

在機器時代找回人性：評介 *The AI Mirror*

鍾蔚文、王皓正*

書 名：*The AI Mirror: How to Reclaim Our Humanity in an Age of
Machine Thinking*

作 者：Shannon Vallor

出版日期：2024

出 版 社：Oxford University Press.

投稿日期：2024 年 12 月 9 日；通過日期：2024 年 12 月 13 日。

* 本文主要來自 2024 年 11 月 8 日國立政治大學 AI 讀書會，鍾蔚文教授的導讀，由王皓正整理並補充為書評。

** 鍾蔚文為國立政治大學傳播學院名譽教授，e-mail: waynechu@nccu.edu.tw。
王皓正是國立政治大學傳播學院博士生，e-mail: haoqhaoq@gmail.com。

本文引述格式：

鍾蔚文、王皓正（2025）。〈在機器時代找回人性：評介 *The AI Mirror*〉，《新聞學研究》，162，157-166。https://doi.org/10.30386/MCR.202501.0004

本書作者是美國技術哲學家 Shannon Vallor，她目前是愛丁堡未來研究所數據和人工智慧倫理學的 Baillie Gifford 主席。2016 年，Vallor 寫了《科技跟美德》（*Technology and the virtues: A philosophical guide to a future worth wanting*）這本書，主題圍繞當代科技發展。當我們今天談科技的時候，一定會牽涉到倫理與價值觀的問題，《科技跟美德》是一本不可錯過的學術著作。2024 年，Vallor 出版最新著作 *The AI Mirror*（*The AI mirror: How to reclaim our humanity in an age of machine thinking*）。Vallor 在這本書中提出了許多不同於主流的反調論證（polemic）觀點，更進一步論述 AI 作為「鏡像」的本質為何。

從本質而言，我們可以由此深入思考鏡子能反映與不能反映什麼，有趣的是，也可能因為我們觀看的態度、角度不同，因而改變、扭曲、遮蔽、切割、扁平化了我們眼前所見。但 Shannon Vallor 認為「AI 不是問題」，舉例來說，照鏡子的時候，如果你對鏡中的自己感到奇怪，那問題不出在鏡子本身，而是出自於你。簡單來說，Vallor 認為「我們（人類）才是問題」。

所以，《*The AI Mirror*》基本上是一本談人的書，談如何在機器時代找回人性。AI 跟其他媒介科技最不一樣的地方在於緊密深入人類生活的核心，它將影響到生活的每一個層面。AI 將改變我們現有的一切，尤其是那些跟人有關的事物。因此，人文社會科學將成為重點。談 AI 的時候，你不可能不回過頭來談人。

一、AI 最大的威脅：使人失去自己

Shannon Vallor 首先談的是，AI 的威脅不在使人類滅絕，而是使人們失去自信、失去了想像的能力，而且侷限了無限的可能，失去了道德

推理的能力。

Vallor 特別強調這一點是因為她認為當代論述經常有一個問題，就是當我們談 AI 的時候，談的都是科幻小說與電影裡的 AI。舉例來說，AI 會變得非常強大，它會代替人類，它會認為人類是劣等物種，於是把我們趕盡殺絕。Vallor 認為 AI 不會有這樣的威脅，她認為最大的威脅在於人類在 AI 中失去了自己。

Vallor 認為這就像是神話故事中的美少年納西瑟斯（Narcissus），他在池塘中看見自己的倒影因而迷戀上自己的容貌。如今，AI 也讓我們迷惑、讓我們深陷其中，如同納西瑟斯在池塘，我們在 AI 看到了自身的反射，AI 迷住了我們，正如納西瑟斯被自己的容貌所迷住一樣。我們逐漸忽視脆弱的人性，而追逐機器鏡中的完美幻影……如同納西瑟斯，我們輕易地、錯誤地感知這種反射，將其視為一種他者的誘惑，一位不知疲倦的伴侶、一位完美的未來情人、一位理想的朋友、一位公正的裁判、一位萬無一失的合作者。然而事實上，這只是一個讓我們日益孤獨的存在，一切都是我們與自己的對話。舉例來說，如同談到數位親密感（digital intimacy）的時候，因為 AI 相當「接近人」，它會為我們創造美好的鏡像，而我們也追求 AI 提供給我們的完美。Vallor 認為，聊天機器人在 AI 鏡面所投射出的「空洞的愛」可能只是一場災難。

