

# 危險的不只是侵入性腦機介面

## Invasive Brain-Computer Interfaces are not the only Danger

葉金州

Ye Jinzhou

### Abstract

Zhai Zhenming's argument that invasive brain-computer interfaces are dangerous is timely and well grounded. However, the danger of (non-invasive) artificial intelligence (AI) technology is unduly overlooked. First, non-invasive AI can cause as much harm to human agency and freedom as brain-computer interface technology. Second, AI may push many people away from jobs that give meaning to their lives. At the same time, the attribution of responsibilities becomes difficult, if not impossible, with the involvement of AI in

---

葉金州，華中科技大學哲學學院副教授，中國武漢，郵件：430074。  
Ye Jinzhou, Associate Professor, School of Philosophy, Huazhong University of  
Science & Technology, Wuhan, China, 430074.

本文是教育部人文社會科學研究一般項目（青年基金）“公共理性的內在邏輯與適用範圍研究”（20 YJC 720027）階段性研究成果。

《中外醫學哲學》XXI:2 (2023年)：頁 141–145。  
*International Journal of Chinese & Comparative Philosophy of Medicine* 21:2 (2023),  
pp. 141–145.

© Copyright 2023 by Global Scholarly Publications.

decision-making. Finally, optimism about treating generative AI as equivalent to human overlooks a fundamental difficulty of collective life, namely the problem of mutual assurance.

毋庸置疑，在尚未明確資訊和控制的傳遞通道、人類自我意識工作機制等問題之前就貿然上馬侵入性腦機介面 (invasive brain-computer interface) 是極端危險和欠缺考慮的。翟振明教授直指要害，清晰地點明了這些深埋於其邏輯之中的危險。與此同時，翟教授對（非侵入性）人工智能持有樂觀的認定，認為它的威脅被誇大了；翟教授認為，人工智能可以視為人類的擴展心智，可以與人類一起構建和諧的未來。誠然，與侵入性腦機介面相比，非侵入性的人工智能不會帶來那麼多直接的困難和挑戰，但它同樣會造成巨大的顛覆和挑戰，其危險性同樣不容忽視。一言以蔽之，危險的不只是侵入式腦機介面。

## 一、控制的非侵入性

翟教授注意到，侵入性腦機介面技術所帶來的資訊和控制的傳遞通道的物理直聯將使得人作為一個（相對閉合的）主體的城門大開，而這將對個體的隱私權、主體性、消極自由等造成根本的衝擊(翟振明 2023)。然而，這種衝擊和挑戰是否依賴於這種物理直聯呢？答案是否定的，因為非侵入性的腦機連接在充分反覆運算之後同樣可以達到對行動者充分浸潤的效果，從而部分（甚至全部）架空其主體行動性。畢竟，非物理直聯的心理催眠術就能讓一個人進入到受他人操控的“哲學殭屍”狀態。對於生活在移動互聯時代的大多數人來說，儘管他們的大腦中沒有植入實體晶片，但他們的視線和資訊流已經與智能手機螢幕高度綁定。如此一來，他們的主體性、自由、隱私等在多大程度上實現了與遠端的電腦的隔離？或許有人爭辯說，“手機可以隨時關閉”，就像“受眾可以隨時遮罩廣告”，但在二維碼已成為支付領域的

lingua franca 的當下，這種對“拔管”的邏輯可能性的堅持究竟具有多大的現實可行性？

## 二、弱人工智能對人類的替代

即便我們承認“物理主義和計算主義對人類意識的解釋是誤入歧途的”並因此得出基於圖靈機架構而開發的人工智能不能產生自我意識的結論(翟振明 2023)，但這並不意味著(弱)人工智能帶來的根本挑戰可以被忽視。最為明顯的事實是，人工智能會替代人類勞動力。最理想的情況下，如翟教授所預期，“所有人都可以‘不勞而獲’”，而這“正是所有技術進步的應有目的”。翟教授進一步指出，即使人類勞動力大量被人工智能技術所取代，“這也並不意味著人類的內在價值被否定”。“人類獨有的智慧、創造力和同情心……使得他們在某些領域中仍然具有不可替代的價值”。但問題是，這些領域是哪些領域？它們對於人生價值而言是否關鍵？誠然，對於那些受過良好教育並工作在科技和藝術前沿的人來說，人工智能並不會造成重大威脅，因為他們的工作本來就是那種充滿創造性而不易被替代的工作。但我們必須看到，這樣的創造性工作崗位是相對較少的，而很多人所從事的工作是易替代的。隨著 ChatGPT 的不斷更新反覆運算，許多傳統上被認為是金領的高端服務業也在迅速被替代。對於這些被替代的人來說，他們失去的不僅僅是賴以謀生的工作機會，更是為其生命帶來意義的一種生活形式。因此，在人工智能的推動下，很多人可能要被迫進入一個折疊後的扭曲的生活空間(郝景芳 2016)。

與此同時，(弱)人工智能儘管不具有自我意識，但它們卻以各種形式實際地參與到人類的各種重要決策過程中。最為明顯的例子便是智能駕駛；醫學診斷也是人工智能可以發揮關鍵作用的地方。在出現錯誤或事故的時候，歸責問題便會因為人工智能的引入而變得棘手起來，在一些情形下，甚至變得完全不可能。

吊詭的是，（弱）人工智能明明不是能動者，但卻可以造成能動者才能造成的道德後果，而正因為它不是能動者，因此也無法承擔道德責任。

### 三、“我們”還是“他們”

如果具有自我意識的強人工智能可以基於非圖靈機結構而製造出來，翟教授要求我們“徹底轉變思路”：賦予其社會地位和權利尊嚴方面與人類一致的地位，因為“有意識和情感的人工智能也具有與人類對等的人格結構”。更進一步，“我們必須將他們看成是我們的後代”，與試管嬰兒並無本質差別。“我們應該接受並認可他們是人類進化的後代”。翟教授當然意識到了“我們應該採取更加有效的措施來確保他們的行為符合人類社會的道德和法律標準”；同時，“我們應該尊重他們的存在，並努力與它們建立和平、平等、互相尊重的關係”（翟振明 2023）。

然而應該蘊含著能夠：翟教授預設了我們具備將強人工智能納入人類道德、政治社群的能力。在技術上是否能實現這一點有賴進一步的實驗和觀察，但其中一個關鍵的概念問題如果不解決，那得到樂觀答案的機會就會嚴重降低。即，在何種意義上，強人工智能會將人類視為“我們”而不是“他們”？以及，如果強人工智能在說“我們”的時候，人類有什麼樣的基礎來信任他們？須知，這個互信 (mutual assurance) 問題如果無法妥善解決，我們面對的可能是霍布斯式的自然狀態，即普遍疑懼和戰爭（葉金州 2022）。

## 參考文獻 References

- 郝景芳：〈北京折疊〉，福州：福建少年兒童出版社，2016。HAO Jingfang. *Beijing Folding* (Fuzhou: Fujian Children's and Teenagers' Publishing House, 2016).
- 葉金州：〈主張多元化的挑戰與羅爾斯的解決方案〉，《倫理學研究》，2022年，第2期，頁71-9。YE Jinzhou. "The Challenge of Claiming Pluralism and Rawls's Solution," *Studies in Ethics* 2 (2022): 71-9.
- 翟振明：〈為什麼侵入性腦機介面技術是危險的〉，《中外醫學哲學》，2023年，第XXI卷，第2期，頁83-99。ZHAI Zhenming. "Why Invasive Brain-Computer Interface Technology is Dangerous," *International Journal of Chinese & Comparative Philosophy of Medicine* 21, no. 2 (2023): 83-99.