



Distr.: General
1 July 2024
Chinese
Original: Arabic/Chinese/English/
Russian/Spanish

第七十九届会议

暂定项目表* 项目 98(ss)

全面彻底裁军：致命自主武器系统

致命自主武器系统

秘书长的报告

摘要

本报告综述了会员国和观察员国依照第 78/241 号决议所提交答复的要点，但不影响其各自立场。本报告包括定义和特性描述；挑战、关切和潜在益处；各国的审议意见；今后的步骤；秘书长的意见和结论。

* A/79/50。



目录

	页次
一. 导言	6
二. 背景	6
三. 定义和特性描述	6
四. 挑战、关切和潜在益处	8
五. 各国的审议意见	12
六. 今后的步骤	14
七. 秘书长的意见和结论	18
附件一	
收到的答复	19
A. 会员国和观察员国	19
安道尔、阿根廷、多民族玻利维亚国、巴西、哥伦比亚、哥斯达黎加、古巴、智利、 多米尼加共和国、厄瓜多尔、萨尔瓦多、危地马拉、洪都拉斯、墨西哥、尼加拉瓜、 巴拿马、巴拉圭、秘鲁、葡萄牙、西班牙、乌拉圭和委内瑞拉玻利瓦尔共和国	19
阿根廷	20
澳大利亚	21
奥地利	22
保加利亚	27
加拿大	29
智利、哥伦比亚、哥斯达黎加、多米尼加共和国、厄瓜多尔、萨尔瓦多、危地马拉、 哈萨克斯坦、尼日利亚、巴拿马、秘鲁、菲律宾、塞拉利昂和巴勒斯坦国	32
中国	33
哥斯达黎加	35
古巴	38
埃及	39
斐济	40
芬兰	42
法国	43
德国	45

希腊	48
危地马拉	48
洪都拉斯	49
印度	50
爱尔兰	52
以色列	56
意大利	58
日本	59
基里巴斯	62
卢森堡	64
马拉维	66
墨西哥	67
荷兰王国	70
新西兰	72
挪威	75
巴基斯坦	77
菲律宾	80
大韩民国	82
摩尔多瓦共和国	83
俄罗斯联邦	85
塞尔维亚	88
塞拉利昂	90
新加坡	92
西班牙	93
斯里兰卡	94
瑞典	96
瑞士	98
大不列颠及北爱尔兰联合王国	100
美利坚合众国	102

巴勒斯坦国	104
B. 欧洲联盟	106
附件二	
从国际和区域组织、红十字国际委员会、民间社会、科学界和工业界收到的答复	108
A. 红十字国际委员会	108
B. 民间社会	111
非洲青少年痴客	111
大赦国际	114
军备控制协会	116
第三十六条	119
TEDIC 协会，大赦国际巴拉圭，Heinoi 研究中心，巴拉圭人权协调委员会，胜利和种子民主基金会	123
人工智能和数字政策中心	126
民政研究所	129
国际终止童妓组织危地马拉分会、秘鲁促进裁军组织、女权主义人工智能研究网络拉丁美洲分会、Anderson Henao 和 Jesús Martínez	131
生命未来研究所	135
日内瓦安全政策中心	136
人权观察	138
国际机器人武器控制委员会	142
机构间研究所	144
国际防止核战争医生组织	147
拉丁美洲和加勒比人类安全网、和平与民主基金会、技术与社区、国际政治研究中心、布宜诺斯艾利斯大学	148
挪威人民援助会和地雷行动(加拿大)	151
PAX 组织	153
新西兰奥特亚罗瓦和平运动和停止杀手机器人组织	154
芬兰和平联合会、芬兰和平委员会、百人委员会、生命技术、妇女促进和平协会、出于良心拒服兵役者联盟	157
和平教育艺术沟通研究所	160
化剑铸犁促进会	161

“保护”组织	163
“安全地带”组织	166
“阻止杀手机器人”组织.....	168
“阻止杀手机器人”青年网络.....	172
可持续和平与发展组织.....	175
国际妇女争取和平与自由联盟.....	177
非洲虚拟星球、国际妇女争取和平与自由联盟津巴布韦分会、文化青年运动教育基金会	180

一. 引言

1. 大会关于致命自主武器系统的第 78/241 号决议第 2 段请秘书长征求会员国和观察员国对致命自主武器系统的意见，特别是如何应对此类系统在人道主义、法律、安全、技术和伦理方面提出的相关挑战和引起的关切，以及人在使用武力中的作用，并向大会第七十九届会议提交一份反映收到的所有意见的实务报告，同时在附件中载列这些意见，供会员国进一步讨论。在同一决议第 3 段，大会又请秘书长征求国际和区域组织、红十字国际委员会、民间社会、科学界和工业界的意见，并将这些意见以收到的原文列入上述报告附件。本报告根据这些要求提交。

2. 2024 年 2 月 1 日，裁军事务厅向所有会员国和观察员国发出普通照会，提请各国注意大会第 78/241 号决议第 2 段，并征求各国对该事项的意见。裁军厅还向同一决议第 3 段所述实体发出普通照会和信函，提请它们注意该段，并征求它们对此事项的意见。截至 2024 年 5 月 25 日收到的意见抄录于本报告附件。此日期后收到的任何意见将以呈件所用语文在裁军厅网站上发布。

3. 本报告第二至第五节综述了从会员国和观察员国收到的答复的要点，但不影响其各自立场。秘书长的意见和结论载于第七节。

二. 背景

4. 各国指出，包括人工智能在内的快速技术进程可以推动经济增长，改善人类福祉，并有助于实现可持续发展目标。同时，新兴技术也可能对国际和平与安全带来挑战，引发对人在战争中的作用的疑问。各国认为，需要特别关注致命自主武器系统带来的独特挑战。

三. 定义和特性描述

形成定义

5. 各国指出，对于自主武器系统或致命自主武器系统，目前还没有国际商定的定义。几个国家指出，就定义或一般特性描述达成一致意见，可能有助于今后的工作。这些国家指出，在制定具体的禁止规定时，例如在就一项具有法律约束力的文书进行谈判的过程中，可以就定义或一般特性描述达成一致意见。有几个国家认为，取得进展和开始就具有法律约束力的文书进行谈判，并不需要确切的定义。

6. 各国在提交的材料中对“自主武器系统”和“致命自主武器系统”的提法各不相同。一些国家认为，“致命”一词是系统使用致命武力能力的重要指代。另一些国家则认为，武器系统的致命性取决于武器系统的使用而非设计。几个国家反对使用“致命”一词，认为该词在国际人道法中没有依据，这几个国家特别指出，致命性是武器使用方式的一种效果。还有国家指出，非致命性使用武力也可能导致违反国际人道法。在本报告中，根据大会第 78/241 号决议的术语，“致命自主武器系统”一词的使用不影响各国在该词使用方面的偏好。

7. 几个国家强调了人工干预程度的重要性，特别是在目标的识别、分类、拦截和打击方面。这些国家强调，名义人工投入是不够的。名义人工投入指的是不会对自主选择或打击目标的功能产生实质性影响的输入或行动。

8. 各国提供了致命自主武器系统的若干工作定义和一般特性描述。一些国家借鉴了红十字国际委员会的工作定义，该定义指出，“自主武器系统”是指启动后无需人工干预就能选择和打击一个或多个目标的武器系统。

9. 有国家认为，致命自主武器系统的一个特点可能包括将人工智能纳入目标选择和武力使用。还有国家认为，致命自主武器系统不需要包含人工智能。

10. 几个国家建议，某些自主或自动防空和导弹防御系统不应视为致命自主武器系统，因为这些系统具有防御性，且用于侦测和攻击目标的算法是确定性而非概率性的。他们指出，这种系统已经使用了几十年，没有引起任何法律争议。

人的控制

11. 有国家指出，人的控制、人实际控制、适当的人的判断和人的参与等概念已得到广泛讨论，有必要进一步讨论和澄清这些概念。

12. 许多国家强调，在使用武力方面必须保持人的控制。几个国家强调，在系统的整个生命周期，特别是在使用期间，必须保持这种控制或参与。这些国家指出，人的控制或参与的确切性质将取决于系统的具体能力及使用背景。这些国家认为，人的控制对于确保遵守国际法、特别是国际人道法以及确保责任和问责尤为重要。还有国家认为，强调人实际控制是没用的，不宜采用单一标准来描述人对致命自主武器系统使用的控制。

13. 各国考虑了人的控制的必要要素，其中包括人类保留以下方面：

- 充分的信息，包括关于武器系统能力和行动背景的信息，以确保遵守国际法
- 在国际人道法所要求的范围内作出判断的能力
- 限制任务和目标类型的能力
- 对使用时限、地理范围和规模加以限制的能力
- 重新定义或修改系统目标或任务的能力
- 中断或停用系统的能力

14. 各国就实现必要程度的人的控制而应采取的各种措施提出了建议，包括：

- 为人机交互创建直观界面
- 制定程序，确保致命自主武器系统得到测试、评估、验证和核查
- 对致命自主武器系统进行充分的法律审查
- 对所有与致命自主武器系统互动的人员进行适当培训

- 确保致命自主武器系统的可预测性、可靠性和可解释性

15. 有国家指出，需要进一步讨论维持必要程度人的控制的措施，而就这方面的良好做法进行交流可以是很好的第一步。

四. 挑战、关切和潜在益处

16. 各国指出，致命自主武器系统引起了一些关切，包括从人道主义、人权、法律、安全、技术和伦理角度的关切。各国呼吁全面解决这些关切。

17. 几个国家指出，这些风险涉及所有国家和社会各阶层，并可能对最弱勢的群体产生尤为严重的影响。有国家表示关切：这些风险可能对全球南方国家造成过分的严重影响。

人道主义考虑

18. 有国家指出，致命自主武器系统的开发和使用引起了根本的人道主义关切。还有国家指出，人将继续首当其冲地承受武装冲突的后果。还有国家认为，人道主义关切本身不足以限制和禁止某些类型的武器系统。

19. 各国提出的其他关切涉及致命自主武器系统的环境影响，特别是与开发和运行此类系统有关的能源成本和随之而来的碳足迹。

法律考虑

20. 各国回顾，国际法完全适用于致命自主武器系统。各国确定了以下适用法律文书和法律领域：

- 《联合国宪章》
- 习惯国际法
- 国际人道法
- 国际人权法
- 国际刑法
- 国家责任法

21. 各国强调了遵守国际法的重要性。几个国家指出，致命自主武器系统对国际法特别是国际人道法、国际人权法和国际刑法的遵守带来挑战。有国家指出，没有任何国际法律文书专门管制或禁止致命自主武器系统。有国家认为，各国之间就致命自主武器系统和国际法自愿交流看法，可能是有益的。

22. 各国指出，武器、战争手段和方法的选择，包括致命自主武器系统的选择，必须符合国际法特别是国际人道法。各国强调，必须遵守攻击中的区分原则、相称原则、军事必要原则和预防原则，并满足避免过分伤害或不必要痛苦的要求。

23. 几个国家强调，任何武器，包括致命自主武器系统，如果不符合国际人道法，事实上就已在禁止之列，不得使用。几个国家呼吁进一步具体说明国际人道法的规则和原则如何适用于致命自主武器系统。

24. 几个国家强调，需要人的控制来确保国际人道法得到遵守。还有国家表示，人的控制本身不是目的，但可以是在不同情况下履行各种国际人道法义务的一个相关概念。几个国家强调，必须考虑到整个武器系统的运行背景以及特点和能力。

25. 几个国家指出，致命自主武器系统无法复制人类在武装冲突中使用武力所需的微妙决策。有国家强调，必须保持人对生死攸关决定的控制。几个国家认为，可预测性、可靠性、可理解性和可解释性是确保国际法得到遵守的根本。

26. 为了确保遵守国际人道法，各方提交的意见认为各国应：

- 评估平民的存在
- 限制系统可以攻击的目标类型
- 限制武器系统运行的时限、地理范围和规模
- 制定接战规则
- 采取技术保障措施，如自毁和自失效

27. 几个国家强调了马顿斯条款¹的重要性，认为在没有人的控制的情况下使用致命自主武器系统可能违反该条款。

28. 几个国家提到，必须报告可能涉及违反国际人道法的涉及致命自主武器系统的事件。有国家呼吁普遍加入 1949 年日内瓦四公约及其附加议定书。

29. 为了确保遵守国际人道法，各方强调了军事法律顾问的作用。几个国家还强调，必须向国家武装部队人员提供有关国际人道法的适当培训和指导。

30. 有国家指出，没有保障措施的技术发展增加了以不符合国际人道法的方式使用致命自主武器系统的风险。在这方面，会议强调了负责任的创新的重要性。

31. 有国家提到《联合国宪章》中阐明的诉诸战争法，包括禁止使用或威胁使用武力，或以与联合国宗旨和原则不符的任何其他方式，侵害任何国家的领土完整或政治独立，以及自卫权和以和平手段解决冲突的要求。有意见认为，这些义务适用于所有国家活动，包括与致命自主武器系统有关的活动。

32. 各国强调，必须确保根据适用的国际法追究责任，并强调，人对军事行动中武器包括致命自主武器系统的效果负有责任并接受问责，这种问责不能转移到机器身上。有国家指出，这一责任适用于武器系统的整个生命周期。

¹ 例如见《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》，序言部分第五段。

33. 有国家认为，致命自主武器系统的设计不得阻止承担责任或接受问责。还有国家认为，只有通过人的控制才能实现问责，这要求授权使用武力的人能够预测和解释使用武力的效果。

34. 几个国家强调了指挥责任的重要性。根据指挥责任，指挥官对其指挥下发生的所有使用武力行为负责，而不论指令是由人类下属执行还是由机器执行。致命自主武器系统的使用可能会破坏这一原则，有国家对此表示关切。

35. 有国家指出，个人应按照纪律程序、国内刑法和国际刑法承担责任。有国家表示关切，致命自主武器系统可能进一步破坏追究冲突中非法暴力行为人责任的努力。此外，几个国家指出，国家对国际不法行为包括非法使用武器系统负责。

36. 几个国家对致命自主武器系统能否遵守国际人权法表示怀疑。有国家指出，国际人权法在使用武力方面的要求比国际人道法严格。有国家特别表示关切，自主选择目标功能的使用可能是任意的，这会违反国际人权法。有国家认为，人的控制的要求是遵守国际人权法的关键。有国家建议进一步讨论国际人权法对致命自主武器系统的设计、开发和使用的适用性。

37. 几个国家强调，必须按照《1949年日内瓦四公约第一附加议定书》第36条的要求，对武器、战争手段和方法进行法律审查。有国家指出，致命自主武器系统可能的不可预测性给进行此类法律审查带来了挑战。有国家建议，法律审查应包括人机互动的各个方面以及如何在培训中处理这些方面。有意见认为，法律审查本身不足以解决致命自主武器系统引起的关切，需要有具体的规则。有国家提到2023和2024年在澳大利亚举行的两次关于自主武器系统法律审查的专家会议。

38. 有国家指出，没有关于如何进行法律审查的规定，也没有公布这些审查结果的要求。几个国家强调，自愿交流关于致命自主武器系统法律审查的信息和做法是有益的。

安全考虑

39. 几个国家提到致命自主武器系统对国际和平与安全构成的风险。有国家强调必须考虑到所有国家的国家安全关切。

40. 几个国家指出，致命自主武器系统的使用可能是一个不稳定因素，它降低使用武力的门槛，可能使冲突的频率和强度增大，并引发人道主义危机。还有国家对致命自主武器系统的扩散可能产生的破坏稳定影响表示关切。一些国家提到致命自主武器系统的不可预测性造成的升级风险，包括机器对机器互动的可能性、战争速度的加快、使用国军事伤亡风险的降低以及不对称战争。各国对致命自主武器系统可能成为军备竞赛的目标表示关切。有国家认为，致命自主武器系统不应被用来谋求绝对的军事优势和霸权。

41. 几个国家表示关切，致命自主武器系统可能配备核武器或其他大规模毁灭性武器，这可能破坏战略稳定，增加核战争的风险。

42. 各国对致命自主武器系统的扩散在恐怖分子和犯罪集团等非国家行为体方面造成的后果表示关切。几个国家还对国内执法人员使用致命自主武器系统表示关切，因为这可能会引起人权关切。

43. 有国家认为，与致命自主武器系统等新兴技术有关的扩散风险可在多边出口管制制度中加以处理，但强调，不得对与这种系统有关的技术的转让施加不适当的限制。

技术考虑

44. 几个国家对致命自主武器系统易受一些技术风险的影响表示关切，这些风险包括：

- 恶意网络活动
- 硬件和软件异常和故障
- 基于不正确信息或不正确解读的信息做出决策

45. 有国家认为，将人工智能应用于致命自主武器系统的关键功能可能会增加意外结果的风险。

46. 各国还对致命自主武器系统的算法偏见表示关切，这种偏见加剧了现有的力量不平衡，对边缘群体的影响尤其大，并导致对冲突地区妇女儿童的附带伤害。各国强调，必须严格记录所用数据集，进行全面测试和审查，并就这一问题进行培训，提高对该问题的认识。

伦理考虑

47. 几个国家对机器程序取代人类判断表示关切，因为这些国家认为，伦理和道德考虑是关于致命自主武器系统讨论的根本。这些系统和其他机器被认为缺乏同理心、同情心和道德推理能力。有国家认为，与使用武力的决定有关的伦理责任要求由人来根据具体情况作出基于价值的判断。几个国家认为，以人为攻击目标，特别是授权机器作出剥夺人生命的决定是有悖伦理的。

48. 各国回顾，《世界人权宣言》和《公民及政治权利国际公约》承认人权源于所有人与生俱来的尊严，并强调所有国家都有责任确保人的尊严不可侵犯。几个国家表示关切，使用致命自主武器系统可能造成尊严丧失和非人化，从而可能导致不合理的暴力和平民伤亡。

49. 有国家指出，伦理方面的关切往往导致采取法律限制措施。几个国家认为，解决伦理关切最有效的办法是适当适用国际人道法。

潜在益处

50. 有国家认为，自主武器系统可能会带来正当的军事利益，包括：

- 提高安全性和效率

- 改善对国际人道法的遵守，改善对平民的保护，通过提高准确性等减少附带损害风险
- 减少国防人员的风险
- 避免由操作人员的精神或身体状态以及他们的道德、宗教和伦理倾向造成的错误
- 更有效地利用劳动力

五. 各国的审议意见

《特定常规武器公约》

51. 各国提到在《特定常规武器公约》范围内审议致命自主武器系统问题。有国家认为，该公约的结构——有可能就附加议定书进行谈判——非常适于处理新兴技术引起的问题。许多国家认为，致命自主武器系统领域新兴技术问题政府专家组是讨论这一专题的适当论坛。

52. 几个国家指出，政府专家组确保了在人道主义关切与各国在这类武器方面正当自卫利益之间取得平衡。这些国家还指出，所有相关行为体，包括在这一领域拥有最先进能力的国家，都派代表参加了专家组。

53. 政府专家组汇集了各领域的专家，包括具有法律、军事、技术和外交专门知识的专家，几个国家对此表示欢迎。但有几个国家吁请专家组进一步加强包容性，反映更多样化的观点。

54. 几个国家指出，政府专家组迄今的审议取得了进展并实现了意见的融合。人们列举的最显著的成就包括致命自主武器系统领域新兴技术问题政府专家组通过的 11 项指导原则² 以及围绕两级方法的意见融合。³ 几个国家对进展缓慢表示遗憾。

55. 几个国家欢迎《特定常规武器公约》缔约国 2023 年会议通过的任务授权，即政府专家组的工作应继续下去，并向 2026 年《公约》缔约国第七次审议大会提交一份实务报告。⁴ 几个国家吁请专家组避免进一步拖延，并在此期间取得切实成果。

56. 展望未来，有国家呼吁政府专家组在以往商定意见的基础上开展工作，同时考虑到已提出的多方面建议。还建议专家组充分考虑到大会第 78/241 号决议。几个国家建议专家组将重点放在澄清现行法律特别是国际人道法如何适用于致命

² CCW/MSP/2019/9，附件三。

³ 见致命自主武器系统领域新兴技术问题政府专家组 2023 年会议报告，第 21 和 22 段。可查阅 [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW-GGE.1-2023-2_English.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW-GGE.1-2023-2_English.pdf)。

⁴ 见 CCW/MSP/2023/7，第 20 和 26 段。

自主武器系统。其他建议的重点领域包括：加强对特性描述和定义的理解，人机合作，安全和负责任地开发和使用致命自主武器系统的技术和治理标准，禁止和管制。

57. 几个国家建议设立一个技术专家组，定期向缔约国通报致命自主武器系统的最新发展情况，并协助各国采取适当措施应对具体挑战。

58. 关于政府专家组的工作成果，几个国家表示支持就《特定常规武器公约》致命自主武器系统问题议定书进行谈判并通过议定书。另一项建议是，政府专家组应编写一份关于在致命自主武器系统方面适用国际人道法和最佳做法的声明。

59. 几个国家认为，在政府专家组之外的平行进程中审议致命自主武器系统问题可能会适得其反，导致各自为政。不过，有几个国家认为，在其他论坛上进行讨论，特别是如果这类讨论支持政府专家组的工作的话，那么这样做在包容性和与相关专题建立联系方面可能具有优势。

大会

60. 几个国家指出，鉴于大会成员的包容性，大会审议致命自主武器系统问题很有价值。在这方面，有几个国家对大会第 78/241 号决议的通过表示欢迎，并希望该决议将加快国际社会对致命自主武器系统的反应。几个国家指出，本报告将使所有会员国和观察员国都能够发表意见，并可作为一种资源，供政府专家组等使用。

61. 有国家认为，致命自主武器系统问题应由裁军审议委员会及在《未来契约》的范畴内处理。还有国家表示，大会可以考虑开始就一项关于致命自主武器系统的具有法律约束力的文书进行谈判。

利益攸关方

62. 几个国家强调，所有利益攸关方，包括国际组织、民间社会、学术界和工业界的代表，在关于致命自主武器系统的讨论中作出贡献，是十分重要的。几个国家呼吁酌情加强与利益攸关方的合作。有国家认为，关于致命自主武器系统的具法律约束力文书的谈判应在最具利益攸关方包容性的论坛上举行。

区域和国际会议

63. 几个国家呼吁加强区域合作和伙伴关系，推进关于致命自主武器系统的共同立场。在这方面，他们欢迎关于致命自主武器系统问题的国际和区域会议，包括在哥斯达黎加、卢森堡、菲律宾、塞拉利昂及特立尼达和多巴哥举行的会议。几个国家提到 2023 年 3 月伊比利亚美洲首脑会议通过的特别公报以及在奥地利举行的自主武器系统国际会议。

军事领域的人工智能

64. 几个国家指出人工智能是实现自主的关键因素，并提到人工智能在军事领域的使用既可能产生积极影响，也可能带来消极影响。有国家指出，人工智能已经在战场上使用，呼吁各国承诺在军事领域负责任地使用人工智能。

65. 几个国家欢迎这方面的举措，包括军事领域负责任人工智能进程、全球人工智能治理倡议、关于在军事上负责任地使用人工智能和自主的政治宣言和《布莱切利宣言》。有国家认为，与人工智能治理有关的举措应由联合国主持，以确保包容性和有效性。还有国家认为，这些举措不应损害致命自主武器系统领域新兴技术问题政府专家组的工作。

六. 今后的步骤

66. 几个国家强调，迫切需要通过多边讨论解决致命自主武器系统引起的关切。几个国家强调，协商一致原则在关于国际安全的谈判，包括关于致命自主武器系统的谈判中，具有重要意义。

67. 几个国家呼吁进一步制定致命自主武器系统治理规范和业务框架。虽然有几个国家呼吁加强国际法律框架，并在致命自主武器系统方面进一步制定具体规定，但其他国家认为，现有法律框架足以解决新的军事能力包括致命自主武器系统的问题。

68. 许多国家表示支持两级方法，即：应禁止其使用不符合国际法的致命自主武器系统，而对其他系统则进行适当管制。有意见认为，禁令不应被视为阻碍技术创新的一种方式，也不应阻碍将技术用于和平目的或正当防卫目的。

具有法律约束力的文书

69. 几个国家呼吁就一项关于致命自主武器系统的具有法律约束力的文书进行谈判，以便：

- 根据《特定常规武器公约》序言，继续编纂和逐步发展适用于武装冲突的国际法规则
- 澄清国际人道法对致命自主武器系统的适用并促进实施国际人道法
- 填补国际法特别是国际人道法的空白

70. 呼吁就一项具有法律约束力的文书进行谈判的国家就文书内容提出了几项建议。他们指出，除其他外，该文书应：

- 纳入国际法包括国际人道法、国际人权法、国际刑法和其他相关文书的相关规范和原则
- 促进对这一问题采取综合办法
- 确保遵守相关国际法
- 通过不同的国家措施防止管制分散
- 解决与致命自主武器系统相关的人道主义风险和基本伦理问题
- 解决致命自主武器系统以人为目标的相关问题

- 促进所有国家平等参与

71. 几个国家提到，秘书长呼吁到 2026 年缔结一项具有法律约束力的文书，禁止在不受人控制或监督的情况下运作以及其使用不符合国际人道法的致命自主武器系统，并对所有其他类型的自主武器系统进行管制。这些国家还提到秘书长和红十字国际委员会主席以及包括特别报告员在内的联合国系统其他行为者 2023 年 10 月发出的呼吁。

72. 几个国家认为，国际人道法现有规则已经足够，没有必要制定一项新的具有法律约束力的文书。另一种意见认为，在制定新的法律文书之前，需要就现行法律特别是国际人道法如何适用于致命自主武器系统达成共同理解。有国家认为，有必要进一步澄清这样一项具有法律约束力的文书的效用。其他国家指出，关于致命自主武器系统的文书可以具有政治约束力，也可以具有法律约束力。

73. 有国家认为，制定负责任地开发和使用致命自主武器系统的规范和标准是最佳的前进方向。还有国家认为，就致命自主武器系统相关行为规则进行谈判为时过早。几个国家认为，自愿举措有助于分享最佳做法和建立规范，但只能作为就具有法律约束力的文书开始谈判之前的临时措施。

74. 展望未来，几个国家呼吁采取包容性、多学科的办法，并争取尽可能广泛的国家集团的支持。还有国家呼吁就致命自主武器系统采取区域和次区域行动。几个国家强调，在讨论致命自主武器系统时必须考虑到性别平等视角。有国家建议，在通过关于致命自主武器系统的文书之前，应宣布暂停开发这些武器系统。

禁止的范围

75. 有国家认为，关于致命自主武器系统的任何文书都应重申国际人道法中已经存在的相关禁令。此外，许多国家呼吁禁止完全不受人控制的致命自主武器系统和运行不符合国际人道法的致命自主武器系统。

76. 各国介绍了致命自主武器系统的几个特点，这些特点使此类系统的运行不符合国际人道法，并会使这些系统：

- 固有滥杀滥伤性
- 无法区分战斗人员和平民
- 旨在对平民或民用物体使用武力
- 无法确定攻击是否会对民用物体造成与预期军事利益相比过分的附带损害
- 具有造成过度伤害或不必要痛苦的性质
- 造成无法可靠预测、预期、理解或解释的影响
- 效果无法限制和控制

77. 几个国家呼吁禁止旨在直接以人为目标的致命自主武器系统。

管制的范围

78. 几个国家呼吁对不受上述禁令约束的致命自主武器系统进行管制。

79. 有国家指出，对致命自主武器系统采取的措施取决于系统的特点、运行环境和使用者。还有意见指出，自主是有范围的，可能必须根据致命自主武器系统在这个范围内所处的位置，为这些系统制定一系列措施。

80. 各国审议了这种管制的目的，包括：

- 确保始终保持人的控制，特别是对关键功能的控制
- 确保人的指挥和责任链
- 确保在系统的整个生命周期内充分遵守国际法，包括国际人道法

81. 为了确保充分遵守国际人道法，各国提议采取一些措施，包括：

- 控制或限制系统可攻击目标的类型，包括将目标限定为军事目标
- 限制使用期限、地理范围和规模
- 确保由人批准任何使用武力的决定
- 通过自毁、自失能和自失效机制等方式，确保任务参数的任何变化，如运行时限、地理范围和规模，都得到人的批准
- 限制攻击次数
- 确保操作人员尽职尽责地执行国际人道法的原则和要求
- 确保充分的可预测性和可靠性
- 确保进行法律审查，以确定新的武器、战争手段和方法符合国际人道法

82. 各国还提出了其他措施，包括：

- 确保进行严格测试，评价和评估武器系统在各种使用情况下的性能
- 确保对致命自主武器系统进行登记、跟踪和分析
- 确保进行风险评估
- 确保对决策人员和操作人员进行充分培训
- 促进采取风险缓解措施和保障措施，确保可靠性并防止失误、滥用、转用和放弃人的特权
- 处理环境影响

拟列入文书的其他问题

83. 几个国家审议了除禁止和管制外关于致命自主武器系统的文书的其他内容。这一文书可包括：

- 序言部分提及致命自主武器系统领域新兴技术问题政府专家组通过的 11 项指导原则
- 定期更新技术附件
- 在国内法中执行文书规定的义务
- 各国定期报告文书执行情况
- 审查文书执行情况的监测程序
- 调查和纠正任何涉嫌、报告或记录的违反文书行为的措施
- 防止致命自主武器系统扩散和未经授权使用的措施
- 能力建设

风险缓解措施

84. 各国强调，风险评估和缓解措施应成为任何武器系统新兴技术整个生命周期的一部分。有国家指出，这种评估和缓解措施应考虑到现有法律义务，并尊重国家管辖权和能力。几个国家呼吁确定风险缓解措施。

85. 各国提议采取一些风险缓解措施，包括：

- 保留负责任的人指挥链
- 有效的生命周期管理
- 综合测试
- 防止未经授权的使用，包括非国家行为体未经授权的使用
- 操作人员培训
- 操作人员持续监测致命自主武器系统的运行，确保终止的可能性
- 保障措施，如数据保护
- 记录翔实的安全程序

86. 有国家建议，可以在自愿的基础上分享风险缓解做法的经验教训。

其他措施

87. 有国家认为，技术标准化对于确保致命自主武器系统符合国际法至关重要，电气和电子工程师协会标准 7007-2021 可作为有益的参考。有国家指出，透明度和建立信任措施可在解决与致命自主武器系统有关的关切方面发挥作用。几个国家强调，国际合作和能力建设十分重要，可以促进各国有效参与有关致命自主武器系统的多边讨论，并促进减少风险。

七. 秘书长的意见和结论

88. 本报告收到的众多意见证明了各国、国际组织和民间社会对致命自主武器系统的浓厚兴趣。从人道主义、人权、法律、安全、技术和伦理角度来看，致命自主武器系统可能产生有害影响，这一点已得到广泛承认。人们普遍担心，这些武器系统有可能极大地改变战争，并可能损害甚至侵蚀现有的法律框架。

89. 人们普遍认识到，人的控制对于确保责任和问责、遵守国际法以及合乎伦理的决策至关重要。因此，我重申，需要采取紧急行动，维护人对使用武力的控制。给予机器剥夺人生命的能力和自由裁量权，在政治上是不可接受、伦理上是令人厌恶的，应该为国际法所禁止。

90. 人们强烈地感到，国际社会在这个问题上采取预防行动的时间已经不多。因此，我再次呼吁到 2026 年缔结一项具有法律约束力的文书，禁止在不受人控制或监督的情况下运行以及其使用不符合国际人道法的致命自主武器系统，并对所有其他类型的自主武器系统进行管制。机器自主地以人为目标，是一条绝不能逾越的道德底线。

91. 在这方面，我注意到致命自主武器系统领域新兴技术问题政府专家组目前在《特定常规武器公约》范围内的任务。我吁请《公约》所有缔约国努力工作，尽快完成这一任务。我还吁请所有其他国家作为观察员参加专家组的工作，并紧急考虑遵守《公约》，以便参与审议和决策。我回顾，《公约》缔约国重申需要继续编纂和逐步发展适用于武装冲突的国际法规则。我吁请各国利用一切机会在这方面取得进展。

92. 我鼓励大会继续审议致命自主武器系统问题。大会几乎实现了成员的普遍性，实务讨论范围广泛，因此为包容性讨论和审议致命自主武器系统与其他国际和平与安全事项之间的关系提供了平台。我建议大会继续充分了解在《特定常规武器公约》的主持下在这方面所做的努力。各国可考虑就本报告提出的事项举行非正式磋商。将于 2024 年 9 月 22 日和 23 日举行的未来峰会也是一个重要的机会，可以为到 2026 年缔结一项关于致命自主武器系统的具有法律约束力的文书创造势头。本报告可作为政府专家组工作以及今后关于致命自主武器系统的任何进程的参考。

93. 尽管需要专门针对致命自主武器系统制定新的禁令和管制规定，但必须充分执行适用于这些武器系统的现有国际法，包括国际人道法和人权法。为此，我吁请所有尚未批准日内瓦四公约及其附加议定书以及核心国际人权文书的国家批准这些文书。所有国家都应严格遵守国际人道法和国际人权法规定的义务。

附件一

收到的答复

A. 会员国和观察员国

安道尔、阿根廷、多民族玻利维亚国、巴西、哥伦比亚、哥斯达黎加、古巴、智利、多米尼加共和国、厄瓜多尔、萨尔瓦多、危地马拉、洪都拉斯、墨西哥、尼加拉瓜、巴拿马、巴拉圭、秘鲁、葡萄牙、西班牙、乌拉圭和委内瑞拉玻利瓦尔共和国

[原件：西班牙文]

[2024 年 5 月 24 日]

第二十八届伊比利亚美洲国家元首和政府首脑会议，2023 年 3 月 25 日，圣多明各
关于自主武器的社会和人道主义影响的特别公报

伊比利亚美洲国家元首和政府首脑齐聚多米尼加共和国圣多明各，出席第二十八届伊比利亚美洲首脑会议：

a. 新兴技术对国际和平与安全构成了具体挑战，并对人类在战争中的作用提出了新的问题。需要全面了解武器系统自主性的影响及其对人实际控制的影响，以充分评估伦理、法律、人道主义和安全方面的影响。

b. 随着新武器技术的不断开发和在武装冲突中的使用，有必要在国际法，包括国际人道法和国际人权法的指导下，根据人道原则和公众良心的要求，制定新的禁令和条例；区别原则、必要性原则和相称性原则；严重违反国际人道法行为的个人责任的重要性；以及伦理观点。

c. 根据国际法，包括《联合国宪章》所载的宗旨和原则、国际人权法和国际人道法，呼吁国际社会通过发展和加强国际法律框架来应对这些威胁，以防止流离失所、平民伤亡和民用物体被毁的影响。

d. 至关重要的是，要保持人实际控制，以防止战争进一步非人化，并确保个人问责、国家和非国家武装团体的责任以及受害者的人权。

e. 伊比利亚美洲国家在《特定常规武器公约》框架内积极参加了致命自主武器系统领域新兴技术问题政府专家组的讨论。有一批国家提出了未来规范性框架的内容和《特定常规武器公约》框架下的议定书六草案。本地区的一些国家也加入了 70 个国家在 2022 年 10 月联合国大会上发表的关于自主武器的声明。

f. 此外，2023 年 1 月 24 日在拉美及加勒比国家共同体第七届国家元首和政府首脑会议上发表的《布宜诺斯艾利斯宣言》强调了 2014 年 1 月在哈瓦那举行的第二届国家元首和政府首脑会议上签署的《宣布拉丁美洲和加勒比为和平区宣言》。

g. 联合国秘书长、联合国法外处决、即决处决或任意处决问题特别报告员、残疾人权利问题特别报告员、联合国当代形式种族主义问题特别报告员、红十字国际委员会、民间社会，科学界和学术界对武器系统的自主性表示关切，并一再呼吁各国谈判一项关于这方面问题的具有法律约束力文书。

h. 在 2023 年 2 月 23 日和 24 日举行的拉丁美洲和加勒比自主武器的社会和人道主义影响会议公报中，本地区的代表认识到武器系统自主性带来的风险和挑战，以及制定具有法律约束力的国际文书的必要性，建立禁令和法规，以保证人实际控制。

伊比利亚美洲各国代表赞同该公报，承认需要采取进一步行动，以便：

1. 开展合作，促进紧急谈判一项具有法律约束力的国际文书，其中载有关于武器系统自主性的禁令和规定，以确保遵守国际法，包括国际人道法和国际人权法，以及伦理观点，并防止武器系统自主性造成的社会和人道主义影响；

2. 承诺积极参与，以推进共同立场并推进此类谈判；

3. 继续并加强与国际组织、民间社会组织、学术界和其他相关利益攸关方的合作和伙伴关系，以利用其相关专门知识和支持；

4. 各国之间继续进行讨论，以拟订可在《特定常规武器公约》和其他多边论坛框架内推广的可能建议。

阿根廷

[原件：西班牙文]

[2024 年 5 月 24 日]

阿根廷认为，首先必须认识到致命自主武器系统在遵守国际人道法、保护人的尊严和人权、伦理关切、国际稳定与安全以及维持和平方面构成的严重风险和挑战。阿根廷致力于防止自主武器的军备竞赛，因为这种武器的开发可能加剧国际安全的脆弱性，扩大技术差距。

《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》是讨论这一专题的适当论坛，因为其目的和存在理由是促进国际人道法的逐步发展，并使人们能够不断评估技术和武器的新发展。这个论坛具有相关的经验和代表性，因为它已就这一问题进行了 10 年的讨论，达成了一些基本协议，并包括该部门技术发展水平最高的国家。然而，我们认识到，在其他论坛进行的讨论可能有益于网络安全、劳工、运输和知识产权等其他讨论议题的普遍性、联系和协同作用。

致命自主武器系统监管方面的任何发展都应遵循使用武力时的区分原则、相称原则和人的责任原则。作为将人工智能应用于自主武器系统的一般管理规则，应在国防需求和人道主义保护之间取得平衡。

阿根廷认为，在开发、制造、使用、拥有和转让这类系统方面，必须保持禁止和管制之间的区别。关于管制，一般原则应该是对自主武器系统的关键功能保持人实际控制。此外，重要的是要有足够的知识和信息来了解致命的自主武器系统，评估这种系统的运作，并避免算法偏见的发展。

阿根廷认为，各国应致力于更新其管制军事和敏感材料，包括无形技术出口的制度，以防止致命自主武器系统的扩散。

2018 年和 2019 年，阿根廷提交了两份文件(CCW/GGE.1/2018/WP.2 和 CCW/GGE.1/2019/WP.6)，其中强调了根据《1949 年 8 月 12 日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》(《第一议定书》)第三十六条进行武器评估的义务。与致命自主武器系统有关的措施可包括对评估已获得或开发的系统提出更严格的要求，以确保其符合国际法，包括国际人道法，并制定各国共享的共同评估标准。正如该文件所强调，必须促进国际合作和援助，以弥补各国开展评估的技术能力方面的差距。在这方面，交流最佳做法、开发人力资源和技术转让至关重要。

阿根廷作为 13 个国家(阿根廷、哥斯达黎加、厄瓜多尔、萨尔瓦多、危地马拉、哈萨克斯坦、尼日利亚、巴拿马、秘鲁、菲律宾、塞拉利昂、乌拉圭和巴勒斯坦国)的一部分，在 2022 年 7 月致命自主武器系统问题政府专家组会议上提交了关于《特定常规武器公约》议定书六的提案。2023 年 5 月，在 2023 年举行的致命自主武器系统问题政府专家组第二次会议上，专家组提交了议定书六的修订草案，其中包含更新和修订，并考虑到 2023 年 3 月会议的讨论和建议。更多的国家加入了该集团，包括智利、哥伦比亚和多米尼加共和国。

阿根廷认为，重要的是促进制定以致命自主武器系统为重点的条例，并考虑到所有国家不同的经济和技术能力，以避免采取将新技术的使用和惠益局限于技术更先进国家的做法。

考虑采取的任何可能措施都不应限制或剥夺各国获得、开发、研究、获取、生产、转让或使用自主技术的权利，包括人工智能的发展。

最后，为了推动可持续和公平的国际安全标准，促进有效监管致命自主武器系统的工作应包括具有约束力的文书以外的文书，如政治宣言、行为守则、市场规则和限制、系统架构、方案拟订基准和共同的军事理论。

澳大利亚

[原件：英文]
[2024 年 5 月 24 日]

澳大利亚感谢有机会根据大会 2023 年 12 月 22 日通过的关于致命自主武器系统的第 78/241 号决议，向秘书长关于致命自主武器系统的报告提交意见。

在澳大利亚提交的意见中，¹ 我们就“如何应对此类系统从人道主义、法律、安全、技术和伦理角度提出的相关挑战和引起的关切，以及人在使用武力方面的作用”提出了看法，同时承认致命自主武器系统的潜在好处。

澳大利亚对报告的意见在以下标题中概述了我们对致命自主武器系统的看法：

- 法律考虑
 - 适用国际人道法
 - 进行法律审查
 - 人的参与和控制
 - 利用技术支持国际人道法
 - 风险缓解措施
- 法律考虑
- 问责考虑
- 安全和技术考虑
 - 网络安全漏洞
 - 扩散危险
- 无意偏见
- 澳大利亚国防战略
- 在负责任地将人工智能和自主用于军事目的方面的补充努力
 - 军事领域的负责任的人工智能
 - 《关于在军事上负责任地使用人工智能和自主的政治宣言》

奥地利

[原件：英文]

[2024年5月24日]

奥地利欢迎有机会根据大会第 78/241 号决议提交其意见，供联合国秘书长审议。

2024年4月29日和30日，奥地利在维也纳召开了题为“十字路口的人类：自主武器系统和管制挑战”的国际会议。

¹ 该报告可查阅 <https://meetings.unoda.org/ga-c1/general-assembly-first-committee-seventy-ninth-session-2024>。

主席对会议的总结反映了自主武器系统在人的控制、问责和遵守国际法，包括国际人道法和国际人权法的性质方面的根本挑战。该系统还反映了它们对人的尊严提出的伦理关切，以及人道原则和公众良知要求所产生的道德要求。

与红十字国际委员会一样，我们认为当前的国际法律框架完全适用于自主武器系统，但还不足以应对所有这些挑战。武器系统自主性提高的复杂性对国际法的适用和遵守提出了前所未有的挑战。达成一项具有法律约束力的文书还需要解决更广泛的问题，如人道主义风险和与自主武器系统有关的基本伦理问题。

我们强调，相关法律框架不仅应涉及国际人道法，而且还应涉及国际人权法和国际刑法，因为自主武器系统极有可能在武装冲突之外使用，相关技术中的双重用途十分普遍，私营企业在设计和开发此类系统方面发挥着重要作用，因此这些法律框架具有相关性。

自主武器系统对和平与安全构成全球风险，包括向非国家武装团体扩散。在安全竞争加速的背景下，自主武器系统可能成为军备竞赛的对象。在这方面，它们可能降低战争的门槛，导致意外的升级。

一个基本的人道主义、法律、安全和伦理关切问题是，人类有可能失去对使用武力的控制，而人类将继续承受武装冲突的后果。

自主武器系统可能在确定攻击责任方面带来挑战。自主武器系统向非国家武装团体和其他暴力行为体扩散，可能会加剧武装冲突和执法中的安全风险。

这些令人深感关切的问题和风险涉及所有国家和社会各阶层。对自主武器系统进行管制的必要性符合人类维护和平与安全的利益。技术部门和工业界应成为这一努力的合作伙伴，努力使法律明确。

所有相关利益攸关方，包括国家、联合国系统、国际和区域组织、技术部门和工业界、学术界和民间社会，都应发挥作用，并应密切参与管制自主武器系统的多边努力。

人类正处于十字路口，必须团结起来应对管制这些武器的挑战。这可能是我们这一代人的“奥本海默时刻”。来自各个领域的专家一直在警告，不受管制的自主武器系统竞赛会对人类造成巨大风险和严重后果。国际社会必须努力应对管制这类系统的挑战。到目前为止，国际社会的努力与这一发展的速度和意义不相称。为此，迫切需要坚定的政治领导。

奥地利坚决支持联合国秘书长和红十字国际委员会主席 2023 年 10 月 5 日的联合呼吁，敦促会员国启动谈判，制定一项新的具有法律约束力的文书，明确禁止和限制自主武器系统，并在 2026 年前完成此类谈判。奥地利强调这一问题的紧迫性，并强调必须从法律、伦理和安全角度解决自主武器系统引起的深刻关切。奥地利还坚决支持在即将发表的文件《未来契约》中处理自主武器系统问题。

鉴于自主武器系统的技术进步速度以及联合国内部已经就此问题开展的工作，2026年是一个适当的目标日期。任何进一步的拖延都将对这一文书的效力产生负面影响。

奥地利议会于2024年4月17日通过了一项所有党派的动议，敦促政府继续在多边层面努力谈判和通过一项具有国际法律约束力的文书，以管制自主武器系统。

奥地利鼓励谈判一项具有法律约束力的文书，其中包括禁令和条例。会造成无法充分解释或预测或充分控制的影响的自主武器系统是不可接受的，将违反国际人道法，因此必须加以禁止。自主武器系统以侵犯人的尊严和价值以及人道原则或公众良心要求的方式选择和使用人作为目标，是不可接受的，必须予以禁止。所有其他自主武器系统都应受到管制，以确保人类对其使用进行切实的控制。关于奥地利对这一问题的看法的更多细节载于2023年和2024年提交致命自主武器系统问题政府专家组的工作文件。

人类对自主武器系统的切实控制可以通过几个条件的组合来实现，包括：(a) 对武器系统的功能理解；(b) 对武器可以使用和正在使用的背景进行充分评估；以及由于这些因素，(c) 可能需要在持续时间、地理区域、交战次数和目标类型方面加以限制。人实际控制还意味着，操作人员能够在法律和道德基础上评估有意使用武力的可预见影响。在自主武器系统的研究、开发、获取和使用过程中，必须不断审查和重新评估在满足上述条件方面对其功能可能作出的任何改变和修改。这应该包括技术方面，例如机器学习和系统功能所基于的任何数据集。

这一监测进程应纳入一个适当的多层次国际管制框架，该框架要求定期审查禁令和积极义务的执行情况，以确保人类对自主武器系统保持切实的控制，并确保法律规则和伦理原则在其设计、开发和使用中得到保护。

另一层法规涉及与基于人工智能的系统有关的措施，例如确保数据的完整性、质量和准确性，防止算法偏见，防止自动化偏见以及充分培训所有相关级别的人员。还必须确保此类武器系统的安全，特别是在网络安全、人工智能特有的脆弱性和扩散风险方面。

对使用武力及其后果的问责不能转移到机器或算法上。国际人道法和国际人权法的规则是针对个人和集体的。作为法律要求的问责可以通过人实际控制来实现。特别是，问责要求授权使用武力的人能够解释和预测其可预见的影响。有效的治理是避免缺乏问责的必要条件。

最后，自主武器系统的特点是，一旦启动，就可以选择目标并施加力量，而无需进一步的人为干预。我们认为没有必要继续使用“致命”这一限定词，因为我们认为这不是描述自主武器系统的合适特征。它在国际人道法中找不到任何理由，因为国际人道法不仅保护平民免遭死亡，而且也保护平民免遭伤害。

十字路口的人类：自主武器系统和管制的挑战，维也纳，2024年4月30日：主席总结

迄今为止，阿尔巴尼亚、亚美尼亚、奥地利、比利时、智利、哥斯达黎加、塞浦路斯、刚果民主共和国、格鲁吉亚、匈牙利、爱尔兰、哈萨克斯坦、基里巴斯、列支敦士登、马达加斯加、马拉维、秘鲁、摩尔多瓦共和国、圣马力诺、斯里兰卡赞同上述总结。

2024年4月29日和30日在奥地利维也纳召开的“十字路口的人类：自主武器系统和管制挑战”会议，¹ 聚集了各国、联合国代表、国际和区域组织、学术界、工业界、议员和民间社会，讨论与自主武器系统有关的影响和挑战，以及解决这些问题的可能途径。包括来自144个国家的1000多名与会者出席了会议。

会议通过高级别政治小组、专家讨论以及各国和其他利益攸关方的发言，进行了内容丰富的意见交流。以下各点仅代表会议主席的总结，并不影响各国和其他利益攸关方的立场。各国和其他利益攸关方在会议期间发表的声明可在会议网站上查阅。

自动武器系统一旦启动，就可以在没有进一步人为干预的情况下选择目标并施加武力，这引起了法律、伦理和安全方面的担忧。根本性挑战涉及人类控制的性质、问责以及此类系统与国际法(包括国际人道法和国际人权法)的总体兼容性。该系统还反映了它们对人的尊严提出的伦理关切问题，以及人道原则和公众良知要求所产生的道德要求。

自主武器系统对和平与安全构成全球风险，包括向非国家武装团体扩散。承诺速度优势的自主武器系统可能不允许人实际控制，并有可能破坏国际安全。一场危险的自主军备竞赛迫在眉睫。这些风险涉及所有国家和社会各阶层，对弱势群体影响尤为严重。

新技术为增进人类福祉带来了巨大希望。它必须为人民赋能，而不是使他们丧失人性。我们如何管制技术的危害也关系到我们从这些技术中获益的能力。在使用武力时，必须以人为主导。把目标选择和生死决定权交给机器是一个关系到我们所有人的问题。它深刻挑战了我们的全球社会契约和《联合国宪章》。

在《特定常规武器公约》框架内设立的政府专家小组正在进行的宝贵讨论、哥斯达黎加、卢森堡、特立尼达和多巴哥、菲律宾和塞拉利昂组织的关于自主武器系统的区域会议，以及关于军事领域人工智能更广泛问题的相关进程，都表明了处理自动武器系统问题的必要性。

¹ 见 <https://www.bmeia.gv.at/en/european-foreign-policy/disarmament/conventional-arms/autonomous-weapons-systems/2024-vienna-conference-on-autonomous-weapons-systems>。

人们一致认为，应明确禁止根据国际法不能使用或在伦理上不可接受的自主武器系统。所有其他自主武器系统都应该受到适当的管制(所谓的两级方法)。联合国秘书长和红十字国际委员会主席强调，迫切需要谈判并通过一项国际法律文书，明确禁止和限制自主武器系统，并在 2026 年前完成谈判。采取这种行动的预防窗口正在关闭。许多国家和其他利益攸关方支持这一呼吁。