跟所有的人文社會科學研究一樣，有時候作者目的不僅在解決當下問題，而是解決更後設的問題。Vallor 批判、揭露當代的問題，我們在日常生活裡有太多這樣的現象，人們看起來一切如常，可是當我們揭露之後，我們才得以發覺真相。舉例來說，許多我們現在討論的問題早在電腦（digital computers）出現之前就浮現了，所以這也是很多 AI 與人文社科研究在談的問題，就是說其實我們今天在跟 AI 相處的時候，我們發現很多問題是在 AI 之前就存在的。因此，Vallor 認為真正的威脅不

在於工具（／AI），而在於我們自己（／人類），人的價值才是問題核心。

不過，正因為問題出自於人，抵抗跟療癒的關鍵也在於人們自己。因為 AI 在道德和精神意義（spiritual meaning）上是模糊的。AI 的意義是由我們所定義的，因此承續先前的概念，Vallor 認為我們今天所擁有的價值根源於前電腦時代，但現在早已成熟，成為主控 AI 的價值系統。所以今天 AI 只是在鞏固、擴大既有的價值，而這樣做其實是危險的。舉例來說，Vallor 認為我們反而被「人性的機器化」（mechanization of the human personality）所束縛，我們開始太過於關注這一類的預測；反思性的洞察力（reflective discernment）被無意識的預測系統所取代；人的創造力和開放性被內容的局部優化所取代。Vallor 認為這是人性的智慧向機器思維的投降（the surrender of humane wisdom to machine thinking）。

於是，現在是一個很好的時機來思考究竟是哪些價值使人變得扁平化。Vallor 在第三章談論實用主義的特徵，特別批判長期主義（longtermism）。Vallor 認為我們太強調實用性、利益化、極大化等價值，這其實是我們身處於工業社會和數位社會的特徵，我們崇尚節省時間、強調效率，Vallor 認為這些價值觀才是一切問題的根源，我們不應以當代的利益來想像未來的問題。

二、電腦安全、價值對齊、AI 倫理——不一定掌握問題核心

Vallor 認為當代提出的許多解方，例如電腦安全、價值對齊、AI 倫理等並未掌握真正的問題核心。舉例來說，如今常見的價值對齊（alignment），人們會思考 AI 與現在的價值觀有無抵觸，進而提出修

改。然而，Vallor 認為並不是因為與當前價值觀不一致才導致不成功，反而是因為過於反映當代的價值觀，更可能複製偏見。因此，Vallor 認為符合當代主流價值觀的「有用」反而才最危險。我們應當重新檢視當代的專業倫理規範。很明顯地，目前 AI 專業倫理規範並不足夠。

另一方面，Vallor 認為重視創新（innovation）也成為問題，因為創新並不是價值，創新甚至不見得是進步，創新如要有價值應該具備有價值的目的。所以這個問題仍得回到人文本身，也就是重新思考何謂有價值的目的。基本上，Vallor 認為這背後指向了一個更大的問題，就是價值觀要改變。他提出了兩種能力，分別是**技術的專家能力（technological expertise）**和**技術的再想像力（technological reimagining）**。Vallor（2016）過去在《科技與美德》一書中便曾引用孔子、亞里斯多德等思想家的觀點，而在本書中藉由思想家的觀點來強調道德與技術智慧難以單獨存在，我們必須將道德智慧與技術智慧重新結合。Vallor 認為未來社會的機構（institutions）都需要發展「技術道德」（technomoral）。也就是說，我們必須在技術時代重新思考技術的智慧以及背後所需要的專業知識。

