ICRAC intervention on The Human Element in the Use of Force

Aug 4 2021, CCW UN GGE LAWS

Chair,

This statement is read on behalf of the International Committee for Robot Arms Control (ICRAC), who are observing the meeting online but cannot be present in the room due to COVID-19 restrictions.

ICRAC has been very pleased with the positive energy in the room and the increasing focus on the problems with autonomous weapons under IHL.

We wish to underline our continued support for a functional definition of AWS as weapons which have autonomy in their critical functions of target identification, selection, and the decision to apply force. As a previous speaker put it very well, autonomy is a capability and not a category of weapons. The past several years of discussion have made this very clear, and this functional definition provides sufficient clarity for the CCW to proceed onwards to the negotiation of an international legal instrument that can establish norms of appropriate behaviour and drive operational approaches in a way that allows us to retain meaningful human control and protect our shared humanity.

It is important to keep in mind that we do not need to include autonomy in the critical functions of weapons in order to add safety features such as warnings to operators, or automatic termination of attacks if civilians or civilian objects are detected. Autonomy in the critical functions and the addition of safety features are independent concerns and regulation of AWS would not ban the addition of safeguards, as long as human engagement remains the core of the decision-making process.

We also reiterate our position on the fundamental unpredictability of the application AWS. Autonomous weapons systems, even if targeting only military objects, can be unpredictable in at least three ways.

First there is an inherent unpredictability in any autonomous system sent to achieve a goal. An example provided by the leading AI company Deep Mind at the CCW, is that if a machine learning algorithm is given an objective function of fishing a lake, it may drain the lake rather than fish in the conventional sense. While this is an extreme example in an AI system, it is borne out in many civilian applications that result in algorithmic bias. It is easy to see how this matters in conflict: when an AWS is sent to target a military group, its methods may unpredictably involve harming civilians.

Second, an autonomous system is unpredictable when operating within a dynamic environment such as in armed conflict. Autonomous weapons rely on sensor readings to provide information to an algorithm that controls their behaviour. If sensors encounter unanticipated circumstances in the environment - circumstances that their algorithms have not been trained on or programmed for, then we should think about them as analogous to a ricocheting projectile. This can also occur when there is an error in the sensor readings.

Third, if we do not stop the continued development of autonomous weapons systems (as defined by the ICRC), it will only be a matter of time before a number of enemy AWS will

encounter each other. This creates another potentially devastating problem of unpredictability. There are many real world examples of major problems created by financial algorithms clashing.

For example two simple price setting algorithms from Amazon market place sellers interacted over a $50 book in 2011. Neither of the sellers noticed for a week and in that time the price of the book had gone up to $23,698,655.93. No one bought it and no one was harmed. However, imagine two or more complex algorithms interacting on high speed autonomous weapons. Without any knowledge of the other algorithms, there is no way to tell what would happen. Any such interaction will be potentially very dangerous and cannot be predicted in advance. Therefore, compliance with IHL cannot be guaranteed, ….and strategic risks, including risks of accidental, algorithmically triggered war, loom large.

Thank you Mr Chair.