



ICRC

# 红十字国际委员会关于自主武器系统的观点

2016年4月11日

《某些常规武器公约》  
框架下关于致命自主武器系统的专家会议

2016年4月11-15日，瑞士日内瓦

# 引言

随着人们进一步了解当前及潜在的技术能力、武器系统自主性的军事目的以及由此在遵守国际人道法和伦理方面引发的问题，红十字国际委员会 (ICRC) 等各方对自主武器系统的看法也在不断发展变化。

过去的三年中，专家们在《某些常规武器公约》的框架内以及在 ICRC 和其他组织主持召开的会议上所进行的讨论，对增进这种理解至关重要。

2016 年 3 月 15-16 日，ICRC 在瑞士韦尔苏瓦召开了主题为“自主武器系统：增强武器关键功能的自主性带来的影响”的第二次专家会议。<sup>1</sup>20 个国家的代表、以个人身份参会的专家、联合国以及民间组织代表出席了此次会议。ICRC 即将发表本次会议的总结报告。

与此同时，为促进《某些常规武器公约》框架内正在进行的讨论，本文件依据最近一次专家会议上的讨论，并从 ICRC 的视角，着重阐述了有关自主武器系统的几个关键问题。

## 一、定义

ICRC 将自主武器系统定义为：“任何在关键功能上具有自主性的武器系统。即无需人类干预就能独立选定目标（即搜索或探测、识别、追踪、选择）并发起攻击（即使用武力、消除威胁、损害或摧毁）的武器系统。”

这个广义的定义包含了一些现有的武器系统，其优点是能在现实世界中对武器技术进行考量，从而评估某些现有武器系统在法律和伦理上的可接受性，以及哪些新技术的发展可能会在国际人道法、人道原则和公众良心方面引发问题。

一些人倾向于采用狭义的定义，主要涉及确定最为关切的武器系统的类型。此外，一些狭义的定义还区分“高度自动化”和“自主”武器。不过，许多专家都强调，从技术角度看并不存在这样明显的差异，核心的法律和伦理问题仍然是相同的。

ICRC 认为，以一个广泛的范围作为起点十分有益，因为从现有的在关键功能上具有自主性的武器系统得到的经验，可以为当前讨论新技术提供借鉴。

## 二、现有武器中的自主性

现今确实存在正在使用中的无需人为干预即可选择和攻击目标的武器系统。经操作人员启动后，这种武器系统可以通过其传感器和程序设计选择一个目标并发动攻击。

广义上说，它们是指用于保护车辆、设施或地区免受导弹、火箭、迫击炮或其他投射物来袭的**反器材防御武器**，包括导弹或火箭防御武器和车辆“主动防护”武器。要具备对这些武器以及对武力使用进行有效控制的能力，似乎需要有某些运行上的限制，包括：对所执行的任务

---

<sup>1</sup> For a report of the first meeting see: ICRC (2014) *Autonomous weapon systems technical, military, legal and humanitarian aspects*, Report of an Expert Meeting held 26-28 March 2014 (published November 2014), <https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf>

进行限制（即防御来袭投射物的单一功能）；对攻击目标进行限制（即主要是物体和车辆）；控制运行环境（例如对自主运行的地理区域和时间范围进行限制）；以及人为干预以关闭该武器（即停止其运行）的程序。

某些进攻性武器系统，包括某些导弹和鱼雷，发射后在选择和攻击目标方面也有一定程度的自主性。在这些武器中，很多是被发射至某一目标区域，然后板载的传感器和程序就会接管，自主地选择和攻击该区域内特定的目标物体或人员。还有一些武器则在行动的时间和/或空间上有更多自由，因此具有更大的自主性。这些武器尤其包括长时间在一个广阔的地理区域内搜索目标的巡飞弹，以及长时间在水下保持静止但可以自主发动攻击的封装鱼雷武器。

导弹的发展趋势似乎也是在飞行的时间和空间上增加自主性。有迹象表明，未来的发展还可能包括这些武器系统对所处环境的适应性日益增强。能否对这些武器以及对武力的使用实现有效控制可能取决于许多因素（正如上文讨论过的防御系统），包括：对所执行的任务进行限制；该系统区分目标的能力；控制运行环境，例如在时间和空间上加以限制；人类与该武器系统连通的能力，例如关闭该系统的能力。后者对于水下系统来说尤其困难。