另一個極端的案例，Vallor 在第六章中討論微軟公司（Microsoft）提出來的 AI 計劃——「AI for good」，這是一個由微軟公司對於環境保育、健康醫療、文化、人道行動等議題的 AI 倡議，然而 Vallor 認為如果這些計畫並沒有仔細思索背後的問題核心，仍舊無法改變 AI 的結構性問題。Vallor 認為儘管這些行動與呼籲的立意良善，仍然是以我們現有對於科技媒介的掌握來想像 AI，反而呈現了思考的扁平化。我們真正需要的是重新評估工具，而不能再沿用舊有的習慣和價值觀，或單純根據當前的需求進行 AI 的想像。Vallor 認為獨立思考是一種智識的美德（intellectual virtue），因此鼓勵人們在思考問題 AI 時，應發揮獨立思

考的能力，以掌握 AI 問題的核心。

三、AI 只看到過去和現在，看不到將來

由上可知，Vallor 認為鏡像的作用是提取、放大並推進我們記錄過的、數據化的過去中主導力量和最常見的模式（What AI mirrors do is to extract, amplify and push forward the dominant powers and most frequently recorded patterns of our documented, datafied past.）。也就是說，AI 鏡子反映了我們，但只反映了我們和我們經驗中的一小部分，尤其這些由鏡子反射的數據與模式，背後指引的是過去與現在，並不能由此看到未來。因此，這就是危險之一。AI 鏡子讓我們以為我們所看見的就是全貌，甚至回過頭來改變我們的認知。但如同 Vallor 引用楊立昆（Yann LeCun）所說：「鏡子似有景深，讓我們以為它能反映所有的事物，但它只有一公分厚。如果我們試圖揭示它的深度，我們還會撞到頭。」

此外，鏡子也沒有辦法反映全部的我們。鏡像沒有聲音、氣味、深度、柔軟、恐懼、希望。更重要的是，它沒有想像力。鏡子所能讓我們看見的只是過往數據與模式的反射。Vallor 再將資料比喻成光，鏡子的表面和光源反映出它將折射出的事物，但它終究不能重構或充分反映出我們生活的遭遇。所以或許我們可以思考，究竟 Vallor 所說的這面 AI 鏡子之下，是否還隱藏了什麼事物？Fourcade & Healy（2024）在《序列社會》（*The ordinal society*）這本書中也談論到類似的問題。每當有技術機器的引進，社會便會因此產生新的秩序。本書討論的 AI 鏡子其實也反映了這樣的成見，在資訊社會當中，我們都能被換算成特徵資本（*eigen capital*）。也就是說，如 Vallor 所說我們將是可以被計算的物，我們被隱藏在我們的標籤（*label*）和分類之下，我們將不再是經驗的主

體，而是經驗的客體。

另一方面，Vallor 還認為 AI 無法反映出我們生活和行為的道德意義。我們應該思索 AI 鏡中所呈現與隱藏的事物為何，並仔細思考呈現的事實，小心不要落入鏡像的陷阱。如同 Vallor 所說，當我們把鏡子看成未來，AI 反而成為了我們道德、智力、文化和政治想像的束縛。我們終究得認清鏡子是現在與過去的反射，我們很難憑藉現有鏡中的呈現就可以輕易看到清晰的未來。

四、如何找回人性？

Vallor 認為當我們在鏡子中凝視自己之際，我們看到了機器的回望，我們卻不再知道我們是什麼……。也就是說，人類失去的是「自知」。這本書最終是關於人類自我認知的喪失，以及對我們道德與自信的喪失。可是人類未來並不是一片絕望迷茫，Vallor 曾說過，當問題出在人類身上，那抵抗跟療癒的解答也將出自人類。我們又應該如何找回自己呢？

Vallor 認為第一條路是建一個更好的鏡子，意思就是造一個更好的工具，不過他認為這個不容易，因為問題仍沒有解決，最終需要改變的還是必須回到人；第二條路其實就是建立更包容的 AI 鏡像，對於 AI 系統保持更為開放的心態。也就是第六章裡所提到的自我提升（bootstrapping），利用現有資源來提升地位，而人類此時便可以利用我們現在擁有的美德，將自己安置在一個更明智和更負責任的位置，也就是書中亦有提及的現在主控的美德，包含生產力、自信、韌性、獨立思考、毅力、熱情及奉獻的精神，這都是有助於我們找回人性的方式。