小组讨论中出现了以下几点，会议主席认为这些观点对于未来全面解决自主武器系统的禁令和法规至关重要：

- 国际人道法和国际人权法的规则是针对个人和集体的。问责是法律制度的核心，不能转移到机器上。有效的治理是避免缺乏问责的必要条件。
- 人类决策者在知情的情况下进行符合道德的参与是我们管理使用武力的法律框架的基础。它使人们对所发生的结果负责。
- 关于使用自主武器系统的法律判断需要了解武器系统、可能使用这些系统的情况及其预期结果。
- 无法被充分理解或局限于特定背景的系统无法受人控制，因此不符合合法使用和问责。
- 自主武器系统风险的严重性要求我们澄清现有法律规则的应用，并制定明确的禁令和法规，以保护使用武力时的人为因素。
- 自主武器系统提出了与国际人权法、人类尊严和基本伦理原则有关的问题，伦理应该成为推动我们做出回应的力量。
- 伦理一直是创造和发展关于战争行为的法律的主要引擎。
- 瞄准目标人群是一个最紧迫的伦理问题。此外，通过区分某些人群和其他人群来发挥作用的自主武器系统可能会在其构建的数据集和算法中出现偏见问题。
- 这些偏见问题反映了社会结构，可能过于严重影响已经被边缘化的群体，并导致严重的错误。
- 武器系统自主程度的提高也对国际和平与安全提出了严重挑战。
- 距离和非人化，降低使用武力的政治门槛的风险以及升级风险，包括机器对机器的互动，引起了进一步的关注。
- 自主武器系统可能在确定攻击责任方面带来挑战。自主武器系统向非国家武装团体和其他暴力行为体扩散，可能会加剧武装冲突和执法中的安全风险。
- 在民事领域，各国和各区域制定了新的法律和管制制度，以解决人们的关切，即在不诉诸人类解释的情况下伤害人民的自主决定有可能损害人类尊严。

- 国际人道法中的马滕斯条款承认，法律可以根据社会关切和公众良心的要求而发展，因此与自动武器系统问题特别相关。
- 人工智能有可能从根本上改变我们与技术的关系，无论是好是坏。自主武器系统的挑战也是一个机会，通过管制，为人工智能在社会中的作用建立共同的规范和护栏，以防止最负面的结果。
- 我们有责任采取行动，制定保护人类所需的规则。

这几点强调了我们所处的关键时刻。人类正处于十字路口。

这是我们这一代人的“奥本海默时刻”，地缘政治紧张局势可能导致重大科学突破走上一条对人类未来非常危险的道路。我们必须听取专家们的警告，表现出这一挑战要求我们具备的政治领导力和远见。

国际局势困难的事实并不能免除我们应对自主武器系统挑战的政治责任。这要求我们在国家和区域机构、联合国实体、国际组织、民间社会、学术界、科技部门和工业界建立伙伴关系。

奥地利将向联合国秘书长提交这份总结，供其审议 2023 年联合国大会关于“致命自主武器系统”的决议所授权的报告。我们敦促所有国家和利益攸关方利用该决议提供的机会，向联合国秘书长提交意见。我们鼓励所有国家和利益攸关方考虑将这些观点纳入其观点。我们还请各国赞同该总结。

本主席摘要申明，我们坚定地承诺与所有感兴趣的利益攸关方一道，加紧努力，制定一项管制自主武器系统的国际法律文书。

保加利亚

[原件：英文]

[2024 年 5 月 25 日]

保加利亚欢迎有机会根据大会 2023 年 12 月 22 日通过的关于致命自主武器系统的第 78/241 号决议，提交其意见供联合国秘书长审议，该决议请秘书长就“如何应对此类系统从人道主义、法律、安全、技术和伦理角度提出的相关挑战和引起的关切，以及人在使用武力方面的作用。”

保加利亚一贯在《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》的框架内，参与审议致命自主武器系统领域的新兴技术。自 2017 年成立致命自主武器系统问题政府专家组以来，保加利亚一直积极和建设性地参与讨论，根据技术、军事影响以及法律和伦理考量，分享其对自主武器系统相关问题和概念的看法和想法。

保加利亚清楚地认识到澄清自主武器系统国际管制的紧迫性，于 2023 年 10 月加入成为大会关于致命自主武器系统的第 78/241 号决议的共同提案国，绝大多数国家通过该决议强调国际社会迫切需要应对自主武器系统构成的挑战。

一. 如何应对此类系统从人道主义、法律、安全、技术和道德角度提出的相关挑战和引起的关切

保加利亚坚决支持两级办法，认为这种办法提供了一个合理和有利的框架，应对此类系统从人道主义、法律、安全、技术和道德角度提出的相关挑战和引起的关切。

两级方法要求区分：(a) 完全不受人类控制的自主武器系统和负责任的指挥系统；(b) 具有自主功能的自主武器系统，需要制定条例，以确保这类系统遵守国际法，更具体地说，遵守国际人道法。

国际人道法的适用是两级办法的核心，其目的应是禁止不能确保遵守这种法律的武器系统。另一方面，对自主武器系统的管制要求其开发和使用过程完全符合国际人道法及其关键原则，特别是攻击中的区分原则、相称原则和预防原则。与此同时，这种管制应包括在自主武器系统的整个生命周期内采用和执行某些政策和措施。

保加利亚与丹麦、法国、德国、意大利、卢森堡和挪威一道，提交了关于在两级办法基础上可能采取的具体措施和政策类型的指导意见。¹

二. 人类的作用

人的控制是遵守国际人道法和自主武器系统的伦理可接受性的核心。为确保此类武器系统的使用完全符合国际人道法及其主要原则，并避免任何问责漏洞，必须在自主武器系统的研发、验证、部署和使用的各个阶段，主要是在瞄准周期，保持人的控制。

在自主武器系统生命周期的不同阶段，人的控制可以以不同的方式进行。为保证遵守国际人道法而对自主系统进行的人的控制程度和类型取决于作战环境的复杂性、武器系统本身的复杂性及其能力，以及武器系统的预定用途和需要执行的任务。

选择目标的过程需要对冲突环境和要实现的军事目标进行复杂的评估，以核实攻击是否符合国际人道法的规范和原则。在授权使用武力之前，必须确保自主武器系统服从负责指挥系统中的上级军事当局。

适用合法使用自主武器系统的要求并在授权之前评估此类系统使用武力的合法性是一个复杂的过程，需要人类：

- 对使用中的系统的能力和功能有深入和透彻的了解，对自主武器系统预期运作的冲突环境有适当的态势感知，对军事目标有足够的情报，从而提高系统性能和行动的可预测性和可靠性。

¹ 保加利亚、丹麦、法国、德国、意大利、卢森堡和挪威提交的工作文件，可参阅 [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2024\)/CCW-GGE.1-2024-WP3.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/CCW-GGE.1-2024-WP3.pdf)。

- 在每次任务和行动之前，核准和验证接战规则，规定时间和范围限制，并根据具体行动规划过程中详细说明了国际人道法规范和要求，界定有控制和有比例地使用军事力量。在任何情况下，使用武力都必须受到控制，并限于指挥官的最初意图和实现军事目标所需的效果。

三. 两级办法的运作

保加利亚认为，《特定常规武器公约》提供了一个适当的国际框架，供广大利益攸关方就致命自主武器系统领域新兴技术的各种层面的问题和概念，如技术、军事影响、国际法和伦理进行讨论、交换意见和分享专门知识。

作为《特定常规武器公约》2024年的主席，保加利亚致力于作出一切必要努力，促进政府专家组的工作，推动关于管制自主武器系统的审议，并履行其进一步审议和以协商一致方式制定一套文书要素的任务。

在这项努力中，政府专家组得益于几年来深入、健全和有充分根据的讨论，以及以前通过的报告和几十份工作文件，这些文件鼓励各国在共同理解的基础上再接再厉，就基本的实质性问题寻求更多的共同点。基于禁止和管制的两级方法的趋同可被视为专家组近年来工作的成就之一。

保加利亚将继续在政府专家组内开展积极主动和建设性的工作，努力促进在禁止和管制的基础上进一步统一两级办法。在一项文书中澄清现行法律中两级办法的内容及其对自主武器系统的适用，是有效管制的一个条件。²

关于致命性自主武器系统文书的形式，保加利亚随时准备支持在两级办法的基础上开始谈判和通过《特定常规武器公约》附加议定书(六)。

加拿大

[原件：英文]

[2024年5月24日]

大会在2023年12月22日通过的关于致命自主武器系统的第78/241号决议中请会员国和观察员国提供对致命自主武器系统的意见，内容特别包括如何应对此类系统从人道主义、法律、安全、技术和伦理角度提出的相关挑战和引起的关切，以及人在使用武力方面的作用，并向大会第七十九届会议提交一份反映收到的所有意见的实务报告，供会员国进一步讨论；

在大会第七十八届会议上，加拿大欣然地支持了第78/241号决议，并看到该决议获得了另外163个国家的支持。我们认为，这项决议是在大会第七十七届会议上发表的关于致命自主武器系统的联合声明的适当进展，加拿大也支持该联合声明。

² 致命自主武器系统问题政府专家组任务规定下的文书的综合提案，见 [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2024\)/CCW-GGE.1-2024-WP3.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/CCW-GGE.1-2024-WP3.pdf)。

加拿大赞赏第 78/241 号决议的措辞通过在日内瓦和纽约的非正式协商得到完善，并认为该决议可以进一步阐明算法偏见的重要性，例如在致命自主武器系统的开发阶段对性别的偏见。

加拿大赞赏该决议强调了遵守国际人道法的必要性，并加强了双轨办法(管制和禁止)，该办法主要是在根据《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》设立的致命自主武器系统问题政府专家组内讨论的，加拿大积极参与了该专家组的工作。加拿大希望看到专家组充分考虑第 78/241 号决议，因为该决议强调专家组是“处理与致命自主武器系统有关的各种问题的中心和独特论坛”。

大会第 78/241 号决议也代表了在联合国范围内讨论致命自主武器系统的广泛兴趣，即将发表的文件《未来契约》中提到自主武器系统进一步强调了这一点。该决议认识到新技术和新兴技术为保护冲突中平民带来的巨大希望，同时就致命自主武器系统对全球安全以及区域和国际稳定的潜在负面后果和影响表示关切。加拿大同意，在今后的讨论中应考虑到这些更广泛的地缘政治因素。

技术的发展速度要求在关于致命自主武器系统的国际政策讨论中取得相应的进展。本答复提供了加拿大的意见，其基础是其他国家本着合作和推动关于这一专题的国际讨论的精神所做的备受赞赏的工作。需要进一步讨论，以审查致命自主武器系统等新兴技术带来的潜在挑战，特别是人类在使用致命武力方面的作用、人机互动以及适用于致命自主武器系统领域新兴技术的开发、部署和使用的规则和原则。

加拿大认识到，处理可能的禁止问题有各种选择，包括可能制定一项具有法律约束力的文书。尽管如此，现阶段尚不清楚一项新的文书将试图填补现有国际框架中的哪些空白，因此也不清楚一项新的条约可能会带来什么结果。

加拿大认为，有关致命自主武器系统的所有讨论都必须立足于确保遵守现有国际法，包括国际人道法。这意味着武器系统必须始终保持一定程度的人的参与，问责和责任必须由人承担。在没有一定程度的人为控制和问责的情况下可以运作的武器系统不符合国际人道法。

作为下一步，应寻求共识，就武器系统遵守国际人道法需要什么样的“人的参与”达成共识。通过进一步讨论和完善对这一概念的解释，各国可以就“完全自主的武器系统”(即缺乏适当的人的参与，因此不符合国际人道法)与伴有适当的人的参与的自主因此能够保持符合国际人道法的武器系统之间的界线应划在哪里达成一致意见。

加拿大认为，这一概念与政府专家组制定的 11 项指导原则中的指导原则(c)密切相关。在使用武力时必须保持适当的人的参与。此外，《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》(第一议定书)明确规定，武装袭击核设施违反国际人道法。加拿大首先通过对所有新武器进行国家法律审查，确保武器系统符合加拿大的国际法律义务，来确保这一必要因素。我们还确保在武器的整个生命周期内严格遵守国际人道法。

另一个需要审查的领域是，如何将第三十六条的审查与确保在使用武力时保持适当的人员参与联系起来。从加拿大的角度来看，使用武力的责任在于人，而不是机器。将人的参与从完全自主武器系统生命周期的一部分中排除，将使决策责任难以追查，这也是加拿大反对这种系统的原因之一。军事指挥官要对在其指挥下发生的所有使用武力行为负责，无论使用武力的是其下属还是机器。我们认为“参与”一词既包括人的判断，也包括人的控制；我们认为，武装冲突从根本上说仍应是人类的事业。分享良好做法将是这方面合乎逻辑的第一步。

各国还应进一步审查致命自主武器系统的可预测性和可靠性概念。各国似乎对如何将可预测性措施适用于这类系统有不同的解释。在使用现有常规武器的情况下，军队必须合理和客观地预测使用武器会造成什么后果。尽管如此，负责的军队采取一系列方法来提高可预测性和减少附带损害，以确保武器系统的使用只符合国际法。

加拿大认为，武器系统的自主性必须源于高度的可信度，以便不可预测性足够低，并且可以通过类似于使用常规武器时管理的适当措施来减轻任何风险。加拿大将支持广泛的协商一致原则，其中包括：

- 禁止基于致命自主武器系统领域新兴技术的武器系统，这些武器系统不能按照国际人道法使用。
- 为了遵守国际人道法，致命自主武器系统领域的新兴技术必须保持适当的人的参与水平。
- 应更新国家政策、理论、指令和程序，以确保新技术的使用符合适用的国际法。
- 利益攸关方应参与自愿交流信息，介绍对致命自主武器系统领域的新武器和新兴技术进行法律审查的情况。

加拿大的一个主要关切仍然是，在武器系统自主功能的开发和规划中，可能存在无意或有意的偏见。我们感到关切的是，完全自主的武器系统可能不符合与妇女与和平与安全议程有关的原则。在加拿大与土著和民间社会伙伴就其注重女权外交政策举行的对话中，与会者提出了一些与致命自主武器系统有关的关切，包括对冲突地区妇女和儿童的附带伤害问题，以及自主武器系统可能加剧现有权力不平衡和偏见的风险。

此外，许多国家表示希望讨论围绕武器系统自动化的道德和伦理影响。加拿大致力于推动关于致命自主武器系统的国际讨论，并考虑到所有这些关切。

最后，加拿大谨感谢民间社会和国际组织为帮助推动关于致命自主武器系统的讨论所作的宝贵贡献(包括研究论文、数据介绍和发言)。

智利、哥伦比亚、哥斯达黎加、多米尼加共和国、厄瓜多尔、萨尔瓦多、危地马拉、哈萨克斯坦、尼日利亚、巴拿马、秘鲁、菲律宾、塞拉利昂和巴勒斯坦国

[原件：英文]
[2024 年 5 月 24 日]

概述

鉴于《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》(通常称为《特定常规武器公约》)的目标和宗旨，其中包括“继续编纂和逐渐发展适用于武装冲突的国际法规则”，由 16 个想法相同的《公约》缔约国组成的小组继续呼吁所有缔约国在致命自主武器系统问题政府专家组会议期间迅速应对自主武器系统构成的风险和挑战。¹

16 国集团重申，自主武器系统的出现及其对不断变化的战争格局的影响要求制定具有法律约束力的规则和原则，包括关于在没有有效的人为控制或不遵守国际法的情况下开发、生产、拥有、获取、部署、转让和使用自主武器系统的条例和禁令。

在此背景下，13 个代表团首次提出并向 2022 年 7 月政府专家组会议提交了关于自主武器系统的议定书六草案。2023 年 5 月，13 个代表团在小组第二届年度会议期间提交了一份经修订的议定书六草案，其中考虑到 2023 年 3 月举行的会议的讨论和建议，进行了一些更新和修订。² 此后，13 个代表团的小组又得到了另外 3 个代表团的支持，即智利、哥伦比亚和多米尼加共和国。

结构

议定书六草案旨在解决自主武器系统引起的伦理、法律、人道主义和安全关切。它载有关于自主武器系统的禁令和条例，可作为谈判一项具有法律约束力的文书的基础。议定书六草案还强调，如果自主武器系统的自主功能被设计用于在人类无法有效控制的情况下进行攻击，则应禁止这种武器系统。³

议定书草案包括序言和 8 项条款：(a) 一般规定；(b) 特性描述；(c) 禁止事项；(d) 法规；(e) 审查武器；(f) 风险缓解；(g) 遵守情况；(h) 缔约方的协商。这些条款突出说明以下几点：

¹ 书面评论呼吁就自主武器系统制定一项具有法律约束力的文书，可查阅 <https://documents.unoda.org/wp-content/uploads/2022/05/20220309-G13-Written-Submission.docx>。

² 更具体地描述人实际控制，纳入了致命自主武器系统问题政府专家组 2023 年 3 月会议上表达的立场；改进关于禁止在人实际控制门槛之外的系统的措辞，反映了在同一次会议上交换的意见；专门用于法规的附加部分(从而将禁止与管制分开)纳入了以美国为首的共同提案国在同一次会议上提出的条款草案内容。

³ 议定书六草案，第 3 条，可查阅 [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_WP.6_2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.6_2.pdf)。

- 首先，该提案提供了“自主武器系统”和“人实际控制”的简单描述。
- 议定书六草案规定了一系列禁止设计、开发、生产、拥有、获取、部署、转让或使用无法在人实际控制下使用的自主武器系统，包括不能以无法预测、解释、预期、理解或追踪的方式操作的自主武器系统。
- 该议定书还制定了管制措施，以确保对自动武器系统进行有效的人工监督、干预和停用。它强调人类限制目标类型、持续时间、地理范围和使用规模的能力。它规定必须制定明确的程序，让人类了解并有权控制自主武器系统。严格的测试和对数据处理复杂性的限制也是为了保证可理解性、可解释性和可预测性。
- 最后，议定书草案载有一个定期审查和修正机制，以确保议定书在不断变化的技术和安全环境中保持相关性和有效性。

议定书六草案是一个重要步骤，它为解决自主武器系统带来的挑战提出了前进的道路，并为开始谈判一项关于这一问题的具有法律约束力的文书提出了一个建议蓝图。议定书请所有有关各方进行建设性对话，以进一步发展和加强这一提议。

建议

(a) 认识到需要新的具有法律约束力的规则和原则来防范自主武器系统带来的风险和挑战，因为人们越来越普遍地认识到，现行国际人道法并不能解决这类系统提出的人道主义、法律和伦理问题；

(b) 申明决定使用武力的责任和问责必须由人来承担，因为这种责任和问责不能转移给机器；

(c) 申明基于情境的人的判断和控制对于确保自主武器系统的使用符合国际法，特别是国际人道法是至关重要的。因此，人实际控制必须确保人能够对攻击效果的可接受性做出道德和法律判断。它还必须确保人类使用者对攻击的影响负有法律和道德责任。

(d) 呼吁尽快启动关于具有法律约束力的条约的谈判，以管制自主武器系统。上述谈判进程应具有包容性，涉及所有利益攸关方，并保持多学科的做法，纳入国际人权法、《联合国宪章》规定的义务、其他裁军公约、国际刑法、伦理考量、国际武器贸易条例和环境条例等要素。

中国

[原件：中文]

[2024年5月23日]

“致命性自主武器系统”(LAWS)是科技发展和新军事革命的产物，既关乎国防安全，也事关人道主义。国际社会对 LAWS 问题的重视和理解程度不断上升。从维护全人类共同价值和利益出发，有效管控人工智能引发的安全、法律、伦理、人道主义等风险，是各国安全与发展的需要和义不容辞的责任。

中国始终以建设性态度参与相关国际讨论和治理进程,积极贡献治理方案。2017 年以来,中国参加了《特定常规武器公约》(CCW)框架下历次 LAWS 政府专家组会议,在 CCW 平台提出了《中国关于规范人工智能军事应用的立场文件》。2023 年 10 月,中国国家主席习近平宣布提出《全球人工智能治理倡议》。《倡议》强调,发展人工智能应遵守适用的国际法,各国尤其是大国对在军事领域研发和使用人工智能技术应该采取慎重负责的态度,要确保人工智能始终处于人类控制之下。

中国认为,各方应从防止新军备竞赛的角度出发,遵循平等安全、共同安全与普遍安全的原则处理 LAWS 问题,反对借助 LAWS 谋求绝对军事优势和霸权。中国主张:

一、《特定常规武器公约》(CCW)平台是讨论 LAWS 问题的合适场所。在国际安全领域坚持协商一致原则十分重要。CCW 框架下讨论持续取得进展,LAWS 相关政府专家组推动达成了“11 条指导原则”等重要成果,发挥了建设性作用。各方立场在诸多议题上出现一定程度趋同,2023 年专家组达成实质性报告,体现出 CCW 平台的独特价值。在此背景下,贸然“另起炉灶”推动到其他平台讨论 LAWS,恐冲击现有进程,扩大国家分歧,使多年凝聚的成果和共识付诸东流。

二、定义特征问题是制定任何切实可行管控措施的基础和关键。LAWS 涉及的问题非常复杂,讨论角度越是发散,各方的关切与分歧就会越多,越难以取得实质性进展。应优先讨论定义,争取先就定义要素、技术特征等达成一致理解,在此基础上有针对性地制定措施,谈判制定具有法律约束力的国际文书。不应脱离“致命性”讨论 LAWS,否则将背离人道主义军控的初衷,混淆 LAWS 与“自主武器系统”(AWS)的概念,使问题复杂化,进一步迟滞讨论进程。

三、应结合技术发展趋势,对 LAWS 所涉法律适用等问题充分研究论证。《日内瓦公约》及其两个《附加议定书》等国际人道法完全适用于包括 LAWS 在内所有武器系统。但当前技术水平下,现有国际人道法是否足够应对 LAWS 引发的挑战仍存在很大的不确定性,比如 LAWS 是否具有在战场环境中的区分能力和作出比例性决定的能力等。鉴此,各国应开展负责任的创新,并采取有效预防性措施,使致命性自主武器系统领域新技术的利用最大限度地减少对平民伤害,践行人道主义宗旨和目标。

四、“分类分级”管理是应对 LAWS 问题的合理路径。笼统地推出禁止或限制措施或将损害各国正当国防能力,甚至损害各国和平利用相关技术的权利。中方认为,应考虑从“可接受的”和“不可接受的”等角度对 LAWS 进行分类,对“不可接受的”部分可考虑效仿《关于激光致盲武器的议定书》,达成法律文书加以禁止;对“可接受的”部分采取必要的风险减缓措施。中方曾就“不可接受的 LAWS”提出五大特征,中方愿与各方继续探讨。

中国将继续秉持开放、包容、互鉴精神,在 CCW 框架下同各方保持沟通交流,推动就 LAWS 问题形成更多共识,共同维护和促进国际和平安全。

哥斯达黎加

[原件：西班牙文]

[2024年5月24日]

作为一个中立国家和非武装民主国家，哥斯达黎加重申致力于全球和平与安全，尊重多边主义、《联合国宪章》和国际法。这一承诺体现在该国担任2023年拉丁美洲和加勒比自主武器社会和和人道主义影响会议的东道国，并积极参与发起大会关于致命自主武器系统的第78/241号决议的核心小组。为响应秘书长的呼吁，哥斯达黎加在此表明其支持谈判一项具有法律约束力的文书以禁止自主武器系统的立场。

需要一项具有法律约束力的文书

“自主武器系统”一词比“致命自主武器系统”更受欢迎，因为红十字国际委员会和一些国家都认为，致命性取决于武器系统的使用，而不是其设计。这类系统提出了重要的伦理和法律问题，特别是因为它们无法像人类一样可以根据具体情况作出决定的能力有限。这些行动必须符合国际法，包括国际人道法和国际人权法。

目前的国际监管框架并不能完全解决自主武器系统带来的法律挑战。因此，哥斯达黎加主张全面禁止其关键功能缺乏人实际控制的自主武器系统，包括其功能无法预测、解释、预期、理解或跟踪的系统。

哥斯达黎加支持执行一项具有法律约束力的文书，该文书应规范自主武器系统，采用多学科方法，并纳入国际法规则，包括国际人道法和国际人权法、国际刑法和《武器贸易条约》等其他文书，以及商定的伦理考量。这一办法力求确保在所有阶段都有人实际控制，并提高透明度和可解释性，以消除这类系统运作中的偏见，避免有系统地复制对历史上弱势群体歧视的评估。

哥斯达黎加对自主武器系统的考虑

a. 人实际控制作为中心支柱

哥斯达黎加认为，不允许维持人的能动性，包括保留人对使用武力的判断和干预的自主武器系统不符合国际法原则。

例如，确保人实际控制需要能够根据环境调整目标，在必要时停用系统并限制系统的自主性。此外，应规范目标的范围、选择和区别，同时考虑到与时间和空间有关的方面。必须能够理解和解释系统的运作，以维护法律要求和问责，追究责任并追溯解释其行动。

b. 分责和问责

哥斯达黎加关切地注意到，在使用自主武器系统方面的分责和问责不明确，因为这可能削弱责任归属，并助长违反国际法，包括国际人道法和国际人权法的行为不受惩罚的现象。必须建立明确的机制，确保这些系统作出的决定和采取的行动可以追踪，并可归因于人类决策者，从而确保问责和对国际规范的尊重。

c. 非国家行为体获取和使用自主武器系统

哥斯达黎加特别关切非国家行为体获取和使用自主武器系统的后果。在国际和区域一级，各国共同努力打击贩毒和有组织犯罪，谴责助长恐怖主义和准军事团体的活动。然而，这种系统的发展可能为犯罪集团提供使其武库多样化的新手段，对区域稳定与安全构成重大威胁。这种系统可以用来保护毒品运输，控制领土，攻击警察和紧急部队，甚至平民。

这些系统所需的专业化可能导致网络犯罪集团的扩散，因为自主武器系统容易受到黑客攻击。这可能导致操纵其已列入方案的功能和目标，使其行为难以预测，并助长破坏区域稳定和国际社会促进安全与和平努力的犯罪活动。

d. 推进技术促进和平

自主武器系统的扩散威胁全球和平，哥斯达黎加对此类系统引发武装冲突的可能性深感担忧。这种系统使战争能够在没有人类军事干预的情况下进行，这种现象可能导致违反国际法的任意攻击。此外，这种系统的发展引发了军备竞赛，违反了《联合国宪章》和促进和平的重要国际条约所倡导的裁军和不扩散原则。

哥斯达黎加主张为保健、教育和气候变化行动等和平目的利用新兴技术，反对开发助长武装冲突的技术。

e. 伦理考量

将生死决定权交给自主武器系统是不可接受的，因为这样做会使人类沦为单纯数据点，并违背了基本的伦理原则。哥斯达黎加促进尊重人的尊严和普遍权利的技术发展。

结论

- 必须在联合国框架内确定一项不限成员名额的任务授权，以启动关于自主武器系统的具有法律约束力的国际文书的谈判。
- 关于这一文书的谈判应侧重于禁止和管制，以确保人实际控制，从而避免战争进一步非人化，并确保个人问责和国家责任，从而促进遵守国际法，包括国际人道法和国际人权法，解决伦理问题，防止自主武器系统的社会和人道主义影响。
- 哥斯达黎加强调，各国都对自主武器系统构成的风险和挑战感到关切，并强调，鉴于其在伦理、法律、人道主义和安全方面的影响，这种系统必须是可预测的、可理解的和可解释的。
- 应该从多学科的角度处理这一问题，确保尊重国际法的基本原则，纳入有关可预测性、相称性和可解释性的明确标准，以及伦理考量和环境条例等。
- 谈判进程必须具有包容性，确保所有有关国家的真正代表性，并允许包括民间社会、学术界和私营部门在内的所有利益攸关方参与。

拉丁美洲和加勒比关于自主武器的社会和人道主义影响会议公报

a. 新兴技术对国际和平与安全构成了具体挑战，并对人类在战争中的作用提出了新问题。需要全面了解武器系统自主性的影响及其对人实际控制的影响，以充分评估伦理、法律、人道主义和安全方面的影响。

b. 随着新武器技术的不断开发和在武装冲突中的使用，有必要在国际法，包括国际人道法和国际人权法的指导下，根据人道原则和公众良心的要求，制定新的禁令和条例：

c. 根据国际法，包括《联合国宪章》所载的宗旨和原则、国际人权法和国际人道法，呼吁国际社会通过发展和加强国际法律框架来应对这些威胁，以防止流离失所、平民伤亡和民用物体被毁的影响。

d. 至关重要的是，要保持人实际控制，以防止战争进一步非人化，并确保个人问责和国家责任。

e. 拉丁美洲和加勒比国家在《特定常规武器公约》框架内积极参加了致命自主武器系统领域新兴技术问题政府专家组的讨论。有一批国家提出了未来规范性框架的内容和《特定常规武器公约》框架下的议定书六草案。本地区的一些国家也加入了 70 个国家在 2022 年 10 月联合国大会上发表的关于自主武器的声明。

f. 此外，2023 年 1 月 24 日在拉丁美洲和加勒比国家共同体第七届国家元首和政府首脑会议上发表的《布宜诺斯艾利斯宣言》强调了 2014 年 1 月在哈瓦那举行的第二届国家元首和政府首脑会议上签署的《宣布拉丁美洲和加勒比为和平区宣言》。

g. 联合国秘书长、联合国法外处决、即决处决或任意处决问题特别报告员、残疾人权利问题特别报告员、联合国当代形式种族主义问题特别报告员、红十字国际委员会、民间社会，科学界和学术界对武器系统的自主性表示关切，并一再呼吁各国谈判一项关于这方面问题的具有法律约束力文书。

h. 该区域的代表认识到武器系统自主性带来的风险和挑战，并认识到有必要制定一项具有法律约束力的国际文书，规定禁止和管制措施，以保障人实际控制。

拉丁美洲和加勒比国家的代表承认需要采取进一步行动，以便：

1. 开展合作，促进紧急谈判一项具有法律约束力的国际文书，其中载有关于武器系统自主性的禁令和规定，以确保遵守国际法，包括国际人道法和国际人权法，以及伦理观点，并防止武器系统自主性造成的社会和人道主义影响；

2. 承诺积极参与，以推进共同立场并推进此类谈判；

3. 继续并加强与国际组织、民间社会组织、学术界和其他相关利益攸关方的合作和伙伴关系，以利用其相关专门知识和支持；

4. 各国之间继续进行讨论，以拟订可在《特定常规武器公约》和其他多边论坛框架内推广的可能建议。

哥斯达黎加埃雷迪亚，贝伦河畔

2023年2月23日和24日

古巴

[原件：西班牙文]

[2024年5月25日]

古巴支持紧急通过一项具有法律约束力的国际文书，禁止制造、拥有和使用完全自主的武器，并管制包括无人驾驶飞机在内的半自主武器的使用。

在《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》下就致命自主武器系统进行的讨论考虑到了对《公约》缔约国而言十分重要的法律、伦理、人道主义和安全层面。

我们注意到2018年至2024年举行的致命自主武器系统问题政府专家组会议的成果。

然而，我们认为，讨论应有助于在达成一项具有法律约束力的文书方面取得进展，该文书不仅包括禁止完全自主武器系统和管制半自主武器，而且还包括在使用武力时保持人实际控制的总体要求，并禁止所有不可预测或不遵守具有约束力的义务时阻止追究责任的系统。我们希望，政府专家组在获得为一项文书拟订一套内容的新授权后，将在这方面取得切实成果。

使用自主致命武器不能确保遵守或符合国际法规则和原则，特别是国际人道法，包括其区分原则和相称原则。

使用完全自主武器将妨碍有效评估国家对国际不法行为的责任。自主武器的性质本身使得在发生涉及使用这类武器的非法行为时，很难或无法将责任归于国家或个人。机器无法取代人类在战争中做出最重要的决定，在任何情况下都无法做出关乎人类生命的决定。

自主性和杀伤力水平是指导自主武器禁止或监管的关键因素：自主性和杀伤力越大，监管框架就越严格。

埃及

[原件：英文]

[2024年5月23日]

完全自主的武器对国际和平与安全构成严重威胁，并提出了必须解决的若干人道主义、法律和伦理关切，以确保此类武器符合国际法和国际人道法，特别是人道原则、必要性原则、相称性原则和区分原则等基本原则。

联合国及其裁军机制是制定必要的国际规则和规范框架的唯一有效和包容性平台，特别是因为技术发展的速度继续远远超过国际一级的必要管制。

埃及高度重视推动联合国在这方面的努力，并坚决支持大会第 78/241 号决议，该决议是朝着在自主武器系统问题上促进多边主义迈出的良好的第一步，并为联合国所有会员国和观察员国直接参与确定如何应对相关挑战和关切提供了平等机会。

同样，埃及欢迎联合国秘书长在《新和平纲领》和未来首脑会议的范围内编写的政策简报中对这一问题的重视，并支持在首脑会议的结果中列入关于这一重要议题的明确指导。

埃及还认为，为裁军审议委员会 2024-2027 年周期建议的第二个议程项目是一个很好的机会，可以提出商定的原则和切实可行的建议，以建立必要的规范性框架。裁军审议委员会是联合国裁军机制的审议支柱，在委员会本周期重点讨论这一专题，将是推进在《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》所设政府专家组框架内进行的讨论的极好机会。

虽然埃及承认上述致命自主武器系统问题政府专家组内的相关讨论的重要性，但它感到遗憾的是，进展仍然很小，没有取得任何具体成果。这主要是由于一些国家继续错误地认为，可以保持在这一领域的绝对主导地位或优势，因此它们抵制任何旨在制定规则的努力，这些规则将限制或禁止恶意使用不符合国际人道法的这类技术，或制止这类技术的过度武器化。

在这方面，埃及关于自主武器系统的立场的主要内容如下：

(a) 埃及认为，采取两级办法，即禁止完全自主的武器和管制人工智能的其他军事应用，是最现实和有效的行动方针。

致命自主武器系统的迅速发展令人十分震惊，需要采取紧急行动，包括按照联合国秘书长在其《新和平纲领》提案中的建议，通过具有法律约束力的规定禁止完全自主武器。

当然，这种禁止规定必须以“完全自主的致命武器”一词的明确定义为基础，专门针对那些可以在没有任何的人实际控制或干预或没有有效核查的情况下运作的武器。

(b) 关于制定人工智能军事应用准则和规则的多项国际举措表明，国际社会对相关风险的认识日益提高。然而，显然需要精简这些举措，并将其置于联合国

的保护伞之下，以确保其包容性和有效性。此外，此类举措必须避免引入任何武断的国际监督机制或实施任何类型的歧视性出口管制。最重要的是，这些举措不应被视为取代缔结一项具有法律约束力的禁止完全自主武器的目标。

(c) 人的责任原则的中心地位怎么强调都不过分。一种能够自行启动、自主选择 and 攻击目标并在没有人的直接控制或监督下夺去人的生命的武器所带来的风险和关切是完全不可接受的。因此，自主武器系统必须继续处于人实际控制和监督之下，以确保从国际法，包括国际人道法的角度对人类负责和问责。

无论是哪种类型的武器系统，把杀人的决定权交给机器都是不符合伦理的，也是对国际人道法的严重违反。

即使一个算法可以被编程来确定什么是国际人道法下的合法行为，它也永远不能被编程来确定什么是符合伦理行为。还需要确保国家对其使用自主武器造成的非法或不法行为承担责任。

此外，有必要确保在军事领域使用人工智能技术和机器学习技术的所有软件、算法和设计在其整个生命周期内仍然受到人为修改和可解释性原则的约束。

(d) 埃及主张采取一种平衡和现实的方法，通过这种方法，我们都可以继续受益于人工智能应用提供的新机会，同时以现实和有效的方式应对相关挑战。在这方面，我们强调必须建设发展中国家的能力，缩小这一领域的巨大差距。

斐济

[原件：英文]

[2024 年 5 月 25 日]

斐济共和国政府欢迎有机会就大会 2023 年 12 月 22 日通过的关于致命自主武器系统的第 78/241 号决议提交意见和建议，供联合国秘书长审议。该决议请各国和利益攸关方就如何应对此类系统从人道主义、法律、安全、技术和伦理角度提出的相关挑战和引起的关切提供意见。

挑战和关切

作为一个太平洋小岛屿发展中国家，我们对与自主武器有关的严重伦理、道德、法律、问责、环境和安全问题深感关切，特别是因为此类问题影响到全球南方和太平洋国家。自主武器系统的开发和使用构成必须紧急应对的重大风险。

自主武器系统既危及平民，也危及战斗人员，有可能以不可预测的方式使冲突升级。此类系统对遵守包括国际人道法在内的国际法提出了挑战。使用机器作出生死攸关的决定可能会破坏有关敌对行动和保护平民的规则。这种以机器程序取代人类判断的做法引发了基本的伦理问题。

此类武器显然有可能扩散到全球、被非国家行为体获取并用于犯罪活动。太平洋岛屿具有独特的脆弱性，因此认为这是一个重大关切问题。自主武器落入非国家行为体手中的风险尤其令人担忧，因为这可能导致不稳定和暴力加剧。

从历史上看，军事技术对环境造成了持久破坏。开发和自主武器可能会延续这一趋势。暴力的自动化减少了对武器使用后果的监督，可能导致重大环境危害。越来越多的证据表明，训练人工智能和机器学习模型会产生大量碳足迹。训练和操作自主武器系统的潜在能源成本令人担忧，研究表明，训练人工智能可能会大大加剧气候变化。当前，减少碳排放已成为全球优先事项，开发和自主武器可能会进一步破坏我们的地球。

自主武器系统中的算法偏见是一个重大问题，尤其是对历来被边缘化的群体而言。这些系统可能会延续种族、性别和其他偏见，导致对某些群体造成过度伤害。依靠传感器提供的数据来施加武力可能会将系统性偏见植入自主武器的决策过程。人工智能在警务和刑事判决等民用领域的应用证据表明，算法偏见对边缘化群体的影响尤为严重。

问责问题也至关重要。不能轻易将自主武器行动的责任归咎于操作人员，因为他们无法直接控制武器何时、何地以及对谁施加武力。在武装冲突的背景下，战争的迷雾已使问责变得更加复杂，而自主武器将进一步破坏追究暴力行为实施者责任的努力。这种缺乏问责的情况将使为受害者伸张正义变得更加困难。

此外，允许机器夺走人的生命会使个人失去人性，沦为由传感器和算法处理的数据点。这种将暴力机械化的做法有损于人类尊严和伦理原则。

应对挑战

为应对致命自主武器系统所带来的挑战，我们呼吁立即开始谈判，以制定一项关于自主武器的具有法律约束力的文书。该文书应：

- (a) 禁止以人为目标的武器系统；
- (b) 禁止无法在人实际控制下使用的武器系统；
- (c) 对所有其他武器系统进行管制，以确保对其进行有效控制、了解并限制其使用区域和持续时间；
- (d) 包括防止扩散和未经授权使用自主武器的条例；
- (e) 解决环境影响问题，确保军方对碳排放负责，并立即对环境和人类生命采取适当补救措施。

大会是进行这些谈判的理想论坛，可确保代表组成具有多样性和包容性。大会会员国包括来自世界各地的国家，这确保小国和发展中国家的意见得到倾听。与此同时，我们认识到，有必要在所有适当的多边论坛继续进行对话，因为这些论坛可以相互促进。现在应该超越致命自主武器系统问题政府专家组，转向一个目标更高、行动更快、更能包容非《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》缔约国的国家以及国际组织和民间社会的论坛。现有国际人道法不足以应对自主武器带来的挑战。

我们与 109 个国家一道，支持制定一项具有法律约束力的文书，以解决自主武器问题。联合国秘书长在其《新和平纲领》中紧急呼吁各国在 2026 年前通过

一项条约，禁止和管制自主武器系统。2023年10月，秘书长和红十字国际委员会主席米里亚娜·斯波利亚里茨呼吁各国“紧急制定关于自主武器系统的新国际规则，以保护人类”。

我们感谢有机会就应对人类面临的这一严重威胁分享我们的看法和建议。国际社会必须迅速果断地采取行动，建立强有力的法律框架，防止自主武器系统的扩散和滥用。保护人类尊严、确保追究责任和保护环境必须成为这些努力的重中之重，包括对人类生命和环境采取适当的补救措施。全球和平与安全的未来取决于我们采取集体行动，反对不受限制地开发和部署自主武器。

芬兰

[原件：英文]

[2024年5月23日]

芬兰谨提及2024年2月1日的普通照会，该照会涉及大会2023年12月22日通过的关于致命自主武器系统的第78/241号决议，其中大会请秘书长就“如何应对此类系统从人道主义、法律、安全、技术和道德角度提出的相关挑战和引起的关切，以及人在使用武力方面的作用”征求会员国的意见，并提出下文概述的国家意见。

通过关于人工智能和自主(包括致命自主武器系统)军事用途的国际原则或条例，对于确保遵守国际法、加强安全并减少潜在冲突风险至关重要。与此同时，有必要使符合国际法的国家自卫能力得以发展。

芬兰支持就原则或条例，如关于开发和部署致命自主武器系统的国际协定进行谈判。该文书可在政治上或法律上具有约束力。芬兰的目标是，在不预先判断其性质的前提下，与尽可能多的国家，包括开发、生产和使用此类武器的国家，缔结一项国际文书，并使该文书在实践中产生真正的影响。芬兰积极参与在《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》框架下开展工作的致命自主武器系统问题政府专家组的工作，并支持将其作为继续讨论未来文书的论坛。此外，确保安全和负责任地将人工智能和自主用于军事目的的其他举措也可能有所助益，但不得与政府专家组的目標相悖。

芬兰强调，在开发和部署致命自主武器系统时，必须遵守国际法，包括国际人道法。人类保留有关使用武力的决定权至关重要。现行国际法禁止不符合国际人道法规则及其相称、区分和预防的基本原则的致命自主武器系统，因此不应使用、获取或开发此类系统。然而，新技术的快速发展需要加强国际管制，以明确国际人道法如何适用于致命自主武器系统。管制对于开发和符合国际法的新技术也十分必要。

因此，应采用禁止和管制相结合的两级方法对致命自主武器系统加以管制。该办法力求：

(a) 禁止没有任何形式的人的参与且不受负责任的人的指挥系统约束的自主武器系统(即能够自行设定目标或修改和(或)执行这些目标，而无需人对其初始方案或任务框架进行验证的系统)；

(b) 对具有自主¹特征或功能的所有其他武器系统的开发和使用进行管制，确保其符合国际法规则和原则。该框架应包含相关规定，以保证有适当的人的参与水平和问责、测试和操作具有透明度和可预见性以及致命自主武器系统的开发、生产、采购和部署过程中进行法律审查并采取风险缓解措施。管制不应针对具体技术本身，而应针对这些技术的任何不符合国际人道法的应用。这也使管制能够经受住时间和武器技术发展的考验。

法国

[原件：英文]

[2024年5月23日]

法国认为，为充分应对致命自主武器系统提出的挑战和引起的关切，应坚持两级方法，明确区分无法持续保证遵守国际人道法的致命自主武器系统和其他致命自主武器系统。具体而言，法国认为：

(a) 应禁止根据国际人道法不得开发和使用的致命自主武器系统。属于该类别的武器系统包括：(一) 本质上具有滥杀滥伤性质的系统；(二) 影响无法加以限制、预测或控制的系统；(三) 可造成过度伤害或不必要痛苦的系统；(四) 完全不受人的控制和负责任的指挥系统约束的系统(即“完全”自主致命武器系统)；

(b) 对于军事指挥可在特定行动框架内分配与高级功能(识别、目标选择和交战)相关任务的计算和执行的致命自主武器系统(即“部分”自主致命武器系统)，应在其整个使用周期内通过采取适当措施对其进行管制，以减轻潜在的法律、伦理、技术和安全挑战。

这一方法可以从根本上确保国际人道法将继续完全适用于所有武器系统，包括致命自主武器系统的潜在开发和使用。该方法还考虑到人类在使用武力方面的作用，特别是保持人的控制以及对人类负责和问责的重要性。使用武力目前是而且必须始终是人的指挥和控制系统的固有责任，特别是在违反国际人道法的情况下。在这方面，人的指挥和控制系统必须保持就使用致命武力作出关键决定的能力。

¹ 自主应理解为以自给自足和自我管理的方式执行既定任务的能力。这包括对任务和所需子任务进行自我规划的自由。在评估致命自主武器系统是否符合国际人道法时，背景评估至关重要。使用致命自主武器系统必须有一个具体和有限的任务框架(目标设定、空间和时间限制、对任务参数的任何重大修改的人的批准)，并保证武器系统将以可预见的方式行动。如已通过其他方式确保遵守国际人道法，人实际参与并不总是需要持续的人机连接。

在确定需要何种及何种程度的人的控制时，应考虑一系列因素，包括整个武器系统的作战环境、特点和能力。法国认为，人充分控制要求：(a) 人类对武器的部署和使用作出知情决定；(b) 人类基于对潜在目标、拟使用的武器的能力和特性以及部署武器的作战环境的了解，有足够的信息来确保以符合国际法的方式使用武力。人的控制可采取各种形式，并可在武器生命周期的各个阶段实施。特别是，在国际人道法所要求的范围内，人类必须：

- 根据其角色和责任级别，充分了解此类武器系统的运作方式及其对环境的影响和可能与环境产生的相互作用，并能够预测和解释此类系统的行为。
- 根据具体情况和背景，确定和验证使用规则、交战规则和分配给此类系统的任务的精确框架，这些任务在时间、空间和确定的目标方面都受到限制。
- 维持在系统部署期间监测其可靠性和可用性以及批准对任务参数的任何重大修改的能力。
- 能够判断是否符合国际人道法的规则和原则，特别是攻击中的区分原则、相称原则和预防原则，从而对使用武力作出关键决定。

这种两级方法应作为在《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》框架内构建关于致命自主武器系统的未来文书的基础。《公约》作为国际人道法的一项重要文书，力求在军事必要和人道主义考虑之间取得平衡，因此它仍然是应对致命自主武器系统所引发的伦理和法律问题的最适当论坛。此外，《公约》汇集了从事人工智能军事应用开发的国家，其独特的结构能够开展谈判并通过更多的附加议定书，从而提供了应对新武器技术所需的灵活性。最后，由政府专家和国际组织、民间社会和学术界代表组成的致命自主武器系统问题政府专家组汇集了技术、军事、法律和外交方面的专门知识，从而为审议致命自主武器系统领域新兴技术的各个方面提供了一个独特的框架。

因此，法国认为，必须继续进行政府专家组的工作，全面完成其目前的任务，即制定一套文书要素，并在将于 2026 年举行的下一次《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》缔约国审议大会上提出其结论。法国认为，根据两级方法，该文书应包含以下要素：

(a) 序言部分，该部分除其他外，重申《特定常规武器公约》的目标和宗旨、缔约国 2019 年商定的 11 项指导原则以及国际人道法对此类系统的适用性；

(b) 适用范围，该范围对致命自主武器系统进行界定，排除不符合该定义的系统，并明确区分完全不受人的控制和负责的指挥系统约束的致命自主武器系统和具有自主高级功能的致命自主武器系统；

(c) 禁令，其中列出禁令所涉武器系统的特性并重申各国在任何时候都有责任遵守适用的国际法规定的义务；

(d) 条例，该部分对开发和具有高级功能决策自主的致命自主武器系统提出限制和要求。此类措施可包括法律审查；风险评估、保障和缓解措施(以确保此类系统的可靠性，防止故障、滥用、转用和放弃人类特权)；适当的培训(确保决策人员和操作人员充分了解系统的影响以及可能与环境产生的相互作用)；保持人的充分控制(以确保人类将始终界定和验证使用规则、接战规则和分配给该系统的任务的精确框架，并就使用武力作出关键决定)以及人的指挥和责任系统(以确保责任和问责)；

(e) 机制，以促进缔约国之间进行协商与合作，并鼓励它们在自愿基础上就为执行文书所载要求而采取的国家措施交流最佳做法。

关于致命自主武器系统文书在《特定常规武器公约》框架内可采取的形式，法国支持谈判和通过一项《公约》附加议定书。

此外，可在《特定常规武器公约》框架内设立独立技术专家委员会，定期向缔约国通报致命自主武器系统领域的新技术发展，帮助对这一本质上具有前瞻性且不断发展变化的问题保持高度警惕，并随时间推移视需要帮助各国在《公约》框架内采取更多适当措施，以应对尚未确定的具体挑战。

德国

[原件：英文]

[2024年5月23日]

一. 导言

与自主武器系统有关的问题和挑战是当今军备控制领域最紧迫和最复杂的问题之一。自主武器系统的开发和使用的潜在挑战包括：在武装冲突中使用此类系统时可能会产生意外后果；对于此类系统是否符合国际法，特别是国际人道法的要求存在质疑以及问责和对国际安全与稳定存在潜在负面影响的问题。与此同时，需要技术进步，包括两用技术方面的进步，以提高数据处理和决策的深度和速度。开发和部署具有自主功能的军事人工智能和武器系统固然有其益处，因为人类决策并非万无一失，但武器系统的自主也伴随着高风险，并且在未按照国际法，特别是国际人道法开发、列装和部署此类系统的情况下会构成人道主义挑战。

德国高度赞扬为促进关于致命自主武器系统的讨论所作的区域努力，例如哥斯达黎加、卢森堡、菲律宾、特立尼达和多巴哥以及塞拉利昂所作的努力。此外，关于“十字路口的人类：自主武器系统和管制挑战”的维也纳会议为交换意见和进一步寻求共识提供了宝贵的论坛。德国继续积极参与荷兰和大韩民国发起的“军事领域的负责任的人工智能”进程，并积极参与落实美国发起的《关于在军事上负责任地使用人工智能和自主的政治宣言》。德国认为，这些举措发挥了重要作用，因为它们深化了这一领域的讨论，并有助于制定负责任的行为规范。

二. 原则和工作假设

主要原则

德国的立场以致命自主武器系统问题政府专家组 2019 年确认的 11 项指导原则和北大西洋公约组织(北约)2021 年商定的以下负责任使用原则为基础：(a) 开发和使用人工智能应用的合法性；(b) 确保军事系统中人工智能的设计和运行可被问责的人的责任；军事领域人工智能应用的可解释性和可追溯性；人工智能和自主系统整个生命周期的可靠性、安全、保障和稳健性；适当的人机交互和减少偏见的可管理性。

致命自主武器系统

德国的理解是，致命自主武器系统是指一旦启动就能够识别、选择、跟踪目标并对其施加武力，同时其行动不受人的控制和负责任的指挥系统约束的系统。此外，此类系统不允许进一步的人为干预，能够自行设定目标或修改其初始方案或任务框架，而不可能进行任何人为验证。

在这种情况下，德国认为，某些自卫系统，如爱国者导弹系统可以而且必须能够在启动后以自动模式运作，无需人为干预，而不会构成任何与国际人道法有关的挑战或伦理挑战。此类系统一旦启动，就能在时间紧迫的情况下(如攻击火炮等速度极快的目标或在导弹防御情况下)自动探测、选择和攻击目标，无需进一步的人为干预。上述系统不应被视为自主系统，而应被视为自动系统，因为它们用于探测和攻击目标的算法为确定性算法，不同于自主系统基于概率计算的算法。