还有其他的杀伤人员武器也能够自主地选择和攻击目标，例如用于保卫设施和边界的所谓“哨兵”武器。不过，现今使用中的这类系统在发动攻击时明显仍处于人为遥控之下。

总之，现有武器系统中选择和攻击目标方面的自主性受到上文所述的运行参数的限制。此外，现有武器系统的技术特征和性能以及使用时的运行参数让这些武器系统的使用效果具有一定程度的可预测性。与迄今为止的情况相比，如果自主武器系统被用来执行更为复杂的任务或部署于不断变化的环境，这种可预测性就会消失。

### 三、新技术和未来的自主武器

尽管很难预见自主武器系统的未来发展，但很明显的是，存在许多增加自主性的军事动机，例如有助于：增加机器人/无人武器系统或平台的机动性；维持这些系统在“通讯中断”的环境中的正常运行；缩短从识别目标到发起攻击之间的决策时间；提高远程操作系统的性能；以及减少操作人员的数量，而增加机器人/无人武器系统的运行数量。非政府武装团体对于发展自主武器系统可能还有其他的动机。

民用机器人技术的总体趋势是朝着有监督的自主性的方向发展，即机器人系统的自主性日益增强，同时操作人员仍对其进行监督而且往往能够进行干预。

机器人系统的自主程度与运行过程中的人为干预水平有关，涉及人为干预的程度与进行干预的阶段。即便如此，通过程序设计，机器仍可以并且确实能够有效做出人类授权其做出的决定，无需有“意识”或拥有类似于人类的智力水平。

许多方面的发展可能使自主性日益增强的武器系统变得更加不可预测。这些发展包括：日益增加的机动性，这意味着该武器系统在较长一段时期内会面临更加多变的环境；日益增加的适应性，例如可以根据环境的变化自行设置目标或改变自身功能的系统（例如防护自身免遭攻击的系统）甚或纳入学习算法的系统；以及在一个自组织群中多武器系统之间日益增加的相互

影响。除了降低该武器系统的可预测性，这些发展还可能在确保有效的可靠性测试方面带来相关问题。

例如，能够自行设定目标，甚或“学习”并调整其功能的自主武器系统，就具有不可预测的性质。高度机动的自主武器系统在具体攻击的目标方面也产生了可预测性问题，尤其是某个系统在一个广阔区域内活动或发动多重攻击时更加难以预测。

#### 四、增强武器系统的自主性所带来的法律和伦理影响

显然，规制敌对行动的国际人道法规则适用于武装冲突各方，更具体地说是适用于人类战斗员和武装人员，他们有义务遵守国际人道法并对违法行为承担责任。这些义务不可能转移给一台机器。然而实际上，仍然存在的一个问题是需要对武器系统的自主性需要施加何种限制才能确保其遵守国际人道法。

人类控制可以采用多种形式并体现在自主武器系统“生命周期”的不同阶段，包括：1) 该武器系统的研发（包括编程）；2) 该武器系统的部署和使用，包括指挥官或操作员决定使用或启动该武器系统；以及3) 该武器系统的运行，该系统在此阶段选择和攻击目标。

显然，在该武器“生命周期”中的研发和部署阶段，人类可以实施控制。而正是第三个阶段，该武器运行时——它自主地选择和攻击目标时——在法律、伦理和军事作战方面产生了一个重要问题：前两个阶段的人类控制是否足以克服最后阶段较少或没有人类控制所产生的影响（另见下文第五部分）。

自主武器系统能否以符合国际人道法的方式使用，对这个问题的评估可能取决于该武器系统的具体技术特征和性能以及预期和可能的使用环境。某些技术特征及其与不同运行参数的相互影响可能会严重影响这项评估，包括：

- 该武器系统执行的任务；
  - 该武器系统攻击的目标类型；
  - 该武器系统运行的环境；
  - 该武器系统机动的空间范围；
  - 该武器系统运行的时间范围；
  - 该武器系统的适应性，即它调整自身行为以适应环境变化、自行确定功能并设定目标的能力；
  - 该武器系统的可靠程度，即容错能力和应对故障或黑客攻击的能力；
- 以及
- 人为监督并干预以关闭该武器系统的可能性；

该武器系统的这些技术特征和性能与其使用时的运行参数结合在一起，对于确定该武器可预见的效果——换言之，使用该武器结果的可预测性——从而确定其使用是否符合国际人道法规则来说至关重要。

确实，如果某一武器系统的效果完全或部分不可预测，那么其部署就会带来违反国际人道法的严重风险。这种风险可能太高以致不允许使用该武器，或者为缓解这种风险而要求限制甚

或取消该武器的自主性。在这方面，上表所列的最后一项因素——人为干预——可以被视为一种风险缓解因素。该武器系统可预测性降低所造成的风险水平可能受许多因素影响，例如该武器的使用环境。

