

Convention on Certain Conventional Weapons (CCW)
Group of Governmental Experts on Lethal Autonomous Weapons Systems
27-31 August 2018, Geneva
Statement of the International Committee of the Red Cross (ICRC)
under agenda item 6(b)

Thank you, Mr Chair.

The ICRC is encouraged to hear the emphasis and general agreement among delegations that human control must be maintained over weapon systems and the use of force.

The ICRC would like to take this opportunity to recall several points we raised in our statements to the April Group of Governmental Experts on the legal and ethical basis for human control. Further we would like to offer some insights from a small meeting of technical experts we convened in June 2018 on the technical aspects of human control.

Firstly, on the **legal basis for human control**, the ICRC is clear that it is **humans who apply international law and are obliged to respect it**. International humanitarian law rules on the conduct of hostilities require that those who plan, decide upon and carry out attacks make certain judgements when launching an attack.

It follows that **human combatants will need to retain a level of control over weapon systems** and the use of force so that they can make context-specific legal judgements – of distinction, proportionality and precautions – in specific attacks.

Legal concerns will arise where the design and/or the use of a weapon system with autonomy in its critical functions prevents the commander or operator from making the necessary judgements required by international humanitarian law.

Secondly, in the ICRC's view, ethical considerations also demand human control over weapon systems and the use of force. From an ethical perspective, (minimum/sufficient/effective/meaningful /appropriate) human control would be the **type and degree that preserves human agency and upholds moral responsibility** in decisions to use force.

Mr Chair,

In June 2018, **ICRC convened a small meeting of independent experts** in civilian artificial intelligence (AI), robotics and autonomous systems to better understand the technical aspects of human control. A report will be published in due course but **two broad themes emerged**:

The first theme was that **all autonomous robotic systems**, which operate based on interaction with their environment, **raise questions about human control and predictability**. The greater the complexity of the environment and complexity of the task, the greater the need for human control and the less tolerance of autonomy.

Humans can exert some control over autonomous systems through “human on the loop” interaction with them. However, this is not a panacea due to human-machine interaction problems, such as automation bias, over-trust in the system, or lack of operator awareness of the system state at the time of intervention.

Further, **quantifying the level of predictability** (and reliability) **required to ensure** (minimum/sufficient /effective/meaningful/appropriate) **human control, or judgement, is very difficult**. Testing also raises unique challenges since it is not possible to test all possible environmental inputs to an autonomous system.

However, **setting boundaries – or operational constraints** – in the operation of an autonomous robotic system – for example, on the task, time-frame of operation, scope of movement over an area, and operating environment – **can contribute to increasing predictability**.

The second key theme that emerged from discussions was that **AI - and especially machine learning – algorithms bring a new dimension of unpredictability** to autonomous robotic systems. This is due both to the lack of transparency (“black-box” nature) in how the algorithms function, and the changing of their functioning over time. Not only are such algorithms unpredictable, but they **can also introduce bias**, whether by design or through bias in the data used to “train” (develop) the algorithms.

This has **reinforced the ICRC’s concerns about the unpredictability of machine learning algorithms, and the consequences if applied to weapon systems** and the critical functions of selecting and attacking targets.

Mr Chair,

In closing, we would like to draw attention to the comments of the ICRC’s President published today, which stressed the importance of retaining human control over weapon systems and the use of force. He notes that, in some ways, **recognising the need for human control is the easy part. The more difficult question is: how much and what type?** It is this issue where the ICRC believes this Group of Governmental Experts should focus its work over the coming days, weeks and months.