

BSI 機器人倫理標準 前瞻人機共存時代



隨著科技的發展，機器人的應用不只是在提升製造業的生產力水準，數位資訊與自動化技術不但催生第四代工業革命，更讓機器人逐漸走入個人、醫療、照護等領域，衍伸出道德倫理層面的問題。

艾西莫夫 (Isaac Asimov) 曾在他的科幻小說中提出「機器人三定律¹」，超過半世紀後的今日，機器人倫理不再只是科幻小說題材，也不是簡單的三條法則可以涵括。BSI英國標準協會於是召集了機器人相關領域的製造業、工程界及學術界專家、安全專家、科學家、倫理學家、哲學家等專業人士，制定出第一個機器人倫理標準 [《BS 8611 機器人與機器系統的倫理設計與應用指南》](#)。

據《衛報》報導²，在牛津大學召開的「社交機器人和人工智慧」(Social Robotis and AI) 大會上，西英格蘭大學機器人學教授Alan Winfield表示「此標準是將倫理道德價值融入機器人和人工智慧的第一步」，「它基本上指出該如何對機器人做道德風險評估」。

¹ 機器人三定律：1) 機器人不得傷害人類，或不作為使人類受到傷害。2) 機器人需服從人類的命令，除非違反第一定律。3) 在不違反第一及第二定律的情形下，機器人要保護自己。

² Hannah Devlin, theguardian, ["Do no harm, don't discriminate: official guidance issued on robot ethics,"](#) 20160918

BS 8611 指南開宗明義提供了基本準則：機器人的設計不應該以殺害或者傷害人類為唯一或首要目的；責任人應該是人類而非機器人；任何機器人都應該有負責人，而這個人應該為機器人的行為負責。

標準中還提及一些較具爭議性的議題，例如人與機器人建立情感連結是否妥當，尤其當機器人是專為陪伴老人和小孩而設計時。另外也暗示一些問題，例如機器人可能缺乏對文化多樣性或多元化的尊重。對此，雪菲爾大學機器人和人工智慧教授 Noel Sharkey 指出，這種現象已經出現在警用科技上，例如用於機場可疑人員示警的技術已被證實有種族偏見。

Winfield 教授則解釋，由於深度學習系統 (deep learning system) 以網路上的資訊為基礎，而這些資訊本身帶有偏見。以至於這些系統吸收了人類的偏見，例如偏好白種中年男性。這樣的現象可能引起的風險包含，未來的醫療應用診斷系統可能較不適用於女性和少數族群。而目前也有相應的例子，像是語音辨識軟體對女性聲音的辨識能力較差；或是臉部辨識程式在辨識黑人和白人臉孔時出現明顯的效能差異。

其他標準中提及的問題包含，對機器人過度依賴；或是出現能開發或修改行動程序的機器，引發無法想像的後果等等。

BSI 主管 Dan Palmer 表示：「運用機器人與自動化技術，使流程更有效率、更靈活及更有適應性，是生產力提升的重要因素。為了讓機器人與自動化技術的使用被接受，最重要的是要能夠識別和處理倫理問題和危害，例如違背人性和對機器人的過度依賴。這個新的指引能協助機器人及自動化系統的設計者與使用者更好地適應新的工作領域。」●



BS 8611:2016 機器人與機器系統的倫理設計與應用指南

目的： 機器人倫理指南，讓機器人能融入現今的社會道德規範中。

適用對象： 機器人和機器人設備的設計者和管理者。終端機器人產品若能依此標準的指引開發製造，一般大眾也將受益。

功能：

- 提供鑑別潛在道德危害 (ethical harm) 的指引。
- 提供機器人設計和應用的安全設計、防護方法及資訊的補充指引。
- 對目前不同形式的機器人制定安全規定，包含工業用、個人照護及醫療用機器人。

[購買標準 >](#)

相關新聞：[超越「機器人三定律」？英國標準協會制定機器人道德標準](#) | [三定律 2.0？英國標準協會發布機器人倫理指南](#) | [英國正式頒布機器人道德標準：不許傷害、欺騙和令人成癮](#)