

《生命伦理线》 12.8.2019

中文大学生命伦理学中心总监

区结成 医生

人工智能武器谁负责？

去年，我听到生物伦理学家朋友正在研究有关人工智能(AI)武器系统的责任和控制问题，初时感到有点惊讶。虽然我知道，生物伦理学的范围确实包括与崭新科技有关的各种道德伦理争议，但我以为它所关注的议题，大体上应属于民间社会的范畴，而军事战争并不在其内。然而我也立即想起，上个世纪后半叶生物伦理学兴起，也是与二战期间犯下的对人类的罪行有关：纳粹德国的医生和科学家对集中营里的人群进行了可怕的活人实验，包括儿童。崭新技术能影响生命、冲击社会既有的价值坐标和秩序，也就带来了新的伦理问题，AI 武器系统当然也是其中之一。

今年4月30日，史丹福大学国际安全与合作中心（简称 CISAC）举办的一场专题讲座，主讲者之一 Paul Scharre 是军事顾问专家也是美国在阿富汗战争的过来人。他观察，新型武器系统的设计在不断演进，「我们见到战争演变的方式之一，是真人的角色被推离战场边界——不仅是人身不需在现场，在认知层面也不需要——越来越多的战斗攻击决策是由这些以机器的高速计算系统作出的。」遥控无人机执行攻击任务是「人身不需在现场」的例子，那尚需要由人操作判断，而一旦 AI 发展至可以自主操作判断，取代人脑，「对于机器代替我们做出生死抉择能感到安乐吗？」他问。

依程序射杀一个小孩？

有人认为，AI 代人作军事决策也可以是很合逻辑的。在不少具体的任务上，人的可靠性和决策质素早已被 AI 超越。例子横跨各种实用实战范围：围棋、某些医学诊断工作、金融投资程序买卖，某些范围的法律咨询等。完全可以想象，在极端的战事状态，人脑会被认知盲点、好胜情绪、固执、残忍、嗜血等非理性因素左右，酿成惨剧。AI 没有情绪、固执、残忍等人性，如果预先设定合理合法的演算规则，它的决策不能比人脑优胜吗？

Scharre 并不认同。他举了一个发生在阿富汗战役的真实事件为例。有一次，他和小队被派到阿富汗和巴基斯坦边境寻找塔利班军人。很快，他们发现一个看来只有 5 岁或 6 岁以小女孩带着一群山羊，在他们的位置附近徘徊。他们意识到这是一个陷阱：小女孩通过无线电为塔利班军队报告美军的位置。依照合法的战争规则，一旦确

认目标是敌人的作战成员，己方就可以开枪歼灭以解除威胁。但对于 Scharre 和他的团队来说，射杀一个小孩子不是一个选项。他们知道这样做在道德上是错误的。如果由一个 AI 程序自主决定，结局会否不一样？

去年 4 月，红十字国际委员会在日内瓦举行的《常规武器公约》政府专家组会议，针对的就是所谓的「杀手机器」(killer robots)的问题，呼吁各国和团体禁止此类武器。委员会认为，人类必须维持对武器系统和使用武力的清晰控制，以确保能遵守国际法和满足道德上的关注，因此各国必须紧急致力制定规则，限制杀伤力武器系统的机器自主。最重要的是，人作出的决定，最终有人要为一切后果负责。AI 自主的武器系统，由谁负责？

审视 AI 自主战争的伦理

国防科技大学作者赵云、张煌有一篇原载于「中国社会科学网」的文章〈智能化战争的伦理审视〉对 AI 武器系统自主作战的伦理问题有很好的分析。这可以分三个范畴讨论：

「开战正义」：指涉诉诸战争的权利和合法性的问题。智能空间（包括网络路空间）的主权与全球公域边疆有争议，而有能力发动智能化战争的主体数量大大增加，非国家的组织崛起，技术雄厚的大公司和恐怖组织都可以轻易使用。当技术变得低成本、武器可使战争做到接近「零伤亡」和「软杀伤」，战争制造者更少顾忌发动战争的成本，从而导致小型战争更加容易爆发。AI 系统必然含有「先发制人」的设计，一旦计算好后果可以接受便可以开战。

「交战正义」又称战争行为正义，上面 Scharre 和他的团队面对阿富汗小女孩的决定就是例子。一般而言，在战争进行过程中参战各方必须遵循的伦理规则包括「区别性」原则和「相称性」原则，避免无区别性的滥杀，以及使用与情境和目标不相称的毁灭性武力。机器人虽然有人工智能，但却没有怜悯或悔恨等情感，杀人不会没有任何的愧疚感，因此，战争的目的可以沦为为屠杀而屠杀。

「终战正义」：这或者可以更准确地称为「战后正义」。清算战争罪行、厘清战争责任是「战后正义」的首要体现（其他方面包括承担战后重建的义务问题等）。相对于传统战争中人类就是唯一的责任主体，在使用自主无人智能武器上面，谁要为违反人道规约的作战决策负责？这在目前难以定论。责任分配是一大困境：武器操控者、军火商、设计 AI 的工程师、选择武器的采购员、战地指挥员、相关维修保养人员乃至 AI 武器本身，一大串相关的人和系统，全都可能有责任，但也就是说，追究责任的难度极大。

这些议题也可以横向思考：AI 系统代替人类自动决策，衍生的伦理问题并不限于武器杀伤摧毁而无法问责。自主系统用于医疗决策、教育(例如大学收生)，以至管理社会秩序(例如内地以大数据建立「社会信用计分制」)，背后都有严肃的问题需要思考。