三. 人的控制是确保负责任地开发和具有自主功能的武器系统的关键要求

德国认为，上文定义的致命自主武器系统不符合国际法，特别是国际人道法的要求。

德国认为，在评估具有自主功能的武器系统的可接受性时，保持人的控制框架是关键要求。任何武器系统，包括具有自主功能的武器系统的开发和使用时，都必须以负责任的指挥系统和人的控制为中心。生死攸关的决定必须由人类做出。

所需人的控制水平取决于作战环境以及武器系统的特性和能力。可以通过人的控制框架来保持对此类系统的控制。

人的控制框架不仅包括人，还包括为决策做准备和提供支持的其他措施和要求(即设计中的控制和使用中的控制)。这一框架需要：

- 减少无意偏见的数据分析方法和具体措施；以人工智能为特征的算法以伦理规范为基础，以避免强化和加剧现有的不平等结构。
- 对参与使用武器的操作人员进行培训，并使他们有能力预测和解释所操作系统的行为，从而实现问责。
- 根据《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受难者的附加议定书》(《第一议定书》)第 36 条制定接战规则，规定指挥人员

和操作人员使用系统的范围，并制定武器测试程序，从而确保可预测性。

- 确保系统在行动持续时间和地理区域、攻击目标类型和行动规模等既定限制范围内运作的负责的人的指挥系统。
- 所部署系统的自毁和自失能等技术保障措施，以确保系统在允许的参数范围内运作。
- 对可能涉及违反国际人道法行为的事件进行报告的程序。

在此背景下，人机交互的设计至关重要，因为它需要确保武器系统始终服从于部署和操作这些系统的人类。人类必须始终能够控制武器系统，并且必须始终是这种交互中的基本要素，承担总体责任。

四. 前进道路

有效应对致命自主武器系统领域新兴技术相关挑战的两级方法

德国与越来越多的政府专家组成员一样，仍然相信两级方法是确保遵守上述原则的最适当方法。鉴于在这一方法的基础上取得了大量共识，德国支持在以下建议的基础上取得进一步进展：

- 首先，同意在政府专家组框架内制定一项具有法律约束力的文书，禁止开发、列装或部署不符合国际人道法并根据事实本身被禁止的致命自主武器系统。
- 其次，同意在政府专家组框架内制定一套关于具有自主功能的武器系统的条例，以确保在任何时候都保持人的控制。

《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》关于开发和^{使用致命自主武器系统的附加议定书}

德国支持制定一项《特定常规武器公约》附加议定书，禁止不受人的控制和负责人的指挥系统约束的致命自主武器系统。

作为包容性论坛的致命自主武器系统问题政府专家组

国际法，特别是国际人道法，以及相关的伦理观点指导着政府专家组的工作。专家组汇集了《特定常规武器公约》缔约国的专家，包括拥有技术先进武装部队的国家的专家。此外，还有来自学术界和民间社会的专家。德国高度赞赏这些专家对讨论作出的宝贵贡献。在专家组框架内进行的讨论取得了重要和实质性共识，并继续促进进一步建立共识。

德国相信，政府专家组的包容性做法对于取得有效成果十分重要，并将继续致力于积极支持专家组的工作。如法国和德国在 2021 年提交的联合工作文件中所述，为更深入地了解技术发展情况，可选择设立一个技术专家组。

与此同时，政府专家组要想履行其职责并取得切实成果，就不能继续拖延。因此，德国是大会 2023 年通过的关于致命自主武器系统的第 78/241 号决议的首批

共同提案国之一，并欢迎秘书长的报告，该报告考虑到了所有观点，对收集现有观点非常有价值。德国认为，这将大大有助于确定更多的共识领域，并重新推动在《特定常规武器公约》框架内进行的重要讨论。

希腊

[原件：英文]
[2024 年 5 月 23 日]

将人工智能和机器学习等新兴颠覆性技术纳入武器系统并加以使用，引发了政治、法律、伦理和伦理方面的严重关切。从政治角度看，开发和使用致命自主武器系统可能会降低武装冲突的交战门槛，从而影响相关决策。

完全自主武器系统是否符合国际人道法仍令人怀疑。特别是，此类武器是否符合国际人道法的基本原则和要求，如针对战场环境不断变化的目标区分、相称和预防措施的原则和要求，令人严重关切。

希腊支持许多国家在日内瓦致命自主武器系统问题政府专家组审议期间所强调的两级方法。

有鉴于此，我们要重申，希腊坚信，《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》仍然是继续讨论禁止和限制使用包括致命自主武器系统在内的常规武器问题的适当论坛。《公约》确保在人道主义关切和军事需求之间取得必要的平衡。

希腊欢迎过去一年在这一问题上取得的进展，包括大会于 2023 年 12 月 22 日通过关于致命自主武器系统的第 78/241 号决议、《关于在军事上负责任地使用人工智能和自主的政治宣言》、致命自主武器系统问题政府专家组在日内瓦开展的工作以及最近于 2024 年 4 月 29 日和 30 日在维也纳举行的题为“十字路口的未来：自主武器系统和管制挑战”的国际会议得到广泛参与。

此外，人工智能技术的快速发展仍是我国最为关切的问题之一。

希腊支持将人工智能应用于军事目的必须充分遵守适用的国际法，特别是国际人道法及其核心原则和要求，如区分、相称和预防的原则和要求，并充分遵守国际人权法。

希腊将积极参与今后对这一复杂问题的审议，并参与制定一项国际文书，以确保自主武器系统充分遵守国际人道法。

危地马拉

[原件：西班牙文]
[2024 年 5 月 23 日]

危地马拉认为，必须考虑到 1949 年日内瓦四公约，这些公约规定了限制战争野蛮性的主要规则，并载有关于惩罚战争罪的章节，其中明确规定，为惩罚此类罪行，必须追究个人的刑事责任。

在这方面，危地马拉认为，建立一个具有致命能力但不受人控制的系统，侵犯了生命权，违反了国际人道法和日内瓦四公约，因为由于一个抽象的实体，如自主操作(或使用人工智能)的武器不可能承担刑事责任，因此使用此类系统将无法追究责任。

洪都拉斯

[原件：西班牙文]

[2024年5月24日]

人工智能、自主武器系统以及世界在其管制方面面临的挑战

目前，对这些每天都在出现的技术和活动进行控制十分具有挑战性，但召开会议无疑极为重要，也是一个良好的开端。还应举办研讨会、课程和各种活动，以确定实施控制的手段并制定适当的条例，以期实现控制并提高公众的认识，从而能够适当和负责任地使用这些技术工具。

至关重要的是，必须考虑到不受约束地使用自主武器系统方面的所有必要条例，同时不轻视世界范围内的创新以及近年来在实施和使用技术和人工智能方面取得的重大发展。虽然不可否认，新技术带来了影响流程和程序的创新，并在某些方面对各国有利，但要想对一个事物加以控制，则该事物必须遵守限制不当使用的特定规则。必须对此加以管制，人工智能也不能例外，因为由于世界在不断变化和发展，人工智能既可能产生积极影响，也可能产生消极影响。

洪都拉斯认识到，必须在所有必要的禁令和条例中遵守国际法、国际人道法和《联合国宪章》的原则，以确保对现有和未来的自主武器系统采取平衡和负责任的方法。此外，还需要提高对适当使用人工智能的认识，从而避免不当使用此类工具并限制其消极影响。

至关重要的是，必须能够确定并在协商一致的基础上做出最佳决定，制定必要的条例，以保持人类对所有自主武器系统的全面控制，因为此类系统可能使所有国家或区域受益，也可能对其产生其他影响。应从最佳角度对此进行分析，并始终以人权为根本基础，同时避免侵犯人权。如果没有确保对自主武器进行控制的法规，人工智能可能导致失去对有区别地使用武力的控制，使公众面临风险并侵犯他们的权利。

现代的技术安全极易受到可能怀有不良企图并采取不良行动的黑客的攻击，从而对这些系统，包括基于人工智能的控制系统和现有自主武器控制系统造成破坏，这种情况将危及当前技术系统的自主决策。

此外，系统并不完美，因为它们具有脆弱性并容易出现机械故障和操作系统故障。因此，可以预料此类系统会出现异常，这可能会对人类的完整性以及所有国家或区域产生影响。

新技术系统大有可为，但必须始终由人来控制，原因是在使用这些技术方面所做的决定和采取的举措必须以维护人权为重点，努力确保人民的福祉，从而避

免他们被非人化，因为我们以何种方式对技术造成的损害进行管制，将对社会的未来以及我们是否能从这些技术中获益产生深远的影响。因此，在技术的支持下实施人的控制至关重要，这样，此类技术就可以支持和简化现有自主武器系统和基于人工智能的武器系统中使用的所有流程和程序，而不会对社会造成损害。自主武器系统一旦启动，就会选择一个或多个目标并使用武力，只需极少的人为干预，这一问题引发了法律、伦理和安全方面的重大关切。自主武器系统也对和平与安全造成全球风险。

非国家武装团体威胁会篡改作为自主武器系统基础的新技术的安全性，以便任由其利用这些技术对各国进行攻击，从而危及人民的身心健全。

如前所述，管制对于保持对新技术和自主武器系统的全面控制至关重要。这需要紧急发挥果断的政治领导作用，以及反映这种发展重要性的远见，因为人类已经走到了将生死攸关的决定交给机器的十字路口。

最后，作为联合国创始会员国，洪都拉斯认识到技术的发展及其脆弱性。因此，洪都拉斯正在采取基本的网络安全措施，培训人员并适应新的机遇和相关威胁。虽然洪都拉斯拥有基本的自主武器系统，但仍不断了解世界各地正在出现的新技术，因为鉴于重大的科学和技术发展虽有助于整个世界但可能对人类的未来构成令人担忧的威胁，洪都拉斯非常重视对这些技术保持持续了解，也非常重视确保普遍公认的原则得到尊重。

印度

[原件：英文]

[2024 年 5 月 24 日]

印度承认，致命自主武器系统领域的新兴技术带来了挑战和关切。印度致力于应对这些挑战和关切，积极和建设性地参与了在《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》框架内对这些事项的审议，并自 2016 年以来参加了致命自主武器系统问题政府专家组。

印度还认为，新兴技术可对减少贫困和改善所有人的生活产生变革性影响。对发展中国家来说，这一点尤为重要。应避免对此类技术进行污名化。

印度认为，在任何时候都必须遵守武装冲突法。将致命自主武器系统领域的新兴技术应用于军事目的必须符合国际人道法。

在不轻视致命自主武器系统领域新兴技术所产生的诸多合理法律和伦理问题的同时，印度也支持这样一种观点，即总体而言，新兴技术有可能切实改善对国际人道法的遵守。

《特定常规武器公约》是讨论致命自主武器系统领域新兴技术相关问题，以期在军事需求和人道主义任务之间取得平衡的适当论坛。《公约》能够并已经汇集了相关的利益攸关方。它还成功汇集了法律、军事和技术方面的相关专门知识。各方的广泛参与、若干国家提交的工作文件数量以及全球对这些讨论的关注进一步证明了政府专家组的审议工作仍然具有现实意义和重要性。

印度认为，政府专家组在提出概念和词汇，以体现致命自主武器系统领域新兴技术的一些难以表述的特性及其可能产生的影响方面，取得了重大成就。这些成就包括专家组提出了 11 项指导原则，为应对致命自主武器系统构成的潜在挑战提供了基本框架，并达成了谅解。这些谅解载于专家组和《特定常规武器公约》缔约国通过的协商一致的报告中。专家组 2023 年届会报告(CCW/GGE.1/2023/2)中所载的结论尤为重要，其中除其他外指出，在不影响专家组今后继续以国际法，特别是《联合国宪章》和国际人道法以及相关伦理观为指导开展工作的情况下，专家组得出结论认为：

- (a) 国际人道法继续充分适用于致命性自主武器系统的潜在开发和使用；
- (b) 基于致命性自主武器系统领域新兴技术的武器系统如果不能按照国际人道法使用，则不得使用；
- (c) 需要对基于致命性自主武器系统领域新兴技术的武器系统进行控制，以坚持遵守国际法，特别是国际人道法，包括在攻击时遵守区分、相称性和预防的原则和要求。

政府专家组还得出结论认为，各国必须确保在基于致命性自主武器系统领域新兴技术的武器系统的整个生命周期内，遵守其根据国际法，特别是根据国际人道法承担的义务。在必要时，各国除其他外应：

- (a) 限制这类系统可以打击的目标类型；
- (b) 限制武器系统运行的持续时间、地理范围和规模；
- (c) 为操作人员提供适当的培训和说明。

政府专家组还得出结论认为，根据国际法规定的国家义务，在研究、开发、取得或采用新的武器、作战手段或方法时，必须确定使用上述新的武器、作战手段或方法是否会在某些情况或所有情况下为国际法所禁止。在这方面，鼓励各国自愿交流相关最佳做法，同时铭记国家安全考虑或关于专利资料的商业限制。

必须承认这些谅解和协议，将其作为进一步工作的基础，在不预先判断其性质的前提下，将其纳入任何可能达成的文书的要素，并作为政府专家组今后工作的主要重点。

印度认为，政府专家组的工作应继续优先加强对致命自主武器系统领域新兴技术的特性和定义的理解。需要强调的是，在如何实际理解和解释与其开发和部署有关的概念要素方面必须实现统一。

印度支持考虑采取适当行动，应对和减轻与致命自主武器系统领域新兴技术相关的风险，关于减轻风险的协议应考虑到现有法律义务，并尊重国家管辖权和权限范围以及相关国家能力。

透明度和建立信任措施对于解决与致命自主武器系统领域新兴技术有关的关切至关重要。

各方还必须承认，会员国的安全现实各不相同。在讨论可能采取的措施时，应考虑到所有各方的关切，努力找到共同点。更重要的是，对致命自主武器系统领域新兴技术的审议不应导致规范领域的支离破碎。各论坛根据其现有任务规定正在开展的工作不应有所重复，这一点很重要。必须避免为制定平行任务或平行规则创造条件。

爱尔兰

[原件：英文]
[2024年5月24日]

导言

爱尔兰所提交的这份意见与关于“十字路口的人类：自主武器系统和管制挑战”的维也纳会议的主席总结相符。

自主武器系统带来严重风险，包括：

- 人道主义风险(如无法区分平民和战斗人员)
- 法律风险(如确保遵守国际人道法，并根据国际人道法进行问责和承担责任)
- 安全风险(如助长军备竞赛)
- 技术风险(如故障和算法偏见)
- 伦理风险(如丧失尊严和非人化)。

爱尔兰积极参加关于自主武器系统的主要论坛和致命自主武器系统问题政府专家组，并参加了哥斯达黎加(2023年)、菲律宾(2023年)和塞拉利昂(2024年)开展的区域举措。¹

爱尔兰的立场侧重于：

- 功能特性描述
- 国际法
- 人的控制和问责
- 伦理

¹ 见 <https://conferenciaawscostarica2023.com/>; <https://dfa.gov.ph/dfa-news/dfa-releasesupdate/33834-philippines-calls-for-indo-pacific-voices-to-address-lethal-autonomous-weapons-systems-risks>; <https://mofaic.gov.sl/sierra-leone-hosts-first-regional-conference-on-autonomous-weapons/>。(2024年5月24日查阅)。

- 偏见
- 多边解决方案。

功能特性描述

自主是武器功能的一个特征，而不是一种特定武器类型。因此，重点应该是对将自主纳入武器系统加以禁止或管制，而不是根据具体技术特性寻求固定的定义。在技术迅速发展的背景下，后一种方法不太可能仍然符合目的。

爱尔兰采用红十字国际委员会的工作定义如下：

“自主武器系统”是指在启动后无需人的干预即可选择和攻击一个或多个目标的武器系统。²

以下是引起严重关切的功能特性的不完全清单：

- 无需人的干预，即可完成一个目标选择周期，最终意图为施加武力
- 无需人的干预即可切换到致命模式
- 无需人的干预即可重新确定任务或目标
- 无法中断或停用自主模式。

国际法

国际法，包括《联合国宪章》、国际人道法、国际人权法和国际刑法，完全适用于自主武器系统的开发和使用。

遵守国际人道法需要人类根据具体情况作出以价值观为基础的判断(这是自主武器系统无法替代的)，并考虑以下因素：

- 认知局限性(缺乏常识和人的判断力)
- 认识论局限性(根据有偏见、不完整或不恰当的数据做出判断)
- 算法偏见。

根据国际人道法不得使用的自主武器系统事实上已被国际法禁止。³ 然而，需要制定新的规则和准则，以明确国际人道法对自主武器系统的适用。

人的控制和问责

爱尔兰对自主武器系统采取以人为本的方针。

² 红十字国际委员会就自主武器系统提交的材料，可查阅 https://www.icrc.org/sites/default/files/wysiwyg/war-and-law/icrc_submission_on_autonomous_weapons_to_unsg.pdf。

³ 致命自主武器系统问题政府专家组，国家发言，2023年3月。

遵守国际人道法要求人类为自主武器系统作出决定并对其负责，并且对此类系统实行全面控制，即在其整个生命周期内保持干预、中断和停用这些系统的能力。⁴ 开发者和制造商也应承担这一责任，不能转移给机器。

必须注意：

- 背景因素：自主武器系统是否表现出足够的态势感知能力，并能正确理解作战环境。
- 技术因素：
 - 自主武器系统在识别、选择和攻击目标时是否具有足够的可靠性和可预测性。
 - 是否存在足够的环境限制，以确保在规划阶段作出的决定在整个执行阶段都得到遵守。
- 人的控制的形式：
 - 是否保留了人对自主武器系统关键功能(如识别、选择和攻击目标)的控制。
- 人的控制程度是否能防止在未经人为验证的情况下重新确定自主武器系统的任务，并允许中断或停用这些系统。

自主武器系统在其整个生命周期内都需要人的指挥和控制系统。在部署或使用此类系统之前，必须告知指挥人员和操作人员此类系统的任何新特性、功能和(或)参数，并对其进行培训。

伦理

治理框架必须考虑使用自主武器系统在法律和伦理上是否可以接受，包括认识到：⁵

- 伦理考量，包括人道原则和公众良心的要求，对于确定自主武器系统是否可接受和对其进行何等限制(包括必须保持的人的控制程度)至关重要。
- 将人对施加武力决策的控制、能动性、责任和(或)意图让渡给算法和计算机控制的程序，引发了基本的伦理问题。
- 施加武力决策的伦理责任要求人类根据具体情况作出以价值观为基础的判断，而将这种判断授权给自主武器系统在法律上不可行，在伦理上也不可取。

⁴ 见欧洲联盟在大会第七十八届会议期间第一委员会一次会议上关于常规武器的发言，可查阅 https://estatements.unmeetings.org/estatements/11.0010/20231023100000000/6geoPnJ6HnKK/M8SpW99Xe5MF_en.pdf。

⁵ 这些观点由红十字国际委员会等机构提出。见 <https://www.icrc.org/en/document/autonomous-weapons-icrc-submits-recommendations-un-secretary-general>。

- 必须在部署自主武器系统的人的理性与使用这些系统的直接后果之间建立直接联系，并清晰明确地界定对人类负责和问责的范围，以防止现有伦理标准受到削弱。

偏见

以下为爱尔兰向政府专家组联合提交的材料要点。⁶

算法和相关机器学习有可能重复、放大或助长不公正的偏见

首先，基于数据的系统会复制现有的不平等。2016年的一项关于旨在评估刑事司法系统累犯可能性的计算机程序的研究⁷发现，该系统对不同种族和性别的累犯预测并不均衡。

其次，现有的数据集和算法偏向于白人男性，这意味着，以有色人种女性为例，经过图像⁸和声音识别训练的机器学习程序能够识别她们的可能性大大降低。⁹

关于1988年至2021年期间在不同经济部门部署的133个有偏见的人工智能系统的公开信息进行审查¹⁰后发现，44.2%(59个系统)存在性别偏见，25.7%(34个系统)同时存在性别和种族偏见。

正如联合国裁军研究所¹¹最近的一份报告所显示，机器学习中的偏见在军事领域造成的后果更为严重。有色人种女性被错误识别的比率可能更高，使她们面临不同的风险，或者由于平民男性在战争中的传统角色，自主系统可能会将他们错误地归类为战斗员。

建议包括但不限于：

- (a) 进行全面测试和审查，以查明和纠正潜在的偏见；
- (b) 严格记录自主武器中使用的数据集，以提高可追溯性和透明度，并提供有关动机、收集过程和使用建议的信息；
- (c) 对照基准测试算法模型，评估其针对不同性别、年龄和种族的运行情况，并在与用于训练机器学习模型的数据集不同的情景中对其进行测试；

⁶ 奥地利、比利时、加拿大、哥斯达黎加、德国、爱尔兰、卢森堡、墨西哥、巴拿马和乌拉圭联合提交的材料(CCW-GGE.1-2024-WP.5)。

⁷ 见 www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing。

⁸ 见 <http://gendershades.org/>。

⁹ 见 <https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases>。

¹⁰ 见 https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity。

¹¹ 见 <https://undir.org/publication/does-military-ai-have-gender-understanding-bias-and-promoting-ethical-approaches-in-military-applications-of-ai/>。

(d) 对测试和使用该系统的人员进行全面培训，提高其对该问题的认识。

多边解决方案

爱尔兰支持制定关于自主武器系统的具有法律约束力的规则和准则。爱尔兰支持政府专家组正在开展的重要工作，以及与专家组相辅相成并有助于采取包容性做法的平行举措。我们注意到红十字国际委员会、民间社会、学术界、工业界和大会的重要作用。

爱尔兰支持两级方法，即彻底禁止某些自主武器系统，并对所有其他自主武器系统进行管制。

在以下情况下，必须禁止自主武器系统，以确保遵守国际人道法：

- 自主武器系统不符合国际人道法，包括关于区分、相称、过分伤害和(或)不必要痛苦的规定。
- 自主武器系统无法充分预测、控制或理解。
- 自主武器系统可以在没有必要程度的人的控制的情况下选择和攻击目标。

多边管制应适用于所有其他自主武器系统，包括：

- 就打击目标类型设限
- 就持续时间、地理范围和使用规模设限
- 确保维持人的控制。

以色列

[原件：英文]

[2024年5月24日]

以色列注意到大会 2023 年 12 月 22 日通过了题为“致命自主武器系统”的第 78/241 号决议，并谨根据该决议第 2 段，向秘书长将提交大会第七十九届会议供会员国进一步讨论的报告提交本国意见。

以色列认识到致命自主武器系统的潜在优势和挑战。在这方面，以色列谨指出，基于致命自主武器系统领域新兴技术的武器系统可同时满足军事需求和人道主义考虑，并可用于确保遵守国际人道法。

以色列认为《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》非常重要，并承认其作为在具有军事、法律和技术专门知识的相关利益攸关方参与下重点讨论国际人道法的论坛的独特性，该论坛有能力在任何有关常规武器的审议中在军事需求和人道主义考虑之间取得适当平衡。这些特点使该《公约》在常规武器领域取得的成就在实地产生最大的实际影响，在讨论致命自主武器系统问题时也至关重要。

关于致命自主武器系统，应当承认在《特定常规武器公约》框架内取得的进展，同时也应认识到需要在致命自主武器系统问题政府专家组内进一步深入讨论，以取得切实的协商一致成果。最近讨论的建设性气氛使以色列更加相信，《公约》仍然是应对致命自主武器系统领域新兴技术带来的挑战和机遇的最适当论坛。

政府专家组有一项强有力的任务，即在不预先判断文书性质的前提下，进一步审议并以协商一致方式制定一套文书要素以及其他可能的措施，以解决致命自主武器系统领域的新兴技术问题。这一基于共识的论坛有各国、民间社会代表和专家的广泛参与，促进了包容性。缺乏这些特点的外部举措不仅有可能造成重复，也有可能造成各自为政。至关重要的是，不能破坏在《特定常规武器公约》范围内取得的重要成果和在其框架内继续开展的对话，包括为此采取不同举措，以制定有关致命自主武器系统的政策措施。

此外，以色列的立场是，对这一专题的任何特性描述都必须继续侧重于致命自主武器系统(而不是自主武器系统)，这也是政府专家组自成立以来一直关注的重点。

必须强调的是，现有国际法，特别是国际人道法，完全适用于致命自主武器系统，以色列认为，这些法律为今后在武装冲突中使用致命自主武器系统提供了充分的法律框架。

以色列认为，在讨论致命自主武器系统这一专题时，必须重点讨论现有国际人道法对致命自主武器系统的适用问题，同时考虑到作战环境。如果在讨论致命自主武器系统的合法性时只关注武器的能力，而不考虑打算使用该武器的作战环境或在这一过程中人的投入类型和水平，就会产生问题。此外，不应把关于执行取决于具体情况的国际人道法规则的讨论与关于武器合法性本身的讨论混为一谈。

至关重要的是，一方面要区分国际人道法的主要规则，包括对武器的绝对禁止、目标选择规则和其他法律规则，另一方面要区分本身并非国际人道法主要规则的法律或实践问题。这包括控制、责任、可预见性、可预测性和可靠性等问题。虽然其中一些概念在国家责任法或国际刑法等其他法律领域具有重要意义，或在执行国际人道法方面具有实际意义，但将其视为国际人道法的规则，甚至在使用这些概念时制定禁令，在许多法律和实践层面上都会产生问题。例如，国际法中没有关于在使用武器时人的控制的规定。人的控制本身并不是目的，而可能是在不同情况下履行国际人道法规定的各种义务的一个相关概念(其他概念也可能如此)。

最后，考虑到技术的不断发展，必须认识到，我们对致命自主武器系统领域新兴技术的理解和期望可能会随着时间的推移而改变。因此，有必要认真慎重地处理这一问题，以对致命自主武器系统进行认真负责的讨论。

意大利

[原件：英文]
[2024 年 5 月 24 日]

意大利是大会第 78/241 号决议的共同提案国，并欢迎有机会根据上述决议第 2 段提出意见，以纳入秘书长将向大会第七十九届会议提交的报告。

自 2013 年以来，意大利先是在非正式专家会议框架内，后来作为致命自主武器系统问题政府专家组本身的参与方，积极参加了在《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》主持下发起的关于致命自主武器系统的国际辩论。

致命自主武器系统属于一个庞大的类别，该类别包括作战、武装冲突和武器技术的新发展，以及人工智能和机器学习等相关概念。随着时间的推移，就武器系统自主的进步及其相关影响进行的具体和结构化的讨论取得了实质性进展。

意大利认为，《特定常规武器公约》是迄今为止处理与武器系统的开发和使用的现有问题和新问题的最适当论坛。许多缔约国已加入《公约》，其中包括人工智能军事应用的主要开发国和生产国。此外，《公约》通过缔约国代表以及国际组织、专门机构和民间社会组织代表等，汇集了外交、法律和军事领域的专门知识。《公约》也是审查武器系统是否符合国际人道法的最佳论坛。

意大利欢迎将致命自主武器系统问题政府专家组的任务期限延长至 2026 年，并致力于推动关于制定未来文书要素的讨论。该文书应规定明确的禁令和条例，以便最终作为《特定常规武器公约》的附加议定书获得通过。事实上，意大利完全致力于实现《公约》的精神和目标，正如《公约》序言所强调，各缔约国认识到“竭尽全力努力促进朝向严格有效国际监督下全面彻底裁军进展的重要性”。这对各国而言都确实是一项重大挑战，必须成为激励我们工作的源泉。

正如 2019 年核可的指导原则所回顾，包括国际人道法在内的国际法完全适用于致命自主武器系统，该指导原则指出，“国际人道法仍完全适用于所有武器系统，包括致命自主武器系统的潜在开发和使用”。国际人道法对冲突各方在攻击前和攻击期间提出了要求。这些要求包括，攻击必须有所区分，允许将合法目标(战斗人员和军事目标)与非法目标(平民、非战斗人员和民用目标)区分开来；必须具有相称性，禁止进行可能附带使平民死亡、受伤害和受损害等与预期军事利益相比损害过分的攻击；遵守在攻击中采取预防措施的要求，其中要求时刻注意避免殃及平民人口、平民和民用物体。

除这些原则外，《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》(《第一议定书》)第 35 条重申，作战方法和手段不是无限制的，这一概念的实施在第 36 条作出了规定，该条规定缔约国有义务确保不使用非法武器。武器审查进程是确保在开发和获取新武器期间履行这一义务的机制。就致命自主武器系统而言，审查过程中存在一项内在的挑战，这与基础技术的复杂性以及自主系统在不断变化的实物环境和作战环境中可能存在的不可预测性有关。

意大利认为，尽管自主武器系统并不存在法律真空，但仍需进一步制定有关此类系统的规范和操作框架。这可以通过采用两级方法，制定禁令和条例来实现。根据这一方法，无法按照国际人道法开发和使用的致命自主武器系统当然将被禁止。另一方面，可在完全遵守国际人道法的情况下开发和使用并在关键功能方面具有决策自主的系统将受到管制。更具体地说，根据国际人道法的现有规定，本质上具有滥杀滥伤作用或其影响无法加以限制、预测或控制，或造成过度伤害或不必要痛苦的致命自主武器系统，将属于被禁止的类别。这也包括完全不受人的控制和负责任的指挥系统约束的武器系统。

其他致命自主武器系统属于被管制的类别。此类系统包括可通过考虑其现有能力、对操作人员进行适当测试和培训(以评估其可靠性、可理解性和可预测性)或限制目标类型以及运行的持续时间、地理范围和规模来评估其是否符合国际人道法的系统。此外，在关键的“高级”功能中，操作人员应能进行干预。要将一个系统视为不完全自主的系统，从而不将其禁止而是进行管制，只需操作人员至少保留最后的关键功能，即决定是否对先前确定和选定的目标施加武力。

意大利认为，人的因素对于致命自主武器系统的整个生命周期，即其设计、开发、生产、部署和使用至关重要。如果目标是确保遵守国际人道法，那么就应保持适当程度的人的判断和控制，以确保根据国际人道法追究责任。根据国际人道法，只有人类才能被追究责任，而绝非机器。人类必须在整个指挥和控制系统中，根据不同的责任级别，充分了解武器系统的特性和运作。所施加的控制应采用各种参数，以限制分配给致命自主武器系统的任务和目标类型，强制规定必须有停用和故障安全机制，并根据对局势和军事背景的评估，对致命自主武器系统的运作设定时间和空间限制。

关于致命自主武器系统的定义和特性描述，意大利认为，一个系统的杀伤力不应仅由其内在特性决定，还应由其使用的背景及其意外或附带影响决定。在《特定常规武器公约》本身范围内，议定书一没有具体说明武器的功能，但提到永久失明是武器的预期影响或已产生影响。《禁止或限制使用燃烧武器议定书》(议定书三)将燃烧武器称为造成烧伤的武器。

最后，我们支持今后建立一个合作与协商机制，以促进缔约国之间就文书的运作进行交流，并分享有关减少风险和法律审查措施的信息和最佳做法。

日本

[原件：英文]

[2024年5月24日]

日本认识到，至关重要的是努力阐明有关致命自主武器系统的规范和操作框架，同时考虑到此类系统与现有国际法，特别是国际人道法的关系，以应对新兴技术的进一步发展。

对于将新兴技术应用于军事目的应进行全面审查，充分了解其风险和益处，并考虑到人道主义考虑因素和安全角度。日本重视以负责任的方式将新兴技术应用

于军事目的，坚持“以人为本”的原则，确保可靠性和可预测性。日本还支持这样一种观点，即国际人道法的义务适用于所有武器系统，包括致命自主武器系统，如果武器系统不能按照国际人道法使用，则不得使用，其他武器系统应受到必要的限制，以确保遵守国际人道法。日本认识到，在使用武器系统时，不能将人的责任转移给机器，必须采取措施，确保以符合国际人道法规定的各国义务的方式，在负责任的人的指挥和控制系统下运作武器系统，明确责任归属。

一. 关于致命自主武器系统主要专题的意见

1. 特性

日本认识到，目前关于自主武器系统的讨论所涉及的武器系统是那些一旦启动就能识别、选择并以致命武力攻击目标，而无需操作人员进一步干预的武器系统。

2. 国际人道法的适用

日本不会研究、开发或运作国内法或国际法，包括国际人道法不允许使用的武器系统。

要确保武器系统的使用符合国际法，特别是国际人道法，就需要人的参与，因为根据国际人道法，可以被追究责任的是人类。如果武器系统完全在无人参与的情况下运行，那么在攻击目标时就可能难以确定对其使用后果负责的指挥官或操作员。此外，缺乏人的参与可能导致系统无法按照用户的意图运作，进而可能对平民造成意想不到的伤害。

日本认为，完全在无人参与的情况下运作的具有致命武力的自主武器系统不一定达到能够确保遵守国际人道法的阶段。相反，日本认为，确保有人的参与的武器系统可以减少人为错误，提高劳动和人力效率，从而为安全带来重大益处。因此，日本表示不打算开发“完全在无人参与的情况下运作的具有致命武力的自主武器系统”。

有鉴于此，日本认为，国际社会绝不当允许开发和符合下列条件的自主武器系统，因为此类武器系统可能造成不可接受的后果：

- 可造成过度伤害或不必要痛苦、本质上具有滥杀滥伤作用或无法按照国际人道法使用的武器系统。
- 完全在无人参与的情况下运作的具有致命武力的自主武器系统，此类系统不具备适当程度的人的判断，也无法在负责任的人的指挥和控制系统内运作。

除不能按照国际人道法使用的武器系统外，其他武器系统应受到限制，或应采取与这些系统有关的减少风险措施，以确保其使用方式符合国际人道法。日本认为，在管制措施方面，特别重要的是控制或限制这些系统可攻击的目标类型，以及其运行的持续时间、地理范围和规模，包括通过自毁、自失能和自失效机制。此外，有必要提高此类系统使用效果的可预测性和可靠性，这是作出适当程度的人的判断的先决条件。

3. 人的参与

人的参与最重要的方面是确保指挥人员和操作人员能够按照他们的意图操作武器系统。为此，他们必须充分了解所使用的武器系统，并确保人的参与能够使其作出适当程度的判断，从而确保在负责任的人的指挥和控制系统下进行适当的操作。

4. 责任和问责

根据 11 项指导原则(CCW/GGE.1/2019/3, 附件四, (b)分段)中确认的原则、致命自主武器系统问题政府专家组 2019 年会议报告(CCW/GGE.1/2019/3, 第 17(c)段)和政府专家组 2022 年会议报告(CCW/GGE.1/2022/2, 第 19 段), 日本认识到, 武器系统应在负责任的人的指挥和控制系统内运作, 应确定责任归属, 并确保人的责任或使人类对机器的行动负责的手段。

5. 风险评估和缓解措施

正如 11 项指导原则(CCW/GGE.1/2019/3, 附件四, (g)分段)所确认, 风险评估和减小风险的措施应成为任何武器系统新技术的设计、发展、测试和部署周期的组成部分。

尤为重要的是, 应采取尽量减少偏见、监测通过自学更新的能力、引入保障措施以避免意外后果以及开展关于使用人工智能和其他技术的教育等措施。

日本将实施培训方案, 以确保适当使用人工智能技术。

6. 法律审查

在确定利用自主武器系统领域新兴技术的新武器系统的使用是否为适用于有关国家的国际法所禁止时, 各国必须根据《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受难者的附加议定书》(《第一议定书》)第 36 条对此类武器系统进行法律审查。

二. 关于今后讨论的看法

日本重申, 在今后讨论过程中应产生的成果所需要的不是具有法律约束力的规定, 而是有效的规则。在充满挑战的安全环境中, 以协商一致方式制定规则是确保这些规则有效的关键。在这一过程中, 必须考虑到在人道主义考虑因素和安全角度之间寻求平衡的观点, 以及确保所有利益攸关国家, 包括拥有先进技术能力的国家参与的观点。为此, 日本坚决支持继续在《特定常规武器公约》框架下进行讨论。此外, 日本欢迎最近关于负责任地将人工智能用于军事目的的讨论。日本希望这些讨论将补充和加强政府专家组在《公约》框架下的讨论。

最后, 日本重申, 它将积极和建设性地促进制定关于使用新兴技术的武器系统的国际规则。

基里巴斯

[原件：英文]
[2024 年 5 月 24 日]

概述

基里巴斯欢迎有机会根据大会 2023 年 12 月 22 日通过的题为“致命自主武器系统”的第 78/241 号决议提交其意见供秘书长审议，大会在该决议中请秘书长就如何应对自主武器系统从人道主义、法律、安全、技术和道德角度提出的相关挑战和引起的关切以及人在使用武力方面的作用征求意见。

基里巴斯是太平洋区域的一个小岛屿国家，该区域是一个和平的区域，不是自主武器的主要生产者。然而，基里巴斯仍然高度关切自主武器的存在及其对环境和平民造成严重伤害的能力，特别是考虑到核武器与自主武器之间的联系。

核层面

1957 年至 1962 年期间，基里巴斯共和国经历了在莫尔登和基里蒂马蒂群岛进行的 33 次核武器试验。因此，基里巴斯对在核决策中可能使用自主系统感到震惊，最近的许多出版物也提到了这一点。

基里巴斯还震惊地回忆起关于自主机器发动核打击的流行电影。应避免发生任何核武器拥有国让自主机器控制其核武器的情况。因此，基里巴斯要求核武器拥有国绝不允许自主武器系统控制其指挥和控制系统，并要求它们禁止核武器。

共同提出奥地利决议

在此背景下，基里巴斯成为奥地利向大会第一委员会提交的关于致命自主武器系统的决议的共同提案国。

出席在维也纳举行的“十字路口的人类”会议

在“十字路口的人类：自主武器系统和管制挑战”会议上，基里巴斯认真听取了学术界、民间社会和红十字国际委员会的专家和代表关于自主武器系统的发言。在这方面，基里巴斯完全同意主席总结中的观点，即自动武器系统一旦启动，就可以在没有进一步人为干预的情况下选择目标并施加武力，这引起了法律、伦理和安全方面的担忧。因此，基里巴斯完全赞同主席的总结。

要求禁止自主武器系统

基里巴斯对自主武器系统同样感到关切，因此呼吁制定一项新的具有法律约束力的条约，以禁止这些武器。基里巴斯的立场基于西非国家经济共同体《弗里敦公报》、拉丁美洲和加勒比国家《贝伦公报》和加勒比共同体《西班牙港宣言》所反映的区域各国表达的普遍看法。

该条约可在联合国大会主持下谈判，或在《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》缔约国之间谈判。

纳入国际法

新的具有法律约束力的文书应强调所有国家在任何时候都应遵守国际法，包括国际人道法、国际人权法和《联合国宪章》。

定义自主武器系统

新条约应明确界定何为自主武器系统。在这方面，基里巴斯同意红十字国际委员会的观点，即自主武器系统是指“设计目的为在启动后无需人为干预即可选择和攻击一个或多个目标的武器系统”。

“无需人为干预”这一表述可进一步加以阐述。正如红十字国际委员会所解释，这一表述可界定如下：

在最初由人启动后，通过测量热、光、运动、形状、速度、重量或声学或电磁信号等现象的传感器接收到的环境信息；根据通用的“目标描述”，如某类军用车辆的形状、红外或雷达“特征”以及速度和方向等，触发施加武力。就这些目的而言，“人为干预”应理解为不包括对目标选择或交战的自主功能无实质性影响的人为投入或行动。

禁令

基里巴斯强烈赞同红十字国际委员会的观点，即一项新的具有法律约束力的条约应规定禁止：

开发、生产、以其他方式获取、储存或保有，或者直接或间接向任何一方转让，或使用任何自主武器系统，而该系统的设计或性质或使用方式使人类使用者无法：(1) 了解、预测和解释[自主武器系统]在任何正常或预期使用情况下将如何运作，特别是哪些情况或条件将触发该系统施加武力；(2) 按照[国际人道法]的要求，预测和限制[自主武器系统]在所有这些情况下的影响。

基里巴斯还赞同，该文书应禁止杀伤人员自主武器系统。具体而言，正如红十字国际委员会所解释，该文书应：

规定在任何情况下都禁止开发、生产、以其他方式获取、储存或保有，或者直接或间接向任何一方转让，或使用任何具有以下特征的自主武器系统：

- 其设计或使用方式是在一人或多人出现、接近或接触时触发，或
- 其目标描述以其他方式代表一人或多人。

包容性

如果国际社会开始就禁止自主武器系统的新条约进行谈判，我们必须邀请学术界、民间社会、宗教团体和青年组织的代表分享他们的意见和观点。他们可以为各国带来新的想法，我们必须鼓励他们献计献策。

结论

鉴于当前世界各地的地缘政治局势，我们必须保护平民的生命。我们不能让武器无法控制，因为它们可能造成滥杀滥伤，无法区分平民和非平民。因此，拟议的条约应涉及一系列有关自主武器系统的禁令，这些武器系统具有人工智能，阻碍人类用户理解和预测系统的行为。

我们不能允许任何可能伤害平民的武器出现。

卢森堡

[原件：英文]

[2024 年 5 月 25 日]

卢森堡一直积极参与有关致命自主武器系统的各种国际举措，并于 2023 年 4 月主办了一次关于自主武器系统的会议，旨在提高公众对这一专题的认识，并提供一个讨论风险和平台的平台。

2022 年 7 月，卢森堡政府成立了致命自主武器系统部际工作组。该工作组的任务是就卢森堡对致命自主武器系统的立场和后续行动提出建议，以便在国际一级进行管制之前，在国家一级确定原则、规则和控制机制。

在分析现有定义的基础上，工作组提出了以下自主武器系统的定义：

“能够在无人干预的情况下识别、选择目标并针对目标部署武力的武器系统。”

考虑到自主武器系统可用于致命和非致命目的，而且使用武力不一定要造成致命后果才违反国际人道法，卢森堡的立场是将其称为“自主武器系统”，而不是“致命自主武器系统”。然而，根据大会第 78/241 号决议所用术语，卢森堡在这份意见中将继续使用“致命自主武器系统”的提法。

描述致命自主武器系统的一个关键因素是其由人工智能实现的适应能力，这是其能力以及相应的伦理和法律辩论的固有组成部分。然而，由于这些技术仍在发展之中，自主这一特征不足以提供准确的定义。因此，卢森堡认为，更有效的方法是界定人为干预程度，而不是试图量化自主程度。

此外，卢森堡的立场是，就关于致命自主武器系统的具有法律约束力的文书开始谈判并不需要一个共同的定义。

一. 应对致命自主武器系统带来的挑战和关切

卢森堡认为，两级方法是应对致命自主武器系统带来的挑战和关切的适当方式，因为这需要：

(a) 禁止开发和使用无法按照国际法，特别是国际人道法部署；本质上具有滥杀滥伤作用，其影响无法加以限制、预测或控制；可造成过度伤害或不必要痛苦；完全不受人的控制的致命自主武器系统。

(b) 对所有其他致命自主武器系统进行管制，以确保其符合国际法，特别是国际人道法。为确保这一点，各国应承担一系列积极义务，在此类系统的整个生命周期内采取适当措施，以减轻潜在的法律、伦理、技术和安全挑战。这些义务可包括对致命自主武器系统的使用保持人实际控制，以及除其他外，对武器进行法律审查以遵守国际人道法，进行适当培训、采取缓解措施和其他保障措施。

两级方法确保国际法，特别是国际人道法，将继续完全适用于致命自主武器系统，包括从其开发到使用。人在使用武力中的作用是两级方法确保遵守国际人道法的关键因素。因此，如果致命自主武器系统的运作不受人的控制，没有一个负责任的指挥系统制定的精确和有限的任务框架，则将被禁止，因为无法确保遵守国际法和国际人道法。

最后，禁止和管制致命自主武器系统的目的不应被视为试图阻碍技术创新，但必须确保国际人道法对其开发和部署进行规制。

二. 人类在使用致命自主武器系统中的作用

人的控制的概念对于确保致命自主武器系统的使用符合国际法和国际人道法至关重要。必须考虑两个关键原则，即适当的人的控制和人的责任。

关于适当的人的控制，可采取以下措施。

- 致命自主武器系统一旦启动，应根据国际法和国际人道法以及指挥人员和操作人员的意图，以可预测的方式行动。
- 充分了解系统的运作模式至关重要。操作人员应能充分预测系统在不同作战环境中的功能和影响，以避免不符合国际人道法的交战。操作人员应能够事后理解系统所采取的行动。
- 必须有能力制定精确的部署框架，遏制与系统环境不必要的相互作用，防止出现导致违反国际法和国际人道法的危急情况。
- 应制定国际法和国际人道法，以纳入合规法律审查，评估可靠性和可预测性，并引入认证程序。
- 在部署期间和有关使用武力的关键决策阶段，必须保持人对遵守国际法和国际人道法的情况的判断。操作人员应尽可能熟悉与系统有关的所有环境和技术因素，以确定其使用是否可能违反国际人道法原则。
- 致命自主武器系统不应能够就任务参数或任务的空间或时间框架作出自主决定。
- 如果致命自主武器系统完全自主行动，违背操作人员或指挥人员的意愿，后者应能停用该系统(“终止开关”)，以防止违反国际法和国际人道法。

应该永远保持人的责任，绝不能将其转移给机器。可以采取各种措施，包括以下措施。

- 从考虑将致命自主武器系统纳入军事武库之时起，就为使用此类系统制定精确的理论和程序。这些理论和程序应适应有关系统的技术特点。
- 对决策者和操作人员进行充分培训，使其充分了解致命自主武器系统的影响及其与环境的相互作用。
- 由人负责作出部署决定，并确定和验证致命自主武器系统的运作、使用和作战规则。
- 进行部署后审查，以评估遵守国际人道法的情况。
- 制定报告违反国际法和国际人道法行为的程序。
- 对被指控严重违反国际人道法的行为人进行纪律处分和起诉。

最后，历来受到边缘化的群体或歧视受害者在科学、技术、工程和数学领域的代表性不足，可能会在人工智能系统中造成严重偏见。生成式人工智能和机器学习功能可能会放大这些偏见。因此，对有关致命自主武器系统的新兴技术的各个方面采取对性别问题有敏感认识的交叉方法至关重要。

三. 运作和建议

卢森堡认为，致命自主武器系统带来的挑战最好在《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》的框架内加以解决。应讨论制定一项文书的要素，有可能作为《公约》的附加议定书，以确保致命自主武器系统符合国际人道法，并保持充分的人的控制和问责。

致命自主武器系统问题政府专家组应完成其目前的任务，即为可能拟定的文书制定一套要素，并在 2026 年举行的下一次《特定常规武器公约》审议大会上提出其结论，推动关于致命自主武器系统附加议定书的谈判。《公约》汇集了正在开发人工智能军事应用的国家，并通过谈判新附加议定书的能力，为解决致命自主武器系统问题提供了一个灵活的框架。专家组由来自政府、国际组织、民间社会和学术界的专家组成，汇集了审查致命自主武器系统新兴技术所需的技术、军事和法律专业知识。

保加利亚、丹麦、法国、德国、意大利、卢森堡和挪威在 2024 年政府专家组第一届会议期间提交的联合工作文件介绍了在《特定常规武器公约》框架内制定文书以解决致命自主武器系统问题的更多要素。¹

马拉维

[原件：英文]
[2024 年 5 月 25 日]

马拉维共和国常驻联合国代表团谨告知，马拉维共和国从人道主义、法律、安全、技术和伦理角度对有关自主武器系统的挑战和关切问题极为关切。因

¹ [CCW/GGE.1/2024/WP.3](#)。

此，马拉维共和国谨呼吁尽快就一项关于自主武器系统的具有法律约束力的文书开始谈判。谈判应在最具包容性的论坛上进行，对所有利益攸关方开放，该论坛应最有可能设定雄心勃勃的目标并取得成功。大会应考虑在 2024 年开始谈判，同时所有适当的多边论坛继续开展对话，以鼓励产生相辅相成的效果。

马拉维共和国认为，需要有一项条约禁止从根本上不可接受的自主武器系统，并管制其他自主武器系统的使用。人对攻击进行实际控制应该是监管框架的核心要素。人类用户必须能够作出法律和道德判断，并能够对任何违反法律和道德规范的行为承担责任。这意味着，出于法律和伦理原因，也必须禁止针对人的自主武器系统。所有其他自主武器系统都应通过限制和积极义务加以管制，确保人对武力使用的实际控制。在制定一项关于自主武器系统的具有法律约束力的文书时，应纳入对人权、道德、偏见的观点以及边缘化和受影响群体的观点。马拉维共和国随时准备与所有国家合作，紧急制定一项法律文书。

墨西哥

[原件：西班牙文]

[2024 年 5 月 24 日]

迫切需要对自主武器系统制定禁止和管制措施，因为此类系统对国际和平与安全带来风险，并与充分遵守国际法(特别是国际人道法)规则相抵触。

墨西哥推动并共同提出了题为“致命自主武器系统”的大会第 78/241 号决议，因为大会是主要多边政治论坛，其会员国几乎包括所有国家，关于这一事项的讨论之前分散在不同论坛，且鉴于此类系统构成多重挑战，应予以重视。

墨西哥强调支持拉丁美洲和加勒比自主武器的社会和人道主义影响问题会议(2023 年，哥斯达黎加)的公报，并赞同主题为“十字路口的人类：自主武器系统和管制挑战”的国际会议(2024 年，维也纳)主席的最后报告。

1. 风险和挑战

自主武器系统在开展敌对行为和保护平民及未参与武装冲突的人方面带来多重法律、伦理、人道主义和安全风险与挑战。

- 伦理角度：机器做出生死决定的风险令人担忧。
- 安全角度：此类系统可能导致不对称战争，将降低使用武力的门槛，增加意外冲突和冲突升级的风险，并可能加速军备竞赛。
- 人道主义角度：开发和使用时不受人实际控制、不能完全遵守国际人道法规则和原则的自主武器系统存在潜在风险。

取消人的控制不仅会导致人的判断被技术或行动指标所取代，而且还会导致无视人的评估以及在复杂和不断变化的情况中确定使用武力的合法性所必需的基于具体情景的定性判断。

2. 可能的定义

虽然尚未就自主武器系统的定义达成一致意见，但这种定义对于在管制方面取得进展并非至关重要。

这种定义不应过于具体，以免排除将需要受到管制的未来系统，也不应过于笼统或模糊，因为这可能会产生灰色地带，使一些自主武器系统能够以滥杀滥伤的方式运作。

2013年，法外处决问题特别报告员指出，自主武器系统是指任何具有自主性的武器系统，根据其设计，系统一旦启动，即可选择目标并对目标使用武力，而无需人为干预。

基本要素包括：(a) 自主性，即在选择平民和其他人等军事目标并对其使用武力方面有独立决策和不受控制地开展行动的能力；(b) 在选择和使用武力的关键功能时使用基于人工智能的技术；(c) 人为控制有限或没有人为控制。