随着自主武器系统变得更加复杂或者在运行过程中被给予更多行动自由，其使用后果可能会愈加难以预测。例如，在一个简单（有序）环境中针对某一类型的目标执行单一任务、位置固定且运行时间有限的一个自主武器系统（例如现有的某些导弹和火箭防御系统），对它的法律评估可能得出的结论是，存在可接受的可预测性水平，能够让操作人员负责并进行追责。但是，对于在一个复杂（混乱）环境中针对不同类型的目标执行多重（或可调整的）任务而且在广阔区域内机动和/或长时间运行的自主武器系统，得出的结论可能就会非常不同。

就依据国际人道法进行评估而言，法律上并不区分进攻武器和防御武器，因为二者都构成法律意义上的攻击。尽管如此，从军事作战或伦理视角来看，区分防御和进攻武器系统可能具有更重要的意义。

根据《第一附加议定书》第36条对新武器进行法律审查的义务对于保证一国武装部队能够以符合其国际义务的方式开展敌对行动十分重要。在对发展或取得的任何自主武器系统进行法律审查时，各国需要谨慎考虑遵守国际人道法所面临的上述挑战。像所有武器一样，在关键功能上具有自主性的武器是否合法取决于其具体特征，以及考虑到这些特征，它能否在所有预计和可能加以使用的情况下都以符合国际人道法规则的方式被部署。要具备进行这种审查的能力，需要充分了解该武器的性能并预见其效果，尤其是通过测试的方式。然而，如果自主武器系统变得更加复杂，或者在运行过程中被给予更多行动自由而变得更加不可预测，那么这种效果就可能变得越来越难以预见。

至于国际人道法的“目标打击规则”（例如比例原则和攻击时预防措施规则）在武器审查时如何加以考虑，也产生了问题。如果是武器本身具有目标打击功能，法律审查就会要求该武器具备非常高的可靠程度，能够以遵守国际人道法的方式运用那些功能。

审查自主武器系统合法性面临的另一个挑战是，在评估这些武器的性能以及使用期间可能存在的风险时，缺乏测试和评价的标准方法和方案。这就产生了如下问题：被测试武器的可靠性（例如故障率或应对网络攻击的能力）和可预测性如何？何种程度的可靠性和可预测性被认为是必要的？在评估一个自主武器系统是否能在预计或可能的使用环境中表现符合预期时，法律审查程序就面临着上述以及其他的现实挑战。

各国为保证自主武器系统符合国际人道法而对新武器进行国内法律审查，对于这种审查的充分性，尽管存在不同的观点，但尤其是考虑到各国落实的程度较低，这一机制对于各国确保遵守国际人道法来说仍是一项至关重要的措施。无论如何，加强国内法律审查程序的努力，在国际层面都应被视为对《某些常规武器公约》框架内讨论的补充和相互加强。

一些人还担心，使用自主武器系统可能会导致发生违反国际人道法行为时的“问责真空”。其他人则认为，从来没有这种真空，因为总会有人参与到部署该武器的决策中，责任可归于该决策人。

根据国际人道法和国际刑法，由于对自主武器系统的控制具有局限性或其不可预测性，导致难以对参与武器程序设计与部署的人员就严重违反国际人道法的行为追究责任。由于机器

可独立地选择和攻击目标，这些人不可能具备归责所要求的明知或意图。程序设计人员不可能知道下一阶段可能部署该武器系统的具体局势以及在该局势中可能发生的违反国际人道法的行为。另一方面，故意给一个自主武器系统编写程序令其实施战争罪的设计人员当然要承担个人刑事责任。同样，若一个指挥官决定以非法方式使用自主武器系统，例如在人口稠密地区部署无法区分平民和战斗员的杀伤人员自主武器系统，他就要为此承担责任。此外，若一个指挥官在明知其无法预测某一自主武器系统的性能和效果时仍决定部署，在这种情况下，其部署该武器的决定就被认为是鲁莽的，他可能要为此后发生的严重违反国际人道法的行为承担刑事责任。总体上，只要有人类参与部署该武器的决定且责任可归于该人，就不可能存在问责真空。

根据国家责任法，一国可能要为使用任何自主武器系统所造成的违反国际人道法的行为负责。确实，根据规制国家责任的一般国际法，国家要为其国际不法行为承担责任，例如该国武装部队使用自主武器系统实施了违反国际人道法的行为。一国如果使用了部署前尚未加以测试或审查或者测试或审查不充分的自主武器系统，也要承担责任。

