度學習（deep learning），就可能朝向將可完美性視為是實際的（actual），而非虛擬的（virtual）。「去失誤」作為「預設」的環境，前者的複雜性要大於後者。如果從複雜性觀點來看，系統藉由自身之運作以化約來自環境的複雜性。在此運作中，系統是簡化了環境的複雜性，卻無法處理或涵蓋其環境中所有的複雜性。假如環境的複雜性總是高於系統之複雜性此一假設是可接受的，那麼無論系統如何運作，或提高其自身之複雜性，在實際上都無法處理環境中的所有複雜性。如此一來，人工智能若只是在發展一種理解（化約）其環境複雜性的運算系統，那麼其目的可能永遠無法達成。然而，如果將辯證的另一端「（去）失誤」視為系統運作的真值，或許情況會有不同，而這也是人工智能科技未來應可多著墨之處。

## 肆、代結論：「例示／窮盡」作為對人工智能技術之反思

綜上所述，人工智能是在虛擬世界中看真實世界，而相反地，人類智能卻是在真實世界中創造虛擬性，例如符號系統、語言、制度等。在人工智能的眼中——假如其真的看得到的話，真實世界只是一堆數字跟代碼。相對地，感官經驗的數位化資訊在人類智能的眼中，也只是一堆數字跟代碼。人類智能不知道那是什麼，也無法從那些代碼中聞到味道或者看到東西。例如，我們無法從代碼中知道這棵樹的樣子、形狀或者是顏色，甚至是這棵樹給我們的感覺、情緒等。在人類智能的眼中，我們是以「如其所是」的方式來觀看世界，相反地，在人工智能眼中，乃是用「如其所不是」的方式來看世界。也因此，當後者想要「進入」或「看到」真實世界的企圖時，其就必須

---

過去經驗...以辨識出該刺激與其對個體的獨特意義」(Dreyfus, 2007, p. 1151)。Freeman：「脈絡依賴是腦部記憶系統的根本性質，每一個新的經驗必然會微量地改變所有儲存的東西，為的是要讓一個新的條目被整合且完全地被部署在既存的經驗中。這與儲存在電腦中的記憶之性質不同，其每一個項目被分配到一個地址（address）...每一個項目都被劃分於彼此，新的項目並不會改變舊的項目。然而，在大腦中，儲存並沒有界線或分區，...每一個新的狀態變化...開啓了建構一個衝撞並形塑整個意向結構的本地類型」（轉引自 Dreyfuss, *ibid.*）。換句話說，新的東西加入會對原有的結構進行改變，並且重新構成一個新的整體。新的東西不是直接被安排到這個既有結構中的一個區塊，而是在與舊有結構的衝突或互動中，既讓原本的結構有了變化，也讓自己在該結構中獲得位置，形成一個不同於舊有的結構。大腦的認知就類似這樣的運作。

要「窮盡」這些「如其所不是」的事物。只有在窮盡之後，才有可能或認為其是所理解的「如其所是」的世界。然而，人類智能的如其所是，並非是先窮盡才知悉，而是一種「例示」（*illustration*）的方式。我們透過如其所是來建構關於這個世界的知識，透過如其所是來知道何謂如其所不是，而非相反。這兩個路徑，「窮盡／例示」，或許正指出了人工智能的盲點之一。

人類智能是可以在模糊性中做決定，但人工智能——至少就目前為止——卻只能在精確性中做決定。人類智能在做決定的當下，其面對的情境對其而言仍是具有一定程度的模糊性，換句話說，人類並不知道或無法確定該決定是否就是正確的決定，這部分是因為該情境仍對其自身具有模糊性。如果從形式理論角度觀之，人類是透過「例示」（*illustration*）的方式來化約其環境——包括技術——的複雜性，環境具有的是一種潛在的已被窮盡之特性（*exhausted world*）——也就是無所不包的世界。儘管也是化約，但化約的前提或目的並非是追求窮盡其環境中一切，而是「生存」，或系統的運作。生存的內涵或者以Stiegler的話來說，就是一種「持續去除原初失誤的過程」——透過技術將人類的原初失誤持續地外部化。在外部化的同時，人類不僅作為（非屬自然的）「人類」而出現，同時也讓作為（自然的）人類消失。這兩種人類是不同的。當人類因為技術的介入而逐漸轉向以窮盡為其認知上的邏輯——其實是技術的邏輯或本質，或者是以此化約來自環境中的複雜性，此一（非屬自然的）人類也將消失於其中。此時，窮盡將成為系統運作的真值（*exhaustion/illustration*），而其努力地想要化約來自其環境的複雜性——亦即「例示」（*illustration*）。這個轉變對於我們在思考科技與人類、科技與社會的關係之變化是有啟發性的。模糊性逐漸被人工智能所需要能夠維持其運作的環境所消除，也因此形塑了不只是新的自然語言，也可能帶來了新的道德性內涵。