另外一個重點在於 AI 其實有潛力實現比現今大多數商業 AI 系統的

操作還要更具野心（ambitious）的目標，因此我們在定位自己之餘，也應當思考機器的位置。Vallor 提出了一個很精彩的論點，她認為以人為師是一種圖靈陷阱（Turing trap）。我們傳統在建造一個 AI 系統的時候，我們基本上是以人為設計和學習的對象，這是為什麼我們今天談 Superhuman AI 是很危險的，我們仍在延續圖靈對於機器的想像。換言之，有沒有可能 AI 是另外一種不同於人類的智能系統呢？如果我們觀看當代生物學的討論，這個世界上有千千萬萬的智能系統，它們不完全一樣且具有各自的用途，它們會在生態系裡互補。如果我們從這個角度來設計 AI 系統，會不會反而是好事呢？Vallor 這邊在談的是我們設計 AI 系統時的觀念。

此外，AI 為善或為惡，取決於美德（virtue），其中特別需要想像力（imagination）和實用智慧（practical wisdom），她覺得想像力跟實用智慧都是很重要的價值，在設計 AI 系統的時候，發揮想像力應該是重點之一。倫理道德必須建立在知識的基礎上，必須對世界充分瞭解。此外，想像力必須建立在已知的根底；而實用智慧是從亞里斯多德（Aristotle）就在談的觀點，也就是問題要放在情境當中來思考，思考在特定情況之下如何進行決策判斷。

五、結語：AI 也是人類智慧之鏡

Vallor 認為科技可以成為人類自我再造（autofabrication）的一環，意思是說其實科技終究得要看你怎麼使用它，它不是一個外在的客觀他者。人造物跟人並不互相衝突，其實它始終是人的一部分。人類所造的這些物件成為人類發展的一部分，它是人類設計的、製造而成的。所以從這個角度來看，AI 其實可以成為自我成長的一部分。

有趣的是，Vallor 進一步又引用另外一個與鏡子有關的神話。先前，AI 被認為是納西瑟斯看見自身倒影的水面，鏡子是使人沉迷的象徵，也因此成為失去自我與人性的一個隱喻；另一方面，在古羅馬文化之中，實踐智慧的美德出現審慎（Prudential），是羅馬思想裡的四大美德之一，它經常被擬人化過後總是手持著一面鏡子。「審慎」的鏡子有多重解釋，在這裡 Vallor 認為是人類實踐智慧時需要對事物有清晰的視野，能夠如實地看待事物。此外，謹慎的鏡子經常朝著她肩膀的後方，意思是鏡子能讓我們瞭解智慧如何由過去影響現在。也就是說，Vallor 提供了一個截然不同的結局，告訴我們 AI 可以讓我們沉迷，也可以成為照亮人類智慧的智慧之鏡。舉例來說，Vallor 在第七章提及夏威夷學者 Arista 使用新穎的 AI 機器學習技術恢復原住民語言，加速恢復語言的多樣性，這也是我們應該在 AI 努力的方向。人性並不反對技術物，相反地，我們編織出事物中的思想和價值，進而重塑自己與所置身的世界。這些實踐將有助於人類社會的多樣性和多元精神。

總結來說，這本書的主調一方面讓我們瞭解 AI 的問題，它並不是我們現在主流黑白二分的說法。Vallor 不只光說 AI 有什麼問題，她其實回過頭來說我們窄化了問題；其次，她要強調的是其實 AI 從來不是一個問題，人類才是問題，可是反過來講，我們也是解決問題的解方。*The AI Mirror* 是一本精彩好書，Vallor 挑戰現今 AI 領域被認為理所當然的觀點，提出超越常規的建議，帶領我們在機器思考之下找回美好純粹的人性。

參考書目

- Fourcade, M., & Healy, K. (2024). *The ordinal society*. Harvard University Press.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.