墨西哥不认为致命性的特征是适当的，因为在国际法中没有使用该词，其合理性也没有得到证明。

3. 国际法的适用性

所有武器系统的开发、部署和使用都必须遵守国际法，包括《联合国宪章》、国际人道法、国际人权法、国际刑法和国际责任法。

在考虑对自主武器系统可能实行的禁止和管制时，适用国际人道法的一般规则和原则。

战争手段和方法不是无限制的。使用武力必须遵循人道原则和公众良心的要求(马顿斯条款)，这些原则和要求只适用于个人。

4. 国家和个人责任

人类必须因关于使用和部署自主武器系统的任何决定、任何使用武力的行为和任何违反国际法(特别是国际人道法)规则的行为而被追究责任并对此承担责任。责任和问责不能转交给机器。

自主武器系统必须在负责的指挥系统内运作，并有人实际控制。指挥官的存在使得有可能确定关于侵犯行为的责任归属以及犯下战争罪和其他严重罪行的个人的刑事责任归属。

墨西哥认为，违反国际人道法规则和原则的行为也可能引起国家对不可开脱的不法行为的责任。国家机构及其人员以及武装部队成员的行为归咎于国家。国家还应对危及国际和平与安全的任何使用武力(诉诸战争权)行为负责。

5. 未来管制的要素

有必要重申国际人道法的现有禁止规定，并禁止在本质上不能按照国际人道法规定的限制被使用、不能遵守攻击中的区分原则、相称原则和预防原则、也不能被充分理解、预测或解释的自主武器系统。

应禁止具有以下性质的自主武器系统：

- 不区分军事目标与民用物体、战斗人员与平民、现役战斗人员与失去战斗力的战斗人员。
- 无法确定某次攻击是否可能会造成与预期的具体和直接军事利益相比将是过分的意外平民伤亡或民用物体损害。
- 在显然目标不是军事目标或目标受到特别保护的情况下，没有取消或暂停攻击的机制。
- 造成过度杀伤或不必要痛苦。

墨西哥认为，人实际控制是适用国际人道法的一个隐含要素，因此必须界定这种控制的范围。

人参与选择目标和使用武力本身并不消除与国际人道法相容方面的风险。人实际控制应该：

- 确保适当限制运行环境，包括与时间和空间有关的方面。
- 通过充分了解态势，在行动背景已知的情况下核准决策。
- 确保在进行操作时采取预防措施，使得在没有人核查的情况下不能改变任务参数。
- 允许持续监测，以确保在必要时进行干预，特别是确保有可能在运行期间停止和停用系统。
- 确保人的决定是实质性的，而不仅仅是名义上的。

6. 需要一项具有法律约束力的文书

墨西哥赞同秘书长和红十字国际委员会主席的意见，认为有必要通过一项具有法律约束力的文书，该文书应规定对自主武器系统的禁止和管制，其目的是维持人对关键系统活动的实际控制，并包括与执行、监测和问责有关的有效措施。

各项国际条约体现了其他类型文书所没有的责任感和确定性；这些条约是确保所有国家真正平等参与的唯一途径，并避免了国家措施造成的管制碎片化情况。

大会第 78/241 号决议启动的进程将使大会能够指导国际领域对这一问题采取的办法，并推动就一个纳入民间社会、学术界和技术专家的有效监管框架进行谈判并通过该框架。

荷兰王国

[原件：英文]
[2024年5月17日]

荷兰王国欢迎有机会根据大会第 78/241 号决议就如何应对自主武器系统带来的挑战和关切提出意见。

2022 年 6 月，荷兰外交部长在致荷兰议会的一封信中阐述了政府在自主武器系统方面采取的政策立场的依据。¹ 荷兰国防部还正在起草关于这一主题的其他内部政策。

讨论管制自主武器系统问题的主要国际论坛仍然是根据《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》设立的致命自主武器系统问题政府专家组。荷兰致力于在《公约》框架内取得进展，因为该《公约》涵盖了自主武器系统领域的所有相关行为体，而且近年来的情况表明，尽管面临具有挑战性的动态发展，但取得进展是可能的。

荷兰的出发点是，必须负责任地并以符合国际法的方式设计、开发和使用自主武器系统。自主武器系统必须在其生命周期的不同阶段全程保持一定程度的人为判断和控制，以确保遵守相关规定。必须通过一项具有法律约束力的文书，为此最好在《特定常规武器公约》中增加一项新的议定书，明确禁止无法按照国际法(特别是国际人道法)设计、开发或使用的自主武器系统。其他类型的自主武器系统需要得到进一步的管制。

下节进一步概述荷兰的立场，并列出了需要进一步审议的关键问题。

定性

荷兰认识到需要对自主武器系统进行一般性定性。然而，目前国际社会尚未就“自主武器系统”的定义或定性达成一致。

在国家一级，荷兰对自主武器系统的一般性定性如下：“一种在启动后能够在没有进一步人为干预的情况下选择目标并对目标施加武力的武器系统”。如果在启动后需要人工投入以选择目标并对目标施加武力，但这种人工投入仅仅是名义上的，则该武器系统仍被视为自主武器系统。

在自主武器系统这一大类中，荷兰区分了两种自主武器系统，一种是未经人的批准即可改变任务、分配事项或目标(包括授予系统的适用接战规则)的自主武器系统，另一种是不具备这种能力的自主武器系统。第一个次类别涉及在本质上不可预测且在负责的指挥系统之外运作的武器系统。此类系统可能会在未来被禁止。

¹ 另见国际事务咨询委员会和国际公法咨询委员会的报告，《自主武器系统：监管和投资的重要性》(2022年4月11日)。可查阅 <https://www.advisorycommitteeinternationallaw.nl/publications/advisory-reports/2021/12/8/autonomous-weapon-systems>。

国际人道法的适用

国际法、特别是国际人道法的现有规则和原则适用于自主武器系统的设计、开发和使用。由于这些系统具有自主特征，例如使用武力所需的人的判断和控制水平较低，因此必须具体说明哪些国际人道法规则和原则特别适用。关于自主武器系统的设计、开发和使用，从国际人道法的角度来看，区分原则、预防原则和相称原则特别重要。为了遵守这些规则和原则，必须确保在设计、开发和使用的自主武器系统方面有一定程度的人的判断和控制。人的判断和控制的必要程度取决于不同因素，例如使用背景和武器系统的参数。具体说明现有规范应有助于各国履行其义务。

对于不能按照国际人道法设计、开发或使用的自主武器系统，应明确禁止。为确保有效实施和执行，一项新的具有法律约束力的国际文书(例如《特定常规武器公约》的新议定书)应明确界定这一禁止所适用的自主武器系统类型。自主武器系统的不同类别可包括：

- 旨在对平民或民用物体施加武力的自主武器系统。
- 本质上具有滥杀滥伤作用或造成过度杀伤或不必要痛苦的自主武器系统。
- 其设计方式使得人无法合理地预测和控制在不同情况下使用所带来影响的自主武器系统。这包括能够未经人的批准而改变任务、分配事项或目标(包括授予系统的适用接战规则)的自主武器系统。

为确保自主武器系统的设计、开发和使用符合国际法(特别是国际人道法)，应采取不同类型的措施。这些措施应取决于自主武器系统的类型、运作环境和最终用户。² 这些措施还应规定自主武器系统所需的人的判断和控制水平，以符合国际法，尤其国际人道法中的区分、预防和相称义务。以下是可以被纳入但需要各国进一步讨论的一系列措施：

- 对以下方面作出限制：(a) 自主武器系统可以攻击的目标的类型和数量；(b) 时间和空间限度。
- 保持人对部署自主武器系统的地理区域的态势感知。
- 创建一个直观的界面，通过该界面进行人机交互。
- 建立广泛的程序，确保自主武器系统得到测试、评价、验证和核实。
- 对自主武器系统进行广泛的法律审查。
- 对与武器系统互动的操作人员进行培训。

² 斯德哥尔摩国际和平研究所和红十字国际委员会，《如何实施人的控制》，第 26 页，2020 年。

责任

在国家责任方面，可根据国际法，使各国对涉及使用具有自主特性的武器系统的非法行为负责。在自主武器系统生命周期中发挥作用的个人或法律实体可酌情根据国内法承担民事责任，或受到刑事起诉。起诉可能的国际罪行的责任主要由国家法律系统承担，国际刑事法院则充当后盾。

国际人权法

在政府专家组中，荷兰一贯强调，无论是在武装冲突之中还是之外，都必须把人权作为一项规范自主武器系统的设计、开发和使用的法律制度。荷兰注意到，与国际人道法法律制度对作战行动的要求相比，人权法律制度对于执法目的的武力使用作出了更严格的要求。虽然在专家组的讨论中尚未详细论述国际人权法如何适用于自主武器系统的设计、开发和使用以及人工智能在军事领域的使用，但荷兰认识到在相关国际论坛上讨论这一专题非常重要。在国际法范围之外，使用自主武器系统带来了更多风险，例如人道主义、伦理和安全方面的风险。荷兰认为，必须也查明和减轻这些风险。

关于军事领域人工智能的更广泛讨论

由于人工智能是自主武器系统的重要实现手段，因此关于军事领域人工智能的更广泛讨论与关于自主武器系统管制的讨论之间存在明显的相似之处。荷兰认为，关于这两个专题的国际讨论是相辅相成、互惠互利的。

新西兰

[原件：英文]

[2024年5月21日]

本摘要反映了新西兰应 2024 年 2 月 1 日普通照会而提交的国家文件的要点。¹

新西兰寻求对自主武器系统实行具有约束力的国际禁止和管制。具体而言，新西兰寻求全面、具体地禁止不符合国际人道法的系统，例如以用户无法预测或控制的方式运作并且不符合相关伦理要求的系统。与此同时，新西兰寻求对其他自主武器系统进行适当管制，确保在这些系统的整个生命周期中实行充分的人的控制或监督，并充分遵守国际人道法。

自主性有不同的程度，因此可能必须明确制定一系列自主武器系统控制措施。符合伦理和法律要求的自主武器系统可能会带来合法的军事利益，因此新西兰认为，全面禁止自主武器系统目前并不具有建设性，也不会获得广泛支持。

新西兰还支持采取临时措施，例如不具法律约束力的准则、宣言或规范，以此作为达成具有法律约束力的文书的步骤和(或)作为此类文书的实际执行工具。

¹ 完整文件可查阅 <https://www.mfat.govt.nz/assets/Peace-Rights-and-Security/Disarmament/New-Zealand-submission-to-the-UN-Secretary-General-on-autonomous-weapon-systems.pdf>。

国际上尚未就自主武器系统的定义达成一致意见。任何得到广泛遵守的定义都应具有灵活性和技术中立性，以确保它仍然符合目的。

考虑因素

人道主义考虑因素

武装冲突继续产生广泛和跨代的人道主义影响。新西兰感到关切的是，不受限制地开发和自主武器系统可能会降低使用武力的门槛，因而增加冲突和随之而来的人道主义危机的频率和强度。

法律考虑因素

国际人道法完全适用于所有武器系统，包括自主武器系统。² 有些国家担心，是否有可能让所有类型的自主武器系统都在符合国际人道法的情况下运作。例如：

- 目前尚不清楚自主武器系统将如何遵守关于区分受保护人员和物体与战斗人员和军事目标的要求，特别是因为需要人根据具体情况作出判断。一个战斗员可能很快失去战斗力，这需要细致的定性判断能力和根据具体情况作出判断的能力。
- 冲突各方必须确保为追求军事目标而进行的所有攻击与任何附带的平民伤亡或损害是相称的，并采取可行的预防措施避免和尽量减少伤亡或损害。相称性在本质上是人的决策过程，取决于人在考虑攻击的预期军事优势时评估当前和未来可能发生的事件的能力。这需要融合各种形式的的数据，但也需要根据行动经验、法律框架、伦理和其他考虑因素进行复杂的判断过程，包括对整个行动的判断。由于冲突的性质，很难将相称性判断分解为易于复制或可编码的指令。
- 如果在评估是否和如何发动攻击时不能作出适当的相称性判断，这将使确定应采取的可行预防措施的过程变得更加复杂。
- 此外，目前尚不清楚如何可以确保对涉及自主武器系统的违反国际人道法行为追究法律责任。法律的执行取决于是否能让参与冲突的人对其行为负责。如果这一追究人的责任的系统被削弱或破坏，那么国际人道法就会受到损害。

武器审查³ 是关于遵守国际人道法的国家义务的重要组成部分，新西兰支持加强武器审查。虽然这种审查对自主武器系统是必要的，但新西兰认为，单靠这些审查不足以解决关于自主武器系统的关切问题，并认为应辅之以具体的规则和限制。虽然有进行武器审查的一般义务，但没有共同标准，也没有分享审查结果的要求。还有一些技术性问题是，审查如何能可靠地评估自主武器系统可能在任何特定场合下的运作方式。

² 如 [CCW/MSP/2019/9](#) 所申明。

³ 《1949年8月12日内瓦四公约关于保护国际性武装冲突受难者的附加议定书》(《第一议定书》)第三十六条。

新西兰还对自主武器系统如何能够符合国际人权法感到关切。自主选择目标的决定有可能是任意的，这可能违反人权法。新西兰还感到关切的是，用于选择目标和(或)决定使用武力的算法所依据的数据集存在偏见，这可能导致违反国际人权法和国际人道法的行为，具体情况视不同情景而定。这并不排除违反其他法律体系的行为，包括违反国际刑法的行为。根据国际刑法，使用自主武器系统可能构成国际犯罪。

安全考虑因素

自主武器系统有可能成为冲突和维持和平的不稳定因素。例如，部署部队的政治门槛可能变得更低，导致冲突的可能性增加，具体情况取决于所使用的不同类型自主武器系统。

几乎可以肯定的是，自主武器系统将会扩散，并迟早被非国家武装团体获得，用于武装冲突、恐怖主义或犯罪。同样可以预见的是，在一些国家，自主武器系统将流入国内执法部门并被其使用，这可能会引起更多国际人权法领域的问题。

技术考虑因素

新西兰主张采取技术中立的办法。为了确保未来的开发不会被无意中排除在外，谈判可以集中在需要新规则和新限制的自主武器系统的影响和特征上。如果不对自主武器系统施加适当的限制，且它们被证明具有无法控制的破坏性影响，则可能会在更普遍的层面上给人工智能应用的社会认可带来不可弥补的损害。

伦理考虑因素

伦理考虑因素与法律可接受性有关。对特定战争方法的伦理关切往往成为采取法律约束的动力。目前，人们对用包括机器学习在内的算法过程取代人的决策的伦理问题存在不同看法，在人(包括战斗人员)的伦理责任、人性和人的尊严方面也存在相关的悬而未决的问题。

新西兰还认识到，在确保使用高度自主的武器系统时根据具体情况作出判断方面存在一些悬而未决的问题。人根据具体情景作出判断对于执行国际人道法至关重要，例如在任何攻击中对相称性进行评估。目前尚不清楚机器在选择目标或发动攻击时如何应用国际人道法。即使自主武器系统可以作出这种判断，仍然存在一个问题，即如何确保持续就违反国际人道法的行为追究人的法律责任。人们一直就各种相互重叠、有时相互冲突的概念进行辩论，包括人的“充分”和“实际”控制等概念以及“人的适当判断”或“参与”等概念。需要加快这项工作，既要考虑法律因素，也要考虑伦理因素。

多边努力

新西兰一直积极参与致命自主武器系统问题政府专家组，并将继续这样做。专家组是一个有用的论坛，可供分享专门知识，并使《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》的缔约国和签署国与民间社会聚集一堂。然而，新西兰也认为通过大会进行更广泛和更具包容性的讨论是有价值的。大会采用普遍会员制，而《公约》有 126 个缔约国，其中大多数集中在全球北方。

挪威

[原件：英文]

[2024年5月24日]

挪威欢迎有机会根据大会关于致命自主武器系统的第 78/241 号决议提交其意见。该决议请秘书长征求会员国的意见，并编写一份实务报告，说明“如何应对此类系统从人道主义、法律、安全、技术和道德角度提出的相关挑战和引起的关切，以及人在使用武力方面的作用”。

挪威认为，至关重要的是，任何武器系统，包括具有自主功能的武器系统的开发和使用，都必须充分遵守各国根据国际法(特别是国际人道法)承担的义务。这在总体上适用于人工智能的任何军事应用，而不仅仅适用于自主武器系统的开发和使用。虽然这类系统也引起重大的道德关切，但各国目前解决此类关切问题的最有效方法是确保妥当地适用国际人道法。

挪威参加《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》政府专家组的主要目的是帮助明确说明如何将现有的相关国际法规则适用于自主武器系统。挪威认为，讨论表明，需要有一项文书明确说明现有国际法如何适用于自主武器系统。挪威仍然不知道作出这种明确说明的文书的具体形式和法律地位将是如何，但我们认为，根据《特定常规武器公约》制定一项关于自主武器系统的议定书将是一个理想的结果。

挪威赞同采用两级方法，即既禁止特定自主武器系统，同时也管制其他此类系统的使用。要使禁止切实有效，就必须采取具有约束力的形式。从这个意义上说，挪威支持制定一项禁止特定自主武器系统的具有法律约束力的文书。

挪威与保加利亚、丹麦、法国、德国、意大利和卢森堡一道，向政府专家组提交了一份工作文件(CCW/GGE.1/2024/WP.3)，除其他事项外，呼吁区分以下两类自主武器系统：

(a) 完全在人的控制和负责任的指挥系统之外运作的致命自主武器系统(即能够设定自身目标或在未经任何人工验证的情况下修改其初始方案或任务框架的系统)；

(b) 在关键功能方面具有决策自主性的致命自主武器系统(即在识别、分类、拦截和交战等关键功能方面具有决策自主性的致命武器系统，军事指挥部可在评估情况后，根据其职责在具体行动框架内向这些系统分配与关键功能相关的任务的计算和执行工作)。

关于自主武器系统的讨论应考虑到同时属于(a)和(b)两种情况的系统。理想的情况是，一项文书应适用于这两个类别，但也可以设想一项单独文书，如果这样做有利于谈判。挪威认为，这种区分的核心目标是确保任何被允许的自主武器系统都受到人实际控制。人对自主系统的控制必须是实际和具有相关性的，也就是说人的投入必须足以确保系统的运作完全符合包括国际人道法在内的国际法。

人对(a)类系统的控制不足以满足国际人道法的要求。因此，这种系统必须已被视为是现行法律所禁止的。为明确起见，我们认为未来文书应明确规定这一禁止。此外，虽然属于(b)类的系统在某些条件下可能是被允许的，但其开发和使用必须符合国际法。因此，被禁止和被允许的系统都可能属于这一类别。在下列情况下，此类系统必须被视为是被禁止的：

- 它们本质上是滥杀滥伤的，而且(或)无法区分军事目标和(或)敌方战斗人员与平民、失去战斗力的战斗人员和其他受保护的物体或人员；
- 在使用此类系统的情况下，无法按照国际人道法的要求对其影响进行限制、预期和控制；
- 在本质上会造成过度杀伤或不必要痛苦；或
- 出于任何其他原因，系统不能按照国际人道法运作。

国际社会应紧急展开实质性讨论，明确国际人道法对自主武器系统的适用。如果需要就国际人道法作出新的明确说明，甚至制定新的规则，以确保国际人道法的基本原则充分适用于自主系统，则各国应在未来文书中详细阐述这些规则。迫切需要响应秘书长的呼吁，在 2026 年之前完成关于这项文书的谈判。

由此可见，一些自主武器系统的开发和使用将不会被禁止。被未来文书涵盖但不受其禁止的系统，只有在完全符合国际人道法的情况下才可以被开发和运用。各国应拟订并商定一个框架，使国际人道法适用于这些系统。这一框架应确保此类武器系统的可靠性、可理解性和可预测性。该框架应落实措施，控制、限制或以其他方式影响此类武器系统可攻击的目标类型，以及此类武器系统运作的持续时间、地理范围和规模。应采取措施，在任何时候、任何情况以及在系统的整个生命周期中保持和保留人的控制。

为了能按照国际人道法对系统进行操作，挪威认为首先操作人员必须充分了解武器系统的运作方式、影响以及可能与环境产生的相互作用。人类必须能够预测和解释此类武器系统的行为。人类应始终根据具体情况和背景，确定和验证使用规则、交战规则和分配给此类系统的任务的精确框架，这些任务在时间、空间和确定的目标方面受到限制。人类必须能够在部署系统期间监测系统的可靠性和可用性，并审批任何对任务参数的实质性修改。人类必须能够判断对国际人道法规则和原则的遵守情况，特别是攻击中的区分原则、相称原则和预防原则，从而对使用武力作出关键决定。

一些组织呼吁禁止自主武器系统以人类为目标。挪威认为，自主武器系统以人类为目标这一情况本身提出了一些重要问题，各国应在制定未来文书时讨论这些问题。

《特定常规武器公约》政府专家组仍然是挪威进行这些讨论的主要场合。如果其他论坛的讨论能够促进或以其他方式有助于按照政府专家组的任务规定紧急拟订和通过关于根据《特定常规武器公约》制定的一项文书的建议，那么这些讨论也将是富有成果的。

巴基斯坦

[原件：英文]

[2024年5月3日]

致命自主武器系统的开发、部署和使用带来了人道主义、法律、伦理和安全方面的挑战。

人道主义和法律角度

国际人道法的本质是以人的判断和酌处能力为前提的，而这些能力在本质上对具体情况非常敏感。国际人道法的实施一直依赖于人的能动性。

从遵守和维护国际人道法基本原则(例如攻击中的区分原则、相称原则和预防原则以及军事必要性原则)方面来看，把武器系统关键功能(即目标选择和交战，包括施加武力)的使用权交出去将极其成问题。

国际人道法载有关于战争手段和方法的规则以及关于敌对行为的规则，该法采用“基于效果”的办法来设定特定限制。因此，如果有一种武器系统所产生的影响无法在所有情况下都得到预测、理解或解释，那么这种系统就会带来重大挑战。此类武器系统将不具备所需的控制力，无法根据国际人道法的规定对效果实行限制。

致命自主武器系统的使用还引起了与责任和问责有关的其他法律挑战。在选择目标和交战方面缺乏人的决策会引起一些问题，即国家对使用致命自主武器系统实施的非法行为或国际不法行为的责任、问责和归责问题。

伦理角度

关于将杀死人类的决定权交给机器这一问题，已有人提出了一些伦理方面的关切。使用致命自主武器系统减少了在战斗情况下表现同情或进行伦理推理的机会，在需要同理心、价值判断或对人类情感的理解的复杂伦理决策过程中尤其如此。

安全角度

开发、部署和使用致命自主武器系统对国际和区域两级的安全与稳定造成严重影响。

由于使用致命自主武器系统可能可以消除使用国的人员伤亡风险，因此这些系统更容易被使用，从而降低了施加武力和武装冲突的门槛。

在危机时期，使用武力的门槛过低将极大地破坏稳定，因为这将增加误判风险，引发冲突升级。在这种关键时刻完全去除人的控制可能会导致不可预测的后果和局势的迅速升级。

按机器的速度加快战争进程、压缩决策时间、消除缓解危机的余地是引发一场不可遏制的灾难的成因，在局势高度紧张和曾发生意外发射事件的地区尤其如此。

致命自主武器系统还可能推动不对称的战争手段和方法，包括匿名和秘密行动，例如在外国领土上进行有目标的杀害。

如果没有实行任何实际限制，致命自主武器系统的开发可能引发代价高昂的新军备竞赛，因为其他国家可能会感到必须获得这些技术以保持战略均势或优势。这些事态发展还将增加此类武器系统向非国家行为体扩散的风险。致命自主武器系统也可能容易受到网络攻击。

需要制定关于致命自主武器系统的新国际法律规则

正如致命自主武器系统问题政府专家组于 2019 年通过的指导原则所确认，国际人道法完全适用于所有武器系统，包括致命自主武器系统的潜在开发和使用的。但是，如果仅仅说明这一点便足矣，就没有必要再增加《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》各项议定书所载的具有法律约束力的规则。

就致命自主武器系统而言，需要制定新的国际法律规则，理由如下：

- 需要申明和进一步阐述现有的国际人道法义务，以明确说明这些义务如何适用于致命自主武器系统。
- 需要通过更多的衔接原则，以促进执行现有的国际人道法，并使其得以逐步开发，从而应对与致命自主武器系统有关的人道主义挑战。
- 现有的国际人道法没有充分解决致命自主武器系统带来的人道主义、伦理、法律和安全挑战。
- 需要设立新的国际法律义务，以全面、综合的方式应对重大风险。

一项关于致命自主武器系统的新国际法律文书的要素

应继续根据《特定常规武器公约》开展努力，以便通过订立一项新的议定书来制定国际法律规则。这项议定书将需要明确规定对致命自主武器系统的禁止和限制。这种包含禁止和限制的两级方法也符合《特定常规武器公约》的现有议定书。

巴基斯坦谨提出一项关于致命自主武器系统的国际法律文书的以下核心要素，可以就这些要素进行谈判，并将其作为《特定常规武器公约》议定书六予以通过。

定义

新文书应采取一种功能性做法，在其范围内涵盖致命自主武器系统的一般类别。新文书应包括以下定义：

- “致命自主武器系统”指被设计成在启动后无需人为干预即可选择目标并对目标施加武力的武器系统。

- 使用“致命”一词意味着，根据其设计，有能力施加致命武力的自主武器系统被列入致命自主武器系统的类别。不过，无论使用致命自主武器系统的实际后果如何，无论所施加的武力是否造成致命影响，此类系统都将受到议定书的管制。

禁止

在所有情况下均禁止开发、生产、获取、储存、保留、转让、部署或使用具有以下特性的致命自主武器系统：

- (a) 在没有人为控制的情况下即可作出使用武力的决定，并且无法使人类用户理解、预测和解释在所有使用情况下该系统作出的使用武力决定；
- (b) 就其设计而言，系统的使用无法在所有情况下都符合国际人道法；
- (c) 会产生一些不能在所有情况下都按照国际人道法要求加以限制的影响。

限制

对于不受上述禁止的致命自主武器系统，应采取以下措施：

- (a) 将目标仅限制于在本质上是军事目标的物体；
- (b) 确保在使用的所有阶段，为人类用户提供干预、中断和停用系统的手段；
- (c) 限制系统运作的时长、地理范围和规模，确保在任何时候都符合国际人道法，而且只限于没有平民或民用物体聚集的地区；
- (d) 按照国际人道法的要求，限制交战次数，以便预测和保持控制影响；
- (e) 限制武器系统在未经人类用户批准的情况下自主改变其选定目标参数；
- (f) 确保在所有时候都对系统的运作进行有效的监管和监督。

其他规定

其他规定应包括：

- 根据国际法和关于国家对国际不法行为的责任的适用规定，对致命自主武器系统负责并控制这些系统的人应始终对使用此类武器的后果负责。
- 各国必须制定并确保有效的监督、调查和补救机制，处理与开发、部署或使用致命自主武器系统有关的任何可疑的、被报告或记录的违规行为。

国际法律文书还应包括一个技术附件(正如《特定常规武器公约》下的经 1996 年 5 月 3 日修正后的《禁止或限制使用地雷、诱杀装置和其他装置的议定书》(经 1996 年 5 月 3 日修正后的议定书二)和《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战争遗留爆炸物议定书》(议定书五)的做法一样)，附件应载有最佳做法、准则和其他减少风险的措施，以促进执行该文书的各项条款。这一附件可包括关于国家一级行动的规定，包括政策、试验、培训和审查(包括武器审查和报告)方面的规定。

结论

根据《特定常规武器公约》设立的致命自主武器系统问题政府专家组凭借其掌握的所有工具和迄今为止所做的工作，仍然是制定切实法律规范框架以全面应对与致命自主武器系统有关的挑战和关切问题的最适当论坛。

菲律宾

[原件：英文]

[2024年5月16日]

人工智能和自主武器系统等先进技术的发展正在迅速改变人类的生活和经验。这些新兴技术的即将扩散可能解决很多老问题，但也可能会扰乱政治和社会秩序。因此，菲律宾认为，国际治理结构必须跟上脚步。

为了提供一个空间，从印太地区的角度讨论和考虑在武器系统中使用自主技术的影响，菲律宾共和国外交部与非暴力国际组织(东南亚)合作，于2023年12月13日和14日组织了自主武器系统印太观点马尼拉会议。

菲律宾认为，有必要加紧谈判，争取达成一项强有力、经得起未来考验的具有法律约束力的文书，以应对自主武器系统的威胁。

菲律宾认为，《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》是一个适当的框架，可在此框架内通过制定一项包括禁止和管制自主武器系统的新议定书，应对自主武器系统带来的潜在威胁，包括非国家武装行为体可能获取自主武器系统的问题。

为了防范与开发自主武器系统有关的风险和挑战，需要下列具有法律约束力的规则和原则：

(a) 自主武器系统的定性。在拟订和发展关于自主武器系统的任何规范性和操作性框架以及就框架进行谈判时，不需要对自主武器系统进行确切定义。自主性具有不同的程度，鉴于技术迅速发展，单纯的技术特征可能不足以对自主武器系统进行定性。以下的现行定性方式是一个有用的出发点，该方式的重点是人的因素和人机互动，这对解决归责和责任问题至关重要：

“自主武器系统”是指将自主性纳入选择目标、导向目标、交战和对目标施加武力等关键功能的武器系统。

(b) 禁止。在下列情况下，禁止在任何情况下开发、生产、拥有、获取、部署、转让或使用自主武器系统：

(一) 系统的自主功能被设计用于人实际控制范围之外；

(二) 系统的使用无法符合国际人道法原则或公众良心的要求，包括以下情况：
(1) 无法区分平民、敌方战斗人员与失去战斗力的战斗人员；(2) 具有造成过度杀伤或不必要痛苦的性质或本质上具有滥杀滥伤作用。

(c) 管制。必须落实管制，包括确保以下几点：(1) 在任何具有自主性的武器系统的整个生命周期中，保留人实际控制；(2) 武器系统不依赖于可能延续或扩大社会偏见(包括性别和种族偏见)的数据集；(3) 制定管制措施，确保妥当登记、跟踪和分析自主武器系统。

人实际控制。菲律宾认为，出发点是人类必须始终对使用武力保持实际控制，且人类目标决不能被沦为仅供机器人和算法解读的数据。虽然使人在使用武力中不再占据中心地位可能会产生与问责有关的法律影响，但菲律宾提出的前提超越了法律考虑因素，而是源于人的尊严不可侵犯这一基本前提。在这方面，菲律宾提出了“人实际控制”作为评估自主武器系统开发的标准。

“人实际控制”涉及维护人的能动性，包括维护人对使用武力的判断和干预。这包括以下要素及其他要素：

(a) 重新确定或修改武器系统的目标或任务或以其他方式使其适应环境的能力；根据需要停用、中止、终止或中断系统的运作和使用，并对系统的自行启动功能进行限制的能力；

(b) 限制武器系统使用的范围和规模(包括时间和空间限度)的能力，以及限制系统的选择目标参数和选择目标手段的能力；

(c) 理解和解释武器系统运作的的能力，以便在事后提供满足有关武器系统运作的法律要求和其他要求的解释，包括关于归责和问责的要求。

国际法的适用性。菲律宾申明，习惯国际法、《联合国宪章》、国家责任法、国际人道法、国际人权法和国际刑法等国际法全部均适用于任何自主武器系统的开发、使用和转让。

武器审查。国际法的适用性的一个关键含义是，各国有义务进行国家武器审查。在研究、开发、获取或采用任何新的武器和战争手段或方法时，必须确定其使用是否在某些或所有情况下为国际法所禁止。

伦理考量。菲律宾指出，伦理考量是有关自主武器系统和人工智能军事应用的多边对话的核心。各国有义务在法律论据范畴之外确保人的尊严不可侵犯。人工智能的任何军事应用都不应导致人类生命在武力使用自动化的情况下沦为仅供机器人和算法解读的数据。

对环境和生态完整性的影响。自主武器系统的存在引起了对环境脆弱性的关切，特别是其对海洋环境和生态系统的影响。在涉及自主武器系统的讨论中，必须考虑海洋和生态方面的问题。例如，有人认为由于在水下对人造成附带损害的风险较低，因此自主潜艇战相对更为可取。然而，太平洋岛屿国家的代表在自主武器系统印太观点马尼拉会议上指出，对海洋生态的破坏直接关系到他们国民的生计。

人工智能的其他军事应用。菲律宾认识到，人工智能的军事应用可能会带来好处，特别是在精确性不断提高的情况下。这种精确性可有助于军事指挥官和操

作人员确保遵守国际人道法。在这方面，菲律宾不反对人工智能在自动化相关功能方面的军事应用，例如导航系统、作战管理系统和精确制导等。菲律宾支持在人工智能应用方面的自愿性负责任行为规范，包括武器审查。

战略性贸易管制。实行战略性贸易管制的国家和多边出口管制制度的成员应共同努力，商定致命自主武器系统的功能定义，找到可能用于管制的控制措施。其中涵盖可能成为致命自主武器系统前身的新兴技术，包括但不限于军用级无形技术，例如旨在为进攻性武器自主识别目标的算法和程序。

各国还应考虑在技术上明确区分致命自主武器系统与自主防御系统。特别是，类似于自动防空和导弹防御系统的防御设备不是致命自主武器系统，因为它们只是发射拦截器合法打击已正在接近的敌机和射弹。应把拦截器和此类类似技术与进攻性系统发射的弹药区分开来，因为前者只应对武装冲突，而不会煽动武装冲突。这种区分应使各国能够受益于用于防御的自主系统所提供的保护，同时避免此类系统被用于进攻目的时所具备的滥杀滥伤性质。

菲律宾还提交了其关于自主武器系统印太观点马尼拉会议的报告，作为提交的这些意见的附件。¹

大韩民国

[原件：英文]

[2024年5月24日]

大韩民国认识到新技术和新兴技术的迅速发展为增进人类福祉带来巨大希望，尤其在某些情况下可有助于在冲突中更好地保护平民；然而，大韩民国也意识到人们对自主武器系统可能产生的消极后果和影响的关切。

在这方面，大韩民国非常重视解决致命自主武器系统问题，在我们考虑新技术和新兴技术可能产生的影响时，会在利益与风险之间以及人道主义考虑因素与军事需求之间保持平衡。以下观点是根据大韩民国对致命自主武器系统的现有理解提出的，据大韩民国理解，这种武器系统一旦启动，将能够在没有操作者进一步干预的情况下识别、选择和攻击目标。

大韩民国赞赏《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》缔约国在这些问题上作出的持续努力。我们特别赞扬和欢迎自根据《特定常规武器公约》设立的致命自主武器系统问题政府专家组于2017年召开第一次会议以来所做的宝贵工作和取得的进展。11项指导原则于2019年得到核可，两级方法被纳入政府专家组2023年会议报告，这些是经协商一致达成的重要里程碑。

大韩民国继续强调，政府专家组始终应该是处理与致命自主武器系统有关的各种问题的独特核心论坛，其重点是有效执行适用于所有武器系统的国际人道法。

¹ 秘书处存档，可供查阅。

大韩民国认识到技术的迅速发展，因此为了推动政府专家组的讨论，大韩民国强调缔约国的做法应当是实际、现实和务实的，并以科学为依据。政府专家组的工作应着重于拟订禁止和管制要素，拟订方法包括就需要采取哪些措施确保遵守国际人道法达成基于协商一致的协议。

在这方面，大韩民国申明，如果致命自主武器系统因其具备的性质而无法按照国际人道法使用，例如由于这些系统会造成过度杀伤或不必要痛苦，或本质上具有滥杀滥伤作用，或因其他原因而无法按照国际人道法使用，则应予以禁止。这就需要继续讨论如何管制其他致命自主武器系统，以确保充分遵守国际人道法。

然而，必须指出，人在一定程度上的参与并非是确保遵守国际人道法的必要条件。相反，在评估遵守国际人道法的情况时，要视具体情况而定，并且攻击中的区分原则、相称原则和预防原则应发挥核心作用。

此外，为了充分应对致命自主武器系统的相关潜在风险，大韩民国认为，将需要引入减轻风险措施，确保在进行攻击时遵守区分原则和相称原则，在攻击中采取预防措施，并确保问责，包括为军事人员提供培训，在武器系统的整个生命周期中保留负责任的人的指挥和控制系统，并制定数据保护等适当的保障措施。我们还认为，分享减轻风险做法方面的经验教训以及关于致命自主武器系统的国家框架、战略和政策的信息，将能促进这一努力。[CCW/GGE.1/2023/WP.4/Rev.2](#) 号文件所载的联合工作文件中明确说明了大韩民国的基本立场。

大韩民国认识到正在开展各种国际举措，例如军事领域负责任人工智能峰会和《关于在军事上负责任地使用人工智能和自主的政治宣言》，因此大韩民国强调所有相关国际努力应继续与其他举措相辅相成、协同合作。

摩尔多瓦共和国

[原件：英文]

[2024 年 5 月 24 日]

根据其《宪法》第 8 条第(1)款，摩尔多瓦共和国承诺遵守《联合国宪章》和它加入的条约，并在得到一致承认的国际法原则和规则的基础上与其他国家建立关系。

这项原则载于摩尔多瓦共和国关于国际条约的第 595/1999 号法律，其中第 19 条规定：“根据条约必须遵守原则，应秉持诚意执行国际条约。摩尔多瓦共和国不得援引其国内法的条款作为不执行其加入的条约的理由”。

摩尔多瓦共和国考虑并充分尊重关于致命自主武器系统领域新技术的准则以及国际人权条约和国际人道法的各项条款。目前，摩尔多瓦加入了国际人道法领域的若干国际条约，包括关于武装冲突受害者、战争手段和方法以及国际罪行的条约。

摩尔多瓦共和国通过第 318/1993 号议会决定加入了 1949 年 8 月 12 日日日内瓦四公约及其 1977 年 7 月 10 日附加议定书。

摩尔多瓦共和国通过第 975-XIV/2000 号法律加入了 1980 年 10 月 10 日在日内瓦通过的《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》及其所附以下议定书：《关于无法检测的碎片的议定书》(议定书一)，1980 年 10 月 10 日在日内瓦通过；经 1996 年 5 月 3 日修正后的《禁止或限制使用地雷、诱杀装置和其他装置的议定书》(经 1996 年 5 月 3 日修正后的议定书二)，最初于 1980 年 10 月 10 日在日内瓦通过；《禁止或限制使用燃烧武器议定书》(议定书三)，于 1980 年 10 月 10 日在日内瓦通过；《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》附加议定书(题为《关于激光致盲武器的议定书》的议定书四)，于 1995 年 10 月 13 日在维也纳通过。

现有国际人道法(《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》(《第一议定书》)关于新武器的第三十六条)规定：“在研究、发展、取得或采用新的武器、作战手段或方法时，缔约一方有义务断定，在某些或所有情况下，该新的武器、作战手段或方法的使用是否为本议定书或适用于该缔约一方的任何其他国际法规则所禁止”。

《1949 年日内瓦四公约第一附加议定书》和 1980 年在日内瓦商定的《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》及其议定书一至四(自 2001 年 3 月 8 日起对摩尔多瓦共和国生效)没有界定或规范对致命自主武器系统的禁止。

摩尔多瓦共和国议会通过第 212/2010 号法律批准了 1998 年 7 月 17 日通过的《国际刑事法院罗马规约》。

为履行所承担的国际义务并确保遵守武装冲突法规则，国防部长第 275/2006 号决定核准了关于在摩尔多瓦共和国国民军中执行武装冲突法规则的指示。

根据这一指示，指挥官和首长为完成分配给他们的任务而开展的大部队、军事部队和分队准备工作以及他们在军事行动中采取的行为必须遵守战争法原则。他们还必须确保遵守武装冲突法的规则，促进采取一切可能的预防措施，以避免或在无法避免的情况下尽量减少平民的损失和平民财产的损失。

即使违反武装冲突法的行为可能是由下属犯下的，也不能免除其上级在纪律或刑事方面的责任。在准备和开展军事行动的过程中，各级指挥官和首长有义务考虑在行政活动过程中对施加武力实行某些限制和约束。

第 2327/2009 号总统令核准的《摩尔多瓦共和国武装部队内部服役条例》也载有相关规定，涉及了解和严格遵守关于武装冲突的国际法要求的义务。

此外，在致命自主武器系统领域的新技术背景下，在武装冲突和作战行动中需要尊重国际人道法，这意味着必须适用 1977 年通过的《1949 年日内瓦四公约第一附加议定书》关于新武器的第三十六条和关于攻击时预防措施的第五十七条的规定。

因此，我们同意，国际人道法也适用于这些系统，且使用这些系统的决定必须由个人作出。各国应在构想阶段考虑它们所开发或获取的新武器的合法性。

然而，根据现有国际文书，和平时期的个人和国家以及武装冲突局势中的战斗人员、军事组织和国家有义务不将使用致命武力的权力或能力交给机器或自动化程序，以便在每一种情况下，都应由人来确定致命武力的使用是否正当、符合伦理和合法。

如果自主武器系统在那些操作者的控制或指挥下造成暴行，这可能会削弱关于指挥官控制其下属行动的责任和义务的概念，从而使指挥官免受在其他情况下可能被视为战争罪的指控。

鉴于目前把致命自主武器系统的开发和军事使用用于消灭有生力量的趋势日益明显(如安全理事会一份报告所述，这发生在 2020 年利比亚内战期间，是有史以来第一次)，以及全球面临和平与安全方面的风险，包括扩散风险，也包括非国家武装团体面临的风险，因此有必要创建和审查国际规范框架，限制和禁止致命自主武器系统的军事使用，包括与人工智能相结合的使用，并禁止在核武器、化学武器和生物武器等方面使用致命自主武器系统。

鉴于与人工智能相关的致命自主武器系统的开发和军事使用方面的区域和全球风险和威胁日益增加，我们认为有必要根据《联合国宪章》、国际人道法和各项国际条约的原则，在国际和国家层面界定和规范部分或全部禁止使用致命自主武器系统的情况，从而规范使用致命自主武器系统可能被禁止的情况，包括禁止使用与核武器、化学武器和生物武器等有关的致命自主武器系统。然而，要有效执行这些关于限制和禁止致命自主武器系统的法规，就需要建立国际和国家机制，对致命自主武器系统(包括与人工智能以及核武器、化学武器和生物武器有关的系统)的非法开发和军事使用进行监测、控制和追究法律责任。

俄罗斯联邦

[原件：俄文]

[2024 年 5 月 8 日]

俄罗斯联邦欢迎通过联合国大会 2023 年 12 月 22 日第 78/241 号决议，并根据该决议第 2 段，谨向大会第七十九届会议提交本国对秘书长报告的意见，供会员国进一步讨论。

我们认为，《特定常规武器公约》是处理致命自主武器系统问题的最佳框架。该《公约》缔约国设立了致命自主武器系统领域新兴技术问题政府专家组，该机构的职责是在有关此类武器的人道主义关切与各国合法自卫利益之间保持合理平衡，并在协商一致基础上作出决定。我们认为，将这一问题转交到联合国等任何其他国际平台，将适得其反。

俄罗斯联邦欢迎《特定常规武器公约》缔约国会议批准专家组直至 2026 年的新的讨论任务。我们希望，专家组将能够在考虑到所有缔约国意见的情况下，就提交《公约》缔约国第七次审议大会的实质性结论和建议达成一致。

根据现有国际法，没有就致命自主武器系统达成协商一致的定​​义，因此很难解决这个问题。就致命自主武器系统形成共同的可行理解，将使这一问题和今后关于这一专题的讨论更加明确。

可行定义应：

- (a) 说明属于致命自主武器系统类别的武器类型及其使用的具体关键特征；
- (b) 不限于对致命自主武器系统的现有理解，而是考虑到这种系统在未来可能会如何演变；
- (c) 得到专家界(包括科学家、工程师、技术员、军事人员、律师和伦理学专家)的普遍理解；
- (d) 不得被解释为限制技术发展或破坏和平的机器人技术和人工智能领域的研究；
- (e) 不能仅仅通过描述其功能来定义致命自主武器系统。

应避免将这些武器归类为“坏武器”或“好武器”；换言之，不应根据某一特定国家集团的政治倾向对此类武器进行分类。

我们认为，将致命自主武器系统定义为旨在没有任何操作人员参与的情况下执行战斗和支援任务的完全自主无人操作技术武器，而不是弹药，是正确的。

不应将现有高度自动化的军事系统归入需要进行紧急限制和约束的“特殊”类别。正是这种自动化水平使此类系统能够在动态的战斗情况和各种环境中有效运作，并保证有足够的针对性和准确性，从而确保它们符合国际人道法的原则和规范。

人们普遍认为，国际人道法等现有国际法完全适用于致命自主武器系统。国际人道法包含一些重要的限制，包括：

- (a) 不应滥用或过度使用致命自主武器系统，也不应对平民使用或在不采取保护平民预防措施的情况下使用；
- (b) 致命自主武器系统的任何军事用途都应符合军事需要与所造成损害相称的原则；
- (c) 关于使用致命自主武器系统的适当性、形式和手段的决定，应由规划军事行动和设想使用此类系统(向此类系统分配任务)的情景的人作出。

1977年《1949年日内瓦四公约第一附加议定书》也载有一项重要义务，即在研究、开发、获取或采用新武器时进行法律审查。然而，议定书一第36条并没有确切规定应如何进行法律审查，也没有规定各国义务公布审查结果或向任何人提供这方面的信息。实际上法律审查是由各国在国家一级处理的。因此，也制定某种普遍的强制性机制来进行此类审查，特别是专门审查致命自主武器系统的机制，将是徒劳的。尽管如此，我们不反对在考虑到国家安全和商业机密因素的同时自愿交流履行议定书一第36条规定的义务的最佳做法。

俄罗斯联邦认为，目前没有令人信服的理由对致命自主武器系统施加任何新的限制或约束，或者为处理此类武器而更新或调整国际人道法。为了就致命自主武器系统的某种“行为规则”达成协商一致而展开讨论也为时过早。

人道原则、公众良心的要求和人权组成部分不能被用作对某些类型武器施加限制和约束的绝对和唯一充分条件。

我们相当重视各国普遍加入 1949 年日内瓦四公约及其三项附加议定书。我们认为，应通过忠实执行现有国际法律规范来解决关于致命自主武器系统的关键问题。

我们认为一个重要的限制是人类应该控制致命自主武器系统的运作。因此，此类系统的控制回路应允许操作人员或上级指挥系统进行干预以改变此类系统的运作模式，包括部分或完全停用此类系统。不过，人为控制的具体形式和方法应由各国自行决定，直接控制不一定是唯一选项。

可以通过以下方式对这些系统进行控制：

- (a) 提高系统的可靠性和容错性；
- (b) 限制目标类型；
- (c) 限制系统运作的时间范围、系统的地理覆盖范围和使用规模；
- (d) 立即采取干预措施并停用系统；
- (e) 在实际运作环境中对系统进行试验；
- (f) 允许已成功掌握使用此类系统程序的人员操作(控制)这些系统；
- (g) 监控单个元件和整个设备的制造；
- (h) 监控单个元件和整个装置的拆除和处置。

我们认为，将某些国家提倡的“人实际控制”和“人的参与形式和程度”等概念纳入讨论是不适当的，因为这些概念一般不具有法律意义，只会导致讨论政治化。

俄罗斯联邦认为，根据国际法，国家和个人(包括开发商和制造商)在任何时候都要对决定开发和使用致命自主武器系统领域的新技术承担责任。使用这些系统的责任在于给系统指派任务和下令使用系统的官员。要使用致命自主武器系统，该官员应具备与系统的运作和操作有关的必要知识和技能，并应负责就使用系统的适当性作出决定，并规划使用系统的形式和手段。

众所周知，致命自主武器系统在执行指定任务时可能比操作人员更有效，并且可能降低失误可能性。此类系统能够大幅减少使用武器对国际人道法的负面影响，这些影响与操作者的失误、其精神或身体状态及其伦理、宗教或伦理信仰有关。按照国际人道法使用此类系统可以确保更准确地将武器瞄准军事设施，并有助于减少对平民和民用物体造成意外打击的风险。

评估与使用致命自主武器系统有关的潜在风险和减轻这些风险的措施，应成为设计、开发、测试和部署任何类型武器系统新技术过程的一部分。

可通过以下方式尽量减少与此类系统有关的风险：

- (a) 确保有效的生命周期管理；
- (b) 在生命周期的所有阶段进行全面测试，包括在接近现实生活的环境中进行测试；
- (c) 确保系统的可靠性和容错性；
- (d) 制定准备工作标准；
- (e) 确保最大限度地防止未经授权的访问；
- (f) 培训操作人员；
- (g) 在收集和处理信息时优先使用人工智能技术，以支持军事决策；
- (h) 为操作人员连续监视此类系统的运行情况提供便利，并使系统能够在操作人员的指挥下紧急终止战斗任务；
- (i) 防止此类系统落入非国家行为体手中，它们可能会把系统用于非法目的。

我们认为，促进各国就致命自主武器系统和国际法专题自愿交换意见将是有益的。

俄罗斯联邦请秘书长在根据大会第 78/241 号决议第 2 段提交的实务报告中考虑到上述提议，并将本文件列入该报告附件。

塞尔维亚

[原件：英文]
[2024 年 5 月 18 日]

塞尔维亚政府与联合国一样，对致命自主武器系统的开发感到关切，塞尔维亚政府认为应禁止这些系统，或限制其生产和监测其使用情况。

现有数据表明，操作武器系统的人工智能无法达到与人类相似的智能和意识水平。在人类道德和伦理缺失的情况下，应该问一问，有什么能防止过度使用致命武力和造成不必要的破坏和死亡，并应该提出关于附带损害的问题，即导致无辜生命丧失的人所承担的道德和刑事责任是否将被免除，以及哪些措施和武器应该是自主的。致命自主武器系统旨在最大限度地减少己方的损失和资源消耗，最大限度地增加敌方的损失和资源消耗，再加上道德和伦理规范被摒除，这可能意味着在未来核导弹将配备此类系统。

在这方面，系统确实有可能被黑客入侵，并可能出现大规模生产所固有的软件和硬件异常情况。在这种情况下，死亡和毁灭是唯一必然产生的结果，而且几

乎不可能确定责任。因此，塞尔维亚政府对这类系统的破坏性潜力及其对和平及区域和全球安全的负面影响表示严重关切，理由如下：

- 可能会降低冲突的门槛，部分原因是人们相信因系统具有较高准确性，军事人员伤亡人数将减少；
- 出现军备竞赛，特别是在一些国家拥有此类系统而另一些国家没有的不对称情况下；
- 非国家暴力行为体、犯罪集团和恐怖主义集团拥有这些系统，造成无法预见的后果；
- 由于识别不足，这些系统规避现有的国际公约、控制和管制；
- 成群的一次性无人飞行器，无论其规模大小，都可能造成大规模伤亡。单独一个人就可以进行这种攻击，而防御则是困难和复杂的；
- 不正确或被误解的信息可能会导致作出错误决策和选择错误目标。

