

# 人本 AI

林從一\*

## 一、前言：這個世界從來不是單單從理性規劃而來的

我們的世界之所以長這個樣子，有非常複雜的原因，有理性的部分、有欲望的部分、有毀滅的衝動、有遊戲的心情、有同情、有著看好戲、有自然而然隨遇而安的、有天注定的、有美學式的、有教育式的……。我認為，理性僅僅是形塑世界的一個小力量，其他非理性的因素才是大力量。

重點是如何讓各式各樣理性的、非理性的力量「如其所是」、「各安其位」卻又相互滋長的存在一起。這條安身立命、同情共感、大家一起旅行的路應該如何鋪？這是政治、社會與修養同有的課題。

理性可能僅僅是一種鋪路的小工具，欲望、毀滅、遊戲、同情、隨遇而安、美感、教育等可能是更有用的工具，甚至是檢驗鋪路成效的標準。舉例來說，「看起來美的社會」或「讓人覺得幸福的社會」或「自然而然隨遇而安的社會」或「充滿遊戲趣味的社會」似乎比「理性的社會」讓人覺得更值得追求。<sup>1</sup>

更根本的說，人與人類社會是多樣與複雜的，充滿偶然性、不可預測性，同時人類具備對既有規範批判、放棄和重新想像的能力，簡言之，人類歷史是對未來開放的。

一個人本 AI 應該尊重人類的開放性、差異性，以及差異性所要求的同情共感（可理解性）。如果人本 AI 真是不可能的，例如 AI 內含一種科技決定論，那麼我們就必須從開放性、差異性與可理解性的角度批判 AI。

## 二、AI 與集體偏好效益主義

AI 所需的巨量數據一開始就涉及專家判斷，無論是專業領域專家或是生活領域專家，所謂生活領域專家也就是一般人，善於過平常日子的一般人。AI、

---

\* 華梵大學校長

<sup>1</sup> 林從一（2017）。《最值得過的人生：哲學爸爸給女兒的大學禮物》，臺北市：平安文化，頁 171-172。

機器學習、深度學習的最基本模式與成果是這樣的：

1. 針對一個問題，例如這是視網膜病變嗎？這適合這工作嗎？這人或公司具有這項貸款的信用嗎？這人犯罪的再發率？
2. 關於這個問題的巨量的案例，專家給出「是」或者「不是」或者「或然性的」的答案，然後將這些帶著專家基本判斷的巨量案例餵給機器。除此之外，專家不給機器其他判斷，特別是不給基本判斷的理由。
3. 經過大量的學習後，機器的演算法會產出比專家正確率更高的判斷。<sup>2</sup>

AI 追求最大效益途徑，AI 是一種實現人的某種功利目的的工具，因此，AI 涉及的主要理性概念是工具理性 (Instrumental rationality)，又稱效率理性、功用理性、目的理性、技術理性、科學理性。相應的，AI 涉及的道德原則是效益主義 (utilitarianism)，而且，既然涉及巨量資料，這種效益主義是一種「集體偏好效益主義」(collective preference-based utilitarianism)。

「集體偏好效益主義」主張道德上正確的選擇是那些可以極大化完成多數人共有偏好的選擇。本文將指陳「集體偏好效益主義」的 AI 帶給「人本 AI」的幾個挑戰，以及一個符合人本 AI 的想像案例。

### 三、工具理性獨大

AI 決策中如何考量價值理性，譬如慈悲、社群感與共善，避免工具理性獨大？根據「集體偏好效益主義」，道德上正確的選擇是那些可以極大化完成多數人共有偏好的選擇。

明顯的，效益主義本身就是一種道德原則或價值的窄化。智慧城市設計極大化了人們的交通便利性，但不一定增加了社會韌性與社會信任感。高度仰賴 Fintech (金融科技) 的國家財政操作，縱然能極大化 GDP，也可能因此忽略社會正義與個人尊嚴。效益主義所涉及主要的價值是幸福 (happiness)，幸福固然重要但不是唯一，社會韌性、社會信任、社會正義是無法化約成幸福的，價值是多面向的。

此外，追求可計算性也導致價值窄化。計算需要相當程度的形式化與量化，然而，並不是所有價值可以輕易形式化或量化的。有些價值是高度脈絡依存的 (context-dependent)，脈絡具有高度變動性，價值取向或至少價值詮釋因此也具有高度變動性，以至於抗拒形式化與量化。

<sup>2</sup> 林從一 (2021 年 4 月 6 日)。〈從人類與 AI 的主奴辯證到 X.A.I.〉，國立中央大學哲學研究所演講稿。

另外，「多數人共有的偏好」的群體性反而窄化了社會中的偏好項目，長此以往，甚至會限制住了人們的想像力，限縮社會與個人發展的可能性。例如，社會主義社會中人們的集體偏好與資本主義社會中人們的集體偏好，兩者之間的差異如此的巨大，以至於兩種社會中的人們難以想像彼此社會的價值差異。例如，在美國與中國，許多人無法想像有些北歐人為何認為「有太多錢是不道德的」。

