

《禁止或限制使用某些可被认为具有过分
伤害力或滥杀滥伤作用的常规武器公约》
缔约方政府专家小组

Distr.: general
4 September 2017
Chinese
Original: English

2017 年 11 月 13-17 日，日内瓦

临时议程项目 6

结合《公约》的目标和宗旨，
审查致命性自主武器系统
领域新技术的各个方面

思考文件

主席提交

导言

1. 《特定常规武器公约》缔约方授权政府专家小组(专家小组)结合《公约》的目标和宗旨，审查致命性自主武器系统领域新技术的问题。《特定常规武器公约》专家小组今天是联合国系统唯一正式处理这一问题的政府间论坛，汇集了相关的跨学科见解和专长。为了使《特定常规武器公约》进程具有价值和持续的相关性，重要的是确保继续向专家小组会议提供资金支持，确保所有缔约方和其他利益攸关方、包括业界对讨论做出贡献。
2. 2017 年用于专家小组的时间从 10 天缩短到 5 天，主席认为，通过彻底审查技术领域的发展现状及其在特定军事系统中的实例，专家小组今年可以有一个良好的开端。专家小组还可继续过去三年关于致命性自主武器系统相关法律和伦理问题的有益讨论，包括有关规范自动驾驶车辆等民用自主系统使用的国家法律以及公司方面的新的发展情况。就有关技术、军事效用和致命性/伦理考虑进行透彻的三重审查，将有助于专家小组进入下一阶段的讨论，下一阶段可能侧重于确定方法，如定义和其他概念，并可能将视线缩小到政策路径。
3. 为了促进讨论并协助各代表团准备专家组会议，主席提出了以下三大类问题。这些问题仅仅旨在帮助开始对三类中每个类别进行实质性审查。这些问题既非详尽，也不完全代表有关这一主题的所有意见；特别是某些与政策有关的问题，它们属于下一阶段讨论的问题，目前暂时搁置。



技术

- 哪些技术正在或可能有助于致命性自主武器系统？它们可否被广泛地定性为人工智能/自主系统(AI/AS)？
- 它们正在或可能被用于哪些民用目的？
- 自主系统最好是想象为有形机器人(分立的)还是虚拟机(分散的信息处理系统)？是否有致命性自主武器系统讨论未涵盖的其他技术途径(例如神经形态、人类增强)？
- 从目前 AI/AS 技术来看，自主的各个方面有哪些？与完全自主(即使是理想化的)相比，我们在这些方面(能源自主、计算自主等)位于何处？
- 近几年已经发生了转变(谷歌猫脸探测器和 DeepMind 基于 DQN 算法的 Atari 2600 2D 游戏机、尤金·古斯特曼智能聊天机器人通过图灵测试、Alpha GO 等)，还是有可能从具体应用程序(弱人工智能)转向通用人工智能(强人工智能)，导致最终的“智力爆炸”/“奇点”/“超级智能”(超人工智能)？
- 现有系统如何核实(构建是否正确)和验证(是否构建了正确的系统)？现有和计划中的自主系统是否可以辨认(知道什么和如何知道)？机器能描述它们的学习吗？
- 业界如何处理人和社会的安全问题(例如黑客和隐私)？可以使自主机器绝对不被黑客入侵吗？
- 是否存在针对机器行为的软件/硬件锁，能否防止学习机绕过/更改软件/硬件锁？
- 人工智能的变革性质及其可能普遍存在是否限制了以任何方式讨论致命性自主武器系统，抑或人工智能像过去其他两用技术一样？

军事效用

- 正在或可能部署自主技术的具体领域是什么？后勤？情报、监视和侦察？边防和区域防御？其部署方式是否可能通过任务蠕升或其他方式导致致命性自主武器系统？
- 是否存在更有可能部署致命性自主武器系统的特定领域(如海面和海面下领域)？为什么？
- 致命性自主武器系统对进攻和防御能力的影响，哪个更多？致命性自主武器系统的潜在部署是否可能降低使用武力的门槛？它是否可能加强不对称部署武力或秘密使用武力？是否有与《特定常规武器公约》有关的军备竞赛和稳定性考虑？
- 致命性自主武器系统是否有可能潜在扩散/学习与恐怖分子和其他不法非国家行为者一起行事？
- 是否有甚至在目前《特定常规武器公约》的背景下也可取的具体军事应用(战争遗留爆炸物、排雷等)？

- 潜在的致命性自主武器系统是否能够适应现有军事指挥和控制链？是否有致命性自主武器系统独特的具体理论问题？
- 潜在的致命性自主武器系统是否符合现有与部署、持有、贸易或转让有关的军备透明的概念？

致命性/伦理问题

- 现有或计划的自主系统的法律问责和赔偿责任在何方？是在计划-开发者、合法所有者、用户还是/抑或机器方面？
- 计划或已有规范如无人驾驶汽车、娱乐或同伴机器人和聊天室等自主系统的国家或地区法律的主要特点是什么？¹
- 是否存在可能与致命性自主武器系统讨论相关的现有行为守则或行业标准？² 这些参考文件中什么与《特定常规武器公约》相关？
- 如遗传学等其他领域是否有规范见解？
- 针对人和国家控制的行为制定的国际人道主义法是否可以继续比照适用于潜在的自主机器及通过何种中介机制适用？
- 在发展、测试和部署之前，潜在致命性自主武器系统如何适合现有国家武器审查？
- 在贸易(原产地规则等)和知识产权等其他国际政策领域，是否存在有关潜在自主系统的法律漏洞？
- 机器能否在人的意义上(现象意识、故意、创造性、同情心、演进、带具身智能的自由行为体)真正智能化？
- 何种伦理道德系统或具体伦理价值观可以适用于机器？在不同的道德/神学系统中，意识与人—物质世界互动的不同观点如何在一套针对学习机的普遍适用的禁令中得到调和？
- 迄今为止，与伦理/道德相关的问题集中在吞噬生命的机器上。如果以人机成对协作行动或人类增强的方式呢？
- 这些伦理禁令如何编纂？一台机器能否成为一个道德的行为体？

¹ 例如，大韩民国“智能机器人开发和分销法”，和欧洲联盟议会法律事务委员会的报告(2015/2103(INL))。

² 如电机和电子工程师学会 P7000-7002 “关于人工智能和自主系统伦理考虑的全球倡议”和英国“EPSRC-AHRC 机器人原理”。