

人工智慧哲學與人文教育*

鄭會穎**

* 本文整理自作者在 2023 年 7 月 7 日「政治大學 AI 讀書會夏日論壇：AIGC 對大學教育的影響與衝擊」的發言稿。AI 讀書會係由政治大學傳播學院名譽教授鍾蔚文與政治大學創新與創造力研究中心合作於 2022 年 9 月開始舉辦，並於寒暑假擴大為論壇，邀請不同領域專家針對特定主題交流。

** 鄭會穎為國立政治大學哲學系副教授，e-mail: thycheng@nccu.edu.tw。

本文引述格式：

鄭會穎（2024）。〈人工智慧哲學與人文教育〉，《新聞學研究》，159，167-174。https://doi.org/10.30386/MCR.202404.0009

人工智慧是資訊跟電腦科學的範疇，跟哲學到底有什麼關係？就算跟哲學有關，這跟人文教育有什麼關係？這是我今天的主題。首先，有種說法是說只要有任何現象 X，就都可以有 Philosophy of X，也就是該現象的哲學，因為在那個領域中總有一些預設是未被充分檢驗，卻值得我們去探討的。

人工智慧哲學的理論與實踐

人工智慧哲學大致可以分為兩大區塊，其一是理論，其二是實踐。理論的部分我比較熟悉，可是實踐的部分往往是多數人所關心的。

下面舉一些例子，理論方面，比如說現今的 AI 系統若能成功通過圖靈測試（Turing Test），難道就表示 AI 有思考能力，而有思考能力就可以說它們有意識嗎？我們賦予 AI 語法（syntax），但如果要從裡面產生語意（semantics），便可能發生問題。此外，其實之前發表的前輩們都有提到，思考跟意識的區分也很重要。或許你問現在的 ChatGPT 說：「你有沒有意識經驗？」它可能會說沒有。你問它會不會思考，它可能也會說沒有，但一般人會認為它們有可能有思考能力，因為它會做很多推論。

至於實踐的部分，AI 系統會不會侵犯到我們，它會不會失控？自駕車的 AI 如果導致車禍，那責任要如何歸屬？新的 AI 技術在各個領域出現，新的法律要如何因應？這些是大家比較熟悉、比較關心的題目。因為我不是很擅長，待會就談一些不太一樣的議題，先花幾分鐘從一些理論問題出發，再去談大家比較少注意的實踐問題，並與人文教育做出相關的連結。

心智、意識與人類中心主義的批判

我們在談 AI 系統，或談人類的嬰兒，談人類以外的動物是否有心智的時候，不外乎是考慮這些部分：**第一他們能不能思考**，比如說能做推論；**他們有沒有意識**，比如說他有沒有痛感。光憑行為來看，我們很容易覺得他們既有思考能力又有意識經驗，可是嚴謹的科學態度通常會比較保守。科學家們可能會建議說，我們不該輕易用人類中心主義去設想其他對象的心理世界。

當然這可以有很多意思，有時候是說我們不要那麼自私，我們要幫其他動物著想、其他植物著想。但在這邊的脈絡是說，我們看到虐待動物的時候，我們都說不應該這麼做，主要是說因為牠們好像也會痛，牠們會心情不好。但有時候我們會覺得有一些灰色地帶，就是到越低等的動物，好像牠們也會悲鳴、會呻吟。但根據神經生理的過程，我們不能很直接結論說牠們真正具有意識狀態。哲學裡面我們會去探討動物的心智能力，比如說貓狗是不是有「主人」的概念；我們也探討其他對象包括人類展現出喜怒哀樂的時候，他們是不是跟我們一樣有這些意識經驗。所以這邊還是分兩點：一個是概念，跟思考比較有關係；而意識經驗就跟情緒，就是跟痛感這種比較有關係。

在概念跟思考方面，有一派看法主張概念之間是息息相關的，要賦予誰一個概念能力，就得賦予他很多概念能力。比如說你要有「校園」的概念，你就要有「學校」的概念，要進一步要有「師生關係」等等這些概念，諸如此類。所以我們如果要賦予 AI 系統，或者是人類嬰兒、還有人類以外的動物任何概念，我們似乎不得不承認他們有很多概念，但這似乎是荒謬的，所以這是一種歸謬論證。就是說我們平常都認為說