從自動駕駛技術與人工智能之應用來看，「窮盡」具有兩種意涵，其一是其需要儘可能地透過各種感測器搜集環境數據，儘可能全面且精確地辨識這些來自環境中的訊號。其二是其也需要「窮盡自身」（*exhaust itself*），亦即，窮盡其運算法本身所能提供的判讀與回饋的選項。人工智能的精確性與可信度，必須要結合兩者，並且在兩者中取得一定程度的平衡，才有可能更好地應用在日常生活中，或者可以跟人類智能並存，而非僅是其輔助：

**圖一：Learning process of HI and AI**


資料來源：本研究製圖

本研究之目的不在主張以人工智能為基礎的自動駕駛技術是不可行的，相反地，即便如最高等級的完全自動駕駛技術，也就是無論如何都不需要人類駕駛干預、沒有方向盤、油門或汽車後視鏡等的完全自動駕駛汽車，都是可預期會有實現的一天。然而，本研究嘗試指出的是，目前的自動駕駛技術都是在受到控制的環境中——也就是如同Bostrom 所言之對人工智能而非人類智能友善的環境——進行測試。這些交通工具基本上不需要與人類智能有完全的互動，也不需要處理因為人類智能所可能導致的各種情況，因此或許可以避免這些可能無法解讀的風險。換句話說，人工智能科技若仍是以窮盡為其邏輯，那麼距離要能夠「像人一樣思考」的人類智能水準，可能還有一段路要走。此外，這個實現更可能會是以犧牲人類智能為代價，才有可能達到。這意思是說，人類能夠創造出來的可能是一個沒有人類智能或者人類智能最小化的世界或日常環境。如此一來，人工智能便可暢行無阻，而且可以完美的執行。這樣的世界或日常生活，有可能不再是人類智能優先的世界，而是人類智能如何適應人工智能的世界。機器出錯，我們會責怪人類，而非機器為何無法適應人類智能。例如，為什麼行人不遵守號誌、不走斑馬線、不依照規定來行動等。因為這些地方才有的人工智能可以準確執行的餘地，我們必須遵守號誌、走斑馬線、依照能夠順利讓人工智能運作的規則來行動，如此一來，人工智能才不會出錯、才不會撞上行走的人們、才不會無法辨識人們的行為等。

另外，我們（人類智能）允許「認錯人」這件事，或許是人類智能之所以為人類智能的關鍵元素，也是人類智能學習的基礎。相對地，當我們期待開發出能夠並駕齊驅或甚至是超越人類智能的人工智能時，我們並未同意人工智能也可以犯下跟人類智能相同的錯誤，或至少在實際的運作上，我們乃是努力地期待或想要解決人工智能

總是不那麼精確的部分或總是會出錯的結果。例如自駕車上路後遇到的各種問題，在對各種情況的判斷中，我們不容許其像人類智能一樣擁有犯錯的權利，例如對道路上人類行為的判讀，總是在避免造成「對人類的傷害」，但這樣的「避免」卻是在「不允許犯錯」的基礎上，而非是在「允許犯錯，但可以更小心」的基礎上，而後者即是人類智能在學習上與人工智能的差異，讓其可以「更精確地」完成某些事<sup>26</sup>。換句話說，我們對人工智能的開發與期待，在形式上受到人類智能的限制。我們期待人工智能可以如人類智能一樣，具有解決複雜問題、抽象思考、創造性的能力，但我們卻不容許其有犯錯的權利。當我們想要儘可能地避免錯誤發生時，或許也不可避免地再次落入「窮盡」的思維模式。儘管「窮盡」的本質便是在提供所有可能性，並且從中給予正確、有效、有效率的判斷，但也正是此一窮盡，在很大程度上脫離了人類智能的本質，亦即，對缺陷或錯誤的理解及允許，或許才是其獨特之處。由此也可以看出來，過去六十年人工智能的研究與發展，逐漸朝向模仿自然或仿生的方向前進，例如神經網絡及遺傳演算法等。其背後所透露的是，如果可以理解大腦神經網絡或生物體的運作，便可以透過科技製造或模擬出一個一模一樣的東西。然而，延伸 Stiegler 的觀點來看，如果缺陷經驗是「成為人」的基礎，也正是其雙關「去—失誤／預設」在人類這邊是不斷地在前者（去—失誤）複製該區別，「成為人」的預設便是一種「不斷去—失誤」的過程，也就是「不斷地去失誤」本身就是人類智能的目的。因此，一旦失誤可以真正地被去除，那麼或許也代表人類智能的終結——人類將不再能夠「成為人」，而是其他東西（something else）。

