【作品名称】禁止进攻型全面自主武器（杀手机器人）的必要性及可行性调查和研究

【关键字】AI技术，国家安全，人工智能伦理，禁止性武器公约

【作品简介】

进攻型全面自主武器（lethal autonomous weapons，更通俗的说法为“杀手机器人”killer robots）作为自主性强、危险性高、隐蔽性好的未来武器，已经随着AI及机器人技术的发展，变为即将实现的军事武器。能够自主决策的自动化杀手机器人不但将完全改变未来战争的面貌，而且将对伦理道德、国家安全和人身安全造成极大冲击。

我们小组期望通过调研普通大众对于智能武器的认知情况、国际非政府组织为禁止杀手机器人推动的倡导性活动、政府组织对于智能武器的限制性措施，以及智能武器技术的发展现状及应用前景，来说明禁止的必要性及可行性。

我们打算从以下四个具体的方面开展调研：

1. Technology：在技术层面上，进攻型全面自主武器目前的发展水平、发展成熟的预期时限、在未来战争中的应用及表现；
2. Public perceptions： 民众（或者说公民社会）对于智能武器和未来战争的认知以及价值判断；
3. State security and human rights：智能武器对国家安全和人身安全带来的杀伤力，以及智能武器作为一种主体同人类发生交互可能产生的伦理问题；
4. Laws and bans：主权国家及其组成的国际组织针对禁止智能武器的现有立法情况，以及国际非政府组织为推动立法进行的倡议性民间活动。

此外，现阶段我们针对禁止必要性的一些假设如下：

1. 大规模使用杀手机器人导致战争规模升级，各国家进行军备竞赛，武器可能落入对方手中，甚至落入恐怖分子手中；
2. 研究机构进行快速而不够严谨安全的研究，这种机器人可能出故障，或者由于技术性漏洞被恶意分子操控，造成不堪设想的安全后果；
3. 机器人精准识别军事目标和其他无关人员的技术难度非常高，可能误伤。
4. 未来战争真的只是机器人之间的战争而不涉及人类吗？战争在我们身边发生，没有分离的战争世界。在没有智能武器的时代，战争作为一种极端的暴力形式，是人类获取利益的极端手段。当杀手机器人出现，人们的战争观即便发生极大的改变，战争仍然是人类与人类的冲突，而不可能完全转化为机器与机器的冲突（当然，假如机器也拥有了战争观，那一天人类是否还存在又是一个新的疑问）。而对于人类之间冲突来说，最有效也最极端的手段常常就是直接终止对方的生存机会。

我们对禁止必要性的探讨，事实上是在讨论一种“合法性基础”。一般来说具备了一定的必要性，即得到各方的认同，在某种程度上也会是对可行性的确保。武器研发公司和国家首先要认同禁令，其次禁令一旦在价值层面上不被认同以至于出现了违背，在实际操作上会带来有效的惩罚。针对禁止的可行性，我们目前的初步想法是借鉴禁止某些军事武器的已有成功案例。比如1997年的禁止反单兵地雷使用以及禁止激光致盲武器的研发和使用，都取得了显著的成果。进一步我们再关注和探讨智能武器和核武器、生物武器等之间的区别和联系，从而预防智能武器可能造成的毁灭性后果。