身邊的動物也好、AI 系統也好，好像有一些概念，但是至少上述看法「概念的整體論」，就是主張概念跟概念之間息息相關，所以你如果要賦予某主體一兩個概念，你就必須要說它有很多概念。在發展心理學中，一般認為兩三個月大的嬰兒，在前語言的狀況下，是不被認為有那麼多概念的。所以最後倒過來說他們不具有概念。這邊會有很多問題，比如說最後人類還是會學到概念，而 AI 系統好像也不得不說它有概念，所以這邊只是提出一派說法給大家參考。

AI 意識的倫理學與道德問題

意識方面有一派看法是主張，意識是演化的副產品或副作用，沒有什麼功能。更具體來說，有可能存在像你我一樣的物理複製，而你我是有意識的，但是複製人卻很不一樣；或者是反過來，比如說你質疑我有意識經驗，叫我證明給你看，這其實很困難。因為我只能跟你說更多話，但是有可能是一個很複雜的機器人，並不具有意識，這邊都是一些不一樣的看法供大家參考。至於什麼叫做副產品，因為神經生理學告訴我們，意識的處理是很慢的，如果現在突然有一顆球丟過來，我可能還沒有意會到的時候就閃開了；或是我摸到很燙的物品，還沒有痛感的時候手就彈開了，所以意識是過慢的。這一派的說法認為，意識慢到如果你是靠意識來生存，比如說躲避獵食者的話，那絕對早就滅亡了，我們是用本能而非意識在反應的。所以說這邊有非常多的討論，我只是提出來讓大家參考。因為在發展心理學，在動物行為學都會討論類似議題，因此用到 AI 系統裡面我們可以做類似的考慮。

這樣的懷疑論，在哲學裡面有源遠流長的討論，這裡就不多贅述。我們接下來考慮這個比較少人探討的問題，就是雖然我們目前還不確

定，未來可能也永遠無法確定。但萬一有些先進的 AI 系統具有意識怎麼辦？這是不是賦予它們更多的權利？我們要如何避免它們痛苦？也可以這麼問，我們有義務避免它們痛苦嗎？

注意我們平常談的是，我們怕 AI 很厲害、讓我們痛苦，而這當然也很重要；但是如果它們複雜到一個地步，變得也有七情六欲，那我們是不是賦予它們除了生存權之外，也有免於恐懼的權利？這樣會變得很複雜。比如說考慮素食者與植物的例子，一般來說吃素如果是道德理由的話，不外乎說多數動物具有意識，牠們會感受痛苦，所以不應該變成人類的食物，所以該吃素。這個思路會遇到一個困難，就是萬一植物有意識怎麼辦？有人認為植物你去折斷它，它們可能也會痛，只是無法表達。就算它們的意識不很複雜，如果植物也會痛，那素食者要怎麼辦？這個替代的食物會有很多實踐上的困難。

有些人認為光具有意識並不足以被賦予道德位階，比如假設某 AI 系統有能力意識到藍色，但除此沒有任何其他的意識經驗，那光這個意識經驗就太簡單了，我們似乎不用去在意如何對待那個系統，除非我們可以證明它一看到藍色會很開心，或是看到會很不舒服，不然的話就是無所謂的。關鍵似乎在於意識經驗要夠複雜，要包括像受苦經驗（suffering）或者具有偏好等等，才会有道德位階。現在的 AI 或許為時尚早，可是我們考慮未來可能發展出像照護機器人，為了讓它們接近被照護者的需求，我們需要賦予那樣的 AI 系統判讀表情跟感受情緒的能力。前面鍾老師也有提到，我們久而久之不斷發展，真的不用去顧慮這些 AI 系統發展出意識經驗嗎？其實這個問題也不是很新，比如以前的科幻電影都會談到機器人發展出七情六慾時，要怎麼辦呢？但因為我們近一兩年的討論比較少涵蓋這一部分，因此我想在此強調這個問題。