---

26. 例如：<http://www.sciencemag.org/news/2017/12/are-we-going-too-fast-driverless-cars>

## 參考書目

- Bernecker, S., & Michaelian, K. (Eds.). (2017). *The Routledge Handbook of Philosophy of Memory*. London and New York: Routledge. <https://doi.org/10.4324/9781315687315>
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press. DOI: <https://doi.org/10.1017/S0031819115000340>
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford & New York: Oxford University Press.
- Clark, A. (2013). Re-Inventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind. In M. More & N. Vita-More (Eds.), *The Transhumanist Reader*. West Sussex, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781118555927.ch11>
- Davies, A. (2018). The Wired Guide to Self-Driving Cars. Retrieved from <https://www.wired.com/story/guide-self-driving-cars/>
- Dreyfus, H. L. (1965). *Alchemy and Artificial Intelligence*. Santa Monica, CA: RAND Corporation.
- Dreyfus, H. L. (1972). *What Computer Can't Do*. New York: Happer & Row.
- Dreyfus, H. L. (2006). Overcoming the myth of the mental. *Topoi*, 25, 43-49. <https://doi.org/10.1007/s11245-006-0006-1>
- Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, 171, 1137-1160. <https://doi.org/10.1016/j.artint.2007.10.012>
- Gardner, H. (2011). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Garfinkel, H. (1992[1967]). *Studies in Ethnomethodology*. Cambridge: Polity Press.
- Glannon, W. (2019). *The Neuroethics of Memory: From Total Recall to Oblivion*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781316443712>
- Hancock, P. (2018). Self-Driving Cars and Humans Face Inevitable Collisions. Retrieved from <https://www.scientificamerican.com/article/self-driving-cars-and-humans-face->



inevitable-collisions/

Haugeland, J. (1998). *Having Thought: Essays in the metaphysics of mind*. Cambridge, Massachusetts; London, England: Harvard University Press.

Hickey, L. P. The Brain in a Vat Argument. Retrieved from <https://www.iep.utm.edu/brainvat/>

Luhmann, N. (1995[1984]). *Social Systems* (J. Bednarz & D. Baecker, Trans.). Stanford, California: Stanford University Press.

Malik, K. (2000). *Man, beast and zombie: what science can and cannot tell us about human nature*. London: Weidenfeld & Nicolson.

Marx, K. (1887[1867]). *Capital* (S. Morre & E. Aveling, Trans.). Moscow: Progress Publishers.

Moore, P. (2003). *Being Me: What it Means to be Human*. West Sussex, England: Wiley.

Nilsson, N. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511819346>

Soudi, A. (2018). Driverless cars might follow the rules of the road, but what about the language of driving? Retrieved from <https://theconversation.com/driverless-cars-might-follow-the-rules-of-the-road-but-what-about-the-language-of-driving-88824>

Spencer-Brown, G. (1969). *Laws of Form*. London: George Allen and Unwin Ltd.

Stiegler, B. (1998). *Technics and Time, I* (R. Beardsworth & G. Collins, Trans.). Stanford, California: Stanford University Press.

Todes, S. (2001). *Body and World*. Cambridge, Massachusetts: The MIT Press.

Tomlinson, J. (2007). *The Culture of Speed: The Coming of Immediacy*. London: Sage Publications Ltd. <http://dx.doi.org/10.4135/9781446212738>