塞尔维亚是一个军事中立的小国，并且在 1999 年曾遭受仅稍早一代的类似自主武器系统的攻击，塞尔维亚认为，将这类或更复杂的系统用于军事目的绝对是毫无意义的，原因有二：

- 能够对目标选择做出独立决定的自主武器系统不能完全模仿人类操作员。它们缺乏同理心，这也许是在遵循“效率/效力”原则的情况下决定使用武器和实现战斗目标的方法的最重要特征。
- 自主武器系统无法理解可能产生的后果，无法确定每个后果的严重程度，也无法评价和比较后果。因此，自主武器系统在自主执行军事任务时很可能无法不造成任何附带损害或仅造成最低限度附带损害。

因此，在满足国际人道法(战时法)的基本条件方面，即确定造成人员伤亡的责任方面，伦理问题至关重要。事实上，不满足上述国际人道法条件的武器或手段不应被用于战斗。如果由个人做出决定，并由此导致对目标采取行动，那么从决定者到下达命令者，一直到扣动扳机/按下按钮的人，都有一个明确的指挥责任系统。在使用自主武器系统的情况下，对于在战斗行动中造成的损害，没有规定明确的指挥责任。

出于这些原因，必须严格管制和紧密监测致命自主武器系统的开发、生产和使用。

此外，各项国际公约应对致命自主武器系统的使用进行规范，以便为现代化武器(特别是致命自主武器系统)的使用规定行为道德守则标准，目前这些武器仍处于新兴阶段，尚未被大规模使用。

塞拉利昂

[原件：英文]
[2024年5月22日]

引言

本报告是塞拉利昂政府根据大会2023年12月通过的关于致命自主武器系统的第78/241号决议所述秘书长呼吁提交。本报告借鉴了2024年4月17日和18日在塞拉利昂弗里敦举行的西非国家经济共同体(西非经共体)成员国关于自主武器系统的和平与安全方面问题会议的审议情况和通过的公报(下称“弗里敦公报”),以及朱利叶斯·马达·比奥总统关于迫切需要制定具有法律约束力的规范和原则以应对自主武器系统构成的威胁和挑战的正式声明。

西非经共体成员国派出包括外交部长在内的高级别代表团出席了会议。会议主题是“自主武器系统的和平与安全方面：西非经共体对如何启动具有法律约束力文书谈判进程的看法”。大会于2023年12月22日通过塞拉利昂作为共同提案国的第78/241号决议后,塞拉利昂政府主动提出主办了这次会议。该决议请秘书长就如何从人道主义、法律、安全、技术和伦理角度应对会员国提出的相关挑战和关切问题征求会员国的意见。

意见

2024年4月18日的弗里敦公报阐述了塞拉利昂的立场。塞拉利昂支持秘书长就一项具有法律约束力的文书进行紧急谈判的呼吁,以便根据国际法,包括人道法和人权法,以及关于战争的基本伦理原则,管制自主武器系统。塞拉利昂承诺推动弗里敦公报所强调的西非经共体协调战略,以应对自主武器系统构成的多方面挑战。

塞拉利昂政府强调,必须为自动化武器系统的采购、转让和部署制定次区域问责标准和业务框架,以及维护国际法和保障人权的完善法律参数和行动议定书。我们特别强调,在现有国际条约和联合国正在进行的努力的基础上开始谈判,以达成一项关于自动化武器系统的具有法律约束力的文书,是重要而紧迫的。塞拉利昂斥责自动化武器系统可能导致军备竞赛升级和武器扩散,并颠覆西非经共体长期以来在武器控制、预防冲突和裁军进程方面的工作。

塞拉利昂根据第78/241号决议提交了2024年4月18日弗里敦公报的内容,供纳入秘书长报告。发布该公报的会议是第一次吸引包括塞拉利昂国家元首在内的政府最高层参与的区域会议,因此该公报不仅是塞拉利昂政府的文件,也代表西非经共体成员国就需要制定一项关于自主武器系统的具有法律约束力的文书而商定的立场。弗里敦公报总结了本意见。

认识到新兴技术对地区和全球安全及地缘政治的重大影响;

还认识到这些技术和人工智能如果得到公平开发和共享,就有可能支持经济增长和提高人类生活质量;

认识到如果不公平开发，这些技术可能会加剧现有的社会、政治和经济不平等，并可能助长偏见、不信任和数字非人化；

担心武器系统的自主性会引发严重的伦理、人道主义、法律以及和平与安全问题，包括担心自主武器系统有可能被用作对打击目标的致命武力，而没有实施对维护伦理、法律和人道主义义务至关重要的人实际控制；

意识到我们地区武装冲突、恐怖主义和内部动乱等问题的历史和现状，这些问题因武器贩运和武器系统的国际监管不力而加剧；

担心自主武器系统可能越来越容易复制和替代，从而成为贸易和转让的对象，包括供跨国犯罪网络、恐怖分子和其他非国家武装团体使用；

还关切在缺乏具体国际监管的情况下，这种扩散将对国家、区域和全球和平与安全造成更多威胁，包括军备竞赛和冲突升级的风险，并将破坏通过经济和政治一体化建设和平的区域努力；

因此，认识到需要加强现有的治理机制，包括国际人道法、国际人权法和国际刑法，制定新的具有法律约束力的规则，禁止和规范自主武器系统带来的威胁和挑战；西非经共同体成员国决心

- 支持紧急谈判一项具有法律约束力的文书，以根据国际法，包括国际人道法和国际人权法以及伦理原则，对自主武器系统进行规范，并制定具体的禁令和条例；
- 努力加强区域合作和伙伴关系，并通过包容性和参与性方法，推进关于自主武器系统的共同立场，为制定新的具有法律约束力的文书这一更广泛的全球努力做出贡献；
- 促进与国际组织、民间社会、学术界和其他利益攸关方的建设性对话与合作，以有效地应对自主武器系统带来的挑战；
- 赞同本次会议通过的公报所阐明的努力形成关于自主武器系统的区域共同立场的进程，并承诺为落实这一立场而共同努力。

结论

塞拉利昂认为，弗里敦会议和公报强调必须在自主武器系统管制方面开展全球合作和采取统一立场。会议和公报为未来就一项关于此类系统的具有法律约束力的文书进行谈判制定了战略议程。塞拉利昂希望本次会议的讨论和成果能对全球有关自动化武器系统的对话产生重大影响，并有助于实现联合国设定的到 2026 年制定具有法律约束力国际框架的目标。会议强调了自主武器系统问题的复杂性和紧迫性，以及国际社会应对这些挑战的集体责任。会议确认塞拉利昂和西非经共同体在促进采取统一办法加强全球安全、尊重人权和遵守国际法方面发挥了积极作用。

新加坡

[原件：英文]
[2024年5月9日]

新加坡作为一个小国，一贯支持基于规则的多边体系和联合国的作用。本组织为国际法和国际准则提供了基础。多边机构、制度和法律对所有国家特别是小国的生存至关重要。

因此，关于致命自主武器系统的治理问题，我们认为，国际社会必须进行多边接触，以取得实质性成果。

新加坡申明，国际人道法适用于致命自主武器系统的使用。鉴于武器系统和人工智能技术的自主性迅速发展，我们认为，在多边环境下继续讨论致命自主武器系统的治理问题非常重要。

推进国际人道法

新加坡于 2024 年 3 月成为《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》缔约国。

新加坡加入《公约》，强调了我们国际人道法的承诺。《公约》作为一项框架公约，产生了进一步加强国际法的重要议定书。新加坡致力于继续编纂和逐步发展适用于武装冲突的国际法规则。

在《公约》的范围内，致命自主武器系统问题政府专家组于 7 年前正式召开会议。新加坡认为，专家组是讨论致命自主武器系统治理的适当论坛，因为该论坛体现了法律、军事和技术专家的包容代表性。新加坡支持专家组在 2023 年商定的关于禁止和管制此类武器系统的“两级方法”。我们认为，应禁止使用无法在遵守国际人道法的条件下使用的致命自主武器系统，并支持限制的概念，确保可以按照国际人道法使用这种系统。

由于联合国的一些会员国不是《公约》缔约国，无法参加政府专家组，秘书长呼吁就这种系统发表意见，这有利于联合国全体会员国的参与。新加坡对大会第 78/241 号决议投了赞成票，该决议请秘书长就致命自主武器系统征求会员国的意见。我们认为，秘书长报告可以作为一项重要资源，供专家组在制定致命自主武器系统治理框架时考虑。

新加坡将继续为政府专家组做出建设性贡献，以期在专家组当前任务授权下取得实质性进展，并希望各国在致命自主武器系统治理文书的要素上取得一致。

将人工智能纳入致命自主武器系统

致命自主武器系统不一定包含人工智能，因为这种系统也可以建立在基于规则的编程基础上。然而，当人工智能应用于此类系统的关键功能时，我们必须认识到意外结果的风险。如果人工智能在此类系统中以不可预料的方式行事，则可能由此产生非常严重的影响，例如意外升级、友军误击或对平民的非法伤害。

新加坡致力于在军事领域负责任地开发和人工智能。通过多年与国防技术专家、军事规划人员、国际法专家和政策专业人士的广泛磋商，新加坡制定了关于负责任的军事人工智能的国家原则。这些原则于 2021 年宣布，解决了与军事领域人工智能有关的四个关键关切问题。

a. **负责**。首先，必须解决新兴人工智能行为的风险。人工智能系统必须有明确的预定用途，开发人员和用户都对人工智能系统的结果负责。

b. **可靠**。其次，必须解决人工智能系统输出中的错误或不精确性风险。人工智能系统应经过测试，确保达到适合其预定用途的水平。人工智能系统的设计应尽量减少意外偏见，并产生一致的输出。

c. **强健**。第三，必须解决恶意行为者利用人工智能的风险。人工智能系统在设计时应考虑到网络和对抗性人工智能威胁。为了解决“黑箱效应”，人工智能系统的开发过程应该有充分的文件记录支持可解释性。

d. **安全**。第四，我们必须关注安全关键环境中人工智能故障的风险。人工智能系统的使用，不仅在部署的平台方面，而且对周围的资产和人员都应该是安全的。

关于人工智能和致命自主武器系统的区域倡议

区域倡议对于确保就致命自主武器系统进行包容各方和针对具体情况的讨论至关重要。2024 年 2 月，新加坡共同主办了军事领域负责任的人工智能亚洲区域磋商会。军事领域负责任的人工智能是由荷兰和大韩民国共同发起的一个进程，侧重于通过包容性、多利益攸关方和跨区域对话推进负责任的人工智能规范。在亚洲区域磋商会上进行了一次互动练习，讨论了将人工智能纳入军事行动的挑战和关切。练习中的公开讨论使参与者加深了区域对致命自主武器系统相关人工智能发展的理解。

新加坡还积极参加了菲律宾于 2023 年 12 月举办的自主武器系统印太观点马尼拉会议。与会者审议了武器系统中使用自主功能的影响，为民间社会、工业界、国际法专家以及国防和外交官员发表意见提供了平台。

西班牙

[原件：西班牙文]

[2024 年 5 月 23 日]

西班牙赞同并完全支持致命自主武器系统领域新兴技术问题政府专家组 2019 年 12 月制定的指导原则。指导原则规定，基于新兴技术的致命自主武器系统的可能使用必须符合适用的国际法，特别是国际人道法。

为此，西班牙与其他志同道合的国家一道，赞成基于禁止和管制的两级方法。

根据西班牙的立场，2023 年 7 月，国防部批准了在该部开发、实施和使用人工智能的战略，这一战略符合欧洲联盟和北大西洋公约组织的原则和倡议。关于西班牙倡导的原则，该战略包括以下几点：

- 人工智能应用程序的开发和使用应符合适用的国内和国际法，包括《世界人权宣言》和国际人道法。
- 任何人工智能的开发或使用都应允许明确的人工监督，以确保适当问责和明确责任归属。

斯里兰卡

[原件：英文]
[2024 年 5 月 25 日]

导言

斯里兰卡作为人道主义裁军、人道主义作战方法和不扩散大规模毁灭性武器的长期倡导者，欢迎大会通过关于致命自主武器系统的第 78/241 号决议，大会在该决议中强调，国际社会迫切需要应对自主武器系统带来的挑战和关切。作为该决议的共同提案国，斯里兰卡赞赏该决议为启动包容性对话和提高对这一问题的认识提供了机会，这是应对相关挑战的重要一步。

技术的快速进步彻底改变了现代文明的方方面面。斯里兰卡认识到人类发展日益依赖先进技术。然而，鉴于自主武器系统将对人类尊严构成的根本挑战以及道德和伦理因素，在武器系统中纳入自主性包括人工智能是需要特别考虑的。斯里兰卡一贯呼吁人道主义裁军，支持多边主义。在这一精神的指导下，斯里兰卡积极促进了在日内瓦和纽约就该议题进行多边讨论。

在这方面，斯里兰卡回顾 2021 年 10 月在大会上发表的关于致命自主武器系统的第一份联合声明，该声明得到了包括斯里兰卡在内的 70 个国家的广泛跨区域集团的支持。斯里兰卡欢迎联合国秘书长和红十字国际委员会主席发出联合呼吁，要求到 2026 年紧急制定关于自主武器系统的新国际规则，并欢迎秘书长在政策简报《新和平纲领》中考虑这一问题。

斯里兰卡仍然坚信，通过一项具有法律约束力的条约将是应对自主武器系统所构成严重挑战的最有效方法。本呈件阐述了斯里兰卡对自主武器系统的人道主义、法律、安全、技术和伦理风险以及应对这些挑战的可能手段的看法。

鉴于行动窗口正在迅速关闭，斯里兰卡希望这项决议发起的讨论将有助于加快对这一问题做出全球多边反应。

人道主义考虑

开发和具有自主决策能力的武器系统，特别是在生死攸关的重大决策方面能自主做出决定的武器系统，引起了根本的人道主义关切。在这个循环中去除人的因素，这种可能性令人不安，可能导致在没有人类同理心、伦理和同情心的情况下匿名选择攻击目标。

在这方面，马顿斯条款尤其重要。马顿斯条款是一项由来已久且具有约束力的国际法规则。《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》和《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》序言中提到，马顿斯条款规定，在没有具体法律规定的情

况下，各国有考虑人道原则和公众良心要求的法律义务。如果一个武器系统在没有人实际控制的情况下选择和攻击目标，将人简化为仅仅是数据点，那么它很可能与马顿斯条款的这两个方面背道而驰。

法律考虑

国际人道法完全适用于自主武器系统，这一点是毫无争议的。¹ 因此，相关的问题就是应如何确切适用国际人道法的规定，并确保这些规定得到遵守。

人们一致认为，自主武器系统如果会造成过度伤害、不必要的痛苦或固有滥杀滥伤作用，就不得使用。² 复杂战场环境中的机器算法能否确保在攻击的区别性、相称性和预防措施方面应用精心计算得出的决定，是非常值得怀疑的。而区别性、相称性和预防措施，都是国际人道法的基本义务。此外，决策过程去除人的因素，在责任和问责方面造成了严重的法律挑战。斯里兰卡认为，人实际控制是这方面的内在要求。可预测性、可靠性和可解释性应被视为武器系统中人实际控制的一些关键要素。

安全考虑

自主武器系统的部署可能导致不对称战争，也会带来冲突升级的风险，因为当使用国的军事伤亡风险降低时，国家更有可能卷入武装冲突。此外还有其他一些风险，扩散的威胁、非国家行为者获得这些系统、甚至大规模毁灭的可能性等，使自主武器系统可能在国家、区域和国际各级造成不稳定。由于获取相关技术相对容易，在缺乏约束性法规的情况下，这些武器系统必然会引发代价高昂的军备竞赛。在网络和空间领域，武器系统中不受管制的人工智能的安全问题可能会进一步加剧。

技术考虑

武器系统可能在一定范围内具有自主性，在这个范围内，人参与的程度可能有所不同。斯里兰卡注意到，目前对“自主武器系统”的商定定义缺乏共识，这对就其管制达成共识是一个挑战。然而，自主武器系统的具体技术定义问题不应妨碍开始就制定相关法律限制进行谈判。鉴于技术的迅速发展，对自主武器系统进行一次面向未来和技术中立的工作特性描述，将在这方面提供一个有益的起点。

伦理考量

《世界人权宣言》和《公民及政治权利国际公约》承认，人权源于所有人的“固有尊严”。对斯里兰卡来说，这个辩论的伦理和道德因素就算不是该问题最重要的方面，也是其基本的方面之一。允许一个算法程序来决定人的生死，这对人道原则构成了重大挑战。潜在自主武器系统中的机器学习和数据偏见也是涉及伦理和法律考虑的引起严重关切的跨领域问题。

¹ 《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》，致命自主武器系统领域新兴技术问题政府专家组，致命自主武器系统领域新兴技术问题政府专家组2019年届会报告，CCW/GGE.1/2019/3，2019年9月25日，附件四，“指导原则”，第(a)段。

² 第六次审查会议最后文件(CCW/CONF.VI/11，2022年1月10日)第19段。

前进方向：争取达成一项具有法律约束力的文书

目前，没有任何国际法规则具体禁止或限制武器系统使用自主功能，现有国际人道法原则也没有具体要求维持人的控制。

斯里兰卡欢迎采取自主武器系统使用行为守则和政治宣言等不具约束力的自愿措施，认为这是一种逐渐发展。然而，这些措施在加强关于这一问题的规范框架方面只具有补充价值，因为它们不足以应对该问题所涉严重法律、伦理和安全挑战。同样，没有国际商定的具约束力条款，仅根据 1949 年 8 月 12 日日内瓦四公约第一附加议定书第 36 条规定的现有义务对武器系统进行法律审查是不够的。

鉴于上述严峻的人道主义、法律、安全、技术和伦理挑战，斯里兰卡坚决支持和倡导就一项关于自主武器系统的具法律约束力的文书进行谈判。致命自主武器系统问题政府专家组中越来越多的国家发出了这一呼吁。这种具约束力的限制将是应对这一问题复杂性的最有效办法，限制措施应包括禁止某些不符合国际人道法的自主武器系统，并严格管制其他系统。

加倍努力就自主武器系统的开发和使用达成紧急多边协议，是一个重要的优先事项。因此，斯里兰卡将继续充分致力于建设性地参与有关这一问题的讨论。

瑞典

[原件：英文]

[2024 年 5 月 24 日]

瑞典欢迎有机会根据大会 2023 年 12 月 22 日通过的关于致命自主武器系统的第 78/241 号决议，向秘书长提交意见。

自主武器系统的应用可能从根本上改变武装冲突，有利有弊，弊端是诸如问责和遵守等方面的挑战需要解决，优势是其精确性和可用于保护平民。因此，围绕致命自主武器系统的讨论应考虑到操作、法律和技术方面，同时铭记伦理观点。

瑞典支持所谓“两级方法”，即应区分根据国际法特别是国际人道法不能使用的武器系统和应管制以确保遵守国际人道法和其他适用国际法的具有自主功能的系统。前者应明确禁止，后者则应加以管制，但管制的形式需要进一步思考和拟订，以便其定义和可能的管制得到国际社会的广泛支持。瑞典认为，两级方法是一个可以达成共识的中间立场。

《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》和致命自主武器系统问题政府专家组

《特定常规武器公约》为致命自主武器系统领域的新兴技术问题提供了适当框架，原因有几个。来自若干相关学科的专家以及来自各国、民间社会和工业界的代表的参与提供了丰富的观点。关于今后的努力，需要加强对法律、军事和技术方面人的控制概念的共同理解，这是一项具有挑战性的工作。所有缔约国专家，包括拥有这一领域最先进能力的缔约国的专家，都必须参与这项努力。

虽然在《公约》框架内关于致命自主武器系统的谈判进展缓慢，但在最近几届会议上可以注意到政府专家组内取得了一些令人鼓舞的进展。尽管每年只举行两次会议，但有迹象表明，在若干议题上已取得一致。专家组目前正在讨论未来文书要包括的内容的具体措辞，这可能是取得实质性突破的明确途径。问题不在于讨论的形式，而在于一些会员国缺乏政治意愿。建立得不到主要利益攸关方支持的平行进程和文书是无法解决这一问题的，因为这样做不会促进有效管制，并有可能进一步分裂会员国，破坏国际人道法。我们应当继续以已经取得的实质性成就，如 11 项指导原则和工作组通过的协商一致结论为指导，推进今后的工作。

瑞典强调，政府专家组必须按照其任务规定开展工作。缔约国能够商定何种类型的条例，仍有待观察，但即使是一项可能达不到某些代表团期望的协议，也仍然是向前迈出了一步。

国际人道法和人实际控制

瑞典支持国际人道法完全适用于所有武器系统、包括致命自主武器系统的可能开发和使用的观点。这是一个基本原则。为了使这项原则始终得到维护，至关重要是对国家武装部队人员进行武装冲突期间适用的国际法培训和演练。具有国际法专长的法律顾问在与国际人道法的解释和适用有关的军事决策中发挥宝贵和重要的咨询作用。瑞典欢迎就未来可能的自主武器系统适用现有国际人道法问题继续进行讨论。

必须保留人对使用武器系统决定的责任，因为不能由机器来接受问责。这一点在武器系统的整个生命周期中都应考虑到。军事行动中军事手段和方式的选择必须符合有关细则和条例。在规划军事行动时，军事指挥官及其参谋人员必须考虑和评估平民的存在，以遵守攻击中的区分原则、相称原则和预防原则。不得部署或使用不能履行或不履行国际人道法这些规定的武器。

维护人对使用武力的实际控制是一个关键目标。人机交互可视为确保这种控制所需的一个重要方面。无论是在对武器系统的了解方面，还是在控制这些系统的能力和技能方面，军事决策者和操作人员都需要掌握控制权。所有武器系统都必须可靠、可预测，以便操作人员能够始终确定系统将按照操作人员的意图运作。仍然需要分析各种情况下的人工控制的更精确要求，从实际角度理解并达成一致。

在军事方面，规则、条例和程序构成所有涉及武器的行动的指令等级架构。这些应涵盖组织结构、程序、安全、基本指挥概念、风险控制和必要的培训要求等。所有系统的手册和培训方案都应随条例一起提供。任何复杂的系统都必须有严格的操作规程，包括培训方法和使用程序。

应考虑在武器系统的整个生命周期采取措施确保人的控制。具体措施将视具体情况而定。系统的目标类型以及空间和时间限制可能是重要因素。

审查程序

各国有义务确定国际法是否禁止使用某种新武器。在瑞典，这项工作由监测军火项目国际人道法代表团负责。所有与国防有关的主管部门都必须从速向代表团报告涉及研究、开发、获取或采用武器或作战方法的拟议项目。

在制定规章、程序、手册和培训方案时，需要考虑人机互动及其局限性。《1949年8月12日内瓦四公约关于保护国际性武装冲突受难者的附加议定书》第36条规定，在对新武器进行法律审查时，必须分析确定国际人道法是否禁止使用该新武器。分析应包括人机交互的各个方面以及手册和培训方案如何处理这些方面。

监测军火项目国际人道法代表团根据第36条进行审查时，对武器系统的特点及其计划用途和其他有关方面，如培训方案和操作规程进行审查。如果有疑问或科学上的不确定性，代表团可要求提供进一步的资料或采用进一步的试验方法。然后，代表团将作出决定，批准或不予批准审查的武器系统或方法。代表团还可以提出严格的修改或限制要求，使该武器系统符合国际法的要求。

现有关于若干国家法律审查制度的资料可协助希望审查现有系统的缔约国开展工作。

风险评估和相应的缓解措施是所有先进武器系统开发的一部分。这些系统的采购、维护和使用过程应通过详细的安全程序加以控制。应从不同角度将这些程序记录在安全手册中，从爆炸物和弹药问题到软件质量问题，包括意外偏见的后果。

在非技术语境中描述技术系统是一项具有挑战性的任务。使用通常用于描述人类行为的形容词很容易造成混淆，并有可能对不具备人类素质的技术系统得出不准确的结论。为了避免这种情况，应该只使用严格的技术术语。

虽然和平利用技术不属于《公约》的范围，但可以注意以下情况：在技术开发方面，民用和军事领域的重叠很大，而且重叠面似乎在不断扩大。这就形成了一种相互依赖。一项新技术若用于军事用途，则需要对系统的稳健性和可靠性提出很高的要求。

自动化、自主、人工智能、数字化和计算机化等方面的技术进步通常在军事和民用领域是共同的，尽管往往由民用(商业)利益驱动。对于可能具有危险性的技术系统(民用)和设计为具有危险性的系统(武器)，确保有效控制的挑战几乎是相同的。

瑞士

[原件：英文]

[2024年5月25日]

瑞士欢迎大会通过题为“致命自主武器系统”的第78/241号决议，并欢迎该决议得到的大力支持。瑞士是这项决议的主要提案国之一。

瑞士认为，总体而言，新技术和新兴技术为人类进步带来了巨大希望，并有助于加强人类和国际安全。在军事领域，自主性可以带来优势，包括提高安全性和效率。此外，根据设想，在具体应用中和在某些条件下，这些技术可有助于更好地保护平民和民用物体或避免附带损害。

瑞士虽然认识到人工智能包括武装部队使用人工智能和自主能力可能带来

的益处和机会，但认为迫切需要加强多边努力，更好地了解并及时有效地解决这些发展可能带来的人道主义、法律、安全、技术和伦理问题，特别是与自主武器系统有关的这些问题。

瑞士认为，自主武器系统包括一旦启动，在识别目标、选择目标和对其实施武力的关键功能方面无需人直接干预即可运作的武器系统。

需要就自主武器系统的开发、部署和使用制定适当的规则和限制，确保遵守国际法律义务，坚持伦理要求，并考虑到人道主义因素和与国际安全有关的方面。瑞士尤其认为有必要继续编纂和逐步发展关于自主武器系统的适用于武装冲突的国际法规则。

在这方面，瑞士强调，致命自主武器系统问题政府专家组在《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》框架内在日内瓦开展的工作十分重要。鉴于政府专家组的组成和在十多年的审议中积累的专门知识，专家组在制定这一领域的多边治理措施方面具有独特的地位。瑞士强调，专家组必须完成其任务，最迟到 2026 年审议和拟订一项关于自主武器系统的文书的内容。在这方面，瑞士注意到联合国秘书长和红十字国际委员会主席联合呼吁到 2026 年缔结一项具有法律约束力的国际文书。

瑞士认为，在政府专家组框架内推动的所谓“两级方法”很有价值，并一直支持这种方法。该方法包括对自主武器系统的开发、部署和使用的具体禁止和管制规定。这一方法应成为任何文书的核心：

第一，瑞士积极支持就一项国际文书进行谈判，首先防止出现不符合国际人道法的越来越具有自主功能的系统。这包括：

- 无法按照操作人员或指挥官的意图以足够的可靠性或可预测性履行职能的系统，或可能在规范参数之外运作的系统；
- 因为无法充分限制使用武力的程度和时间等原因，而无法按照国际人道法的规定限制其影响的系统；
- 根据国际法特别是国际人道法和人权法不能使用的系统。

第二，瑞士主张，在原则上根据国际法可以使用的条件下，通过关于自主能力日益增强的武器的条例和措施。这些条例应侧重于减少风险，包括坚持人的控制的积极义务，并着重对武器参数和部署环境加以限制，同时考虑到军事、安全政策和伦理因素。

这一方法的核心在于人的参与或控制的概念，特别是对国际人道法的遵守。瑞士认为，首要任务是制定一个国际框架，确保足够程度和某种类型的人的控制，并强调可以在系统的整个生命周期，特别是在目标选择周期内进行控制。这一点至关重要，因为在使用自主武器系统时，与任何武器系统一样，人类必须确保法律得到遵守。为了确保人对使用武器系统的决定负责，必须在武器生命周期的适当阶段实行或嵌入一定程度的人的控制。

在《公约》框架内向前推进，至关重要也十分紧迫。这一直是并将继续是瑞士的优先事项。我们重申，似乎没有其他多边论坛能更好地处理与可能被认为具有过分伤害力或滥杀滥伤作用的自主武器系统相关的关键方面，并有助于制定和通过相应的国际人道法的规定。瑞士促请《公约》所有缔约国继续在政府专家组内为有效应对这些挑战做出努力，并期待专家组实现三年期任务中概述的目标。

尽管瑞士坚决支持《公约》，但我们认识到，自主性越来越强的武器系统，以及更广泛地说，人工智能的军事使用，特别是军事行动中的人工智能辅助决策，是一个多层面的问题。因此，必须对这一问题采取全面和包容的办法，这个问题除了涉及国际人道法和敌对行为外，还包括国际法、稳定、军备控制、扩散、出口管制和伦理等各个方面。

大不列颠及北爱尔兰联合王国

[原件：英文]

[2024年5月23日]

人工智能正在从根本上改变我们的社会，并将改变我们面临的威胁。联合国认识到，人工智能的使用引起了社会关切，并对既定的军事治理和保障制度构成挑战。

联合王国正在人工智能管制方面发挥主导作用。2023年人工智能安全峰会和《布莱奇利宣言》、英国《国家人工智能战略》、人工智能安全研究所和负责任技术采用部门都倡导安全、合法、符合伦理和负责任地使用人工智能。联合王国任何使用人工智能来增强国防能力的行为都受到该国“雄心勃勃、安全、负责任”政策声明中阐述的国防人工智能伦理原则的约束。

摘要

- 联合王国不拥有完全自主的武器系统，即在没有任何适当人类参与的情况下或在人类负责和问责之外运作的武器，也没有开发这种系统的意图。任何国家都不应开发或部署这种系统。
- 国际人道法和现有的武器系统开发、采购和使用管制框架是管制新军事能力的适当框架。
- 在自主武器系统的开发和使用过程中，人的判断永远是必要的。
- 不可能将责任转移到机器上。无论系统的自主程度如何，都不能免除人对使用一个系统来达到某种效果所负的责任。
- 在国际层面努力制定负责任地开发和自主武器系统的规范和标准，是确保查明和追究任何非法、不安全或不道德使用这些技术的行为并追究责任的最佳途径。

遵守国际人道法

国际人道法适用于敌对行为规划和实施中使用的所有军事能力，包括具有自主功能的军事能力。所有国家都必须遵守国际人道法，没有例外或豁免。我们反对任何削弱或减损国际人道法所提供的基于原则的强健法律框架的企图。

联合国认为，使用具有自主功能的武器，就需要有治理制度使国际人道法在法律和伦理上得到遵守。我们认识到，一些国家和民间社会在呼吁制定具有法律约束力的新规则，其依据是，具有自主功能的武器将带来国际人道法未涵盖的新的战场要素。然而，我们认为，在自主武器方面，国际人道法的适用不存在差距。现有的国际人道法已经对各国开发和采购武器以及战争手段和方法，包括使用先进技术的武器和手段，作出了规定。它是一个技术上无限定的、强健灵活的管制武装冲突的法律制度。

根据国际人道法，冲突各方选择战争方法或手段的权利并非是无限制的。《1949年8月12日内瓦四公约关于保护国际性武装冲突受害者的附加议定书》第36条正强调了这一点。第36条规定，各国义务确定在某些情况下或在所有情况下，国际人道法是否禁止使用某种新的武器、战争手段或方法。

武器系统的自主功能可以且必须以合法和合乎伦理的方式使用。事实上，自主系统有可能支持更好地适用国际人道法，并强化对国际人道法的遵守。它们可以改善证据、分析和决策及时性，因此有可能确保对平民的更大保护。更好的决策意味着符合国际人道法基本原则的更好结果。

问责和责任

规定国际人道法中国家责任以及国际和国内刑法中个人责任的法律框架，不允许将对军事行动的影响所负的责任转嫁给机器。国家对国际不法行为的实施负责，包括对无差别或以其他方式非法使用武器系统负责。国际人道法依赖于指挥官问责制规则，这一规则将人置于使用武力决策的中心。在武器中使用自主功能并未也不能否定人作为法律上负责任的行为者的作用。

人有责任确保自主系统在国防中的使用是建立在对如何治理进行了明确阐述的基础上的。参与决定使用包含自主功能的系统的人员必须了解使用方式和预期效果，并知道他们要对该效果负责。

这些原则适用于系统从概念到部署的整个生命周期。这包括向开发人员提供指导，确定管理开发的技术标准；测试和验收流程；严格的实地测试程序；对使用这些系统的人员进行培训。这种生命周期方法必须包括对系统所作的任何改变、系统使用环境或投入使用后的预期操作环境，以及适当的反馈和报告机制。

系统一旦部署，使用该系统的受过培训的操作人员就须接受问责，对开展战役负有行动或战术责任的各级指挥官也须就其所作决定接受问责。在参与行动的所有军队都必须执行的命令、指示和标准作业程序中，明确规定军事指挥系统和问责措施。

控制

在所有用例中，武器系统包括具有自主功能的武器系统的开发和使用，必须始终有具体背景环境下适当的人的参与。这种参与必须形成有效的人的控制，其程度足以满足我们的政策、伦理原则和国际人道法义务。人的参与的性质将根据能力的性质、行动环境和使用背景(例如目的、环境、可能的威胁、与系统行为相关的风险以及监管环境)的变化而变化。这些因素将决定人的参与的类型和时机，以确保人的参与最大限度地满足军事、法律和伦理目标。

在整个系统生命周期的许多节点上，都必须有适当的人的参与，需要得到授权、有资格、有经验的人作出判断，影响、指导或限制自主系统的行为及其效果。¹

我们反对制造和使用在没有适当人类参与和问责的情况下运作的具有自主功能的武器。我们认为，所有国家都应明确承诺负责任地开发和军事人工智能，包括自主系统，并确保查明非法、不安全或不道德地使用这些技术的行为，追究其责任。

国际办法

建立使用和积极义务的国际规范是展示如何根据国际人道法使用自主武器系统的最佳方式。

政府专家组是审议这一问题的适当论坛。它的任务规定为取得进展奠定了积极的基础。

联合国提议专家组就致命自主武器系统方面国际人道法的适用和最佳实践制定一份权威声明。这一点以及关于一套“自主武器系统条款草案”的提议将表明，有严格的原则管理自主武器的使用，并根据国际人道法进行问责。

专家组应评估：

- 如何根据国际人道法使用自主能力，以及如何在不同的使用环境中实际应用这些能力。
- 整个系统生命周期的规范和人机协作方法。
- 安全、负责任地开发和军事人工智能的技术和治理标准。

美利坚合众国

[原件：英文]

[2024年5月23日]

美国赞赏有机会根据大会第 78/241 号决议提出自己的看法。美国积极参与多边论坛关于致命自主武器系统的讨论。我们鼓励其他国家也这样做。我们还坚决支持国际组织和民间社会发挥作用，观察和促进关于致命自主武器系统的国际讨论。

¹ 联合国 2018 年和 2020 年提交给致命自主武器系统问题政府专家组的工作文件对这种人工智能系统人类参与和控制的生命周期方法进行了描述。

美国仍然认为，在《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》主持下召开的致命自主武器系统领域新兴技术问题政府专家组会议，为推动有关致命自主武器系统的国际努力提供了最佳机会。政府专家组是一个包容各方、透明的论坛，各国和民间社会都参与其中。专家组的工作重点是国际人道法。参与专家组工作的代表团通常包括具有军事、技术、法律和政策经验的成员，这积极促进了专家组的工作。在成员的专门知识的基础上开展了大量工作，反映了非常多样化的参与国的共识。2022年以来，各代表团也提交了许多实质性提案，包括关于具有法律约束力的文书的提案、关于不具约束力的文书的提案以及其他成果。专家组有明确和强健的任务授权，即在不预断文书性质的情况下，以协商一致方式拟订一套文书要素，以及处理致命自主武器系统领域新兴技术的其他可能措施。这一任务明确指出，专家组工作的最终目标是到2026年制定一项文书。在专家组之外不包括所有有关国家或不以协商一致方式运作做出的努力，可能会导致各自为政和方法分歧。

美国对致命自主武器系统的做法首先是承认现有国际人道法已经提供了禁止和限制在武装冲突中使用自主武器系统的适用框架。美国与澳大利亚、加拿大、日本、波兰、大韩民国和联合王国一起提交了一项提案，阐明了如何适用国际人道法，以及如何有效落实与使用自主武器系统有关的要求。这项题为“关于自主武器系统的条款草案——基于国际人道法的禁止和其他管制措施”的提案¹采用了在政府专家组内得到广泛支持的“两级方法”，反映了国际人道法对本质上禁用的武器和没有明确禁止在所有情况下使用的其他武器的使用规定之间的区别。

国际人道法并不禁止在武器系统中使用自主功能，也不禁止使用能够选择和攻击目标的武器。几十年来，各国一直在使用能够选择和攻击目标的计算机和武器，而这并没有带来法律争议。这包括像“宙斯盾”武器系统和“爱国者”防空导弹防御系统这样的系统，以及“发射后锁定”的寻的武器。国际人道法远未禁止具有目标选择和攻击功能的武器，事实上，当这种武器的使用比使用没有这些“智能”功能的武器更精确、更准确，对平民和民用物体的风险更低时，国际人道法的主要宗旨之一则得到了推进。然而，由于操作人员依靠武器来执行目标选择和攻击功能的能力有别，因此在国际人道法如何适用于这些武器的使用方面出现差异，并且在武器系统开发期间或部署之前需采取不同的措施，以支持负责任地使用这些武器。

条款草案首先包含防止就其性质而言无法根据国际人道法使用的自主武器系统的措施。条款草案还解释道，根据国际人道法，使用自主武器系统进行攻击必须符合攻击中的区分原则、相称原则和预防原则和要求。战斗人员依靠自主功能识别、选择或攻击目标时，必须是出于诚意并根据当时可获得的信息，而且必须符合在执行这些原则和要求时的尽职调查要求。条款草案详细说明了在制定阶段和使用期间为确保有效执行国际人道法可采取的措施。条款草案还阐明了管制措施，确保对自主武器系统的使用进行全面问责。

¹ CCW/GGE.1/2023/WP.4/Rev.2, 可查阅 [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_WP.4_US_Rev2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.4_US_Rev2.pdf)。

美国支持使用这些措施和条款草案中所列的其他措施，但美国在其政策中并没有使用“人的实际控制”等术语，并详细解释了为什么注重“控制”会掩盖而不是澄清这一领域的真正挑战。相反，正如国防部第 3000.09 号指令和美国提交政府专家组的工作文件所反映，关键问题是确保机器有助于实现指挥官和武器系统操作人员的意图。实现这一点的办法除其他外还包括，在武器设计、开发和部署过程的不同阶段采取实际步骤，减少意外攻击的风险，并使人员能够对使用武力进行适当程度的人类判断。不应在所有情况下都适用一个固定的、一刀切的人的判断程度。有些功能可能由计算机来执行比由人执行更好，而其他功能则应由人执行。随着武器工程师提高自主功能的有效性，可能会出现更多使用自主功能比手动控制更合适的情况。

美国国防部为履行以负责任的方式开发和新技术和新兴技术的承诺，发布了一项关于武器系统自主性的政策指令(国防部第 3000.09 号指令)以及一系列政策和其他文件，包括《国防部人工智能道德原则》、《国防部负责任的人工智能战略和实施途径》、《国防部 2023 年数据、分析和人工智能采用战略》。美国公布了这些政策和《负责任的人工智能工具包》等相关资源，证明美国在这方面的承诺，并鼓励国际透明度。

美国还寻求通过《关于在军事上负责任地使用人工智能和自主的政治宣言》围绕开发、部署和使用军事人工智能和自主的负责任行为规范建立国际共识。该宣言是补充性的，但独立于政府专家组。美国于 2023 年 2 月发布《政治宣言》，开始围绕负责任行为规范建立共识以确保这些技术在军事上的使用是负责任、合乎伦理的，并加强国际安全。《政治宣言》为就负责任地开发、部署和使用军事人工智能能力进行包容性国际对话奠定了基础。

美国认为，秘书长的报告可促进在政府专家组内就要素和措施达成共识，为在致命自主武器系统方面取得进展作出宝贵贡献。为了实现这一目标，报告应平衡并包括所有会员国的观点。我们感谢秘书长在网上公布我们提交的完整报告。

巴勒斯坦国

[原件：英文]

[2024 年 5 月 25 日]

我们已经进入了一个真正危险的历史时刻，包含人工智能的武器系统正被用来在加沙实施种族灭绝。自动化战争技术领域的专家指出，包括东耶路撒冷在内的巴勒斯坦被占领土正在被当作实验室，用于测试目前在全世界销售的形式复杂的自主武器系统。这些武器的开发和使用对全人类构成严重威胁。我们认为，秘书长必须大胆地向所有国家强调，迫切需要一项具有法律约束力的文书，防止这些系统的法律、伦理、人道主义和安全风险。

然而，与文书的形式同样重要的是框架的实质性政策，它必须能够在实践中实际处理这些风险。对巴勒斯坦国来说，显然必须通过一项定义，防止出现使各国能够绕过商定禁令和条例的漏洞。其中一个不可或缺的组成部分是，需要认识到“名义人工投入”并不等于为界定自主武器系统而进行的干预。

《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》的一些缔约国以及红十字国际委员会广泛同意，自主武器系统的特点是：

“由人类用户启动后，通过处理传感器数据来选择和攻击目标而无需人工干预的系统”。

虽然巴勒斯坦国在对自主武器系统的定义中也使用了“没有人工干预”一词，但我们认为，如果我们在不作进一步澄清的情况下接受“没有人工干预”一词，则可能在定义中造成重大漏洞。从理论上说，一个系统只需在激活后有一次人工投入，就不在自主武器系统框架内。

因此，必须界定我们所说的“没有人工干预”的含义。例如，在系统启动后，要求人类用户按下键盘按钮以执行武力，而不考虑任何伦理或法律后果，这是否等于“人工干预”，从而使武器不属于自动武器系统的范畴？换句话说，人在系统启动后不假思索地点击键盘按钮，是否会将该系统排除在自动武器系统框架之外，从而使载有禁止和管制规定包括需要人实际控制规定的框架根本不适用于该武器系统？

这种解释不仅完全违反直觉，而且更重要的是，它将非常危险。

事实上，许多武器公司都以开发“最低程度必要人工投入”的自主武器为荣。这种系统声称不属于自动武器系统的定义，因为它们在系统启动后需要人工投入（无论多么小）。从对大量具有自主性的武器系统的分析中可以明显看出，几乎所有这些系统都允许人在系统启动后以“名义人工投入”与系统互动。尽管是“名义上的”，但武器设计者和制造商仍能借此避免让系统贴上自主武器系统的标签，他们说，因为可以在系统启动后进行干预，所以系统不在自主武器系统的范围之内。

我们的呈件试图将这一点作为问题提出，通过引入“名义人工投入”的概念来填补漏洞，并促使各国和其他利益攸关方实质性地参与这一领域的思考。

我们将“名义人工投入”定义为在系统启动后和自主过程中由人执行的不会对过程产生实质性影响的投入。如果人工投入没有带来更多信息来选择和使用武力打击目标的决定提供信息，那么它就不会对自主过程产生实质性影响。换句话说，在这一人工投入的过程中，没有进一步的人类伦理和法律反思。这些“名义人工投入”是在远离该系统执行武力的地点和时间作出的，而且人也没有适当考虑到发生违反国际法行为的风险。因此，人工投入是“不假思索的”，其效果就像没有人工投入的自主过程的效果一样。

我们必须对自主武器系统作出适当的定义，使这些系统能够纳入可能的禁止和管制范围。最后，需要明确的是，“名义人工投入”并不等于各国正在讨论的本来合理的定义中所指的“干预”，而且需要明确的是，即使有这种投入，该系统仍应被视为自动武器系统。

巴勒斯坦国提交秘书长供其注意的工作文件进一步详细阐述了“名义人工投入”的含义。

执行摘要

1. 自主武器系统是由人类用户启动后，通过处理传感器数据来选择和攻击目标而无需人工干预的系统。
2. 系统启动后的名义人工投入不等于人工干预。
3. 这些系统带来了一系列法律、伦理、人道主义和安全风险。为应对这些风险，既需要禁止，也需要管制。
4. 必须禁止开发和使用下列自主武器系统：
 - (a) 设计或用于直接以人类为攻击目标；
 - (b) 使用时人无法实际控制。
5. 人实际控制要求自动化武器系统必须满足以下所有要求。这些系统必须：
 - 可预测
 - 可靠
 - 可理解和可解释
 - 可追踪
6. 需要包括积极义务和限制在内的条例，以确保自动化武器系统可以在人实际控制下使用。
7. 这种禁止和管制的结合应该采取具有法律约束力的国际文书的形式。在通过这样一项文书之前，必须暂停开发自动化武器系统。

本呈件完整版本可在 <https://meetings.unoda.org> 上查阅。

B. 欧洲联盟

[原件：英文]

[2024年5月25日]

新兴技术，特别是人工智能和武器系统的自主，正在重塑冲突格局及其对全球安全的影响。欧洲联盟认为，我们必须集体应对这些挑战，制定原则、国际规范和条例，确保负责任地使用这些技术。

欧洲联盟认识到，人工智能既是一种赋能性技术，也是一种颠覆性技术。在军事领域开发和使用人工智能，包括其对全球安全的影响，既带来潜在机会，也有潜在风险。欧洲联盟鼓励进一步参与国际对话与合作，应对军事领域人工智能带来的机遇和挑战。我们支持有关论坛努力交流最佳做法，促进对法律和伦理影响共同理解。我们欢迎最近的国际和区域倡议作出的宝贵贡献。

关于减轻风险和建立信任措施，欧洲联盟认识到数据对基于人工智能的技术发挥的关键作用。还应适当考虑对新兴技术有潜在影响的社会偏见，例如算法中的性别偏见。应采取和执行定制的风险缓解措施，包括在整个生命周期内采取的措施。

欧洲联盟认为，《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》提供了一个适当框架，可以在《公约》的目标和宗旨范围内处理致命自主武器系统领域的新兴技术问题。欧洲联盟仍然致力于在致命自主武器系统问题政府专家组中继续努力，履行任务并在这一问题上取得进展，确保成果除其他外反映遵守国际法特别是国际人道法的必要性，同时考虑到相关的伦理因素。过去的成就，包括 11 项指导原则、以前的成果文件以及已经确定的共同点，应成为协商一致建议的基础。欧洲联盟强调，关于使用致命武力的决定必须由人作出，人必须对其使用的致命武器系统进行控制，并对使用武力的决定负责，确保遵守国际法特别是国际人道法，同时考虑到伦理因素。因此，那些使用致命自主武器系统计划、决定和实施攻击的人必须确保武器系统及其使用方式将维护人作出必要法律判断的能力，从而确保遵守国际人道法。此外，必须在武器系统的整个生命周期中始终保持对人的问责，并应在这方面实行适当措施。

欧洲联盟回顾，各国负有根本责任确保致命自主武器系统领域新兴技术的开发、生产、部署和使用符合国际法特别是国际人道法。在这方面，我们支持所谓的“两级方法”，即区分各国应承诺不开发、不生产、不使用的根据国际法特别是国际人道法不能使用的武器系统，以及需要管制以确保遵守国际人道法和其他适用国际法的具有自主功能的系统。

正如大会关于致命自主武器系统的第 78/241 号决议所指出，国际和区域会议和倡议，包括有欧洲联盟成员国参加的会议和倡议，作出了重要贡献，丰富了关于自主武器系统的国际讨论。

最后，性别平等和增强妇女权能是欧盟的一个重要横向优先事项。我们认为，考虑到性别平等与新兴技术之间的联系，在讨论致命自主武器系统问题时必须纳入性别视角。

附件二

从国际和区域组织、红十字国际委员会、民间社会、科学界和工业界收到的答复¹

A. International Committee of the Red Cross

[19 March 2024]

Summary

Full submission available at: <https://www.icrc.org/en/document/autonomous-weapons-icrc-submits-recommendations-un-secretary-general>.

The International Committee of the Red Cross (ICRC) submits its views to the United Nations Secretary-General, in accordance with General Assembly resolution [78/241](#).

The ICRC's assessment is that the unconstrained development and use of autonomous weapon systems (AWS) pose serious legal, ethical, and humanitarian concerns.²

Need for new, binding international law

While international humanitarian law (IHL) already regulates and constrains the design and use of AWS, States hold different views regarding the specific limits and requirements it imposes. Thus, the ICRC believes that new rules are urgently needed to provide legal certainty and stability, and to address wider humanitarian risks and fundamental ethical concerns.

Guiding principle: human control over the use of force and effects

IHL requires weapon users to be able to anticipate, control and limit the effects of weapons.³

¹ In accordance with operative paragraph 3 of General Assembly resolution [78/241](#), the replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry are included in the original language received. The Secretary-General remains committed to multilingualism as a core value of the United Nations.

² ICRC position paper on autonomous weapon systems, May 2021: <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>; Joint Call by the United Nations Secretary-General and the President of the ICRC, October 2023: <https://www.icrc.org/en/document/joint-call-un-and-icrc-establish-prohibitions-and-restrictions-autonomous-weapons-systems>; ICRC commentary on the guiding principles of the CCW GGE, 2020: <https://documents.unoda.org/wp-content/uploads/2020/07/20200716-ICRC.pdf>.

³ E.g., the rule on proportionality requires weapon users to be able to 'anticipate' effects in the form of military advantage, and the prohibition against indiscriminate attack requires weapon effects to be capable of being 'limited' (Additional Protocol I, Art. 51(4 and 5(b)); Customary IHL Rules 12 and 14).

While IHL obligations do not necessarily demand direct human control over the weapon itself at all stages of its use, they do require human control over the weapon's *effects* in the circumstances of a specific attack. This principle should underpin the drafting and interpretation of a legally binding instrument on AWS, even if "human control" is not explicitly included as a requirement. This calls for a combination of prohibitions and restrictions on the design of the weapon and on its operating parameters.

Preamble

The legal instrument could include preambular paragraphs: reaffirming the need for compliance with existing legal frameworks; acknowledging the range of concerns raised by AWS; reaffirming the need to continue codification and progressive development of IHL; and recalling the protection afforded by the "Martens clause."⁴

Definitions

The instrument should contain an unambiguous definition covering the general category of AWS to which the whole instrument applies. Within that category, certain types of AWS will be subject to specific prohibitions (see next section).

"Autonomous weapon system" means a weapon system that is designed to select and engage one or more targets without the need for human intervention after activation.

This understanding – shared by the ICRC, many States and other actors is not based on a specific technology, but rather on the role of humans in the process of target selection and application of force. Such a functional and technology-neutral approach to the characterization of AWS is essential to ensure that new rules remain relevant in the face of technological developments.

