

AI 倫理的兩面性初探 —人類研發 AI 的倫理與 AI 倫理*

甘偵蓉

國立成功大學人文社會科學中心
E-mail: ganrrec@mail.ncku.edu.tw

許漢

國立中正大學哲學系暨研究所
E-mail: pyhhsu@ccu.edu.tw

摘要

目前 AI 國際倫理準則在規範方向上偏向 AI 研發及使用者，比較缺乏針對 AI 系統本身，但 AI 的自主學習能力，使得 AI 的倫理規範應該涵蓋這兩個面向，本文提出較為含括的哲學倫理觀來說明。這項倫理觀包含了 AI 的自主道德決策能力，如依據其目前及可預期的技術基礎，如與人類相較雖屬於知性與低限度的，但在決策的責任承擔上仍具備分散式行為主體性。最後針

© 中央研究院歐美研究所

投稿日期：108.7.22；接受刊登日期：109.5.8；最後修訂日期：109.4.27

責任校對：吳智偉、黃意函、曾嘉琦

* 本文架構與內容，皆經過兩人多次討論，初稿各自撰寫分工（甘偵蓉與許漢合寫第一、肆、陸節，甘偵蓉先寫第二、伍節，許漢先寫第三節），接下來再整合與修改，審查意見回覆亦是。此文曾在 2019 年 5 月 10 日中央研究院歐美研究所「歐美 AI 的發展與挑戰」跨學門研討會發表，感謝評論人陳瑞麟及當天與會者所有的提問與建議，並作相當實質的修正，而本文之完成也得益於匿名審查人嚴肅的批評，使文章更完善，在此敬表謝意。

對 AI 兩面向的倫理規範，我們提出底線式道德義務作為所有規範的起點。並對於落實 AI 規範的議題，提出個案研發的倫理評估機制設立方向，以為回應。

關鍵詞：國際 AI 倫理準則、AI 倫理的兩面性、含括的倫理觀、分散式行為主體性、底線式道德義務