或許有人認為我們離那一天還很遙遠，等時候到了，我們實際跟那

些 AI 去溝通就可以。可是如果從心理學或者動物研究來看，我們就會知道意識經驗跟主觀報告往往是脫鉤的，有太多例子都是主體具有意識經驗，但因為種種理由而沒有辦法清楚表達出來。比如說心理學實驗，刺激物閃動很快的時候，你覺得似乎看到了什麼但你講不清楚；反過來的例子也有可能，就是說你主觀報告自己有什麼意識經驗，但其實是沒有的，而你也沒有要欺騙任何人。比如說在神經心理學中我們知道，有的病症是患者不幸已經盲了，但是他還真心覺得自己有看到什麼，所以他會堅持說看到了前面有桌子有什麼，你不用攙扶我，但是他會撞來撞去，然後再幫自己找一些理由比如說，因為我剛顧著跟你說話、我沒有認真看路等等；過了好一段時間他們才會真的意會到其實自己已經沒有相關的意識經驗。所以你光問 AI 說「你有沒有意識經驗？」這當然是一個方法，但是沒有辦法確保它是否真的具有或不具有意識。

人文教育與 AI 發展的未來

回到我們的主題人文教育，有一個核心在於不只是「以人為本」，還要避免「以人為本」產生的各種自私，為世界帶來不好的後果。像現在的環保議題、氣候變遷的議題等等都是這樣。對於 AI 系統也不意外，在我們不斷發展技術的當下，人類除了要自我保護以外，我們真的沒有義務去保護被我們製造，創造出來的潛在有意識的系統嗎？其實有些國家已經在做這個事情，或許十年、二十年後，我們就要制定 AI 權利的法律。

因為有另外一派想法，認為只要沒有生命，就不會有生命權，所以上述所有討論都是多餘的。比如說變形金剛裡面的角色，它們受傷、痛苦，而之所以讓我們在意它們，至少有一部分的理由是因為它們有生

命，它們是機械的生命。但那是電影中的設定，跟實際世界的情況是不同的。這一派認為我們不需要擔憂 AI 意識的問題，但我們可以去想想是不是意識一定要預設生命，特別是生物學意義上的生命。因為我們現在正討論 AI 意識，但是我們好像同意它們沒有生物學的生命，所以跟電影裡面設定的就比較不一樣。

生命是意識的必要條件嗎？如果我們使用嚴謹的生命定義，從生物學、無機物、有機物的區分等等，AI 系統至少目前是不會有生命的。但既然我們目前對 AI 系統意識經驗的討論方興未艾，有很多新的看法，表示至少很多人認為生命不是意識的必要條件，有可能有的系統具有意識經驗，但是它們沒有生物學定義上的生命。當然，我們也可以討論生命可以不同的定義。

現今的大國不只比太空，還有比 AI 發展，這些競賽背後還可以粗分為兩派。有一派完全不管後果，說不定也會發展類似昆蟲般大小的殺人機器，這當然目前只是科幻電影裡面的題材，但這種思路是不管後果的。而另一派當然也擔心自己落於人後，他們也會去發展技術，但是這派比較會考慮到人道考量，會注意或者商議我們前述的討論。或許有人會認為後者是偽善，好像環保跟氣候問題，大國訂出的議定書通常都是螳臂當車、緩不濟急，或許裝飾的成分很高。

我個人認為至少從人文教育的觀點來看，即便是有偽善的成分也比什麼都不做來得好，因為多少做一點還是有幫助。這樣做是否影響技術發展，是另外一個難題。比如說我在美國、亞洲或歐洲參加類似的會議就很有這種感覺，雖然很難一概而論，通常歐洲一般會更在意 AI 的權利，他們有時候會被人家笑說，你想太遠、太多了，我們現在要先怕的是人類會被 AI 滅亡，而不是去煩惱它們的權利。但後面這一派就會說，我們要去想動物權、植物權，我們當然也就要想 AI 權。但這樣去

做，或許也會有點影響 AI 技術的發展，這就是各國或各地區不同的發展方式與取捨。

今天我們簡單探討人工智慧哲學跟人文教育的幾種面向，很多題材都沒有涵蓋到，包括傳統上哲學跟語言學對語法語意的討論，這個部分幾乎完全沒有講到，還有當今對 AI 系統過度發展帶來法律難題等等。