"Without the need for human intervention" could be defined as meaning that, after initial activation by a human, the application of force is triggered in response to information from the environment received through sensors, and on the basis of a generalized "target profile."⁵ "Human intervention", for these purposes, should be understood as excluding human inputs that do not materially affect the autonomous functions of target selection or engagement.

Definitions of other concepts including "military objective", "self-destruction mechanism" and "self-deactivating" could be incorporated from existing international agreements.⁶

⁴ E.g., in Additional Protocol I, Art. 1.

⁵ The concept of a "target profile" has been referred to in the GGE, e.g., 2019 Report [CCW/GGE.1/2019/3](#), and could be further defined in an instrument.

⁶ E.g., Additional Protocol I, Art. 52(2); CCW Amended Protocol II; Convention on Cluster Munitions.

Prohibitions*Unpredictable AWS*

The instrument should provide that it is prohibited in all circumstances to develop, produce, otherwise acquire, stockpile or retain, or transfer, directly or indirectly to anyone, or to use any AWS that is designed or of a nature, or used in such a manner that does not allow a human user to both (1) understand, predict and explain how the AWS will function in any normal or expected circumstances of use, in particular what circumstances or conditions will trigger the system to apply force, and (2) predict and limit the effects of the AWS in all such circumstances as required by IHL.

Users of AWS must be able to, with a reasonable degree of certainty, predict the effects of that weapon, in order to determine whether it can be directed at a specific military objective and take steps to limit those predicted effects, as required by IHL. This entails the ability to understand the nature and functioning of the AWS' sensors, the definition of its target profile and the potential effects in the circumstances of use, including any risk of error or malfunction.

Anti-personnel AWS

The instrument should provide that it is prohibited in all circumstances to develop, produce, otherwise acquire, stockpile or retain, or transfer, directly or indirectly to anyone, or to use any AWS:

- that is designed or used in such a manner to be triggered by the presence, proximity or contact of one or more persons, or
- the target profile of which otherwise represents one or more persons.

AWS designed and used to target humans pose real risks of IHL violations. Target selection in an AWS is based on a generalised target profile, which is unlikely to account for the non-exhaustive range of contextual signals indicating that a person is protected from attack (e.g., whether surrendering, wounded, or sick etc).

Further, a prohibition against anti-personnel AWS is necessary due to the unacceptability of such weapons from an ethical perspective, since the killing of a human based on a machine process would undermine the human agency of the person using force and the dignity of the person against whom force is used.

Restrictions

The instrument should provide that, in the use of AWS other than those which are prohibited by the other articles, measures shall be taken to protect civilians and civilian objects, and other protected persons, from the effects of AWS, including:

- restricting targets to only those which are military objectives by nature

- limiting the location, time and situation in which the AWS is operating, including to avoid concentrations of civilians or civilian objects⁷
- limiting the number of engagements that the AWS can undertake
- ensuring, to the maximum extent feasible, the ability for a human user to:
 - effectively supervise, and
 - in a timely manner intervene and, where appropriate, deactivate operation of the AWS
- AWS that do not allow a human user to do so, must be equipped with an effective self-destruction, self-deactivation or self-neutralization mechanism, designed so that the AWS will no longer function as an AWS when it no longer serves the military purpose for which it was launched.⁸

The ICRC is grateful for the opportunity to share the above views and stands ready to assist States in taking effective action to address the risks posed by AWS.

B. Civil society

Africa Teen Geeks

[13 May 2024]

As the rapid advancement of technology continues to reshape global landscapes, the development and potential deployment of autonomous weapons systems (AWS) present significant ethical, legal, and security challenges. This input aims to provide a comprehensive overview of key considerations from industry, scientific, state, and civil society perspectives to aid in the formulation of an international treaty governing autonomous weapons.

Ethical Considerations

One of the foremost concerns regarding AWS is the ethical implications of delegating life-and-death decisions to machines. Ethical considerations include the potential loss of human control over lethal force and the challenge of ensuring that AWS operate within the bounds of international humanitarian law (IHL). Autonomous weapons must adhere to principles of distinction, proportionality, and necessity, yet their ability to make nuanced judgments akin to human operators remains questionable.

⁷ NB similar language found in Additional Protocol I, Art. 51(5)(a) and CCW Amended Protocol II and Protocol III.

⁸ Language of CCW Amended Protocol II.

Legal Implications

The integration of AWS into military arsenals necessitates a robust legal framework. Existing laws, such as the Geneva Conventions, must be examined to determine their applicability to autonomous systems. Additionally, the legal accountability for actions taken by AWS poses a complex issue. The lack of clear attribution for unlawful acts could undermine the enforcement of IHL and erode trust in the international legal system. Establishing clear guidelines for the development, deployment, and use of AWS, along with mechanisms for accountability, is crucial.

Security Concerns

From a security standpoint, the deployment of AWS could trigger an arms race, with nations striving to outpace each other in developing increasingly advanced systems. This escalation could destabilize global security and increase the likelihood of conflicts. Furthermore, the risk of AWS being hacked or malfunctioning poses a significant threat. Ensuring robust cybersecurity measures and fail-safes are in place is essential to mitigate these risks.

Industry Perspectives

The defense industry plays a pivotal role in the development of AWS, driving innovation and technological advancements. However, it also bears a responsibility to ensure that these technologies are developed ethically and safely. Industry stakeholders must engage in transparent and accountable practices, adhering to international standards and norms. Collaborative efforts between industry, governments, and international bodies are necessary to establish guidelines and best practices for AWS development.

Scientific Insights

Scientists and researchers contribute critical insights into the capabilities and limitations of AWS. Ongoing research must focus on enhancing the reliability and predictability of autonomous systems, ensuring they can operate within ethical and legal boundaries. Interdisciplinary research, combining expertise from artificial intelligence, robotics, law, and ethics, is essential to address the multifaceted challenges posed by AWS.

State Perspectives

States have a crucial role in shaping the international regulatory framework for AWS. National governments must engage in multilateral dialogues to harmonize their positions and develop a cohesive approach to AWS governance. This includes committing to transparency in the development and deployment of AWS, as well as supporting international efforts to establish binding legal instruments. States should also invest in research and development to ensure their defense capabilities are ethically and legally sound.

Civil Society Engagement

Civil society organizations (CSOs) offer invaluable perspectives on the humanitarian and ethical dimensions of AWS. CSOs can serve as watchdogs, advocating for stringent regulations and monitoring compliance. Their involvement ensures that the voices of affected populations are heard, and that human rights considerations are central to the discourse. Collaborative efforts between CSOs, governments, and industry are essential to develop a holistic approach to AWS governance.

Recommendations

1. **Establish Clear Definitions and Boundaries:** Develop internationally agreed-upon definitions of AWS and establish clear boundaries for their development and use.
2. **Implement Robust Legal Frameworks:** Ensure that AWS are integrated into existing legal frameworks, with clear guidelines for accountability and compliance with IHL.
3. **Promote Transparency and Accountability:** Encourage transparency in the development and deployment of AWS, with mechanisms for accountability and oversight.
4. **Foster International Cooperation:** Strengthen multilateral dialogues and cooperation to develop cohesive and harmonized approaches to AWS governance.
5. **Enhance Cybersecurity Measures:** Prioritize cybersecurity to protect AWS from hacking and malfunctions, ensuring their safe and reliable operation.
6. **Encourage Interdisciplinary Research:** Support interdisciplinary research to address the ethical, legal, and technical challenges posed by AWS.
7. **Engage Civil Society:** Involve civil society in the discourse to ensure that humanitarian and ethical considerations are prioritized.
8. **Prevent Arms Race:** Implement measures to prevent an arms race and promote the responsible development and deployment of AWS.

The development of an autonomous weapons treaty is a critical step in ensuring that the advancement of military technology adheres to ethical, legal, and security standards. By incorporating insights from industry, scientists, states, and civil society, the international community can develop a comprehensive and effective regulatory framework for AWS. Collaborative efforts and robust dialogue are essential to address the multifaceted challenges posed by autonomous weapons and to safeguard global peace and security.

Amnesty International

[24 April 2024]

This submission highlights the intractable challenges related to the use of AWS in law enforcement contexts in relation to compliance with international human rights law (IHRL) and standards on the use of force. For Amnesty International, AWS are weapons systems that detect and apply force to a target based on sensor inputs, rather than an immediate human command.

International human rights law/standards

The use of AWS in law enforcement threatens to undermine the right to life, liberty and security of person, the right to be free from torture and other cruel, inhuman or degrading treatment or punishment, and the right to freedom of peaceful assembly, among other rights. It would also be incompatible with the key international standards related to the use of force and law enforcement practice: the UN Basic Principles on the Use of Force and Firearms (BPUFF) and the UN Code of Conduct for Law Enforcement Officials (CCLEO).⁹ The use of AWS – whether lethal or less lethal – undermines the principles of human rights-compliant law enforcement, which are based on close, positive and proactive human relationships between law enforcement officers and the publics they serve.

Human agency

Human agency and judgement are required for the lawful use force under IHRL law and standards. The decision to use force, whether lethal or less lethal, can only be made by a human being and human rights law “places a strong emphasis on human reasoning and interaction”.¹⁰ According to BPUFF, human abilities to communicate, de-escalate, persuade and negotiate are central strategies for avoiding or minimizing the use of force. The decision to deploy intentional lethal force comes at an extremely high threshold, and is especially nuanced and complex, taking into account emotional, motivational, psychological and logistical factors to assess the imminence of a threat to life and limb.¹¹ AWS cannot be used to replace indispensable human judgement in the decision to use force; and would not be able to operate in a way that seeks to

⁹ UN Basic Principles on the Use of Force and Firearms by Law Enforcement Officials (UN Basic Principles), adopted on 7 September 1990; UN Code of Conduct for Law Enforcement Officials (UN Code of Conduct), adopted on 17 December 1979 by General Assembly resolution 34/169.

¹⁰ OHCHR, Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons 13–16 May 2014, Geneva, Switzerland by Christof Heyns, Professor of human rights law, University of Pretoria United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions, 13 May 2014, www.ohchr.org/en/statements/2014/07/presentation-made-informal-expert-meeting-organized-state-parties-convention; see also UK Parliamentary Committee, Professor Thompson Chengeta – Written Evidence (AIW0021), <https://committees.parliament.uk/writtenevidence/120291/html/>.

¹¹ UN Basic Principles, Principles 9 and 20.

minimize harm, prevent the loss of life or assess whether the necessity and proportionality of lethal force is still justified at a given moment.

Accountability

The use of force without meaningful human control could create an “accountability vacuum”.¹² This derives from the fact that international human rights law binds states and individuals – not machines - and seeks to hold them to account, and the difficulties in law of ascribing human responsibility when using machines operating outside of meaningful human control.¹³ The use of AWS would weaken the ability of a law enforcement officer or their superiors to make judgements about individual responsibility for the unlawful use of force and to take appropriate precautions to prevent further violations – one element of the provision of remedy for harm suffered.

Dignity

The question of dignity is core to IHRL. The Universal Declaration on Human Rights and the International Covenant on Civil and Political Rights recognise that human rights are derived from the “inherent dignity” of all.¹⁴ The Special Rapporteur on extrajudicial, summary or arbitrary executions has argued that the concept of dignity is inextricably linked to the right to life as “it is the value of life that makes it worth protecting”.¹⁵ The African Charter on Human and Peoples’ Rights General Comment 3 on the right to life “proceeds from an understanding that the Charter envisages the protection not only of life in a narrow sense, but of dignified life”.¹⁶ In relation to law enforcement, Article 2 of the CCLEO states that law enforcement officers “shall respect and protect human dignity and maintain and uphold the human rights of all persons.”

Bias/discrimination

One of the dangers in the development of AWS is the already well-developed use of biometric data in law enforcement. Many police forces have developed

¹² Christof Heyns, “Human Rights and the use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement, Human Rights Quarterly”, Vol. 38, No. 2, May 2016, p. 366.

¹³ Interim report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, [A/65/321](#), 23 August 2010, para. 33

¹⁴ Human Rights Watch, *Shaking the Foundations: The Human Rights Implications of Killer Robots*, IV. Human Dignity, www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots

¹⁵ OHCHR, Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons 13–16 May 2014, Geneva.

¹⁶ African Commission on Human And Peoples’ Rights, General Comment No. 3 On The African Charter On Human And Peoples’ Rights: The Right To Life (Article 4), Adopted During the 57th Ordinary Session Of The African Commission on Human and Peoples’ Rights held from 4 to 18 November 2015 in Banjul, (3), <https://achpr.au.int/en/node/851>

largescale databases linking biometric data to other personal data, including criminal justice system records.¹⁷ It would be a short step for law enforcement to feed this data into an AWS in order to target criminal suspects. The use of biometric markers to identify targets carries enormous risks. The Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression has noted that facial recognition can lead to profiling on ethnicity, race, national origin, gender and other characteristics, which are often the basis for unlawful discrimination.¹⁸

The need for a legally binding instrument to prohibit and regulate AWS

To clarify and strengthen existing international humanitarian and human rights law as it relates to growing autonomy in weapon systems used in both military and law enforcement contexts, Amnesty International advocates for the creation of a legally binding instrument to prohibit some types of AWS and regulate others. Such an instrument would:

- prohibit the development, production, use of, and trade in systems which by their nature cannot be used with meaningful human control over the use of force;
- prohibit “anti-personnel AWS” – that is systems that are designed to be triggered by the presence of humans or that use human characteristics for target profiles;
- regulate the use of all other autonomous weapons systems;
- and include a positive obligation to maintain meaningful human control over the use of force.

Arms Control Association

[24 May 2024]

In Resolution [78/241](#), the General Assembly expressed its concern, *inter alia*, about the “impact of autonomous weapon systems on global security and regional and international stability, including the risk of an emerging arms race [and] lowering the threshold for conflict and proliferation.”

¹⁷ For example, Amnesty International, *Automated Apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT*, 2 May 2023, (Index Number: MDE 15/6701/2023), www.amnesty.org/en/documents/mde15/6701/2023/en/; Amnesty International, Ban the Scan, banthescan.amnesty.org/

¹⁸ Surveillance and human rights: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression to the UN Human Rights Council, UN General Assembly, [A/HRC/41/35](#), 28 May 2019, para. 12: <https://www.undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F41%2F35&Language=E&DeviceType=Desktop&LangRequested=False>

The Arms Control Association shares these concerns about the impact of AWS on international peace and stability. For more than fifty years, the ACA has worked to promote effective measures to reduce nuclear risks through national self-restraint, diplomatic engagement, bilateral and multilateral arms control, nonproliferation, and disarmament, and other forms of international regulation.

Notwithstanding the ACA's primary focus on reducing the dangers posed by nuclear weapons and achieving full nuclear disarmament, we believe that the deployment of autonomous weapons systems and automated battlefield command-and-control (C2) systems pose significant risks to strategic stability, and therefore require strict regulation and oversight. Two ongoing developments, we believe, are of particular concern: the integration of autonomy with nuclear command, control, and communications (NC3) systems, and the use of conventionally armed AWS to target and destabilize nuclear forces.

Threats to Strategic Stability Between Nuclear-Armed States:

The major powers are automating their battlefield C2 systems and equipping them with algorithms for calculating enemy moves and intentions, selecting the optimal countermoves, and dispatching attack orders directly to friendly units for implementation—all with ever-diminishing human oversight. Research by a number of analysts suggests that in future conflicts among the major powers, such systems will contribute to and increase the risk of mutually reinforcing escalatory moves, potentially igniting accidental or inadvertent nuclear escalation.¹⁹

Although none of the nuclear powers are thought to be extending this type of software to autonomously manage their nuclear forces, many states see a potential for and are likely already developing AI algorithms to assist discrete components of their nuclear early warning and launch systems, for example with the interpretation of possible enemy missile launches.²⁰ It is essential that AI software used to support these applications remain physically disconnected from nuclear launch authority to prevent any possibility of an unintended AI-triggered nuclear exchange.

Meanwhile, concern is growing that conventionally armed AWS, in combination with advanced, AI-enhanced autonomous intelligence and reconnaissance systems, might contribute to accidental or unintended nuclear escalation by creating the

¹⁹ See Eric Schmidt, et al., "Final Report of the National Security Commission on Artificial Intelligence," March 2021, <https://cybercemetery.unt.edu/nscai/20211005220330/https://www.nscai.gov/>, and Michael T. Klare, "Assessing the Dangers: Emerging Military Technologies and Nuclear (In)Stability," Arms Control Association Report, February 2023, https://www.armscontrol.org/sites/default/files/files/Reports/ACA_Report_EmergingTech_digital_0.pdf.

²⁰ Alice Saltini, "AI and Nuclear Command, Control and Communications: P5 Perspectives," Report, European Leadership Network, Nov. 2023, https://www.europeanleadershipnetwork.org/wp-content/uploads/2023/11/AVC-Final-Report_online-version.pdf, pp. 16-17.

impression that an attacker is conducting a disarming counterforce strike, aimed at eliminating or degrading the target state's nuclear retaliatory capabilities.

Of particular concern is the potential of loitering AWS to reveal the location of elusive nuclear retaliatory forces, such as mobile ICBMs or ballistic missile submarines.²¹ The fear that an AI-controlled AWS swarm could uncover the locations of a nuclear-armed state's submerged submarines or road-mobile ICBMs could prompt that state to place its weapons on a higher state of alert in a crisis and possibly trigger their unintended or accidental use.

Retaining Human Control

The Arms Control Association strongly adheres to the principle that the decision to use nuclear weapons must always remain the responsibility of a human being, and that such decisions conform with the Laws of War and particularly International Humanitarian Law, which rules out the employment of nuclear weapons particularly in response to nonnuclear threats. The profound legal, ethical, and humanitarian ramifications of any nuclear weapons employment—potentially extinguishing the lives of millions of people and rendering the planet uninhabitable—demand that humans, and never machines, bear the responsibility and moral culpability for their use.

Starting from this premise, and in recognition of the risks of escalation described above, we also believe that any fully autonomous weapons systems or automated battlefield C2 systems operating outside of continuous human supervision when in combat should be prohibited under binding international law and that all other lethal weapons systems featuring autonomy be regulated in order to ensure compliance with international humanitarian law, including by insisting on human responsibility and accountability.

Recommended Actions

In accordance with these basic principles, the Arms Control Association offers these additional recommendations to the Secretary General and the General Assembly:

1. Mindful that the use or threat of nuclear weapons has been deemed “inadmissible” and contrary to international law and the Treaty on the Prohibition on Nuclear Weapons, *the UN General Assembly should call on all nuclear-armed states to commit—either through coordinated action or in a binding agreement—to retain human control over any decision to use nuclear weapons and to insert automated, failsafe “tripwires” in advanced command-and-control systems to disallow action resulting in nuclear weapons employment without human approval.*

²¹ James S. Johnson, “Artificial Intelligence: A Threat to Strategic Stability,” *Strategic Studies Quarterly*, Vol. 14, No. 1 (Spring 2020), pp. 20-22.

Ideally, the nuclear weapons states should themselves take steps toward creating an international norm that recognizes and affirms this principle by issuing unilateral statements that decisions involving nuclear use will always be reserved for human beings. A more ambitious but more effective measure would be a multilateral statement by the P5 nuclear-weapons states that jointly commits to the same norm.

To give effect to this norm, the nuclear weapons states should integrate technical tripwires in all deployed C2 systems that would automatically prevent escalation to nuclear weapons use without human intervention. Critically, this would also mean ensuring that all AI-enabled C2 systems for conventional military operations are carefully and deliberately prohibited from giving instructions to nuclear weapons systems.

2. *The UN General Assembly should call upon on all states to commit to retaining uninterrupted human control over any AWS potentially involved in strategic counterforce missions and to disallow such weapons from falling under the control of AI-enabled decision-support systems that could authorize counterforce missions without human oversight.*

Such commitments are urgently needed because unauthorized, accidental strikes on nuclear forces by loitering autonomous strike systems could give rise to false warning of an incoming strategic attack. Likewise, unauthorized conventional strikes with strategic implications could be undertaken by an AWS strike force that is given erroneous orders by an AI-enabled decision-support system.

To prevent this category of accidental escalation, states should ensure that forces assigned to conventional counterforce missions with strategic implications remain under human control at all times and forego integration with AI systems altogether.

3. *The UN General Assembly should convene an expert body to assess the types and roles of AI algorithms that are used in nuclear command and control systems and the dangers these could pose.* This body should also suggest possible restrictions on the use of AI in nuclear C3 systems and whether there are certain roles within NC3 systems that should never be assigned to algorithms.

Given the rapid pace of research into new AI models and the lack of existing norms and understandings between nuclear powers about their application, the United Nations could play a key role in convening experts to track the technical evolution of these models.

Article 36

[8 May 2024]

Article 36 has worked extensively on the issue of autonomous weapons – including framing the requirement for meaningful human control and promoting the

need for a structure of international legal regulation that includes both prohibitions and positive obligations.

This submission only highlights key points that we consider significant at this stage of the process.

We should recognise autonomous weapons systems as referring to ‘systems’ or ‘processes’, rather than ‘objects’.

Discussions of this issue often talk about ‘autonomous weapons’ as concrete, unified physical ‘objects’ – that is to say, as physical objects that share a recognisable set of characteristics. However, the defining characteristics of autonomous weapons systems are the tied to the relationship of human users to processes of decision-making.

Autonomous weapons systems may function through distinct and widely dispersed physical assets, all of which *might also* function in ways that would not constitute an autonomous weapons system.

Rules therefore need to be focused on human understanding and control over individual attacks and on how such systems are used. There will still be unified physical systems that need to be subject to these rules, but the starting point should be to regulate human understanding and control of the ‘process’.

AI is not a necessary characteristic of autonomous weapons, but it raises distinct challenges.

AI is one of the technical drivers that is making the issue of autonomous weapons systems particularly pressing. However, it is the relationship of human operator(s) to certain decision-making processes that is the defining characteristic, not the technology that is involved.

It is possible to have autonomous weapons systems that do not employ AI and we should not define the boundaries of this issue in relation to AI. However, AI does provide distinct challenges. AI may make it more difficult for the users of systems to have a practical understanding of how their systems work and so to adequately predict outcomes from their use. In certain roles, AI might serve to embed bias from training datasets into the functioning of weapon systems – which is a particular challenge in relation to weapon systems that would target people, or specific groups of people.

A new legal instrument could be short - establishing key overarching rules that provide a structure for evaluating technological developments in the future.

A legal instrument on this issue should focus on the key general rules that promote human dignity and meaningful human control. Rules should include:

- A prohibition on using AWS to directly target people (anti-personnel systems).

- Positive obligations to ensure meaningful human control, including requirements that:
 - Users sufficiently understand AWS they intend to use, including the conditions that would trigger an application of force by the system;
 - Users sufficiently evaluate the context where the system would be used; and
 - Users sufficiently limit the duration and area of system functioning in order to meaningfully apply existing legal rules.
- A prohibition on systems that cannot be used in accordance with these positive obligations, and so are likely unpredictable and incompatible with the necessary human control.

Such a legal structure can then provide a framework under which specific cases can be addressed.

This must be a future orientated instrument against which new technological structures are evaluated as they are developed (including through weapon review processes). Sharing good practices on assessment processes would be valuable multilateral work once the legal instrument has been established.

Regulating autonomous weapons is an important opportunity to limit the negative potential of AI without curbing its wider positive potential.

Adopting this legal treaty should be recognised as a critical action to prevent negative effects from AI in the world. The legal treaty would establish guardrails that prevent the development and adoption of AI functions in some critical roles that undermine human control and human dignity in the use of force. As such, it points to one potential mode for regulating AI more broadly which is to limit its scope of use in specific roles and manifestations. The key to that regulatory mode is not to regulate the AI directly (which is too amorphous) but to establish the obligations for human understanding and action.

Current ‘defensive systems’ (missile defence systems etc) should not be prohibited, but *should* fall within a legal instrument and be used in accordance with its positive obligations (this is in line with current practice).

Some states have raised concerns that ‘defensive’ systems should not fall within the scope of consideration of discussions regarding autonomous weapons. The types of ‘defensive’ systems driving these concerns are broadly ‘anti-missile’ systems that use sensors and computer-directed guns to detect and apply force to incoming weapons (such as missiles, rockets and mortars).

These systems fall within the scope of the consideration because they use sensors to determine specifically where and when force will occur in response to

matching data from the environment against a generalised target-profile. However, such systems would not be considered at risk of prohibition under a future instrument because:

- A. they can be used with meaningful human control, appropriate human judgement etc. The users of such systems can have an effective understanding of how these systems function, including what will trigger an application of force by the system and the location and duration of system functioning can be specifically controlled by the human operator.
- B. they do not target 'people' directly.

Given this analysis, we do not see that anti-missile systems could be prohibited under the two-tier approach.

A prohibition on systems that would target people directly should be a critical moral and societal priority.

The ethical and moral concerns with respect to autonomous weapons are most critical in relation to systems that would target people directly. Allowing systems to be used to harm people on the basis of machine processing is dehumanising and should be considered incompatible with requirements of human dignity. Such systems would also be fraught with legal risks.

Claiming that systems could somehow distinguish combatants from civilians would be a transference to machine functioning of determinations that should be made by a human commander. Furthermore, such mechanisms would likely neglect the obligation to protect soldiers *hors de combat* and may be liable to problems of racial, age and gender bias if built on certain AI processes.

We have an opportunity to prevent the adoption of autonomous systems that target people. The working presumption for future negotiations should be that systems targeting people are unacceptable.

A legal instrument should be developed through an inclusive multilateral process that is open to all states but that cannot be blocked by any one country.

It is urgent to start negotiations and that process needs to start in a forum that can bring in the views of diverse stakeholders and that is open to all states to participate (if they wish) on equal terms. It is not prudent to insist that discussions should only take place in forums where militarised states are consistently allowed to prevent the majority from moving forwards.

Energising international humanitarian law and international commitment to protect civilians requires action in a framework that has the potential to reflect the will of the majority. This issue is too fundamentally important for society to continue to remain constrained by procedural exploitation.

Asociación TEDIC, Amnistía Internacional Paraguay, Centro de Estudios Heñói, CODEHUPY, Fundación Vencer y Semillas para la Democracia

[18 April 2024]

Las organizaciones no gubernamentales paraguayas abajo firmantes, comprometidas con la promoción de una cultura de paz, con los derechos humanos y firmes contra toda forma de deshumanización, señalamos cuanto sigue:

Los sistemas de armas de destrucción autónoma (SADA) son un tema de creciente preocupación en el ámbito internacional, ya que plantean serias cuestiones éticas, legales y de seguridad. Las máquinas están reemplazando a los humanos en la aplicación de la fuerza con consecuencias impredecibles y devastadoras para la humanidad. Estos sistemas, también conocidos como “armas letales autónomas” o “robots asesinos”, son sistemas de armas que están diseñados para seleccionar y atacar objetivos sin intervención humana directa una vez que han sido activados. Utilizan inteligencia artificial (IA) y algoritmos de toma de decisiones para identificar y atacar objetivos.

Es importante reconocer que antes de crear regulaciones, es necesario realizar una investigación exhaustiva sobre los sistemas de armas autónomas, sus capacidades, implicaciones éticas, riesgos y posibles consecuencias. Esto necesita involucrar a expertos en diversas disciplinas, incluidos científicos, ingenieros, juristas, filósofos, expertos en ética y representantes de la sociedad civil.

En algunos países del sur global, donde ni siquiera tenemos regulación en materia de protección de datos personales²², afectados además por el avance permanente de grupos vinculados al crimen organizado transnacional y grupos violentos paramilitares, el despliegue de este tipo de tecnologías plantea riesgos serios para asegurar una gobernanza global de este tipo de sistemas, porque los pisos de conocimiento de los Estados son distintos.

Por eso es esencial una regulación internacional para salvaguardarnos contra los riesgos éticos, legales y de seguridad que plantean los sistemas de armas autónomos. Exigir un control humano significativo y contrarrestar la deshumanización digital, que garantice la responsabilidad y la rendición de cuentas, en cualquier uso de la fuerza, es de suma importancia. La regulación debe basarse en principios éticos sólidos que protejan los derechos humanos, minimicen el sufrimiento innecesario y preserven la dignidad humana.

²² En el caso paraguayo, la Coalición de Datos Personales viene luchando hace años por la adopción de una ley integral de protección de datos personales. <https://www.datospersonales.org.py/comunicado-de-la-coalicion-de-datos-personales-en-respuesta-a-las-publicaciones-y-declaraciones-hechas-en-medios-periodisticos-sobre-el-proyecto-de-ley-de-proteccion-de-datos-personales/>

En el contexto de los países del sur global, hay varias implicaciones y preocupaciones específicas sobre los sistemas de armas de destrucción autónoma:

1. **Desigualdad tecnológica:** Los países del sur global enfrentan desafíos para desarrollar o adquirir tecnología sobre armas autónomas debido a limitaciones financieras, de recursos humanos y tecnológicas. Esto podría crear una brecha tecnológica entre los países desarrollados y en desarrollo, lo que podría aumentar la dependencia de estos últimos en tecnologías militares importadas.
2. **Aumento de la brecha de poder:** La proliferación desigual de SADA podría exacerbar las disparidades en el poder militar entre los países del sur global y las potencias mundiales. Esto podría tener implicaciones en términos de seguridad regional y global, así como en la capacidad de los países del Sur Global para proteger sus intereses nacionales.
3. **Falta de regulación:** Existe una falta de regulación internacional sólida sobre el desarrollo y el despliegue de armas autónomas. Esto plantea preocupaciones sobre el uso indiscriminado o inapropiado de tales sistemas, especialmente en conflictos en los que los países del Sur Global podrían estar involucrados. La falta de regulación también puede exacerbar la carrera armamentística y aumentar las tensiones regionales.
4. **Impacto humanitario:** Los sistemas de armas autónomas plantean serias preocupaciones sobre el cumplimiento del derecho internacional humanitario y los principios de proporcionalidad y distinción en el conflicto armado. Sin la supervisión humana adecuada, existe el riesgo de que estos sistemas ataquen a civiles o infraestructuras civiles de manera indiscriminada, lo que podría resultar en un aumento de las víctimas civiles y daños colaterales.
5. **Dinámicas geopolíticas:** La proliferación de sistemas de armas autónomas podría exacerbar las tensiones geopolíticas y aumentar la posibilidad de conflictos en regiones donde los recursos son escasos o las disputas territoriales son comunes. Los países del sur global podrían enfrentarse a una mayor presión para adquirir y desarrollar tales tecnologías como medida de disuasión o para mantener el equilibrio de poder regional. La falta de transparencia en el desarrollo y despliegue de tales sistemas podría aumentar la desconfianza entre los países y aumentar el riesgo de escalada militar.
6. **Transparencia y Divulgación:** Debería existir un requisito de transparencia y divulgación por parte de los estados y las organizaciones que desarrollen o utilicen SADA. Esto podría implicar la divulgación de información sobre el desarrollo, despliegue y operación de estos sistemas, así como la realización de evaluaciones de impacto ético y legal como momento previo a la implementación de este tipo de sistemas.

7. **Cooperación Internacional:** La cooperación internacional es esencial para abordar eficazmente los desafíos asociados con los SADA. Los estados deben trabajar juntos en la elaboración y aplicación de regulaciones, así como en la promoción de normas y estándares internacionales para mitigar los riesgos asociados con estos sistemas.

Abordar estas preocupaciones requerirá una cooperación internacional sólida y un compromiso con la ética y los derechos humanos en el desarrollo y uso de tecnologías militares avanzadas.

Como organizaciones de la sociedad civil,

- Creemos en la dignidad inherente de todas las personas.
- Creemos en la igualdad y el derecho a controlar nuestras propias identidades libres de estructuras de discriminación por motivos de raza, género, capacidad, sexualidad, estatus socioeconómico y otras etiquetas.
- Respetamos la diversidad de individuos y comunidades, y creemos que esta diversidad no debe reducirse a etiquetas fijas y valores fijos adscritos, sino que debe informar y desafiar activamente nuestro espacio ético compartido.
- Enfatizamos que es importante generar reconocimiento de que somos individual y colectivamente responsables de desarrollar y dar forma a las tecnologías que enmarcan la interacción entre nosotros.
- Creemos en la búsqueda de soluciones no violentas a los problemas y en el uso de la cooperación, la negociación y el activismo para construir las respuestas sociales que se necesitan.
- Envisionamos un mundo en el que la tecnología se desarrolla y utiliza para promover la paz, la justicia, los derechos humanos, la igualdad y el respeto a la ley.
- Asumimos la responsabilidad de cómo nuestras elecciones con respecto a la tecnología cambian las relaciones entre nosotros, individual y colectivamente.
- Promovemos la ley como un proceso social, creado por personas, para personas, y que requiere el compromiso social para garantizar que funcione para ayudar a los vulnerables, no a los que ya son poderosos.
- Promovemos el empoderamiento de las personas como autores y controladores de la tecnología y nos resistimos a la mecanización de nuestras relaciones entre nosotros, como individuos y a través de estructuras políticas.

Por lo tanto,

- **ALENTAMOS** al Estado Paraguayo a negociar un instrumento internacional vinculante sobre sistemas de armas autónomas que rechace la automatización de

la matanza y garantice un control humano significativo sobre el uso de la fuerza.

- **INSTAMOS** a la activación de procesos internos dentro del Estado Paraguayo que justamente permitan un mayor grado de desarrollo y entendimiento institucional sobre los límites y desafíos de las tecnologías digitales para diversos fines, incluido aquellos relacionados a la seguridad interna y externa, y la necesidad base de una ley integral de protección de datos personales.
- **HACEMOS** un llamamiento a los Estados de todo el mundo para que alienten a los gobiernos a iniciar negociaciones sobre un tratado internacional sobre sistemas de armas autónomas.

El posicionamiento favorable de los países ayudará a salvaguardar contra los riesgos éticos, legales y humanitarios que plantea la autonomía en los sistemas de armas y promoverá un mundo más pacífico ahora y para las generaciones venideras.

Center for AI and Digital Policy

[25 May 2024]

We write on behalf of the Center for AI and Digital Policy (CAIDP), an independent global non-profit research organization based in Washington, DC, established to promote a better society – fairer, more just – a world where technology promotes broad social inclusion based on fundamental rights, democratic institutions, and the rule of law. CAIDP serves as AI policy expert to the Council of Europe Committee on AI, the OECD, the European Parliament and UNESCO.

In our input to the U.N. Interim Report, “Governing for Humanity,”²³ we urged U.N. members to commence negotiations on an international treaty to regulate autonomous weapons systems, prohibit LAWS, establish stringent safeguards against the integration of AI-enabled autonomous systems with weapons of mass destruction, and enshrine the principle of human responsibility for the use of lethal force.

Our overarching recommendations under [78/241](#) are as follows:

The U.N. General Assembly must pass a resolution to initiate the process of negotiating and adopting an international legally binding treaty to ban LAWS. This treaty would preserve human responsibility and accountability for any use of lethal force, and enshrine rigorous safeguards prohibiting the integration of AI-enabled autonomous systems with weapons of mass.

Human Rights:

Conflicts in Ukraine and Gaza show the extreme risks to human rights and security posed by the use of AI in warfare. Resolution [78/241](#) affirms that

²³ CAIDP Statement, UN Interim Report, UN Interim Report “Governing for Humanity” (30 Mar, 2024).

international law applies to autonomous weapons systems, particularly the U.N. Charter, international humanitarian law, and international human rights law.

In 2020 Roadmap for Digital Cooperation, the U.N. Secretary-General stated that “Digital technologies provide new means to advocate, defend and exercise human rights, but they can also be used to suppress, limit and violate human rights,” noting with emphasis lethal autonomous weapons and facial recognition.”²⁴

Concerns over killer robots also arose at the 75th U.N. Assembly.²⁵ Later, at the 2022 UN General Assembly, 70 countries endorsed a joint statement:

“We are committed to upholding and strengthening compliance with International Law and International Humanitarian Law, including through maintaining human responsibility and accountability in the use of force.”²⁶

Furthermore, UN Secretary-General and President of the ICRC jointly called on States to

“establish specific prohibitions and restrictions on autonomous weapon systems, to shield present and future generations from the consequences of their use. In the current security landscape, setting clear international red lines will benefit all States.”²⁷

Safety:

Stuart Russell warns that “loitering” AI missile systems can have both autonomous and remotely operated modes, making it difficult to know whether a human or the machine carried out any given attack.²⁸ Further, a single individual can

²⁴ UN Secretary General, Report – Roadmap for Digital Cooperation (Jun 2020), https://www.un.org/en/content/digital-cooperationroadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf); see also UN Secretary General, The Highest Aspiration - A Call to Action for Human Rights (2020) https://www.un.org/sg/sites/www.un.org.sg/files/atoms/files/The_Highest_Aspiration_A_Call_To_Action_For_Human_Right_English.pdf).

²⁵ Stop Killer Robots, 75th UN Assembly (Oct. 30, 2020), <https://www.stopkillerrobots.org/2020/10/un-diplomacy/>.

²⁶ United Nations (UN) General Assembly, First Committee, Joint Statement on Lethal Autonomous Weapons Systems First Committee, 77th United Nations General Assembly Thematic Debate – Conventional Weapons (Oct. 21, 2022), https://estatements.unmeetings.org/estatements/11.0010/20221021/A1jJ8bNfWGIL/KLw9WYcSnnAm_en.pdf.

²⁷ Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems (Oct 5, 2023) <https://www.icrc.org/en/document/joint-call-un-and-icrc-establish-prohibitions-and-restrictions-autonomous-weapons-systems>.

²⁸ Stuart Russell, “AI Weapons: Russia’s War in Ukraine Shows Why the World Must Enact a Ban,” *Nature* 614 (2023): 620-623. doi: 10.1038/d41586-023-00511-5. Loitering systems such as the Shahed, Kargu, Israel’s Harpy drone, Russia’s Lancet missile, and the Chinese Ziyang Blowfish A3.

launch an unlimited number of weapons. This type of weaponry should be classified as weapons of mass destruction.

Waging war through autonomous weapons will not protect military lives.

New AI-based weapons systems are probabilistic, introducing a degree of uncertainty. It remains unclear whether there is meaningful human control of targeting decisions.²⁹ Further, these AI systems can be hacked, increasing the likelihood of accidents and preventable escalation of conflict. Professor Lucy Suchman highlights that the “automation of data analysis under the sign of AI can only serve to exacerbate military operations that are at once discriminatory, in their reliance on profiling and other techniques of prejudicial classification.”³⁰

We urge adoption of fundamental obligations for AI systems set out in the Universal Guidelines for AI: **Obligations of Accountability** refers to ongoing need to assess the risks during design, development, and implementation. **Termination Obligation**³¹ presumes systems must remain within human control. If that is no longer possible, the system should be terminated. **Obligations of Public Safety** requires institutions assess public safety risks that arise from deploying AI systems and implement safety controls.

Ethics:

Algorithms are incapable of determining what is legal or ethical. The principle of human responsibility and accountability must be preserved. “Human life would be devalued if robots take life-or-death decisions, raising moral and justice concerns”³²

In 2023, the U.N. Secretary-General’s New Agenda for Peace called for the prohibition of LAWS. UN has a clear mandate to unite all stakeholders around the collective mitigation of long-term global risks. **Voluntary commitments are insufficient and provide no restraint.**

In “Killer Robots,”³³ Robert Sparrow warns that conditions for just wars cannot be met by autonomous weapons systems and that it “would be unethical to deploy such systems in warfare.”

“The same pressures that are pushing for the deployment of military robots in the first place also push for them to be given control over which targets to attack and

²⁹ CAIDP [Statement](#), On the Occasion of the 75th Anniversary of the Universal Declaration of Human Rights (Dec 10, 2023).

³⁰ Lucy Suchman, “Algorithmic warfare and the reinvention of accuracy,” *Critical Studies on Security*, 8:2, (2020): 175-187, DOI: 10.1080/21624887.2020.1760587.

³¹ Ibid.

³² Russell (2023).

³³ Robert Sparrow, “Killer Robots,” *Journal of Applied Philosophy*, Vol 24, No.1, (2007): 62–77.

when to open fire. Indeed, as AI technology improves, a human operator may prove not merely redundant but positively disadvantageous in such systems [...] time available to make survival critical decisions will often be less than the time required for a human being to make them.”³⁴

Legal:

The main components of a U.N. legally binding instrument should be proportionality, human oversight, human responsibility, risk assessment, and mitigation measures should be among the main components of the Treaty.

CAIDP’s annual AI & Democratic Values Index (AIDV) highlights strong support among democratic nations for limits on LAWS. AIDV Index notes “one of the first AI applications to focus the attention of global policymakers was the use of AI for warfare.”³⁵ More than 60 countries signed the “Political Declaration on Responsible Military Use of AI and Autonomy,” noting “military use of AI capabilities needs to be accountable, including through such use during military operations within a responsible human chain of command and control.”³⁶

Civil Affairs Institute

[20 May 2024]

Introduction: The advancement and deployment of autonomous weapon systems (AWS) in conflicts present critical legal, ethical, humanitarian, and security challenges. It is essential that the international community adopts a legally binding framework to regulate these systems. This document consolidates key perspectives and recommendations for the Secretary-General's report.

The Russian invasion of Ukraine in February 2022 has significantly influenced the landscape of modern warfare, with both sides utilizing drones extensively. This conflict has accelerated the development of potential autonomous weapons, highlighting the urgent need for international regulation to prevent further escalation and ensure compliance with humanitarian principles.

The context of the war in Ukraine is crucial for the way forward for autonomous weapons in Europe and the world. There is a need for global awareness and attention to the potential temptation to use artificial intelligence in military solutions, currently unregulated by international law.

³⁴ Sparrow (2007).

³⁵ CAIDP, AI and Democratic Values Index (2023), <https://www.caidp.org/reports/aidv-2023>.

³⁶ U.S. Department of State, “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy,” Bureau of Arms Control, Deterrence, and Stability (Nov 9, 2023), <https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>.

Legal and Ethical Concerns:

1. **Meaningful Human Control:** AWS should be designed to ensure that humans retain significant control over all decisions involving the use of force. This includes human operators making crucial decisions regarding targeting, engagement, and the context in which force is applied. Systems that do not allow for such control should be banned.
2. **Accountability:** AWS pose significant challenges in attributing responsibility for unlawful actions. Legal frameworks must ensure clear accountability mechanisms for violations of international law, preventing an accountability gap that undermines justice and human rights.
3. **Human Dignity:** Delegating life-and-death decisions to machines undermines human dignity and violates ethical norms. Machines lack the capacity for moral judgment and compassion, which are essential in making decisions about the use of lethal force.

Ethical Context and Asimov's Laws:

The deployment of AWS contradicts the ethical principles outlined by Isaac Asimov in his laws of robotics, particularly the first law, which states that a robot may not harm a human being or, through inaction, allow a human being to come to harm. Implementing AWS, which can potentially cause harm autonomously, starkly contrasts these ethical guidelines and poses severe moral dilemmas.

Humanitarian and Security Risks:

1. **Civilian Harm:** AWS pose a high risk of harm to civilians due to their inability to reliably distinguish between combatants and non-combatants. This increases the potential for unlawful killings and excessive force, violating principles of international humanitarian law (IHL).
2. **Proliferation:** The ease of replication and deployment of AWS raises concerns about their proliferation, particularly to non-state actors and regimes with poor human rights records. This could destabilize regions and escalate conflicts.
3. **Arms Race:** The development of AWS could spur an arms race, leading to increased global instability and lowering the threshold for entering conflicts. The rapid, autonomous decision-making capabilities of AWS could exacerbate crises, making conflicts more volatile and harder to control.

Prohibitions and Regulations:

1. **Two-Tier Approach:** A comprehensive treaty should combine prohibitions on certain AWS and regulations on others to ensure they operate under meaningful human control. Specifically:

- Prohibitions: Ban AWS that autonomously select and engage targets without meaningful human intervention, especially those targeting humans directly.
- Regulations: Ensure AWS that are not prohibited are strictly controlled through design, testing, and operational procedures that guarantee compliance with IHL and human rights standards.

Human Rights Implications:

1. Right to Life: AWS must comply with the right not to be arbitrarily deprived of life, ensuring that force is used only when necessary, proportionate, and as a last resort. The absence of human judgment in AWS challenges compliance with these principles.
2. Non-Discrimination: AWS must be free from algorithmic biases that could lead to discriminatory practices, particularly against marginalized groups. Strong measures are needed to eliminate biases in data and decision-making processes.

Conclusion: The adoption of a legally binding instrument on AWS is urgent to address the myriad challenges these systems pose. The international community must act decisively to regulate AWS, ensuring the protection of human rights and maintaining meaningful human control over the use of lethal force.

ECPAT Guatemala, Perú por el Desarme, the Feminist AI Research Network - Latin America Chapter, Anderson Henao and Jesús Martínez

[25 April 2024]

This is a summary of a document that emerged from a collaborative effort with experts who are members of ECPAT Guatemala (María Eugenia Villarreal); Perú por el Desarme (Gisela Luján); the Feminist AI Research Network - Latin America Chapter (Paola Ricaurte, Mexico/Ecuador, Mariana Díaz and Wanda Muñoz, Mexico); and Anderson Henao (Colombia), Jesús Martínez (El Salvador), experts in the rights of persons with disabilities and international humanitarian law.

The purpose of this summary is to encourage further reflection on both the national and international levels regarding the challenges inherent to autonomous weapons systems (AWS), underlining the urgent need to begin negotiations to establish a binding legal framework in a truly inclusive and representative forum. The full, original text in Spanish with references can be found [here](#).

Our report consists of two sections: 1) Challenges and concerns with respect to AWS from humanitarian, legal, and ethical perspectives; and 2) Ways to address these challenges.

1. Challenges and concerns with respect to autonomous weapons systems

1.1 Autonomous weapons will have a disproportionate impact on groups and populations with marginalised identities and characteristics, particularly women, Afro-descendant/racialised persons, indigenous/native peoples, children, and persons with disabilities, among others.

- There are several examples from the civilian sector that show that not only do emerging technologies pose risks, but that they have already caused damage and violated human rights.
- The use of these technologies in weaponry will likely cause disproportionate damage to the aforementioned populations.
- Understanding the difficulties and the differentiated negative impact of artificial intelligence systems is critical to analysing AWS, since these are the types of problems that could be replicated with the use of AI and emerging technologies in the military sector.
- Additionally, it is important to consider the risk of transferring autonomous technologies to forces of law and order, which could contribute to racial profiling in surveillance, and even to political repression.

1.2 Autonomous weapons will increase the barriers to accessing justice and compensation for victims of violations of human rights and International Humanitarian Law.

- The characteristics of AWS—including those related to the lack of predictability and explainability of emerging technologies and to applications of artificial intelligence, among others— will further hinder accountability, reparations, compensation, and more generally, access to justice, particularly for marginalized groups already face difficulties in this area, and are the most affected by AI bias.
- Remote war already has a disproportionate impact on certain groups. Not knowing when or where an attack will occur nor who might be a target is affecting different groups in different ways, and those effects are exacerbated in persons with a combination of marginalised identities and characteristics.

1.3 Autonomy in weapons systems is increasing and is already being used, a case in point being Israel in Gaza.

- Emerging technologies are already having a specific negative and differentiated impact in conflict zones. Similarly, autonomy in targeting and attack decisions is increasing.

- The most recent and flagrant case is unfolding in the context of the destruction of Gaza by Israel³⁷.
- During the first months of the campaign, the Israeli army used 25,000 tons of explosives on countless buildings, many of which were identified using artificial intelligence. What is more, the Israeli government's use of AI-driven technology has led to attacks against 11,000 targets in Gaza since the beginning of the most recent conflict on October 7, 2023.
- Two highly troubling examples of AI-driven technology are the Habsora (“Gospel”) and Lavender systems, which use AI and automation to identify and generate targets en masse.

2. Ways of addressing the challenges and concerns of autonomous weapons systems

2.1 International Humanitarian Law and International Human Rights Law apply to autonomous weapons systems (AWS), and a legally binding instrument (LBI) specifically addressing autonomy in weapons systems is needed.

- Currently, there is no legally binding international framework specific to such systems that ensures meaningful human control over the use of force. This is a serious legal vacuum for two reasons: a) It allows the development and use of weapons with autonomy in critical functions like targeting and engaging, and b) It hampers victims (affected persons, families, and communities) from seeking accountability, guarantees of non-repetition, and compensation for damages.
- From our perspective, the only credible way to address autonomy in weapons systems is through adopting a new LBI. The fundamental goal would be to regulate the autonomy of weapons systems in keeping with International Human Rights Law, International Humanitarian Law, and International Criminal Law.

2.2 Characteristics of the legally binding instrument needed to respond to the challenges of AWS.

- A LBI on AWS must include clear prohibitions and regulations, aim to maintain meaningful human control over force, and include effective implementation, monitoring, and accountability measures.
- This regulatory instrument must prohibit those weapons systems that: a) would delegate targeting and attack decisions to autonomous functions; b) would target human beings and civilian infrastructure; and c) would profile humans as targets.

³⁷ Report of the Special Rapporteur on the situation of human rights in the Palestinian territories occupied since 1967, Francesca Albanese*: Anatomy of a Genocide: UN Doc. [A/HRC/55/73](#) (Advance unedited version), 25 March 2024.

- Regulations must refer to autonomy in other functions.
- Said instrument must recognize the differentiated and disproportionate impact that these weapons would have on different population groups.

2.3 Characteristics of the forum where said instrument should be negotiated.

- United Nations General Assembly (UNGA) Resolution [A/C.1/78/L.56](#) is clear evidence of the majority opinion —146 States— on the “urgent need for the international community to address the challenges and concerns raised by autonomous weapons systems”.
- This majority voice could only make itself heard in a democratic and participatory space such as the UNGA.
- However, this has not been possible in the CCW —where the topic of AWS has been addressed for more than a decade— among other factors, because that forum allows the exercise of a veto under disguise of consensus.
- It is necessary to shift the deliberations on AWS to other forums, particularly the UNGA, whose rules facilitate more equal participation of a greater number of countries.
- Recent regional meetings (Costa Rica, Trinidad and Tobago, the Philippines, Sierra Leone) show that it is possible to make progress in inclusive forums, and that it is important to create spaces that truly allow, reflect, and value diverse perspectives.
- Furthermore, it is essential that all forums on AWS take specific measures to ensure the meaningful, free, and informed participation of civil society in all its diversity, particularly organisations representative of marginalised groups.
- It is necessary to advocate the inclusion of military and defence topics in UN work on artificial intelligence and other technologies, and in the framework of other Conventions on Human Rights and regional bodies.

Concluding thoughts

Delaying the start of negotiations for a legally binding instrument on autonomous weapons systems, ensuring human control over significant force use and prohibiting systems attacking humans, only benefits highly militarized countries or military industries continuing to develop, test, and deploy such technologies without regulation.