自主武器系统还会引起伦理方面的担忧，值得仔细考虑。在这些担忧中，最为核心的基本问题是，不论自主武器系统能否以符合国际人道法的方式加以使用，人道原则和公众良心要求是否允许机器在武装冲突中不经人员参与即做出生杀予夺的决定。近年来，各国政府、专家、民间团体和公众对于这个问题的争论表明，人们对于任何不受人类控制可自行使用武力的武器系统深感忧虑。

然而始终存在的问题是，从伦理角度出发，人类需要在什么情况下、进行何种程度的控制？人类根据某些参数为一个自主武器系统设计好程序，然后决定将其部署在某一环境中，这就足够了吗？或者人类是否有必要对每一次攻击都做出自己的判断？如果该武器自主地针对某一人员目标使用武力，需要进行哪些伦理上的考量？

## 五、人类控制

在关于自主武器系统的争论中，人类控制的概念已成为首要问题。对于武器系统和武力的使用必须保留人类控制这一观点已达成广泛共识，尽管尚不清楚这是出于法律、伦理、军事作战抑或政策原因，也不清楚什么可以让人类控制“有意义”、“适当”或“有效”。

从 ICRC 的角度看，关注人类在目标打击过程中的作用以及人机交互，而不是纯粹在技术上关注武器系统的“自主水平”，可以为深入理解自主武器系统可能引发的问题提供一个有效的办法。

对攻击进行一定程度的人类控制，是国际人道法的区分原则、比例原则和攻击时的预防措施原则中所固有的，也是保证遵守这些原则所要求的。深入思考这些要求是什么，可能有助于确定一个标准，即对于选择和攻击目标这两项关键功能中的自主性而言，哪些方面根据国际人道法是可接受的。

已经有人提出了许多考量因素，为今后确立这些要求提供了可行的方法，它们包括：

- 该武器系统在预计或可能的使用环境中的可预测性；

- 该武器系统在预计或可能的使用环境中的可靠性；
- 在该武器系统研发、部署和使用期间对其运行的人为干预；
- 对该武器系统运行及其预计或可能使用环境的了解及准确情报；
- 启用后，对该武器系统运行的问责机制。

许多与评估国际人道法的遵守有关的技术特征和运行参数（见第四部分）也是判定对使用武力进行必要人类控制的重要因素，同时也与军事行动或伦理考虑密切相关。例如，对现有自主武器系统的人类控制（见第二部分）在很大程度上是对系统运行进行技术和操作上的限制（例如限制任务和目标，限制空间和时间，对环境进行实际控制，以及人为监督和关闭武器系统的能力）。

军方在武器系统的人类控制方面具有明确的意愿，这既是为了保证遵守国际人道法，也是为了保证指挥官对特定军事行动的控制权。与特定武器系统相关的交战规则是对武器系统和武力的使用实施作战控制的首要方式，因此也是保证对武器系统和武力的使用进行人类控制的一个重要因素。

要确保对武器系统和武力使用保持人类控制就必须划定必要的界限，这要求对于构成人类控制的要素以及人机交互过程进行更深入的思考和了解。正如上文所述，由于对武力的使用和具体的攻击将在何时、何地发生缺乏可预见性，对于能够调整或改变自身功能的自主武器系统已经存在一些担忧，而那些在广阔区域内“猎取”目标的自主武器系统则可能引发关于人类控制和遵守国际人道法的严重问题。

## 六、前景展望

大体上，各国至少可以采用三种方法来处理自主武器系统引发的法律和伦理问题。第一种是加强法律审查和实施国际人道法的国内机制，确保包括自主武器系统在内的任何新武器的使用都能符合国际人道法。

第二种是各国就可能引起法律和 / 或伦理问题的武器系统制定一个关于“致命自主武器系统”的定义，从而为武器系统的自主性设定具体的限制。

第三种方法是各国根据国际人道法和伦理考量（人道原则和公众良心要求）的具体要求制定出衡量人类控制的参数，从而为武器系统的自主性设定具体限制。

第二种和第三种方法均承认，需要在国际上考虑如何对武器系统的自主性加以限制，从而确保守法和伦理上的可接受性。最终这两种方法可能在辨别需要加以规范或禁止的武器系统方面殊途同归。

在 ICRC 看来，侧重于人类控制和人机交互的第三种方法可能是推动《常规武器公约》发展的有效途径，同时还可将加强国内法律审查作为一项相辅相成的举措。人类控制的框架为各国达成共识提供了一个基准，并以此为基础为武器系统的自主性设定标准或限制。这符合各国政府、专家和其他利益攸关方达成的广泛共识，即鉴于相关的法律义务、军事行动要求和伦理考量，有必要对武器系统和武力的使用保持人类控制。