

## 《生命倫理線》 12.8.2019

中文大學生命倫理學中心總監  
區結成 醫生

### 人工智能武器誰負責？

去年，我聽到生物倫理學家朋友正在研究有關人工智能(AI)武器系統的責任和控制問題，初時感到有點驚訝。雖然我知道，生物倫理學的範圍確實包括與嶄新科技有關的各種道德倫理爭議，但我以為它所關注的議題，大體上應屬於民間社會的範疇，而軍事戰爭並不在其內。然而我也立即想起，上個世紀後半葉生物倫理學興起，也是與二戰期間犯下的對人類的罪行有關：納粹德國的醫生和科學家對集中營裡的人群進行了可怕的活人實驗，包括兒童。嶄新技術能影響生命、衝擊社會既有的價值座標和秩序，也就帶來了新的倫理問題，AI 武器系統當然也是其中之一。

今年 4 月 30 日，史丹福大學國際安全與合作中心 (簡稱 CISAC) 舉辦的一場專題講座，主講者之一 Paul Scharre 是軍事顧問專家也是美國在阿富汗戰爭的過來人。他觀察，新型武器系統的設計在不斷演進，「我們見到戰爭演變的方式之一，是真人的角色被推離戰場邊界 —— 不僅是人身不需在現場，在認知層面也不需要 —— 越來越多的戰鬥攻擊決策是由這些以機器的高速計算系統作出的。」遙控無人機執行攻擊任務是「人身不需在現場」的例子，那尚需要由人操作判斷，而一旦 AI 發展至可以自主操作判斷，取代人腦，「對於機器代替我們做出生死抉擇能感到安樂嗎？」他問。

### 依程式射殺一個小孩？

有人認為，AI 代人作軍事決策也可以是很合邏輯的。在不少具體的任務上，人的可靠性和決策質素早已被 AI 超越。例子橫跨各種實用實戰範圍：圍棋、某些醫學診斷工作、金融投資程式買賣，某些範圍的法律諮詢等。完全可以想像，在極端的戰事狀態，人腦會被認知盲點、好勝情緒、固執、殘忍、嗜血等非理性因素左右，釀成慘劇。AI 沒有情緒、固執、殘忍等人性，如果預先設定合理合法的演算規則，它的決策不能比人腦優勝嗎？

Scharre 並不認同。他舉了一個發生在阿富汗戰役的真實事件為例。有一次，他和小隊被派到阿富汗和巴基斯坦邊境尋找塔利班軍人。很快，他們發現一個看來只有 5 歲或 6 歲以小女孩帶著一群山羊，在他們的位置附近徘徊。他們意識到這是一個陷阱：小女孩通過無線電為塔利班軍隊報告美軍的位置。依照合法的戰爭規則，一旦確

認目標是敵人的作戰成員，己方就可以開槍殲滅以解除威脅。但對於 Scharre 和他的團隊來說，射殺一個小孩子不是一個選項。他們知道這樣做在道德上是錯誤的。如果由一個 AI 程式自主決定，結局會否不一樣？

去年 4 月，紅十字國際委員會在日內瓦舉行的《常規武器公約》政府專家組會議，針對的就是所謂的「殺手機器」(killer robots)的問題，呼籲各國和團體禁止此類武器。委員會認為，人類必須維持對武器系統和使用武力的清晰控制，以確保能遵守國際法和滿足道德上的關注，因此各國必須緊急致力制定規則，限制殺傷力武器系統的機器自主。最重要的是，人作出的決定，最終有人要為一切後果負責。AI 自主的武器系統，由誰負責？

### 審視 AI 自主戰爭的倫理

國防科技大學作者趙雲、張煌有一篇原載於「中國社會科學網」的文章〈智能化戰爭的倫理審視〉對 AI 武器系統自主作戰的倫理問題有很好的分析。這可以分三個範疇討論：

「開戰正義」：指涉訴諸戰爭的權利和合法性的問題。智能空間（包括網路空間）的主權與全球公域邊疆有爭議，而有能力發動智能化戰爭的主體數量大大增加，非國家的組織崛起，技術雄厚的大公司和恐怖組織都可以輕易使用。當技術變得低成本、武器可使戰爭做到接近「零傷亡」和「軟殺傷」，戰爭製造者更少顧忌發動戰爭的成本，從而導致小型戰爭更加容易爆發。AI 系統必然含有「先發制人」的設計，一旦計算好後果可以接受便可以開戰。

「交戰正義」又稱戰爭行為正義，上面 Scharre 和他的團隊面對阿富汗小女孩的決定就是例子。一般而言，在戰爭進行過程中參戰各方必須遵循的倫理規則包括「區別性」原則和「相稱性」原則，避免無區別性的濫殺，以及使用與情境和目標不相稱的毀滅性武力。機器人雖然有人工智能，但卻沒有憐憫或悔恨等情感，殺人不會沒有任何的愧疚感，因此，戰爭的目的可以淪為屠殺而屠殺。

「終戰正義」：這或者可以更準確地稱為「戰後正義」。清算戰爭罪行、釐清戰爭責任是「戰後正義」的首要體現（其他方面包括承擔戰後重建的義務問題等）。相對於傳統戰爭中人類就是唯一的責任主體，在使用自主無人智能武器上面，誰要為違反人道規約的作戰決策負責？這在目前難以定論。責任分配是一大困境：武器操控者、軍火商、設計 AI 的工程師、選擇武器的採購員、戰地指揮員、相關維修保養人員乃至 AI 武器本身，一大串相關的人和系統，全都可能有責任，但也就是說，追究責任的難度極大。

這些議題也可以橫向思考：AI 系統代替人類自動決策，衍生的倫理問題並不限於武器殺傷摧毀而無法問責。自主系統用於醫療決策、教育(例如大學收生)，以至管理社會秩序(例如內地以大數據建立「社會信用計分制」)，背後都有嚴肅的問題需要思考。