We consider it unacceptable that a minority of countries can obstruct the start of these negotiations, which already harm those affected by increased autonomy in these weapons, as seen in Gaza.

Future of Life Institute

[23 May 2024]

The Future of Life Institute (FLI) is a global nonprofit working primarily on the governance of emerging technology. The organisation is the UN Secretary-General civil society co-champion for Artificial Intelligence under the UN's Digital Roadmap and is best known for developing the Asilomar AI principles. FLI has long promoted governance of autonomous weapons systems (AWS) since the organisation's founding. FLI is recently most known for having published an [open letter](#), signed by over a thousand leading AI researchers, that sparked a global public debate on the development of advanced artificial intelligence.

Autonomous weapons systems are ethically wrong: machines and algorithms should not make life and death decisions. It is unclear who can be held accountable for potential war crimes, extrajudicial killings, and unlawful use. AWS present tremendous global security risks: they raise the risk of unintended escalation and flash wars and lower the threshold for war. They can proliferate in the wrong hands. They could be used as weapons of mass destruction (swarms), for targeted killings, and can create instability leading to an increased likelihood of nuclear weapons being used.

Over 115 states explicitly back new binding international law. FLI strongly supports the UN Secretary-General and International Committee of the Red Cross' (ICRC) urgent call for states to adopt a legal treaty to prohibit and regulate autonomous weapons systems by 2026. We call upon states to commence treaty negotiations as soon as possible.

FLI, in line with the ICRC's recommendations for a legally binding instrument, supports a two-tier approach, including:

- Prohibitions on unpredictable autonomous weapons systems and anti-personnel autonomous weapons systems.
- Positive obligations towards all other autonomous weapons to ensure the maintenance of meaningful human control, including measures such as:
 - Restricting targets of the AWS to only those which are military objectives by nature.
 - Limiting the location where, time that and situation in which the AWS is operating, including to avoid concentrations of civilians or civilian objects.
 - Limiting the number of engagements that the AWS can undertake.
 - Ensuring, to the maximum extent feasible, the ability for a human user to effectively supervise, and to, in a timely manner, intervene, and, where appropriate, deactivate operation of the AWS.

Geneva Centre for Security Policy

[21 May 2024]

1. Global conflicts are accelerating the development of LAWS, with potentially severe strategic consequences

The past 10 years have brought little by way of international regulatory frameworks on Lethal autonomous weapon systems (LAWS) but have brought us increasingly close to their full realisation. As of today, the technology for a weapon to track, select, and engage a target autonomously after the point of activation, with no further human intervention, already exists.³⁸ Whether or not current highly autonomous and AI enabled weapon capabilities neatly fit within an agreed upon definition of “LAWS”, or whether or not every stage of the weapon’s cycle was completely fully autonomously without human intervention should not detract from the reality that increasingly autonomous weapon systems (with and without AI) are already on today’s battlefields, and are raising many of the legal, ethical, and security concerns posed by LAWS.³⁹

Technological advancements in this space are furthermore being accelerated by a worsening global security environment and the ensuing technological competition it generates. Unfortunately, LAWS are not only technologically possible today, but the opportunities for their use are also multiplying. The battlefields of Ukraine and Gaza are for example both increasing the interest in, and fielding of, autonomous capabilities as well as raising many of the legal and ethical concerns linked to autonomy on the battlefield, posing questions over automation bias, human control and agency over the use of force.⁴⁰

The proliferation of such capabilities and their – mostly dual-use – enabling technologies has resulted in a fundamentally changed strategic environment compared to when international discussions began in 2014, worsening the consequences of inaction with respect to the regulation of LAWS. A wider array of state and non-state actors are now capable of developing and deploying increasingly autonomous capabilities, and increasingly willing to do so. Absent regulation, and a

³⁸ O’Neill, Paul, Cranny-Evans, Sam, Ashbridge, Sarah. “Assessing Autonomous Weapons as a Proliferation Risk: The Future has Not Been Written.” *Royal United Services Institute*. February 2024. <https://static.rusi.org/future-laws-occasional-paper-feb-2024.pdf>.

³⁹ Rickli, Jean-Marc, Mantellassi, Federico. “The War in Ukraine: Reality Check for Emerging Technologies and the Future of Warfare.” *Geneva Centre for Security Policy*. Geneva Paper No. 34. April 2024. <https://dam.gcsp.ch/files/doc/geneva-paper-34-24>.

⁴⁰ Renic, Niel, Schwarz, Elke. “Crimes of Dispassion: Autonomous Weapons and the Moral Challenge of Systematic Killing.” *Ethics and International Affairs*. Vol 37 (3). 2023, pp. 321-343. <https://doi.org/10.1017/S0892679423000291>; Davies, Harry, McKernan, Bethan, Sabbagh, Dan. “‘The Gospel’: How Israel uses AI to select bombing targets in Gaza.” *The Guardian*. December 1, 2023. <https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets>.

global governance framework, the current strategic environment will continue to incentivise and accelerate the development and deployment of such weapon systems, leaving an increasingly small space for regulation.

2. The question of human control

It has become a well-established norm that human control is a necessary component of both LAWS' ethical and safety dimensions, and of their legal compliance, especially with regards to IHL.⁴¹ However, lack of agreement persists behind what is meant by "human control", what it looks like in practice, and what amounts to sufficient levels of it. This lack of clarity is one of the biggest challenges to building effective regulatory frameworks and an important area to elucidate for the ethical, safe, and legally compliant development and deployment of autonomy on battlefields.

The expressed intent by states to ensure human control over LAWS and their effects, could however run counter to the desire to accelerate the tempo of military operations and to increasingly rely on complex algorithms to enable the various functions of LAWS.⁴² Indeed, research has repeatedly shown that humans have a tendency to offload cognitive and moral loads to highly autonomous machines, especially in times of high stress, cognitive workload and compressed timelines.⁴³ It is therefore not unreasonable to envision a near-future where humans can no longer realistically maintain the contextual understanding, cognitive and physical abilities necessary for meaningful control of weapon systems.⁴⁴ As states increasingly turn to complex AI systems to enable LAWS, the capacity for humans to be meaningfully engaged will increasingly be challenged.

It is primordial for states to define the processes, rules, as well as technical requirements, which would enable effective human control over LAWS. This should include a clarification as to what restrictions in design, capabilities, and operational parameters this would inevitably entail, ensuring states cannot resort to instances of so-called "nominal human control". *In fine*, this would allow for the maintaining of

⁴¹ Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 24 May 2023, GGE.1/2023/2, [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_2_Advance_version.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_2_Advance_version.pdf).

⁴² Scharre, Paul. "The Perilous Coming Age of AI Warfare." *Foreign Affairs*. February 29, 2024. https://www.foreignaffairs.com/ukraine/perilous-coming-age-ai-warfare?check_logged_in=1.

⁴³ Schwarz, Elke. "The (im)possibility of meaningful human control for lethal autonomous weapon systems." *Humanitarian Law and Policy*. August 29, 2018, <https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/>; Johnson, James. "Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command and Control Decision-making in the Digital Age", *Defence Studies*, Vol. 23(1), 2023, pp.43-67, <https://doi.org/10.1080/14702436.2022.2102486>.

⁴⁴ Schwarz, Elke. 2018.

moral and ethical agency over LAWS and ensure their safe, legally compliant, development and use.

3. The way forward

The most realistic way of delivering on the UNSG and ICRC president's call to achieve a legally binding document on LAWS by 2026 is through the Group of Governmental Experts (GGE) created by the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW). While consensus requirements have somewhat impeded the work of the group thus far, the group's convergence towards the so-called "two-tier approach" offers the most realistic chance of achieving a legally binding prohibition in the relatively near future.

Alongside the work of the GGE, states must continue to speak, exchange, and pronounce on the broader issue of the use of artificial intelligence in the military domain. Dialogue at the regional, plurilateral, and bilateral level about how humans can remain in control of weapon systems will be important confidence building measures and help in the developing of codes of conduct and guardrails. The UN should support these endeavours, for example through its regional disarmament centres, acting as bridge to the other discussions on AI governance, helping to bring across ideas relevant to the disarmament and international security community.

Resolution [78/241](#) creates the possibility of a new track on LAWS. If the GGE fails to reach consensus on a legally binding instrument, then it is likely that a large majority of states will support starting negotiations in the General Assembly. Creating a parallel process now, whilst the GGE is still working, is unlikely to achieve the desired outcome of a legally binding prohibition ratified by all the major states. For now, the UNSG could consider recommending an annual First Committee thematic session devoted to the military use of AI. He could also consider recommending an eighth cluster for First Committee, entitled 'The use of artificial intelligence in the military domain'. This would allow states to bring other resolutions on the topic, as the issue is broader than just LAWS.

Human Rights Watch

[6 May 2024]

Human Rights Watch appreciates the opportunity to submit its views and recommendations for consideration by the United Nations secretary-general in response to Resolution [78/241](#) on "Lethal autonomous weapons systems" adopted by the UN General Assembly on 22 December 2023.

I. Background

Human Rights Watch is an international nongovernmental organization (NGO) that conducts research and advocacy to uphold human dignity and promote human rights and international human rights law across the globe. In October 2012, Human Rights Watch co-founded the Campaign to Stop Killer Robots, which advocates for the negotiation and adoption of an international treaty to prohibit and restrict autonomous weapons systems.

II. Challenges and Concerns Raised by Autonomous Weapons Systems

This submission is based on and informed by our years of research and advocacy on this issue. Since 2012, Human Rights Watch has published more than two dozen reports on autonomous weapons systems, most in conjunction with the International Human Rights Clinic (IHRC) at Harvard Law School.

In November 2012, Human Rights Watch and IHRC released “Losing Humanity: The Case against Killer Robots,” the first major civil society report to examine the dangers of removing human control from the use of force. This report – and later ones – found that allowing machines to select and attack targets without further human intervention would be incompatible with fundamental provisions of international humanitarian law, including the principles of distinction and proportionality.⁴⁵

Human Rights Watch and IHRC have detailed the significant hurdles to assigning personal accountability to the actions undertaken by autonomous weapons systems under both criminal and civil law.

Autonomous weapons systems would also contravene basic principles of humanity and the dictates of public conscience established by the Martens Clause under international humanitarian law.

Autonomous weapons systems raise serious concerns under international human rights law because they are likely to be used in law enforcement operations as well as situations of armed conflict. They raise concerns under the foundational rights to life and to remedy.

Autonomous weapons systems would undermine the principle of dignity, a legal and moral concept, which implies that everyone has a worth deserving of respect. As inanimate objects, machines cannot comprehend or understand the value of human life or the significance of its loss. Allowing them to make life-and-death determinations thus strips people who are being targeted of their human dignity. In the process of determining whom to kill, autonomous weapons systems boil human targets down to data points.

⁴⁵ Please see our full submission for links to relevant reports by HRW and IHRC: <https://www.hrw.org/news/2024/05/06/submission-United-Nations-Secretary-General-Autonomous-Weapons-Systems>.

Security concerns include the risk of an arms race, the threat of autonomous weapons systems reaching the hands of states or non-state actors with no regard for international law, and a lowering of the threshold to war. Because autonomous weapons systems would have the power to make complex determinations in less structured environments, their speed could lead armed conflicts to spiral rapidly out of control.

III. Ways to Address the Challenges and Concerns Raised by Autonomous Weapons Systems

A Legally Binding Instrument

Human Rights Watch calls for the urgent negotiation and adoption of a legally binding instrument to prohibit and regulate autonomous weapons systems. Clear, strong, and global rules are essential. Only new international law will suffice to deal with the dangers raised by autonomous weapons systems.

A legally binding instrument should be accompanied by national legislation and other measures to implement and enforce the treaty's provisions at the domestic level.

Essential Treaty Elements

Human Rights Watch, IHRC, and others have outlined the essential elements for an international treaty on autonomous weapons systems, following precedent provided in previous disarmament treaties, international human rights instruments, and international humanitarian law, which all offer models for the proposed provisions.⁴⁶

A legally binding instrument should apply to all weapons systems that select and engage targets based on sensor processing, rather than human inputs. While the treaty's restrictions will focus on a narrower group of systems, this broad scope will help future-proof the treaty and ensure that no systems escape review.

The new treaty should include: 1) a general obligation to maintain meaningful human control over the use of force; 2) prohibit weapons systems that autonomously select and engage targets and by their nature pose fundamental moral and legal problems; and 3) include specific positive obligations that aim to ensure that meaningful human control is maintained in the use of all other systems that select and engage targets.

The concept of meaningful human control is fundamental to such an instrument because most of the concerns arising from autonomous weapons systems are attributable to the lack of such human control. The concept of meaningful human control should comprise a combination of components, such as, but not necessarily limited to: 1) Decision-making components, for example, the ability to understand

⁴⁶ HRW and IHRC, *New Weapons, Proven Precedent: Elements of and Models for a Treaty on Killer Robots*, October 20, 2020, <https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots>.

how the system works; 2) Technological components, including predictability and reliability, and 3) Operational components, notably restrictions on time and space in which the system operates.

A new treaty should prohibit the development, production, and use of systems that inherently lack meaningful human control over the use of force. It should also prohibit autonomous weapons systems that target people in order to prevent the use of weapons systems that strip people of their dignity, dehumanize the use of force, or lead to discrimination. It should cover weapons that always rely on data, like weight, heat, or sound, to select human targets. These prohibitions would help protect civilians and other non-combatants in armed conflict, and reduce infringements of human rights during law enforcement operations. They should apply “under any circumstances” to ensure that the provisions cover times of peace and war.

The Way Forward

In terms of negotiating fora, the Convention on Conventional Weapons (CCW) has run its course after providing a forum for useful discussions and the development of support for a legally binding instrument over the years. It is time to step outside of that forum to one that can aim higher, move faster, and be more inclusive of countries that are not party to the CCW as well as of international organizations and civil society. Disarmament precedent shows that stand-alone and UN General Assembly-initiated processes are viable options in which committed, like-minded states, in partnership with other stakeholders, can produce strong treaties in 15 months or less.

The world is approaching a tipping point on this topic as support for negotiating a legally binding instrument on autonomous weapons systems reaches unprecedented levels.⁴⁷ The Stop Killer Robots campaign’s Automated Decision Research project identifies more than 110 countries that have expressed their desire through national and group statements for a new international treaty on autonomous weapons systems.⁴⁸

Human Rights Watch affirms our strong commitment to work with urgency and with all interested stakeholders for an international legal instrument to ban and regulate autonomous weapons systems.

⁴⁷ More than 1,000 representatives from 144 countries and international organizations, industry, academia, and civil society attended the largest international conference ever held on autonomous weapons systems in Vienna on April 29-30. See the chair’s summary: https://www.bmeia.gv.at/fileadmin/user_upload/Zentrale/Aussenpolitik/Abruestung/AWS_2024/Chair_s_Summary.pdf.

⁴⁸ Automated Decision Research, <https://automatedresearch.org/state-positions/>.

International Committee for Robot Arms Control

[24 May 2024]

The International Committee for Robot Arms Control (ICRAC) values the opportunity to submit its views to the United Nations Secretary-General with respect to Resolution [78/241](#) on Lethal Autonomous Weapon Systems (AWS). Founded in 2009, ICRAC is an international committee of experts in robotics technology, artificial intelligence, robot ethics, international relations, international security, arms control, international humanitarian law, international human rights law, and philosophy of technology. We have published extensively on the ethical, legal, technical, and security challenges of AWS, on the question of meaningful human control, and on military escalation at speed. Based on our expertise, we are particularly concerned that military robotic systems will lead to more frequent, less restrained, and less accountable armed conflict. In light of these risks, we call for an international treaty to prohibit and restrict AWS.

We note the threat that AWS pose to compliance with international humanitarian law as well as international human rights, particularly the right to life and the human right to dignity. We thus support calls for a legally-binding instrument to prohibit and restrict the use of AWS, urging the Secretary-General to initiate a forum within the UN General Assembly that can include all States, cover autonomy and automation in the use of all weapons, and address international humanitarian law as well as human rights concerns.

Scope

In accordance with the International Committee of the Red Cross (ICRC), we understand an AWS as one that, potentially after initial activation or launch by a human, selects targets based on sensor data and engages targets without human intervention. We endorse the recommendations of the ICRC for a two-tiered approach that prohibits unpredictable systems and systems that explicitly target humans, while strictly regulating the use of autonomy in all other systems for the command, control, and engagement of lethal force. This includes restrictions on the time, space, scope, and scale of operations of such systems, as well as the types of targets and situations in which they may be used. In particular, we strongly agree that the only permissible targets of such systems should be military objects by nature, and never civilian or dual-use targets, which should always require human judgment.

More discussion is needed on the appropriate forms and regulation of the human-machine interaction in complex command-and-control systems. As computers and AI collect and automatically analyze more and more data, greater clarity is needed on what constitutes meaningful human control in the context of automated target generation/identification and how to ensure respect and responsibility for international law when such systems are used.

Key Challenges to Global Peace and Security

- Uncontrolled Escalation and Missed Opportunities for De-escalation and Diplomacy

The technical characteristics of AWS pose a considerable risk in enabling uncontrolled escalation at speed. Escalation from crisis to war, or escalating a conflict to a higher level of violence, could come about by erroneous indications of attack or a simple sensor or computer error. Mutual interaction between the control programs could not be tested in advance. The outcome of this interaction would be intrinsically unpredictable, and fast escalation is possible and likely. Unpredictable systems will give leaders false impressions of their capabilities, leading to overconfidence or encouraging preemptive attacks. Moreover, automated attacks, responses, and escalations will make it more difficult for leaders to interpret the intentions, decisions, and actions of their adversaries, and will limit their options for response. The overall effect will be to close off avenues and opportunities to avoid conflicts, to de-escalate conflicts, and to find means to end hostilities.

- Moral responsibility

No machine, computer, or algorithm is capable of recognizing a human as a human being, nor can it respect humans as inherent bearers of rights and dignity. A machine cannot even understand what it means to be in a state of war, much less what it means to have, or to end, a human life. Decisions to end human life must be made by humans in order to be morally justifiable. These are responsibilities of unavoidable moral weight that cannot be delegated to machines or satisfied by the mere inclusion of humans in the writing of computer programs. While accountability for the deployment of lethal force is a necessary condition for moral responsibility in war, accountability alone is not sufficient for moral responsibility. This also requires the recognition of the human, respect for the human right to life and dignity, and reflection upon the value of life and justifications for the use of violent force.

- Meaningful Human Control

Much hinges on the degree to which AWS can be *meaningfully* controlled by humans. Scientific scholarship on human psychology suggests that humans experience automation bias, a condition that cognitively hinders their ability to develop sufficient contextual understanding to intervene in fully autonomous systems that operate at speed. In order to safeguard *meaningful* human control (not merely functional control) over AI-enabled AWS, those involved in operating or deciding to deploy AWS should have full contextual and situational awareness of the target area at the time of a specific attack. They must also be able to perceive and react to changes or unanticipated situations that arise; ensure active and deliberate participation in the action; have sufficient training and understanding of the system and its likely actions; have adequate time for meaningful control and have the means and knowledge

required for a rapid suspension of an action. For many AWS this is not possible. Meaningful human control is fundamental to the laws of war and the ethics of war.

Moving Forward: A Treaty to Prohibit and Regulate the Use of AWS

We support calls from States, as well as the UN Secretary-General and the President of the ICRC, for an international legally-binding treaty prohibiting and regulating the use of AWS.

Such a treaty would obligate States to adhere to prohibitions and regulatory limitations for AWS. Codes of conduct and political declarations are not enough for systems that pose such grave risks to global peace and security. The treaty must apply to the automated control of all weapons, requiring meaningful human control in compliance with substantive regulations for the use of force in all cases. It should apply to all military uses of AWS and systems that generate or select targets, as well as to all police, border security, and other civilian applications that automate the use of force.

The treaty should prohibit AWS that are ethically or legally unacceptable, such as systems whose operation or effects cannot be sufficiently understood, predicted, or explained; AWS that cannot be used with meaningful human control; and AWS designed to target human beings.

The treaty should include positive obligations for States to use AWS that are permitted only within the bounds of clearly stipulated regulations that ensure adherence to international human rights and the key principles of international humanitarian law. We believe that an emerging norm around meaningful human control can be articulated and codified through a treaty negotiation in a process that includes all States, civil society, and industry and technical experts. We urge the Secretary-General to advance the creation of such a forum within the General Assembly and look forward to offering our expertise to those discussions.

InterAgency Institute

[25 May 2024]

Introduction

The InterAgency Institute is a digital think-tank focused on global security and local development with a human-centered approach. We have advocated for a treaty banning Autonomous Weapons Systems for over three years. Created by academics from the Global South, it has members with diverse backgrounds in 7 countries, mainly in Brazil, Portugal and Angola.

Appreciating the opportunity to submit this report addressing aspects related to the issues contained in operative paragraph 2 of [A/RES/78/241](#), the Institute hopes to stress the need to create new standards for the prohibition and regulation of

Autonomous Weapons Systems (AWS), as well as to point to the socio-technical dimension of the problem⁴⁹.

Below we list and develop the four main aspects related to AWS that we believe should be taken into account by the Secretary-General's Report.

1. Humanitarian Aspects

The existing obligations of International Humanitarian Law and the International Human Rights System should only serve as an *ad hoc* framework, since no treaty specifically regulates the matter. The new technological paradigm renders the current obligations of international law insufficient in regulating autonomous weapons because many aspects concerning autonomous weapons, such as automation replacing the use of force by human beings, were not part of reality when the main treaties of international humanitarian law were formulated. The possibility of automation in the critical functions of anti-personnel weapons facilitates war crimes, genocide and crimes against humanity, while also making it difficult to attribute responsibility for these crimes.

2. Security Aspects

The proliferation of autonomous weapons has unpredictable consequences for armed conflict. Despite this, externalities have already been identified, such as increasing the asymmetry of armed conflicts, reducing the necessary threshold for violence and the potential for non-state actors to possess these weapons and their uncontrolled proliferation.

The bigger the distance separating the operator and the machine, the more probable the use of violence⁵⁰. In autonomous weapons, those who develop and train the model cannot know what consequences, or actions, the model will follow after being launched. The possibility of autonomy in critical functions reduces the capacity of control by a human supervisor. This means that, even with the possibility of human intervention, the possibility of automating these functions entails security risks and the violation of international humanitarian law; not only because of the possibility of biases in the algorithm, but also because of the aforementioned distance between the person who creates the algorithm (and does not know what implications it will have in the real world) and the deliberate use of these algorithms to persecute people based on their phenotype, gender or clothing.

⁴⁹ The present text was adapted from a [Policy Brief](#) in Portuguese published by the InterAgency Institute in March 2024.

⁵⁰ Galliot, J. (2016). War 2.0: Drones, Distance and Death. *International Journal of Technoethics*, 7(2), 61-76. doi.org/10.4018/IJT.2016070104.

3. Technologic Aspects

Machine Learning (ML) uses input data to train a model to recognize patterns and thus make predictions a human would not be able to⁵¹. ML can be: supervised; semi-supervised; or unsupervised. Supervision concerns the labeling of the data entered during training. Another important division is that of dynamic (or online) and static (or offline) models: while dynamic models are trained frequently (or continuously), static models are trained before use⁵². The externalities of using ML to create target selection algorithms are not restricted to just one of the models, but cut across the different categories. The possibility of training these models with labels that correspond to people can deepen social biases, which means that the technological aspect and the ethical aspect cannot be dissociated. While it is important to highlight that online and unsupervised models exacerbate the social dimensions of the risks identified, they can also occur in offline and supervised models.

4. Ethical Perspective

Technology is not neutral. Technology sorts and ranks scenarios, based on the possibilities of action that it offers, shaping its context, regardless of how they are used⁵³. Digitalization reinforces the tendency to make cost-benefit (utility) analyzes to determine an act's morality, accentuating the tendency of optimizing procedures⁵⁴. In the context of military technologies, this utilitarianism can come with the trade-off of increasing death-tolls and violations of IHL principles. Despite algorithms seeming objective and neutral are built according to arbitrarily established criteria for the desired result, reinforcing biases⁵⁵.

Digital dehumanization is the act of reducing human beings to data, then used to make up decisions/actions that negatively affect their lives⁵⁶. This is a central part of the implications that such weapons have for international security.

Conclusion

A legally binding instrument is urgently needed to regulate autonomous weapons. Anti-personnel weapons capable of automating critical functions (aiming

⁵¹ Google Developers. (n.d.). Machine Learning Glossary. Available at: <https://developers.google.com/machine-learning/glossary>.

⁵² Ibid.

⁵³ Miller, B. (2021). Is technology value-neutral? *Science, Technology, & Human Values*, 46(1), 53-80. doi.org/10.1177/0162243919900965.

⁵⁴ MUL (2024). Week 2: The Ethics of (Digital) Technologies. QMUL Short Courses. Retrieved from <https://qmul.ac.uk/shorts/week-2-the-ethics-of-digital-technologies/>.

⁵⁵ O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

⁵⁶ Stop Killer Robots. (n.d.). Digital Dehumanisation. Available at: <https://www.stopkillerrobots.org/stop-killer-robots/digital-dehumanisation/>

and firing) must be banned. At the core of this debate are the activities - involving decision-making or not - we want to delegate to machines. Concluding discussions on the regulation of autonomous weapons is a step towards ensuring a future in which technology serves to advance global peace and security, while respecting human rights and human dignity. An international treaty banning weapons that operate without human control by 2026, as suggested by the United Nations and the International Committee of the Red Cross, is a crucial step in this direction.

International Physicians for the Prevention of Nuclear War

[24 May 2024]

We welcome the UN Secretary-General's initiative to address the challenges related to lethal autonomous weapons systems and the role of humans in the use of force. Because of our expertise on the catastrophic humanitarian impacts of nuclear weapons, IPPNW is particularly concerned with the danger of nuclear escalation arising from the use of artificial intelligence (AI) and machine autonomy in nuclear command and control.

IPPNW is profoundly alarmed by both the growing risk of the use of nuclear weapons, whether by accident, design or miscalculation, and the simultaneous rapid development and deployment of advanced AI systems in the defense sector. The integration of AI and automation into nuclear command and control would mark a new era of exponential risk to humankind.

We underscore the warnings of experts in the field that an AI arms race is already underway. According to Connor Leahy⁵⁷, "Whoever has the least humans in the loop, they can act the fastest...giving more and more to the machines, until one day we're just not in control anymore. It's a race to the bottom. There can be no winner – and the only loser is humanity."

We recall the incident in 1983 when one man, Soviet Lieutenant Colonel Stanislav Petrov, refused to respond to a false warning of an incoming US ICBM⁵⁸. Within a few fateful minutes, he intervened and, exercising his discretion, broke with "launch on attack" protocol. His reasoned judgment likely saved human civilization.

Almost 80 years into the nuclear era, we have survived not because of wise leaders, sound military doctrine, or infallible technology but because of luck. Recent events have laid bare how incredibly dangerous it is to gamble the future of the world on a hope for indefinite good luck.

⁵⁷ Co-founder and CEO of Conjecture, an AI startup working on controlling AI systems. BBC News. 2 May 2024.

⁵⁸ Stanislav Petrov, 'The Man Who Saved The World,' Dies At 77. NPR 18 September 2017.

An even more dangerous future may be fast approaching. If human agency is removed from nuclear command and control, then the fate of humanity will rest with self-taught machines that will make the snap decision whether to launch on warning.

Turning control of nuclear weapons over to autonomous AI is irresponsible madness. We therefore support efforts to regulate, through national legislation and binding multilateral agreements, the development, deployment and use of AI and autonomous weapons systems in nuclear weapons command and control. As we work toward the universalization of the Treaty on the Prohibition of Nuclear Weapons (TPNW), and the complete abolition of nuclear weapons, it is crucial that humans maintain control over the use of nuclear weapons. The international community must begin the processes of building a permanent firewall between AI and nuclear weapons.

**Latin America and the Caribbean Human Security Network,
Fundación para la Paz y la Democracia, Technology and
Community, Centro de Estudios Política Internacional,
Universidad de Buenos Aires**

[24 May 2024]

This document outlines the views of SEHLAC on the main risks and concerns of autonomous weapons systems in Latin America and the Caribbean, as well as the solutions to address them. SEHLAC is an organization that works for the transformation of regulations, public policies, and practices to implement and strengthen International Humanitarian Law (IHL), International Human Rights Law (IHRL), and Humanitarian Disarmament in the Latin American and Caribbean region.

The members of our network have participated in international discussions on autonomous weapons over the past ten years, seeking to understand the relationship between these emerging weapons and the human security issues affecting the region.

Risks and concerns about autonomous weapons

While there is no consensual definition of what an autonomous weapon is, SEHLAC aligns with the proposal of the International Committee of the Red Cross and defines it as a weapon system that can select and attack a target without significant human intervention.

"Significant" implies that a human must be the actual determiner of the critical functions of targeting and firing. Since a person is not the one making the critical decisions, these weapons are in opposition to the main element of IHL and IHRL: humanity. No element of these international frameworks can be applied to a weapon that makes the decision to take a human life on its own.

Why the concern about the region?

Latin America and the Caribbean is a diverse region that shares cultural, ethnic, and social elements, but also faces common challenges, such as inequality, poverty, corruption, and lack of state presence, as well as others inherited from a colonial past, such as systematic discrimination against historically marginalized groups.

This complexity has manifested in serious security problems, which have resulted in years of social conflicts and, in some countries, armed conflicts, as well as in the promotion of legal and illegal arms development.

The security problems affecting the region today would be directly related to the possible use of autonomous weapons.

Autonomous weapons would facilitate the perpetration of various types of violence in the region:

1. In Latin America, the upward trend of contract killings is worrying. *Sicariato* - paying to have someone killed - is gaining ground in the region and costs vary according to the complexity of the crime. The introduction of autonomous weapons would expand this practice, making contract killings more accessible and quicker. Also, achieving justice would be nearly impossible because these weapon systems' nature makes tracing the perpetrators extremely difficult.
2. Drug trafficking, as the axis of most of the violence that occurs in the region⁵⁹, could be favored by these weapons as these systems would facilitate the surveillance of drug production laboratories, as well as transit and protection of export routes.
3. Wars between gangs and/or armed groups outside the law are another phenomenon of violence in countries such as Haiti, Guatemala, El Salvador, Honduras, Mexico, Colombia, Ecuador, and Brazil, among others. The ease and cost-effectiveness of using autonomous weapons would cause this violence to spread to areas and countries where it is still precarious or non-existent, enabling criminal networks to expand their power and influence.
4. Using autonomous weapons as personal and private property protection, or even as a means of combating rural and urban crime, could lead to massacres and indiscriminate killings, as these weapons would use force against any person or object deemed a potential enemy.
5. Using autonomous weapons for state repression would facilitate systematic violation of human, civil, and social rights of anyone opposing the government,

⁵⁹ De México a Ecuador y de Costa Rica a Uruguay: la violencia del narcotráfico se extiende en América Latina [From Mexico to Ecuador and from Costa Rica to Uruguay: drug trafficking violence spreads across Latin America], José Denis Cruz, January 21st 2024, <https://www.newtral.es/violencia-narcotrafico-america-latina/20240121/>.

making it harder for victims to seek justice due to the technology's lack of traceable responsibility and accountability. This, in turn, would encourage impunity and further undermine confidence in law enforcement institutions.

6. It should be noted that the use of autonomous weapons could increase differentiated impact of violence, mainly on women, people with diverse identities, racialized individuals, and indigenous peoples, among others.

In addition to the dangers posed by autonomous weapons, it is also crucial to address the concern about indiscriminate use of biometric data, both private and public, without adequate treatment and, in some countries, without personal data protection⁶⁰.

The lack of explicit regulations that enable or restrict use of these technologies and the misuse of personal data create serious risks, as public security decisions are made without significant human oversight, compromising Human Rights. The situation is already observed and begins public discussion with the implementation of mass surveillance technologies in Latin America and the world, leading to litigation in several cases to protect civil liberties.^{61,62} Therefore, it is not only necessary to prohibit these technologies, but also to advance towards stricter regulations that protect privacy and personal data of individuals.

How to avoid a humanitarian catastrophe caused by the development and proliferation of autonomous weapons systems?

The only possible way to avoid the damage caused by autonomous weapons systems is to prohibit those systems that select and attack targets without significant human intervention.

At this point, it is essential to emphasize that autonomous weapons are not a specific category of weapons per se. Any type of weapon, from a small drone to a nuclear missile launcher, could potentially be equipped with autonomous capabilities, meaning that the control system can make crucial decisions without requiring human intervention.

It is important to make this point because any negotiation process that makes possible the existence of a new multilateral instrument with prohibitions and

⁶⁰ Sequera Buzarquis, M., Carrillo, E. y Gómez Berniga, L. (2022). Deudas, desafíos y conquistas en la intersección de los derechos humanos e internet [Debts, challenges, and achievements at the intersection of human rights and the internet]. En Codehupy (Ed.), Derechos Humanos en Paraguay 2022 (313-328). Asunción, Paraguay: Codehupy.

⁶¹ ALSur (2021). Reconocimiento facial en América Latina: tendencias en la implementación de una tecnología perversa [Facial recognition in Latin America: trends in the implementation of a perverse technology]. <https://www.alsur.lat/reporte/reconocimiento-facial-en-america-latina-tendencias-en-implementacion-una-tecnologia>.

⁶² Access Now (2021). Tecnología de vigilancia en América Latina: hecha en el exterior, utilizada en casa [Surveillance technology in Latin America: made abroad, used at home]. <https://www.accessnow.org/tecnologia-de-vigilancia-en-america-latina/>.

regulations must be carried out considering that it will apply to autonomous systems applicable to any type of weapon. Thus, these discussions must establish the level of human intervention in target identification and execution of the shot, ensuring that in no scenario is an algorithm the one deciding whom to shoot.

Final thoughts

It is important for States to honor the commitments made in the Belén Communiqué and other regional agreements and declarations that call for urgent negotiation of a legally binding instrument on autonomous weapons systems.

States must act decisively and with political will to address the humanitarian threat posed by autonomous weapons systems. The time for action is now!

Norwegian People's Aid and Mines Action Canada

[24 May 2024]

Mines Action Canada (MAC) and Norwegian People's Aid (NPA) appreciate the opportunity to input in response to Resolution [78/241](#). As members of the Stop Killer Robots (SKR) campaign, we align ourselves with the [submission by SKR](#), but would like to make a contribution in our own capacity.

Our contribution focuses on autonomous weapon systems that target people and makes an analogy to the work on prohibiting anti-personnel landmines. Both MAC and NPA have worked for decades on reducing the suffering caused by anti-personnel landmines and ensuring the implementation and universalization of the Anti-Personnel Mine Ban Convention (APMBC). We have seen first-hand the impact of anti-personnel weapons that are activated by the victim.

Autonomous weapon systems that target humans should be prohibited

As part of a legally binding instrument with prohibitions and regulations, anti-personnel autonomous systems should fall within the category of prohibitions. There are well-developed moral, legal, and ethical arguments for why such weapon systems would be unacceptable, made for example by SKR, the ICRC and others. Anti-personnel autonomous weapon systems represent extreme forms of digital dehumanization, challenging our common human dignity. Such systems are a threat to soldiers and civilians alike, questioning core IHL principles such as distinction, as well as risking direct and systematic humanitarian harm to civilians.

We argue that states have already developed laws and norms on weapon systems targeting people through the negotiation, adoption and implementation of the APMBC. States have, in policy and practice, recognized that victim-activated weapons with an anti-personnel target profile, such as anti-personnel landmines, are particularly problematic.

We would like to encourage states to be reminded about this when developing new law to control autonomy in weapon systems.

We note that many states have agreed to a prohibition specifically on anti-personnel landmines, in contrast to anti-vehicle mines, which are not subject to a ban treaty but are regulated. The difference in target profile made a significant difference to states negotiating the APMBC, which to us illustrates that specifically targeting humans should be a distinct concern when it comes to autonomous weapon systems as well.

We also note a certain reluctance among states to use weapons systems that could target and fire on people automatically in their autonomous modes. While such systems exist and can be applied, there seems to be an existing taboo on leaving life-and death decisions to machines, with states limiting, in practice, their current development and use of such systems. This reluctance may derive from legal considerations and the deliberate application of some basic principles of IHL, or ethical considerations – both of which also need to inform the negotiation of a legally binding instrument on autonomous weapon systems.

Like anti-personnel landmines, autonomous weapons systems cannot account for the potential shifts between combatant and civilian status. Autonomous weapons systems programmed to target a specific sensor profile will be unable to assess if the target is in the process of surrendering, ill, injured or otherwise hors de combat. The transient nature of combatant status is incompatible with weapons that target people based on static criteria such as weight for anti-personnel landmines or other sensor data for autonomous weapons systems.

Though it has been raised in discussions, we have seen no good evidence that it would be technologically feasible for current or future weapons systems to consistently and effectively integrate indications of changes in combatant status into their sensing and calculation processes, and plenty of expert opinion that it would not. In any case, distinguishing between civilians and combatants is a task for deliberative decision making by humans, who are responsible for applying the law: as states have long recognised, legal decisions cannot be left to a machine.

A human combatant will always have the possibility to refrain from using force, even though it would be legal under IHL. On the other hand, an autonomous weapon system would not make any deliberations beyond those programmed into it. A landmine is as crude as it is brutal: it does not see the difference between a combatant, a civilian or an animal. Machines, regardless of their future sophistication, will not share our understanding of what is human, and will only sense the world in a cold and inhumane fashion.

IHL compliance is not the only consideration when we call for a prohibition on autonomous weapons targeting humans. Rejecting the automation of killing in war is

important for profound ethical reasons and will have significance also into the civilian domain.

Therefore, autonomous weapon systems should not be allowed to target people. As the principle of banning anti-personnel landmines shows, more control over the use of force matters and preserves lives and dignity. As an international community, we should opt for more control, not less.

The way to safeguard human control over the use of force is by starting negotiations on a new legally binding instrument that prohibits autonomous weapons systems that are ethically or legally unacceptable, such as those that target personnel, and regulates the use of all other autonomous weapons systems through positive obligations to ensure meaningful human control.

PAX

[23 May 2024]

The urgent need for a treaty on autonomous weapons

Increasingly we are seeing weapons systems with autonomous functions being developed and used in ongoing conflicts. This is extremely concerning in the absence of clear rules and limits that specifically apply to these weapons. Especially as developers and users are pushing the limits of what is acceptable under legal and ethical norms. This demonstrates the urgent need to develop a legal instrument related to autonomous weapons. A treaty should prohibit fundamentally unacceptable autonomous weapons systems, and regulate the use of other autonomous weapons systems. Meaningful human control over the use of force should be a central element of a treaty. This means the human user must be able to make a legal and moral judgement regarding the effects of an attack, as well as being accountable for any violations of legal and moral norms. To ensure this the human user should be able to sufficiently predict the effects of an attack, in order to make a legal and ethical judgement. The human user should also be able to explain the actions of the weapons system and the human role after an engagement has taken place. This is essential to ensure accountability. Based on certain positive obligations can be developed. These could include that the user should be able to:

- Have a functional understanding of how the weapons system works, specifically what will trigger an application of force.
- Have sufficient understanding of the context where the weapons system will be deployed. For example, being aware of the presence of civilians and civilian structures.
- Limit the functioning of the weapon system. For example they should be able to limit the duration and geographical area of operation and the type of targets than can be engaged.

Following from this, autonomous weapons systems that cannot be used in line with these obligations should be prohibited. Furthermore autonomous weapons systems that target people must be prohibited. There is a risk of protected persons being targeted when autonomous weapons systems are used, as peoples' status under the law is fluid. Using target profiles that can trigger an application force, based on predetermined simplified representations of people is dehumanizing. Also algorithm bias could increase the risk of reinforcing existing structures of inequality.

Due to the serious legal, ethical, security and humanitarian concerns related to autonomous weapons, the negotiation of a legally binding instrument on autonomous weapons systems should start in 2024 and be finalized as soon as possible. This should happen in the forum that is most likely to be ambitious and successful in developing a treaty. Currently the General Assembly in New York seems to be the best forum to ensure this. At the same time, it is useful to continue dialogue in all appropriate multilateral fora as they can be mutually reinforcing.

We are at a crucial moment in the history of humanity. We need to do all we can to retain human control over the use of force and prevent digital dehumanisation, for ourselves as well as future generations. The time is now.

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand

[25 May 2024]

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand welcome the opportunity to contribute our views to the UN Secretary-General's report. Our submission briefly outlines our involvement in this issue, and has three sections summarising our position on: a) the urgent need for a new international instrument; b) key focuses of a new international instrument; and c) scope of a new international instrument. The points below are based on discussions with our member and supporting groups about the content of this submission.

Introduction

Peace Movement Aotearoa is the national networking peace organisation in Aotearoa New Zealand, established in 1981 and registered as an Incorporated Society in 1982. Our purpose is networking and providing information and resources on peace, humanitarian disarmament, human rights and social issues; and we have extensive national networks of member and supporting groups and individuals. We are a founding member of the Stop Killer Robots campaign and coordinate the national Stop Killer Robots Aotearoa New Zealand (SKRANZ) campaign.

SKRANZ was launched in April 2013 to support the global campaign, with a specific national focus on urging New Zealand to take national action to prohibit the development, production and use of autonomous weapon systems in New Zealand;

and to take international action to support negotiations on a new treaty to prohibit autonomy in weapon systems.

(a) A new international instrument on autonomy in weapon systems is urgently needed

It has been clear for some years now that rapidly developing technological advances in the use of force and increasing autonomy in weapon systems towards fully autonomous weapon systems pose an unprecedented threat both to humanity and to the foundations of international human rights and humanitarian law, which are based on respect for human life and dignity, protection of humanity in times of oppression and armed conflict, and human responsibility and accountability for harm.

The serious ethical, humanitarian, legal, and security concerns posed by these developments have been discussed for more than a decade within United Nations bodies - including the Human Rights Council, meetings related to the Convention on Certain Conventional Weapons and in the UN General Assembly - as well as in regional and national governmental and non-governmental forums.

Even as these discussions have taken place, some states have increasingly incorporated autonomy into military use of force with disastrous consequences for civilian populations, in ways that many would argue have already resulted in violations of international law. It is apparent that the absence of specific international law on autonomy in weapon systems, and with differing interpretation by some states as to how existing law applies to new technological developments, the risk of proliferation of ever more dangerous and uncontrollable weapon systems is increasing rapidly.

The need for urgency for new international law has recently been highlighted again by military use of AI, for example, Israel's use of AI-powered target suggestion systems in Gaza to make high explosive strikes on numerous targets possible in a short time frame, resulting in indiscriminate slaughter of civilians and systematic destruction of life-sustaining infrastructure. The reality of digital dehumanisation with catastrophic consequences is now very evident, as is the increasing tendency towards the development and use of autonomous weapon systems that will remove any remaining vestige of humanity from war.

A new international instrument on autonomy in weapon systems is necessary to clarify and strengthen existing law, and is already long overdue. The instrument must include both prohibitions and regulations, as outlined below.

Negotiations on a new instrument must begin without any further delay, in a multilateral forum where states can come together to work constructively free from the threat of veto, where the voices of those whose lives have already been impacted by increasing autonomy in weapon systems can be heard, and where UN agencies,

the International Committee of the Red Cross (ICRC), and NGOs are active participants.

(b) Key focuses of a new international instrument

While much of the work around autonomous weapon systems has focused on the issue of meaningful human control over the use of force, it is our view that the key underlying ethical imperative is preventing machines using sensors and digital code from targeting and / or attacking humans. A prohibition on autonomy in weapons systems that are designed or used to target human beings must be the starting point.

Meaningful human control over the use of force does have an ethical component of course, but it is also a practical and legal means to ensure accountability for any autonomy in weapon systems that breach the key dictates of humanitarian law.

(c) Scope of a new international instrument

It is our view that it is not necessary for a new international instrument to prohibit or regulate specific weapons, partly because it would rapidly become outdated and also because it is not specific weapons that are the problem. Instead, it should include overarching rules to establish a framework for evaluating current and future technological developments, while promoting increased compliance with international human rights and humanitarian law.

Such overarching rules would prohibit autonomous weapon systems that are designed or used to target humans, and lay out specific obligations to ensure meaningful human control over other systems: for example, that the human operator/s understand the capabilities and limitations of the system, are able to fully evaluate the context in which the system will be used, and are making mindful firing decisions rather than assuming the technology is accurate - this would act to regulate autonomy in weapon systems. It would be useful to specify that decisions made by states on their assessment of new or altered weapon systems that incorporate autonomous features or functions must be transparent.

We referred above to a recent example of military use of AI, which suggests that the scope of a new international instrument must be wider than only prohibiting fully autonomous weapon systems, that is, weapon systems that make target selection and attack decisions based on the processing of data from sensor inputs autonomously without any human involvement beyond the initial activation. It is increasingly apparent that there is a spectrum of harmful military use of autonomy, ranging from target decision support systems (as some have described systems such as Lavender), data-based targeting systems, generation of target lists by algorithm or AI, sensor-based targeting systems, through to weapon systems that combine these elements and incorporate varying degrees of machine learning to make target selection decisions and attack autonomously.

We note last year's Joint Call by the UN Secretary-General and ICRC President stated "*The autonomous targeting of humans by machines is a moral line that we must not cross*"⁶³, yet that appears to have already happened.

It is therefore our view that a new instrument must cover systems that automate significant decision making in weapons systems, such as target generation, force deployment, and engagement, as well as autonomous weapon systems.

Finally, although we have referred in this submission to military use of autonomy in weapon systems, prohibitions and regulations in a new international instrument must also apply to all coercive agencies of the state, including those used for policing and internal security, for border control, in corrections facilities and in places of detention.

Peace Union of Finland, Finnish Peace Committee, Committee of 100, Technology for Life, Women for Peace, and the Union Of Conscientious Objectors

[24 May 2024]

Finnish civil society organizations Peace Union of Finland, Finnish Peace Committee, Committee of 100, Technology for Life, Women for Peace, and the Union Of Conscientious Objectors appreciate the opportunity to submit our views to the United Nations Secretary-General in response to Resolution 78/241, entitled "Lethal Autonomous Weapons Systems". As members of the international Stop Killer Robots⁶⁴ network, we are committed to advocating for a new international law on autonomy in weapons systems.

This submission provides a summary of our views on this issue, highlights some specific humanitarian, ethical, and security challenges related to autonomous weapons systems, and outlines ways to address these concerns through a legally binding instrument.

⁶³ Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 5 October 2023

⁶⁴ Stop Killer Robots is a coalition of more than 250 non-governmental organizations and academic partners working across 70 countries towards an international treaty on autonomous weapons systems that ensures meaningful human control over the use of force and rejects the automation of killing.

Ethical concerns

Increasing autonomy in weapons systems means that machines are beginning to replace decision making of humans in the use of force. Machines are incapable of making complex ethical evaluations or understanding the value of human life. A machine does not understand that it is killing, and therefore life and death decisions should never be delegated to machines.

As artificial intelligence develops, systems become increasingly difficult to understand or to predict. Humans can't influence the rapid decisions executed by a machine. Moreover, who would be responsible for a possible war crime committed by an autonomous weapon system? Holding a human responsible for the actions of a weapon system they can't adequately control would be problematic, leaving victims and their loved ones without satisfactory explanation or justice in cases of grave injustice. Concerns posed by a lack of responsibility or accountability for violations of law must not be accepted or ignored.

Security concerns

The development and proliferation of autonomous weapons systems pose significant threats to global security and stability. The pursuit of keeping up with advancing military technology will likely result in or trigger a destabilizing arms race. The accelerating arms race increases the likelihood of conflicts and endangers international security. In the worst case scenario, an escalation of the autonomous arms race might lead them to be equipped with nuclear warheads, when the decision to annihilate the world would be left to a machine. This risk of escalation must not be ignored.

The deployment of autonomous weapons systems is likely to lower the threshold to go to war. When risks to human soldiers are perceived to be lower, states may become less inclined to seek political solutions. This would make armed conflicts more widespread and increase civilian casualties, suffering and displacement. It also escalates the problems associated with remote warfare.

The rapid speed and subsequent errors from autonomous decision-making, as well as the weapon's malfunction could increase the risk of conflict escalation, or lead to other unpredictable consequences. Artificial intelligence, for instance, can generate incorrect and potentially harmful information or detect things that are not within their range, leading to unexpected outcomes. In life-or-death situations, the necessary system testing to correct these errors would equate to dangerous human experimentation. AWS are vulnerable to hacking and can be manipulated to turn on friendly forces or behave in other undesirable ways. Risk of misuse and hacking would also pose an existential threat to every human being.

A legally binding treaty is crucial for ensuring a future for humanity

There is an urgent need for the regulation of autonomous weapons systems due to the accelerated advancements in artificial intelligence and increasing autonomy within these systems. Their ability to operate without human control and responsibility raises significant concerns that demand urgent action from the international community. To address these concerns, it is imperative to establish a legally binding international treaty that regulates the development, deployment and use of autonomous weapons systems.

There is growing international support for new rules regulating autonomous weapons systems.⁶⁵ We align ourselves with the view advocating for the implementation of the treaty and want to address the threat posed by the development of autonomous weapons systems. A common approach, often referred to as the two-tier approach⁶⁶, suggests that a treaty should prohibit autonomous weapons systems that are deemed ethically and legally unacceptable. All autonomous weapons systems should be subject to regulation through positive obligations to ensure meaningful human control over the use of force.

We need to act now

Time is running out; artificial intelligence is advancing at an increasing pace, and its military applications are spreading across all branches of weaponry. A treaty banning and regulating autonomous weapons systems would draw a red line for governments and companies as to what is acceptable and what is not.

We need a treaty even if not all states decide to join it. Global instability and the ambitions of a handful of militarized states already investing in the development of autonomous technologies shouldn't dictate our future, or decrease states' willingness to mitigate the serious risks posed by autonomous weapons systems.

The prohibition of autonomous weapon systems should not be left outside the considerations within the broader regulation of artificial intelligence. We must ensure that militaristic objectives or arms industry interests do not dictate AI development, but rather that it happens democratically and with respect for human rights. It is imperative that we draw the line now, before automated decision-making leads to automated killing.

⁶⁵ In addition to numerous scientists, tech workers, religious leaders, and members of civil society, the majority of states support the negotiation of a legally binding instrument. See: <https://automatedresearch.org/state-positions/>

⁶⁶ There is wide agreement that a normative framework on AWS should combine prohibitions and regulations. We align ourselves with a view that AWS that would target people, and cannot be used with meaningful human control are deemed ethically and legally unacceptable, and therefore should be prohibited. Positive obligations should ensure that all other AWS are used in line with legal and ethical norms.