統而言之，「集體偏好效益主義」主張「道德上正確的選擇」是那些可以極大化的完成多數人共有偏好的選擇，在這樣的道德框架下所設計或使用的 AI 會放大人類群體偏見，同時，不利於維持差異性與獨特性。例如智慧城市設計，當我們問什麼樣的公共運輸系統具有最大效益，最有效率的滿足最多數市民的最多交通需求，我們的智慧城市系統便會成為一個集體偏好效益主義的系統。但是，對你方便的不見得是對我方便的，對多數人方便的不見得對少數人方便。

除了在追求集體偏好的極大化效益外，如何一開始就接受差異，這是非常重要的 AI 設計或運用的人本原則。

#### 四、「參與」的價值

哲學家麥金泰爾 (Alasdair MacIntyre) 在其 *After Virtue: A Study in Moral Theory* 中說 “Good life for man is the life spent in seeking for the good life for man.”<sup>3</sup>。他的重點是「美好生活」的不是被動給與的，不是別人或命運賞給你的，不是一個「靜態的結果」，美好的生活 (或生命) 是一個參與的過程，是你花在追求美好生活的那些年歲。換句話說，你必須主動參與在你的美好生活的建構中。

我認為，這個「參與」的價值不僅在美好社會上成立，在 AI 科技的發展上也成立。事實上，在 AI 科技的發展上更需要注意參與的價值，因為，集體偏好效益主義是一種結果論，容易忽略「參與」的價值。

#### 五、新的可能性

集體偏好效益主義是一種目的導向的觀點，但常常受限於舊有、常規目的的規限，開創出新的目的、新的可能性？我認為，這個挑戰可以表述為：除了

---

<sup>3</sup> Alasdair MacIntyre. (2013). *After Virtue*, p.254. A&C Black.

在追求集體偏好的極大化效益，AI 的設計或運用如何一開始就接受差異？如何接受偶然、無常與變化，並向未來開放？我認為，把人的因素以適當的方式帶回 AI 的設計或運用中，可以顯示出一條回應這個挑戰的路。

## 六、人本 AI 的想像案例——光合菌的成長故事

筆者正執行一個農業科技公共化的國科會計畫——《以科技公共化為取徑之永續農業發展》計畫。這個計畫本期有大約三千多位農友參與，下一期計畫將邀請他們開始講「光合菌的成長故事」。

農友的光合菌成長故事首先要符合脈絡化原則。每個農民有自己的故事，這些故事包含幾種知識。首先，監測系統所蒐集了光合菌田間理化資料如日照、氣溫、空氣濕度、土壤溫度、大氣壓力等數據，讓我們知道了許多光合菌田間理化資料。與理化資料一起蒐集的是生物資料，包括光合菌種類、使用分量、方法、作物成長情形、抗病性、抗旱、抗水澇、產量等。在一般模式下，這兩類資料就會被餵給 AI，運算出一些相關性，做出精準預測。但是我們是在脈絡化的原則下，蒐集光合菌的田間理化與生物資料，脈絡化讓資料的可詮釋性變得更高，人性的因素變得更明顯。理化資料如日照、氣溫、空氣濕度、土壤溫度、大氣壓力，生物資料如作物抗病、抗旱、產量，兩種資料都是放在一個個農友的個別實踐脈絡裡一起蒐集的，而每個實踐脈絡有著不同的場域面貌，有著農民個別的生活、生命、夢想與遭遇。

農友各自的「光合菌的成長故事」中有成功、有發現、有失敗、有漏洞、有偶然，甚至資料本身也不必精準，畢竟人的故事充滿缺點與不確定性。事實上，人的故事不僅是歸納已經發生的，更是面向未來的，未來便帶著高度不確定性，面向未來的因素如「希望」是超過去與眼前的，我們所希望的常常不是基於證據、資料與歷史。

沒有不尊重差異的脈絡，沒有不尊重變化的脈絡，光合菌的故事帶來脈絡化，因此接受了差異化、變化、偶然與無常。

脈絡化讓農友回到自己土地、生命與生活，農民也比較容易知道自己如何做出貢獻，為一個涉及 AI 的系統帶來豐富的差異性，使其更能接受變化、向未來開放。

我們的農友不是 AI 產品的消費者，他們因此沒有提供資料的義務，參與社群資訊分享是自願的，因此具有減少乃至於停止提供資訊的權利。整個光合菌

故事系統的設計原則不是極大化資訊，反而是追求在少量的資訊中，就能獲得穩定、豐富、多元的光合菌與人類成長的故事軸線。

另外，也值得一提的是，除了脈絡化原則外，光合菌的成長故事也遵守互助原則（互惠原則）。每一位農民的光合菌成長故事不僅是表現自己，同時也帶著協助他人、形成共善的意圖。

透過述說自己的光合菌故事，農友以保有主體性、差異性的方式，參與了 AI 科技的發展，實質的參與未來社會的發展。