

人工智能時代下
腦機介面的倫理反思
Ethical Reflections on
Brain-Computer Interfaces in
the Age of Artificial Intelligence

呂雪梅 鄧蕊

Lv Xuemei and Deng Rui

Abstract

This paper reviews the important thesis of David M. Lyreskog, who provides a profound interpretation of the complexity and uncertainty of ethical issues associated with brain-computer interfaces in the age of artificial intelligence (AI) and lays the theoretical foundation for an ethical framework for AI-brain-computer interfaces

呂雪梅，山西醫科大學人文社會科學學院教授，中國太原，郵編：030600。
Lv Xuemei, Lecturer, School of Humanities and Social Science, Shanxi Medical University, Taiyuan, China, 030600.
鄧蕊，山西醫科大學人文社會科學學院教授，中國太原，郵編：030600。
Deng Rui, Professor, School of Humanities and Social Science, Shanxi Medical University, Taiyuan, China, 030600.

《中外醫學哲學》XXI:2 (2023年)：頁73-77。
International Journal of Chinese & Comparative Philosophy of Medicine 21:2 (2023), pp. 73-77.

© Copyright 2023 by Global Scholarly Publications.

in terms of autonomy, integrity, identity, and privacy. This paper responds to Lyreskog's main points in terms of four dichotomies, namely bounded and unbounded, controllable and uncontrollable, self and non-self, and open and liberalized, to promote in-depth discussion of the ethical issues associated with AI-brain-computer interfaces.

腦機介面 (BCI) 是改變世界的新興技術之一。廣義 BCI 由腦、機和介面三部分組成。狹義 BCI 只指介面這一設備。腦機介面分為非侵入式和侵入式兩種。非侵入式 BCI 是在頭皮上放置電極，使用“腦電圖”(EEG) 方法檢測大腦區域神經元的平均放電情況。EEG 收集記錄已從單次建模式發展到自我調整誘導式，提高其穩定性和精確性。侵入式 BCI 需要手術將電極植入大腦，因離神經元更近，可獲取更高品質的神經信號。

人工智能-腦機介面技術 (AI-BCI Technology, ABT) 的倫理問題是應用倫理研究中的新問題。它的重要性不僅在彌補研究的空缺，而且在警醒研究者需要改革調整，構建新領域倫理指導框架。只有未雨綢繆方可從容應對。該文章從本體論和分類法視角提出了 ABT 倫理指導框架，其中自主性、精神完整性、同一性、資料隱私與控制為爭論的關鍵；還進一步探索 ABT 在功能增強和兒科人群中的可能性和挑戰；最終得出 ABT 風險與利益並存，具有極大的不確定性，呼籲研究者以開放的心態做好準備，倫理學科的概念和框架很可能會被重新調整。整篇文章觀點明晰，結構嚴謹，推理得當，促進了筆者對相關問題進一步思考，將沿著 Lyreskog 的思路稍作深入討論。

首先，有界與無界。自主性問題在 BCI 中經常被討論，卻容易陷入自相矛盾。根據求助者增強自主性的需求，可以大致分為下列情況討論：

求助者	BCI 干預後	干預成效	爭論問題
無自主性	有自主性	明顯療效	削弱自主性
	有部分自主性	有療效	可能削弱自主性
	無自主性	無成效	有利問題
有部分自主性	有自主性	明顯療效	削弱自主性
	有部分自主性	可能有療效	混淆自主性
	無自主性	反效應	無傷問題

BCI 階段是否削弱自主性問題的追問終將還原為邊界問題即腦機邊界問題，也就是晶片演算法和腦神經控制的邊界和作用程度問題。BCI 是干預治療實現功能修復，還是功能增強到已越界，甚而侵犯了自主性。這個爭論要麼落入演算法漩渦，要麼籠罩神秘色彩。BCI 階段對自主性的削弱是有界的、有限的，甚至是可以預知的。但是 BCI 發展到 ABT 階段，AI 的深度學習已經無需討論腦機邊界問題，與其說是“AI 習得認知與我們理性完美融合” (Lyreskog et al. 2023)，不如說當我們選擇 ABT 時已經將自主性讓渡給了 AI，ABT 腦機已無界，充滿了不透明、不可測、不確定。

其次，可控與失控。精神完整性 (mental integrity) 問題是神經科學技術討論的顯性問題，也是干預性 BCI 討論的首要問題。精神完整性與自主性的概念密切相關，Lyreskog 分析了兩者不同步的情況，筆者延伸探尋了其可能的原因：

干預	完整性情況	自主性情況	可能的問題	問題的原因
ABT	完整性破壞	自主性完好	破壞的部分反而加強其深層的價值觀，恰好吻合	程度深淺問題
	完整性完好	自主性破壞	由於其他原因，導致決策誤導。非ABT也會發生。	外因要素作用
	完整性看似完好	自主性完好	刻意避開理性系統導致個體與心理間的疏離感	喪失了自我決定權的失控感

精神完整問題結合“Alice”的案例，可以看出BCI到ABT，從可控到失控。BCI可控的原因是一般都會有一個“開關”。不論這個開關是在輕症患者自己手中，還是在重症患者的監護人手中，是有一定的可控性。但ABT沒有開關，AI會讓一切看似“合理”的情況可能不合適。患者的心理疏離感是一種失控感。ABT以精神完整性、自主性完好為前提，故意繞開理性決策系統，其實是徒勞的，情緒、信念、欲望等要素的轉變不僅會破壞完整性，也會影響自主性。

再次，自我與非我。同一性問題的討論中，BCI初期可能短暫擾亂人的同一性，長期削弱還是破壞並沒有明確的資料支援。Lyreskog列舉了沒有降低反而增強同一性的案例。雖然目前BCI比換頭術、嵌合體等主題關於同一性研究的資料要少，但這並不能直接說它不重要。Neuralink的目標是全腦腦機介面，最終必將

追溯到意識與思維的起點，“自我”的同一性與確定性尤為重要。如果以最新“擴展心智理論”同一性的概念直接延展到 ABT，可能已遮蔽了一些倫理問題，或者說當選擇 ABT，就選擇了“非我”，在人工智能的環境下重新逐步確定新的“非我的自我”。

最後，開放與放開。Lyreskog 呼籲學者以開放的態度來看待新技術及其應用。這個態度比其在“人類增強”研究中的態度更加包容。但開放態度迎接新技術的進步與變革，並不是放開。在隱私與資料控制的討論中，AI 帶來的風險同時依賴於 AI 來解決，如設計可解釋的、更安全的，甚至誠實守信合道德的 AI。但 ABT 的發展已經超過了現有監管的範疇。ABT 的隱私問題還會涉及到個人神經權、精神隱私權等新問題，需要法律、政策和倫理的進一步規範和保障。

參考文獻 References

- 萊瑞斯科、佐赫尼、辛格、薩烏萊斯庫：〈與機器一起思考：腦機介面技術〉，《中外醫學哲學》，2023 年，第 XXI 卷，第 2 期，頁 11–34。
David M. Lyreskog, Hazem Zohny, Iliana Singh, Julian Savulescu. “The Ethics of Thinking with Machines: Brain-Computer Interfaces in the Era of Artificial Intelligence,” *International Journal of Chinese & Comparative Philosophy of Medicine* 21, no. 2 (2023): 11–34.