PEAC Institute

[8 February 2024]

In response to the request from Ms. Nakamitsu for comment regarding Resolution [78/241](#) “Lethal autonomous weapons systems”, I offer this opinion based on working from 2002 to 2013 for the city of Hiroshima and Mayors for Peace, the campaigning NGO of Hiroshima and Nagasaki. For six years I was chairman of the Hiroshima Peace Culture Foundation, Hiroshima’s organization managing peace programs and international relations.

What the UN decides regarding autonomous weapons systems is irrelevant. As with nuclear weapons, the problem is not the weapons. The problem is the mindset, the attitudes and thinking behind the weapons. The UN was created to prevent war. It has been unable to do that for all the reasons you know much better than I. However, we are entering an extremely dangerous, possibly lethal time in human history.

White people in the Global North have been dominating the Earth and its resources for five hundred years. Now, the US and Europe are rapidly weakening while China and Asia, including Russia, are growing stronger. Whether the society comprises chickens, chimpanzees, or humans, shifts in dominance hierarchies usually involve violence.

Right now, we are witnessing many of the phenomena that preceded World Wars One and Two. The tensions are greatly intensified by a wide array of completely catastrophic weapon systems, by dwindling resources, by escalating independence movements, and by existential threats to Earth’s ability to sustain us.

In this context, the violent competition among the US, Europe, China, Russia, Asia, etc. is the problem. Who wins is irrelevant. As long as the competition remains violent and potentially terminal, humanity will remain utterly unable to respond to any of the global threats we face. The only competition that can possibly affect the outcome, that is, enhance the chances of Homo sapiens surviving into the next century, is the competition between war and peace.

The UN’s mandate is to resolve all conflict through dialogue, discussion, negotiation, treaties, and the pursuit of solutions that are based on truth and that benefit and satisfy all parties. From this point of view, the UN should not trouble itself with the details of weapons or weapon systems. The UN should focus intently and with determination on the propensity of certain UN members to ignore the fundamental meaning, purpose, and function of the UN. The UN needs to stand up for the peaceful resolution of conflict. The UN also needs to stand up for international law, expressing intense criticism and using every tool at its disposal to fight rogue nations and other lawless elements, including members of the Security Council.

Humanity desperately needs the UN to prevent war. That means we need the UN to stand up for international law and for itself as the proper, international forum for

the resolution of ALL international conflict. War and violence of any sort for any reason are beyond the pale, and any people, nation, or organization that resorts to violence is a rogue committing a crime against humanity. Until the UN accepts its duty to confront and forcefully oppose inter- or intra-state violence, the UN and international law itself amount to lipstick on a pig.

The elimination of war is the goal, but nuclear weapons are the place to start. No one wants a nuclear war. The case against nuclear weapons is extremely easy to make, and nuclear weapons are the easiest global problem we face. Nine countries could solve this problem in less than a month if they put their minds to it. A campaign against nuclear weapons would be relatively easy, and when humanity decides to set aside its doomsday machines, it will be taking a giant step toward cooperating for survival. Once humanity decides to cooperate for survival, peace culture is just around the corner. The keys are: 1) identify the selfish, violent enemies of humanity who refuse to abide by Article VI of the NPT and 2) spend millions of dollars informing the global public exactly who is stealing what from them, what the culture of perpetual war is doing to them, what the constant threat of annihilation is doing to them, and what nuclear weapons will actually do to them if anyone is stupid enough to start a nuclear war.

Unfortunately, this process will require the UN to identify and defeat its most powerful enemies, and doing that will be the fight of the century. As evidenced by the recent withdrawal of funding from UNRWA, a genuine effort to eliminate nuclear weapons will put the very existence of the UN at risk. This battle can only be won by communicating loudly and effectively through the media to the global public, and that will require millions of dollars spent year after year on telling the people the truth with the expectation that they will choose universal health and wellbeing over nuclear war and total environmental collapse. The battle will be dangerous and difficult. The UN might go bankrupt and fail for lack of funding, but if humanity fails to abandon violence as a legitimate method of conflict resolution, our species will soon be extinct.

Project Ploughshares

[24 May 2024]

Project Ploughshares, a Canadian peace research institute, has focused advocacy and research efforts on the issue of autonomous weapons for the past nine years. While there have been rapid technological advancements in artificial intelligence (AI) and robotics and these technologies are now being used in conflict zones, international governance frameworks have not kept pace. At the same time, growing geopolitical competition is increasing the likelihood of deployment of technologies that are not ready for complex and dynamic settings; such use could have far-reaching implications for civilians. What is needed now to address the growing use of emerging technologies in warfare is a legally binding instrument on autonomous weapons.

Such an instrument must be future-proofed to respond to evolving—even revolutionary/transformational—uses of technology. In our view, it is crucial that this governance mechanism operates on two tiers: prohibiting certain uses and providing restrictions based on risk categorizations.

This submission does not indicate all the factors that states should take into account in developing this recommended instrument. Instead, we would like to raise several concise key points that address the crucial need to ensure human control over targeting of humans, and protect civilian infrastructure and the environment needed to sustain human life.

- A clear prohibition should be placed on antipersonnel weapons and on autonomy in weapons of mass destruction (WMD), such as nuclear weapons. While WMDs are already prohibited by existing agreements, addressing technological advancements enabling autonomy in their functioning requires further bolstering.
- States should develop categories of risk and concurrent levels of restriction. For example, a high-risk category would be one that impacts the health and safety of civilians. Included in this category would be systems that are less lethal but still capable of immobilizing or causing reverberating effects on civilians and the environment. Moderate-risk systems would be regulated to ensure the transparency of their functioning. Low-risk systems could involve voluntary best practices.
- The proposed instrument should be based on meaningful/necessary human control over weapons systems. Demonstrable human control in time and space is required for all high-risk systems; human control over moderate- and low-risk systems must be clear. Systems that cannot meet the human control requirements of the appropriate risk category should be prohibited. Decision-support systems or target-generation systems can be suitably addressed by ensuring that targeting decisions are subjected to an evaluation of appropriate accountability.
- The instrument should provide the foundation that will support more regulation at various levels, and therefore should focus on crucial aspects necessarily addressed at the global level. A toolkit of governance mechanisms can and will follow such instruments. States already have other forums at which to exchange best practices and voluntary measures. The aim of this new instrument is to provide the necessary legal commitments on human accountability and human-decision making.
- A forum that allows all states parties to participate and includes civil society is critical to address the risks posed by these weapon systems.

Without a global regulatory framework and specific prohibitions on certain autonomous systems, it seems inevitable that ever more autonomous systems will soon be developed and employed by many states and become readily available to nonstate armed groups, posing an unacceptable risk to global stability. Countries need to begin serious negotiations immediately to avoid these consequences.

Protection

[24 May 2024]

تشرف منظمة الحماية من الأسلحة وآثارها بتقديم رأيها كمنظمة مجتمع مدني حول أنظمة الأسلحة ذاتية التشغيل إلى الأمين العام للأمم المتحدة وفقا لقرار الجمعية العامة للأمم المتحدة رقم 241/78.

منظمة الحماية من الأسلحة وآثارها هي منظمة فرنسية تأسست عام 2003 وتعمل في مجال نزع السلاح الإنساني حيث تركز مجال عملها جغرافيا على الشرق الأوسط وشمال أفريقيا، وهي عضو لجنة القيادة بحملة منع الروبوتات القتالة التي تستمد منها وتشاركها الرؤية في هذه المساهمة. ومن ضمن أهدافها حظر وتنظيم عمل الأسلحة ذاتية التشغيل وضمان وجود السيطرة البشرية عليها عبر اعتماد صك دولي ملزم حولها.

الحاجة الملحة إلى قواعد دولية ملزمة جديدة

بعد مرور أكثر من عقد على التفات الآليات الأممية الدولية لخطر الأسلحة ذاتية التشغيل وبدء النقاش حولها وتمدده، ثبت أن التوصل إلى صك ملزم قانونا هو السبيل الفعال الوحيد لعلاج التحديات والمخاوف الإنسانية والقانونية والأمنية والأخلاقية الناجمة عن أنظمة الأسلحة ذاتية التشغيل، حيث إن الاعتماد على القوانين الحالية ليس كافيا في الوقت الراهن ولن يكون كافيا بتاتا في المستقبل، وهذا ما أثبتته نقاشات الدول والخبراء غير الحكوميين في مجلس حقوق الإنسان واتفاقية حظر الأسلحة التقليدية، ما جعل الحاجة ملحة إلى قواعد محددة وقانون واضحة يضمن بوضوح حماية الإنسانية.

المجموعات المسلحة من غير الدول

كما أثبتت النزاعات المسلحة الحديثة خصوصا في الشرق الأوسط وشمال أفريقيا سهولة حصول المجموعات المسلحة من غير الدول على تكنولوجيا الأسلحة بدرجات متفاوتة تكنولوجيا (اليمن نموذجا) ما ساهم في تهديد الملاحة الدولية ومعها السلم والأمن الدوليين. وحصول هذه الجماعات على أسلحة ذاتية التشغيل أيا ما كانت درجتها يعني تهديد شديد بكوارج لا يمكن علاجها، بينما عدم وجود هكذا سلاح من الأساس يجنب العالم ويلات هو في غنى عنها.

الغالبية تدعم التحرك قدما

يزداد الدعم وراء الدعوة إلى إبرام صك ملزم يوميا وفي كل محفل ذات صلة، فالآن هناك أكثر من 115 دولة تؤيد ذلك منهم 19 دولة من الشرق الأوسط وشمال أفريقيا وأعضاء في جامعة الدول العربية

وعقدت عدة اجتماعات إقليمية حول الأسلحة ذاتية التشغيل في أمريكا الجنوبية وجنوب شرق آسيا والمحيط الهادئ وأفريقيا وأوروبا، كما أصدرت العديد من الدول بيانات إقليمية تلتزم فيها بالعمل من أجل التوصل إلى ذلك الصك.

ودعت اللجنة الدولية للصليب الأحمر والأمين العام للأمم المتحدة إلى التفاوض على صك ملزم بحلول عام 2026، وهي الدعوة التي نؤيدها بشدة.

ويدعم ذلك أيضا آلاف العلماء والعاملين والقادة في مجال تكنولوجيا الروبوتات والذكاء الاصطناعي والزملاء الدينيين والبرلمانيون، ومجموعة واسعة من منظمات المجتمع المدني، والرأي العام في جميع أنحاء العالم.

البداية في المفاوضات

يجب على جميع الدول وأصحاب المصلحة الملتمزمين إنسانيا وأخلاقيا بالتوصل الى صك ملزم أن يعملوا الأمل قبل غدا معا بروح الشراكة الحقيقية العابرة إقليمية وأن يتخذوا خطوات ملموسة نحو بدء المفاوضات.

ولا بد من التفاوض على الصك في منتدى يضم كافة الدول والمجتمع المدني، حيث لا يمكن للقلة غير الرغبة في ذلك الصك عرقلة التقدم من خلال حق النقض، أو قواعد الإجماع اللذين يساء استخدامهما كما في اتفاقية الأسلحة التقليدية التي لا يمكن إحراز تقدم فيها.

وكما يخبرنا التاريخ والواقع العملي للاتفاقيات ذات الصلة بنزع السلاح الإنساني التي تم تبنيها، فإن إنشاء ولاية تفاوضية بشأن صك ملزم من خلال الجمعية العامة للأمم المتحدة هو أفضل فرصة للمضي قدما وتحقيق هكذا صك.

وخلال العقد الماضي ونتيجة المناقشات المتعددة في عدة مندوبات ومنها اتفاقية الأسلحة التقليدية، فإن هناك فهم واسع النطاق بين الدول حول المقصود بأنظمة الأسلحة ذاتية التشغيل، كما والحاجة إلى نهج "ثنائي" للحظر والالتزامات الإيجابية لتنظيمها بشكل فعال، والذي يمكن البناء عليه للاتفاق على صك ملزم قانونا.

معاهدة فعالة بشأن أنظمة الأسلحة ذاتية التشغيل

مجال الاتفاقية

إن فهمنا لأنظمة الأسلحة ذاتية التشغيل كأعضاء في حملة منع الروبوتات القاتلة هي أنها تلك التي ترصد وتطبق القوة على الهدف بناء على معالجة البيانات من مدخلات أجهزة الاستشعار. ففي هذه الأنظمة، بعد التنشيط من قبل مستخدم بشري، هناك فترة من الوقت يمكن فيها للسلاح تطبيق القوة على الهدف دون موافقة بشرية مباشرة، ويتم تحديد الهدف المحدد الذي سيتم مهاجمته، والوقت والمكان المحددين للهجوم، من خلال معالجة أجهزة الاستشعار، وليس البشر.

لذا يجب أن يشمل صك أنظمة الأسلحة ذاتية التشغيل جميع الأنظمة التي تقع ضمن هذا النطاق، حيث تتبع مخاوفنا من الأتمتة التي تعمل بها هذه الأنظمة. كما تتمثل مخاوفنا الرئيسية في اندثار السيطرة البشرية على استخدام القوة - الأمر الذي يثير مخاوف قانونية وأخلاقية وإنسانية وأمنية خطيرة - فضلا عن عدم القبول أخلاقيا بتزايد القتل الآلي والتجريد من الإنسانية رقميا.

مركزية الدور البشري

يجب أن يكون الدور البشري هو العنصر الأساسي في الإطار التنظيمي، حيث يجب أن يكون ضمان السيطرة البشرية ذات المغزى أساسا للمعاهدة.

بناء من الحظر والتنظيم

حيث ينبغي أن يحظر الصك أنظمة الأسلحة ذاتية التشغيل غير المقبولة أخلاقيا أو قانونيا وعلى رأسها تلك التي تفقد السيطرة البشرية. كما ينبغي تنظيم استخدام جميع أنظمة الأسلحة ذاتية التشغيل الأخرى من خلال التزامات إيجابية لضمان بقائها تحت سيطرة بشرية حقيقية.

يجب أن تخضع أنظمة الأسلحة ذاتية التشغيل غير المحظورة للالتزامات إيجابية بشأن التصميم والتطوير والاستخدام لضمان السيطرة البشرية ذات المغزى

تتطلب السيطرة البشرية ذات المغزى أن يتمكن الشخص من إصدار حكم أخلاقي وقانوني على آثار الهجوم، والوفاء بالتزامه بتطبيق القانون، ويعني كذلك أن هناك مستخدماً بشرياً مسؤولاً أخلاقياً وقانونياً عن آثار الهجوم.

ولذلك، ينبغي للالتزامات الإيجابية للمعاهدة أن تضمن قدرة المستخدم البشري على توقع آثار الهجوم بشكل كاف، وأنه بعد وقوع الهجوم، ينبغي له أن يفهم آثاره ويكون قادراً على تفسيرها.

ولتحقيق مثل هذه السيطرة ذات المغزى يجب:

أن يكون المستخدم قادراً على الحصول على فهم وظيفي كافٍ لكيفية عمل النظام.

أن يكون المستخدم قادراً على الفهم الكافي للسياق الذي سيتم فيه نشر نظام الأسلحة، وكيف سيعمل في تلك البيئة.

أن يكون المستخدم قادراً على الحد من عمل نظام الأسلحة في الزمان والمكان.

يجب حظر أنظمة الأسلحة ذاتية التشغيل التي لا يمكن التحكم البشري فيها بشكل ذات مغزى

يجب حظر الأنظمة التي لا يمكن استخدامها بحكم تكوينها التقني أو بيئة استخدامها أو أن المستخدم غير قادر على تلبية المتطلبات المذكورة أعلاه بما يتماشى مع المعايير القانونية والأخلاقية، ومن شأنها أن تثير تساؤلات حول الجهة المسؤولة قانونياً وأخلاقياً عن أي انتهاكات للقانون.

يجب حظر أنظمة الأسلحة ذاتية التشغيل التي تستهدف البشر

وحتى لو تم استخدامها مع سيطرة بشرية ذات معنى، فإننا نعتبر مثل هذه الأنظمة غير مقبولة لأنها تحول الناس إلى مجرد أشياء، وهو ما يمثل إهانة شديدة للكرامة الإنسانية.

فأنظمة الأسلحة ذاتية التشغيل التي تستهدف البشر تثير المخاوف الأخلاقية الأساسية، وتمثل نموذج صارخ على التجريد الرقمي من الإنسانية للضحايا المدنيين والعسكريين على حد سواء.

علينا العمل الأمس وليس الآن وقبل غدا

بالإضافة للكوارث الإنسانية والقانونية والأخلاقية لأنظمة الأسلحة ذاتية التشغيل، فإن استخدامها خارج نطاق النزاع المسلح، مثل مراقبة الحدود وحفظ الأمن، من شأنه أن يزيد انتهاكات القانون الدولي لحقوق الإنسان والحريات.

إن التطورات في تكنولوجيا الأسلحة والأحداث التي تشهدها الصراعات في جميع أنحاء العالم تظهر الحاجة الملحة إلى العمل. وقد أظهرت برامج تطوير الأسلحة من قبل دول مثل روسيا والولايات المتحدة وإسرائيل وإيران أن الدول تتنافس بالفعل على التميز العسكري من خلال هذه التقنيات، مع ما يصاحب ذلك من مخاطر جسيمة على السلم والأمن العالميين.

وتظهر التقارير الأخيرة عن استخدام إسرائيل لأنظمة دعم القرار في غزة، رغم كونها ليست أنظمة أسلحة ذاتية التشغيل، كيف يمكن أن يؤدي السعي إلى السرعة من خلال الذكاء الاصطناعي، وتآكل السيطرة البشرية ذات المغزى، وتحويل الأشخاص لنقاط بيانات، إلى إحداث أضرار إنسانية مدمرة للمدنيين يستحيل جبرها.

SafeGround

[25 May 2024]

Secretary-General, in relation to resolution [78/241](#) “Lethal autonomous weapon systems” adopted by the UN General Assembly on 22 December 2023. The resolution requests the Secretary-General to seek views of countries and other stakeholders on “ways to address the related challenges and concerns raised[d] from humanitarian, legal, security, technological and ethical perspective and on the role of humans in the use of force,” posed by autonomous weapons.

SafeGround works to reduce harms of legacy and emerging weapons through outreach and education in Australia and the Pacific and finds it essential autonomous weapons are regulated through new international law to ensure the international community responds to these concerns. This submission highlights the moral, legal and humanitarian imperatives and briefly discusses national, regional and international security considerations. The submission then presents how a legally binding instrument addresses these concerns, including outlining elements such as specific prohibitions and obligations.

(1) Our concerns of autonomous weapons systems: moral, legal, humanitarian imperatives

SafeGround first formed as part of the International Campaign to Ban Landmines, advocating a ban on anti-personnel landmines which are indiscriminate and abhorrent. Autonomous weapons, particularly used against humans, have been called ‘morally repugnant’ by the UN Secretary-General and would not comply with IHL’s three core principles of distinction, proportionality and precaution. Human judgement is essential in applying IHL and the challenges to IHL compliance are one reason why new international law should be adopted to strengthen existing understandings and set new precedents based on legal and moral concerns.

The delegation of the decision over life and death to machines is a key moral risk raised by autonomous weapons. This would cross a moral red line and, along with digital dehumanisation and concerns such as human dignity and bias present an imperative against targeting humans with autonomous weapons.

Furthermore, autonomous weapons pose new humanitarian risks, already increasing autonomy is transforming how wars are fought, and AI-enabled targeting systems used in Gaza are facilitating unprecedented speed and scale of killing. There is a humanitarian need to ensure meaningful human control is maintained over the use of any weapons systems.

(2) Security considerations

Autonomous weapons pose an acceleration of conflict and higher risk of escalation. The proliferation of these weapons, especially if unregulated, and without requirements for meaningful human control will be destabilising. The Indo-Pacific region, for instance, has complex security dynamics and tensions and maintaining meaningful human control over autonomous weapons is important to manage these security risks. Autonomous weapons have implications for arms racing and regulation is an important mechanism for establishing guardrails and delineating what is and is not acceptable.

Based on historical trends, there is also a danger of these weapons proliferating globally, being used by non-state actors, in conflict and for other criminal activities. Misuse by non-state actors, for drug trafficking, illegal fishing, and political unrest, is particularly concerning for our region. These risks are heightened for Pacific small island developing countries, many of which have no military or only small peacekeeping forces.

Autonomous weapons systems also have environmental security risks. The use of landmines, cluster munitions, chemical, biological and nuclear weapons have had hazardous consequences for the natural world and security of impacted communities. Autonomous weapons systems could exacerbate environmental degradation by reducing human oversight when selecting and engaging targets. While the energy cost of training and operating AWS is not fully known, studies indicate that the carbon footprint of AI technologies could substantially contribute to climate change. SafeGround finds it crucial to promote the responsible development and use of any weapons system that takes into account the environmental security impacts.

(3) Addressing concerns through a legally binding instrument.

SafeGround calls for the establishment of a legally binding instrument on autonomous weapons to address the legal, moral, humanitarian and security concerns. New international law is essential to establish new international norms, moral precedents, and legal clarity.

i. Prohibitions

To address these challenges, we call for the prohibition of certain autonomous weapons

- Prohibitions on weapons systems that target people
- Prohibitions on weapons systems that cannot be used with meaningful human control

ii. Regulations

Autonomous weapons should also be regulated to ensure those not prohibited are effectively controlled through obligations related to;

- Understanding the system's functioning
- Limiting target types, context of use
- Limiting the duration, area, scope, and scale of operating

Beyond new legal rules, these limits may also include common policy standards and good practice guidance, which can be complementary and mutually reinforcing. Specific considerations should be made as to how regulatory responses can address proliferation to non-state actors, dual-use and environmental concerns and risks.

We are grateful for the opportunity to share our views and recommendations on addressing issues of autonomous weapons from a regional and civil society perspective. We must take this opportunity to take collective action with urgency, to respond effectively to this challenge to our shared humanity.

Stop Killer Robots

[20 May 2024]

The Stop Killer Robots campaign welcomes the opportunity to submit our views to the United Nations Secretary-General in response to Resolution [78/241](#). Established in 2012, we are a coalition of more than 250 non-governmental organisations and academic partners working across 70 countries.⁶⁷ Towards our vision and goals,⁶⁸ we are calling on states to adopt an international treaty on autonomous weapons systems that ensures meaningful human control over the use of force and rejects the automation of killing.

New rules are needed:

After ten years of international discussions, a legally binding instrument is the only effective way to start to address the humanitarian, legal, security, technological and ethical challenges and concerns that autonomous weapons systems raise. Relying on existing law alone will not be sufficient. The contributions of states and non-governmental experts to discussions have decisively demonstrated this. Specific rules and legal clarity are needed to draw clear lines to protect humanity.

There is now a critical mass of support:

Around the world, momentum continues to build behind the call for a treaty.

- More than 115 states now support a legally binding instrument.⁶⁹

⁶⁷ See www.stopkillerrobots.org/about-us and www.stopkillerorobts.org/a-global-push/member-organisations

⁶⁸ See www.stopkillerrobots.org/vision-and-values/

⁶⁹ See <https://automatedresearch.org/state-positions/>

- States have issued regional communiqués committing to work for a treaty, and convened crucial international and regional conferences in Austria, Costa Rica, Luxembourg, Trinidad and Tobago, the Philippines, and Sierra Leone.
- The International Committee of the Red Cross and the United Nations Secretary-General have jointly called on states to negotiate a treaty by 2026, which we strongly support.
- Thousands of scientists, tech workers, and leaders in robotics technology and AI, as well as religious leaders, parliamentarians, a wide range of civil society organisations, and public opinion in countries across the globe also support a treaty.⁷⁰

Negotiations can and must start:

All states and stakeholders that are committed to achieving a treaty must now work together in a spirit of genuine, cross-regional partnership and take concrete steps towards starting negotiations.

- A treaty must be negotiated in a forum that is inclusive of all states and civil society, and where progress cannot be blocked through veto, consensus rules, or their misuse.
- Because not all states parties to the Convention on Conventional Weapons (CCW) wish to negotiate a treaty, progress cannot be made on this in that forum.
- Establishing a negotiating mandate for a treaty through the United Nations General Assembly, as soon as possible, represents the best opportunity to move forward.
- From work in the CCW, there is now a broad understanding amongst states of what autonomous weapons systems are, and that a “two-tier” approach of prohibitions and positive obligations is required to effectively regulate them. This can be built on to agree a legally binding instrument.

⁷⁰ See <https://www.stopkillerrobots.org/a-global-push/a-shared-movement/> and <https://www.stopkillerrobots.org/the-story-so-far/>

An effective treaty on autonomous weapons systems

Scope:

Stop Killer Robots understands autonomous weapons systems to be those that detect and apply force to a target based on the processing of data from sensor inputs. In these systems, after activation by a human user, there is a period of time where the weapon system can apply force to a target without direct human approval. The specific object to be attacked, and the exact time and place of the attack, are determined by sensor processing, not humans.

- A treaty on autonomous weapons systems must include all systems that fall within this scope. Our concerns arise from the automated process by which such systems function. Our key concerns are at the erosion of human control over the use of force – which raises serious legal, ethical, humanitarian, and security concerns – as well as the ethical unacceptability of increasingly automated killing and digital dehumanisation.

The centrality of the role of humans:

The human role must be the central element of a regulatory framework. Ensuring meaningful human control must be a core purpose of a treaty.

A structure of prohibitions and positive obligations:

A treaty should prohibit autonomous weapons systems that are ethically or legally unacceptable. The use of all other autonomous weapons systems should be regulated through positive obligations to ensure they remain under meaningful human control.

Autonomous weapons systems that are not prohibited must be subject to positive obligations on design, development, and use to ensure meaningful human control:

Meaningful human control requires that a person can make a moral and legal judgement on the effects of an attack, and fulfil their obligation to apply the law. It also means there is a human user who is morally, and legally, responsible for the effects of an attack.

A treaty's positive obligations should therefore ensure that a human user is able to sufficiently anticipate the effects of an attack, and that after an attack has taken place they should understand and be able to explain its effects. To achieve such meaningful control, among other requirements:

- The user should be able to have a sufficient functional understanding of how a system works.
- Systems should be predictable and reliable.

- The user should be able to have sufficient understanding of the context where the weapons system will be deployed, and how it will function in that environment.
- The user must be able to limit the functioning of the weapon system in time and space.

Autonomous weapons systems that cannot be used with meaningful human control must be prohibited:

Systems whose technical configuration or environment of use means the user is not able to meet the requirements above could not be used in line with legal and ethical norms, and would raise questions about who was legally and morally responsible for any violations of the law. They must be prohibited.

Autonomous weapons systems that target people must be prohibited:

Even if used with meaningful human control, we consider such systems unacceptable because they reduce people to objects, which is an affront to human dignity.

- Autonomous weapons systems targeting people raise the most fundamental ethical concerns. They represent an extreme example of digital dehumanisation, for civilian and military victims alike.
- Racialized and historically marginalised populations are already disproportionately harmed by automated decision-making processes – for example, because biases in our societies are reproduced in data used to train algorithms. Autonomous weapons systems targeting people run the risk of biased technologies making determinations related to life and death. This must be prevented through a categorical prohibition.

The need for action now

Autonomous weapons systems change the relationship between people and technology. They disempower, disconnect, and dislocate humans in the use of force. The use of autonomous weapons systems beyond armed conflict, for example in border control and policing, would, furthermore, undermine international human rights law and freedoms.

Developments in weapons technology and events in conflicts around the world are already showing the urgent need for action. National programmes for weapons development by countries such as Russia and the US show countries are already competing for military advantage through these technologies, with grave risks for global peace and security. Recent reports of the use of decision-support systems by Israel in Gaza, though not autonomous weapons systems, show how the quest for

speed through AI, the erosion of meaningful human control, and the reduction of people to data points can contribute to devastating humanitarian harm to civilians.⁷¹

It is vital that states negotiate a treaty to start drawing clear lines for humanity now.

Stop Killer Robots Youth Network

[21 May 2024]

The Stop Killer Robots Youth Network welcomes the opportunity to submit recommendations for consideration by the United Nations Secretary-General in response to Resolution 78/241 on “Lethal autonomous weapons systems”. This is a summary, [find our full submission, including a list of signatories, here](#).

We interviewed young people from around the world to learn about their views on autonomous weapons and what their impacts would mean for young people today and in the future - their words and perspectives are reflected throughout this submission.

“The existence of autonomous weapons would mean we would be more disconnected with humanity.” - Sophie, 16, USA/Tajikistan/Tanzania

As a global youth network working to secure a future free of automated killing, we advocate for the creation of a new treaty on autonomous weapons systems (AWS) — in particular, we insist on a total prohibition of anti-personnel autonomous weapons as we wish to build a world without such dehumanising weapons. While youth will inevitably face the risks of new weapons technologies, we remain underrepresented in the decision-making process and are often sidelined in forums that shape our interests. With escalating conflicts and the rapid deployment of new weapons technologies around the world, there is an urgent need to reinvest in international law as a measure to build trust and achieve sustainable peace and security.

“I worry about autonomous weapons causing harm without human control, leading to more conflicts.” - Jefferson, 25, Ghana

Youth are disproportionately affected by armed conflicts. In 2022, one in six children were living in armed conflict⁷² and must continue to live with the horrendous consequences of these events⁷³. Children can be direct or indirect victims of attacks from means and methods that are not intended to be used on children⁷⁴. In addition,

⁷¹ See <https://www.stopkillerrobots.org/news/use-of-lavender-data-processing-system-in-gaza/>.

⁷² ICRC, *Childhood in Rubble : The Humanitarian Consequence of Urban Warfare for Children*, p. 7, Geneva, May 2023.

⁷³ Save the Children, *Children in Conflicts*, <https://data.stopwaronchildren.org/>.

⁷⁴ ICRC, *Childhood in Rubble : The Humanitarian Consequence of Urban Warfare for Children*, pp. 26-

critical infrastructures necessary for the development of children and youth are being destroyed or disrupted in current armed conflicts⁷⁵, sometimes intentionally. As the means and methods of warfare have evolved over time, recently, with the evolution of automated decision-making in acts of war, civilians have been increasingly impacted by these new means and methods⁷⁶. In Gaza, where artificial intelligence systems have allegedly been used to identify targets, 40% of civilian casualties have been children⁷⁷. With this in mind, we, as youth, are progressively apprehensive about the development of AWS. Since 1945, generations have grown up under the fear of being killed by nuclear weapons. Let us safeguard future generations from growing under the fear of being killed by AWS.

“If a human makes a mistake, there would be consequences for them. But, a killer robot, well, nothing will happen.” - Juline, 16, Belgium

Current conflicts have significantly strained youth confidence in international governance, fostering a sense of disillusionment and frustration. The continuous disregard for international law and human rights and a perceived lack of consequences for such actions have left many feeling hopeless and without adequate international representation or protection.

“Conflicts in recent years are escalating rapidly as never before, and this can be attributed in part to the misuse and proliferation of advanced technology and development of increasingly lethal weapons systems.” - Maria, 23, Lebanon

Vulnerable young people in conflict zones, who often bear the brunt of these incidents, find themselves in a particularly precarious position, with limited involvement in the creation of the conflicts that unfold around them. However, the prevailing trends of marginalisation and discrimination against vulnerable youth are likely to worsen with the advent of an autonomous weapons arms race.

“Just the fact that you even think about creating these kinds of weapons is a problem.” - Daniela, 26, Italy

Should international discussions continue to stall, a new catastrophic arms race may very well be underway. Autonomous weapons systems are a looming disaster in a world increasingly losing its sense of shared humanity. The UN stands at a pivotal moment with both an opportunity and an obligation to ensure that these detrimental

27.

⁷⁵ UNICEF, *Water Under Fire Volume 3, Attacks on water and sanitation services in armed conflicts and the impact on children*, New York, 2021. Global Coalition to Protect Education from Attack, *Attacks on Education and Military Use of Education Facilities in Ukraine in 2022*, February 2023.

⁷⁶ ICBL — CMC, *Landmine monitor 2023*, Geneva, November 2023, p.55.

⁷⁷ United Nations, *Two Thirds of Gaza War Dead are Women and Children, Briefer Says, as Security Council Debates Their Plights*, 22 November 2023, <https://press.un.org/en/2023/sc15503.doc.htm>.

trends do not persist unchecked. Ultimately, young people will bear the brunt of inaction, both now and in the future.

Weapons systems with varying levels of autonomy⁷⁸ are already impacting the lives of those impacted by armed conflict⁷⁹ — these weapons are no longer a future problem. Countries are racing ahead to field these capabilities and in reality, we are only a few software updates away from machines making life-and-death decisions⁸⁰. We are already witnessing the problematic effects of automated decision making and artificial intelligence resulting in ethical issues, including digital dehumanisation, and a lack of compliance with international humanitarian law.

“Young people like me should be able to live in a world free from the fear of being turned into numbers and data.” - Hevelyn, 24, Brazil

Planet Earth and its inhabitants are precious, we must all invest our time and energy to protect them. Technology should be used to promote peace and human rights, not for the creation of dehumanising weapons. What if we could create Autonomous Demining Systems to help communities impacted by landmines? The Anti-Personnel Mine Ban Convention should be a source of inspiration for us at this moment – 25 years ago, the international community agreed upon the ban of anti-personnel landmines, an inhumane weapon which also had disproportionate impacts on children and future generations⁸¹.

We urge states and leaders to act now, to stop the development of AWS and other means aimed at killing and dehumanising people. We have the unique opportunity to act before catastrophe occurs.

“Reaching a treaty on autonomous weapons would signify the beginning of a new era for international cooperation and disarmament” - Valeria, 22, Costa Rica

Using the UN and international law as a beacon of hope, we look to global institutions to advocate for justice and protect the most vulnerable among us. We rely on global leaders to uphold ethical standards, emphasizing the imperative of meaningful human control over weapons.

“A comprehensive treaty would mean that the voices of those who are and would be most affected by these weapons were heard and were taken seriously. It would reinvigorate my trust in the international multilateral system.” - Deborah, 28, Sweden

⁷⁸ Automated Decision Research, Autonomous weapon and digital dehumanisation, November 2022, p.3.

⁷⁹ +972 Magazine, ‘Lavender’: The AI machine directing Israel’s bombing spree in Gaza, 3 April 2024.

⁸⁰ Automated Decision Research, “Weapons systems”, [automatedresearch.org](https://automatedresearch.org/weapons-systems/), n.d. <https://automatedresearch.org/weapons-systems/>.

⁸¹ Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, Preamble, 18 September 1997.

Sustainable Peace and Development Organization

[25 May 2024]

Sustainable Peace and Development Organization (SPADO) welcomes the opportunity to submit views to the United Nations Secretary-General in response to Resolution 78/241. SPADO is a civil society organization based in Pakistan and is a member of the Campaign to Stop Killer Robots.⁸² The organization has actively and productively participated in the discussions on LAWS at the CCW and other such forums. We are committed to raise awareness and advocate for a legally binding instrument on autonomous weapons keeping in view its ethical, humanitarian, moral, legal and security concerns.

The submission outlines our views on the serious and unimaginable implications of autonomous weapons in the context of preserving our shared humanity, global peace and security, accountability and transparency and the urgency to initiate negotiations for a legally binding instrument that prohibit and regulate autonomous weapons without further delay.

Preserving our Shared Humanity:

The debate over killer robots extends far beyond mere questions of national security or ethical principles; it speaks to the very essence of our shared humanity. Allowing machines to arbitrate who lives and who dies risks dismantling the fundamental foundations of our society, relegating empathy and compassion to mere trace of our past. The development of lethal autonomous weapons contradicts the ethical principles and teachings of all faiths which call for compassion, human judgment, and the preservation of life. The development and use of LAWS will lead to actions that contravene the moral values and principles enshrined in our faiths as one human family. This is a question of our shared humanity, which is at stake and that of the preservation of human dignity.

Pope Francis has urged world leaders to establish an international treaty regulating AI, and warned against relying on machines to make moral and ethical decisions that should be left to humans. The most dangerous risk, however, lies in the military development of autonomous weapons. Pope Francis argued that the technology could never be "morally responsible subjects" and warned it could fall into the wrong hands and wreak havoc – potentially killing innocent citizens and upending democracies.⁸³

⁸² Stop Killer Robots is a coalition of more than 250 non-governmental organizations and academic partners working across 70 countries towards an international treaty on autonomous weapons systems that ensures meaningful human control over the use of force and rejects the automation of killing.

⁸³ See https://www.theregister.com/2023/12/15/pope_francis_ai_treaty/.

Global Peace and Security:

We have to look at the issue of autonomous weapons more broadly in the context of global peace and our shared humanity. If one country for instance develops and deploy autonomous weapons then it is for sure that other states will follow and a new arms race in the area of LAWS will emerge. The race will not stop and most probably the Non-State Actors will be in a position to develop and deploy such weapons if not in the near future but for sure in future. Such an arms race will be uncontrollable and unimaginable. We believe that prevention is better than cure and today, we have the time to prevent such an unethical arms race. The arms race will be like the COVID-19 Pandemic, which originated from one location and then affected the whole global community irrespective of race and region whether developed, developing or under developed. Now is the time to stop the future pandemic of LAWS.

The world has already become quite polarized where there is a lack of trust and mutual respect and understanding among states. The brutal wars are underway where parties to the conflict are using every form of lethal force. We are witnessing humanitarian catastrophe and the killing of innocent civilians including women and children on daily basis. In the midst of these conflicts certain states are increasingly investing in autonomous weapons and it is unclear to what extent such weapons have been deployed or used. If the use of such weapons becomes more evident than the already started race in the area of LAWS will become so fast, which will be uncontrollable. Such proliferation will be having catastrophic impacts on the international peace and security.

Accountability and Transparency:

The arguments that the use of autonomous weapons to protect civilians, target legitimate military objects and protect the armed forces raise many fundamental questions and concerns such as how will one limit the scope of the use autonomous weapons and how an autonomous weapon will be capable to identify that the target is legitimate and not a civilian object. There are higher chances that the use of such weapons will proliferate quickly and will raise many ethical and moral concerns, which will compromise the question of human dignity. Autonomous robots would lack human judgment and the ability to understand context. As a result, autonomous weapons would not meet the requirements of the laws of war. Replacing human troops with machines could make the decision to go to war easier, which would shift the burden of armed conflict further onto civilians. There are greater chances that antipersonnel autonomous weapons with deadly effects can be used by individuals against individuals based on their DNAs, facial recognition, biometric data etc. where it will be extremely difficult to find who developed and deployed such weapons.

Urgency is Needed:

States have been discussing the issue of LAWS for more than 10 years in the CCW where a large majority of member states were hoping that the discussions will lead to prohibitions and regulations of autonomous weapons. The global civil society, academia, tech experts, faith leaders and international organizations have been warning and showing their concerns related to the threats of autonomous weapons and asking for taking urgent steps towards a legally binding treaty. However, despite the large majority of states in favor of a legally binding treaty, the discussions didn't culminate into legal prohibitions and regulations on autonomous weapons.

The resolution of Austria on autonomous weapons at the UN General Assembly and the regional conferences in Latin America, Africa and Asia have provided an opportunity to the United Nations to carry forward the discussions into concrete outcomes leading to a legally binding treaty. We believe that the Secretary General New Agenda for Peace is the best way forward to build trust, solidarity and universality, which has also proposed to conclude a legally binding treaty on autonomous weapons by 2026. SPADO is of the view that the current wars and conflicts across the world demand the United Nations to take immediate steps to build consensus among states and initiate the negotiations of a legally binding treaty on autonomous weapons.

In conclusion, we would like to emphasize that there is a dire need to collectively utilize our human wisdom and don't just focus our attention about the present status and scenario of autonomous weapons but to imagine a devastating future scenario of automated killings where no one will be safe when such weapons will proliferate. States have the responsibility to act now as it will then be too late to stop autonomous weapons, which will eliminate human dignity, the value of life and the moral and ethical values, which the global community possess.

Women's International League for Peace and Freedom

[20 May 2024]

Pursuant to UN General Assembly resolution [78/241](#), WILPF is offering these concerns and recommendations to the UN Secretary-General's report on autonomous weapon systems (AWS). More in-depth analysis can be found in [Reaching Critical Will's papers on AWS](#).

International peace and security

The use of force has already become too disengaged from human involvement, through the use of armed drones and weapons operating with artificial intelligence (AI) or autonomous features. AWS further abstract violence from human beings. Algorithms create a perfect killing machine, stripped of empathy, conscience, emotion, judgement, or understanding of human life. AWS would not

hesitate to act; they would not take into account extenuating circumstances, nor challenge their deployment or operational mandate. They would simply do as they have been programmed to do—and if this includes massacring everyone in a city, they will do so without hesitation.

AW risk lowering the threshold for war. They present a perception of “low risk” and “low cost” to the military deploying the weapon. This perception increases the scope for the deployment of weapons into situations and to carry out tasks that might otherwise not be considered possible. Having an amoral algorithm determine when to use force means that we will likely see more conflict and killing, not less.

As seen with armed drones, remote-controlled weapons have made war less “costly” to the user of the weapon. Operators do not face immediate retaliation for acts of violence. While this is attractive to militaries that do not have to risk the lives of their soldiers, it *raises* the cost of war for everyone else. AWS would likely be unleashed upon populations that might not be able to detect their imminent attack and might have no equivalent means with which to fight back. Thus the burden of risk and harm is pushed onto the rest of the world.

War profiteering and global asymmetries

New weapons lead to new war profiteering. The production and proliferation of weapons means profits for corporate CEOs and shareholders. Corporations will be seeking to make money from the development and use of these weapons, and high-tech countries will use autonomous weapons to oppress and occupy others.

Countries of the Global South may not be the ones to develop and use AWS, but they will likely become the battlegrounds for the testing and deployment of these weapons. It will be the rich countries using these weapons against the poor—and the rich within countries using it against their own poor, through policing and internal oppression.

Human rights abuses

Existing military and policing technologies that use AI devalue and dehumanise people, and lead to violations of human rights and international law. AWS will exacerbate this further.

AWS could be programmed to commit acts of sexual violence. Some people who support the development of killer robots have argued that these weapons will be better than human soldiers because they will not rape. But just as sexual violence in conflict is ordered by states and by armed groups using human soldiers, an AWS could be [programmed](#) to rape. It is also important to consider the broader culture of rape in relation to weapons and war. Sexual violence is used as a weapon in conflict, and the risk of this kind of violence is also heightened during and after conflict. War destabilises communities and exacerbates already existing gender inequalities and

oppression of women, LGBTQ+ people, and others who do not conform to societies' gender norms.

AWS will also facilitate gender-based violence, including against men, by exacerbating policies and practice that count all cisgendered men as militants. In armed conflict, civilian men are often targeted (or counted in casualty recordings) as militants only because they are men of a certain age. Exacting harm on the basis of sex or gender constitutes gender-based violence. This erodes the protection that civilians should be afforded in conflict and violates many human rights, including the right to life and due process. It also has broader implications in the reinforcement of gender norms. Assuming all military-age men to be potential or actual militants entrenches the idea that men are violent. This devalues men's lives and increases the vulnerability of men, exacerbating other risks adult civilian men face such as forced recruitment, arbitrary detention, and summary execution.

As can be seen by [Israel's use of AI technologies](#) that generate target lists (Lavender) and target locations (Go Daddy), as well as the use of predictive policing software and border biometric systems in the United States and other countries, AI-enabled technology lends itself to this kind of gender-based violence. Reportedly, the only human checks on Lavender's kill lists are to ensure the targets are men.

Autonomous and AI technologies in weapon systems will further enable police and militaries to target people based solely on their gender, appearance, location, or behaviours, defining whole categories of people as militants, terrorists, or criminals without any due process. AWS could also be deliberately programmed to target people based on gender, race, socioeconomic status, (dis)ability, and sexual orientation. Just as AWS will lower the threshold for armed conflict, they will also lower the threshold for state violence against people. Police forces will be able to send machines to violently suppress protests and to repress certain categories of people, exacerbating discrimination.

In addition, data sets and the training with this data will cause bias. Parameters, boundaries, labels, and thresholds selected in the design phase necessarily exclude and include. This both creates bias and replicates existing bias within data and social structures. We already see examples in related technologies. Facial recognition software struggles to recognise people of colour; voice recognition struggles to respond to women's voices or non-North American accents; images of anyone standing in a kitchen are labeled as women; people's bail is denied because a program deemed a woman of colour more likely to reoffend than a white woman; trans people are surveilled on the basis of the clothing they wear. If such biases are left unchecked, there will be no counteracting human intervention.

Recommendations

The best solution is a **legally binding international treaty to prohibit the development, production, and use of AWS.**

Technology companies, tech workers, scientists, engineers, academics, and others involved in developing AI or robotics should pledge to never contribute to the development of AWS.

Financial institutions such as banks and pension funds should pledge not to invest money in the development or manufacture of autonomous weapon systems.

States, civil society groups, activists, tech workers, and others should also work to prevent AI-enabled technologies from being used by militaries and police forces. It is not just AWS that are problematic, but the overall automation of violence, as well as sensor-derived target detection, algorithmic bias and software-generated kill lists. These must not be normalised, they must be prevented.

AWS are a product of an arms race that derives from the global system of militarism and war profiteering. This system fuels armed conflict and armed violence, human rights abuses, and other violations of international law. It is therefore important to not just ban AWS, but to dismantle the structures of state violence as a whole.

Virtual Planet Africa, Women’s International League for Peace and Freedom Zimbabwe and Cultural Youth Movement Education Foundation

[16 May 2024]

Autonomous weapons systems present risks and concerns to the security of states and communities, particularly in the Global South.

We support the creation of a legally-binding instrument that considers (a) meaningful human control as an integral and intrinsic concept, (b) risk assessment and mitigation, (c) clarification of international humanitarian law (IHL) requirements for AWS, (d) transparency and confidence building measures (TCBM), (e) prohibitions and regulations, (f) regulations on autonomous arms trade, and (g) consideration of AWS challenges in other international law frameworks.

B. Meaningful Human Control

Meaningful human control of systems refers to the making of critical decisions by humans, contextualisation, and rationalisation by a human(s) operator in the application of force. All systems that can make critical decisions on the use of force without meaningful human oversight should be banned.

C. Risks and Mitigation Measures

AWS pose several risks, challenges, and concerns. AWS that can result in excessive or superfluous injury to civilians and civilian objects should not be developed, distributed, or used. Targeting civilian objects like hospitals, schools, inter alia using AWS should be prohibited.

AWS have a high risk of being accessed and used by armed non-state actors. A legally-binding instrument on AWS should consider ways of safeguarding developing states, and would ensure that armed non-state actors do not easily access AWS. This would be done through building a robust framework for the regulation and restriction on the development and trade in AWS.

Algorithmic bias is also a major concern for AWS especially for people in the Global South. AWS are only as good as the data used to train them. The risks of cultural, social, and identity bias presented by AWS can have adverse effects. The proposed legally-binding instrument should include clear and strong measures to ensure that algorithmic bias relating to gender, sex, race, ethnicity, age, class, language, location, and other identity categories can be eliminated before a system is deployed. Protocols should be put in place for the prosecution and punishment for developers and users of AWS with algorithmic bias and that make mistakes based on algorithmic bias.

System failure and possible manipulation by malicious actors is another key concern. Our approach values the control of systems by human operators at all levels in a system's cycle, and machine failure, hacking, or reprogramming by malicious actors can lead to loss of control.

D. International Humanitarian Law

The International Court of Justice Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons opined that all types of weapons, whether from the past or in the future, are governed by IHL. Guiding Principle (a) of the GGE on LAWS affirmed that IHL "continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems." Although IHL is just a bare minimum requirement, the development, distribution, and use of AWS should ensure that AWS fulfil the 'cardinal principles' of IHL;

Distinction: AWS should not target civilians and civilian objects. We suggest the expansion of the interpretation of this principle in the context of AWS to prohibit the targeting of human targets by AWS.

Proportionality: AWS must not cause excessive, superfluous, and unnecessary injury in relation to anticipated military advantage anticipated. To avoid the killing of those deemed as *hors de combat* by AWS, we suggest that AWS should not be used to target or against humans.

Precaution: proper precautions must be made before deployment of AWS to ensure that they do not attack civilians and civilian objects, they do not contain algorithmic bias, they will not cause unnecessary damage/injury, and that they will not disregard humanitarian needs in the use of force.

E. Transparency and Confidence Building Measures (TCBM)

Transparency and Confidence Building Measures (TCBM) should be put in place to ensure that AWS do not lead to an arms race, unintended conflicts, and misconceptions about AWS. These should include Article 36 reviews of weapons, transparent risk assessments at all levels, information sharing and exchange, internationally agreed review processes and procedures, national obligations for safety and risk mitigation, national and regional ethical standards, periodic reports on military expenditure on AWS for all states, and the creation of an oversight body.

F. Prohibitions and Regulations

In line with the ‘two-tier’ approach, the following AWS should be **prohibited**: systems that

- operate without meaningful human control,
- with unpredictable or unexplainable actions and outcomes,
- that cannot distinguish between civilians, civilian objects, and military objects/targets,
- cause unnecessary or superfluous injuries,
- cannot fulfil the requirements of international law, especially IHL,
- cannot be terminated by a human user once deployed,
- contain algorithmic bias and other harmful identity-based risks,
- can learn and change decisions, parameters, or goals after deployment, and
- target humans directly without the input of a human operator.

The following positive obligations should be put in place as regulations:

- Limits should be put in place on the geographical and mission-based operation of AWS,
- The types and nature of targets shall be limited,
- Human supervision must be ensured at all stages of the weapon’s lifecycle,
- Mandatory obligations for training of personnel shall be put in place before a system is used,
- The duration and scale of AWS operations must not be unlimited,
- Clear chains of command and responsibility shall be outlined,

- Frameworks for the prosecution and punishment of individuals, entities, or states who misuse or develop prohibited systems must be established before systems are deployed,
- AWS must be traceable, and
- Human users must be able to terminate AWS at any and all times.

G. AWS Trade and Distribution

The trade and distribution of AWS must be strictly regulated. In addition to the provisions of the Arms Trade Treaty, clear regulations must be put in place to ensure that AWS do not fall into the wrong hands.

States and companies that develop AWS should not be allowed to test them in active combat zones in fragile communities. Security assistance for fragile states and states with high levels of insecurity should only be done with AWS whose effects, effectiveness, and efficacy has already been tested elsewhere. States should not use security assistance as an excuse to test AWS. This should be put in place to ensure the safety and dignity of vulnerable communities.

Industry developers of AWS must have clear guidelines on the development and distribution obligations they have. AWS should be sold only to state parties and reputable registered companies. Rules to prosecute developers who sell to armed non-state actors and terrorist groups shall be put in place.

AWS shall be sold only to, and used only in, states with existing national ethical guidelines and standards to govern the development, testing, trade, and use of AWS. These ethical guidelines and standards must be in line with the proposed international legally binding instrument.

H. Other International Law Provisions

International Human Rights Law, International Criminal Law and other Codes of Conduct shall also guide and support the proposed legally binding instrument.