

反思機器人的道德擬人主義*

何宗興

國立中正大學哲學系
E-mail: tsunghsing@ccu.edu.tw

摘要

如果機器人的發展要能如科幻想像一般，在沒有人類監督下自動地工作，就必須確定機器人不會做出道德上錯誤的行為。根據行為主義式的道德主體觀，若就外顯行為來看，機器人在道德上的表現跟人類一般，機器人就可被視為道德主體。從這很自然地引伸出機器人的道德擬人主義：凡適用於人類的道德規則就適用於機器人。我反對道德擬人主義，藉由史特勞森對於人際關係與反應態度的洞見，並以家長主義行為為例，我論述由於機器人缺乏人格性，無法參與人際關係，因此在關於家長主義行為上，機器人應該比人類受到更嚴格的限制。

關鍵詞： 機器人、人工智能、道德主體、道德擬人主義、反應態度

©中央研究院歐美研究所

投稿日期：108.6.28；接受刊登日期：109.2.9；最後修訂日期：109.2.12

責任校對：張文綺、趙麗婷、范馨文

* 本文是科技部人文社會科學研究中心的計畫成果 (107-2420-H-002-007-MY3-V10701)。另外，我十分感謝兩位匿名審查人，以及以下學者給本研究的建議與提問：謝世民、王一奇、侯維之、蔡政宏、王華、趙之振、嚴如玉、林映彤、陳樂知、張智皓、朱家安。