Statement of the UN Institute for Disarmament Research

at the CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems
12 April 2016
Delivered by Kerstin Vignard, Deputy to the Director

Mr Chairman, distinguished colleagues

These last two years of discussions on LAWS have been rich and educational. We have had the opportunity to interact with expertise on technical, legal, ethical and other aspects. Discussion of LAWS has opened a space for a vast array of issues to be brought to the table—ranging from what sorts of proxy indicators are acceptable for use in targeting, to whether AI will be the end of humanity. Some of these issues are specific to increasing autonomy in weapon systems, and others aren't.

These rich exchanges have been productive for helping the international community to get its bearings on this issue. This third week of informal discussion is particularly important, as it will help to delineate the field of future discussion as we move towards the CCW Review Conference.

So in the spirit of encouraging governments to move the conversation forward, UNIDIR would like to make five observations—the first and last of which are about alternative frames for this issue--the first being a minimal frame and the last a more radical frame shift.

1 AWS is a misnomer

Some of you might recall that UNIDIR's first Observation Paper in 2014 was called "Framing Discussions on the weaponization of increasingly autonomous weapon systems". Framing discussions in a productive way can help set them up for success and focus in on the critical issues.

Thus the first observation is that it is time to take the international discussion on "Lethal Autonomous Weapon Systems" and at a minimum reframe it as "Autonomy IN Weapon Systems". This isn't just playing with words. It acknowledges that varying levels of autonomy might be applied to different characteristics within the same object or weapon system. This frame also allows us in the short term to pivot away from trying to draw--and agree upon--lines between fully versus semi-autonomous or supervised autonomy, or attempts to come up with a formula for determining whether something is autonomous or just highly automatic, and focus in on the functions that when increasing amounts of autonomy is applied to them raises concerns, challenges or where we need to develop shared understandings of how our existing obligations and norms apply.

2 The whole will be greater than the sum of its parts

Many States affirm that autonomy discussions in CCW are not about existing systems. However, there are highly autonomic/autonomous components or features of existing systems--that if in the future are combined in particular ways--may pose unique and new concerns, even if the features or weapons are not problematic in themselves today. This moves the time-line from a far-off future concern to a much more near-term technological possibility.

A dynamic exchange between States on how these existing or near-term features might combine and the acceptability of these different combinations would be useful. Some are perhaps uncomfortable about talking about existing systems. However, I stress that this discussion is different than--and shouldn't be confused with--the categories that States might decide to eventually regulate or control. Understanding and affirming the particular areas that States don't see as problematic will help to clear space and focus on those that might be--or where States might have some uncertainty. So starting from existing, widely accepted, highly automatic systems, as we “dial up” autonomy on different parameters (mobility, time of autonomous operation, etc.—we listed many of these in our first Observation Report), at which point do specific legal, technical,

operational or ethical concerns arise? Mapping these friction points in a more systematic way will bring much more focus to the international discussions. We acknowledge that this would be a complex discussion. However it would also allow, in the absence of a definition, the conversation to be narrowed down to potentially problematic applications of autonomy.

There is a second area where the whole is greater than its parts—increasingly autonomous systems working in concert with other increasingly autonomous systems. We need to not lose sight over how connected/interactive increasingly autonomous features might further attenuate human control or intent.

3 We see a real gap in many policy-makers' understanding in two relevant areas: Learning Systems and Cyber

a) A much deeper understanding of learning systems—their capabilities and their limitations—in necessary in order to ensure that we are developing sound policy. In this regard, the concept of predictability in learning systems needs much greater attention. Unpack what is meant by "predictability"--in the goal? In the subgoals? In the means of achieving goals and subgoals? More discussion on learning systems and legal reviews is also needed.

b) We will eventually need to consider that autonomous weapons are not limited to conventional weapons. At one point States will need to ask whether the concerns we have about the weaponization of increasingly autonomous technologies are also applicable to increasingly autonomous intangible technologies, such as cyber operations, particularly if these can have kinetic effects.

4 Check your blind spots—keep alternative developmental trajectories in sight

There are two ways to develop AWS:

- Building from scratch or adapting a remotely controlled or automatic system and empowering it to select and attack targets autonomously (for example through more capable sensors, programming, or processing power)
- Or weaponizing a civilian autonomous technology.

Most States interested in increasing autonomy are likely to use the first path. And discussions in CCW are likely to predominately focus on preparing for addressing this developmental trajectory. However, it is the second trajectory that will be more challenging to control and respond to--because it is more likely to take us by surprise.

Thus, we could usefully pay more attention to increasingly autonomous technologies in the civilian sector. In contrast to a gradual development of increasing autonomy under military control and oversight, it is possible that an actor takes a completely civilian technology and weaponizes it.

This developmental trajectory isn't limited to conventional forces (also non State actors, terrorists/extremists, criminals, individuals). This trajectory is characterized by:

- **Creativity**--civilian autonomous devices could be "weaponized" and used in novel ways that were "unthinkable" until they happen (prior to 9/11, it was mostly novelists, conspiracy theorists and screenwriters who imagined how commercial aircrafts could be used as a projectile in a mass casualty attack)
- **Being technologically accessible** –off-the-shelf, civilian tech, relatively low cost, difficult to control access to components. Open source Software Library for Machine Intelligence: Tensor flow, etc.
- **Having lower standards of reliability and predictability**—thus posing a much graver threat to IHL and Human rights as they are likely to be much cruder weapons.

This challenge is of course not unique to autonomous weapons--we struggle to effectively respond to IEDs for many of these same reasons.

So while this isn't the immediate focus of the discussions in CCW, one of the most important things we can be doing now is starting to consolidate norms about the weaponization of increasingly autonomous technologies. Norm development is already starting outside the CCW—consider the Open Letter on AI. This is another reason to step up the pace and focus of international discussions.

5 Put the horse before the cart—or in this case the human before the robot

We all agree that there are two sides to the autonomy issue--the technology side and the human side. In discussions on AWS, the tendency is to start with discussion of the technology.

Technology moves quickly, and humans move slowly. Any approach that attempts to predict or regulate the evolution of the rapidly moving fields of technology, computation, robotics and artificial intelligence will always remain--pardon the pun--a moving target.

To put it another way, the conversation has been set up in a reactive way--reacting to technology, whether existing, potential or imagined. We suggest that it shouldn't be the technology guiding the conversation. We need to shift to a proactive conversation on human control/judgement/intent/responsibility etc. And then apply that to technology.

Ultimately the autonomy question is really about what control/oversight do we expect humans to maintain over the tools of violence that we employ. In UNIDIR's first Observation Report in early 2014 we put it as follows "Rather than trying to agree upon rigid categories or definitions of thresholds of autonomy, in the initial stage of discussions, States might consider focusing discussion on identifying the critical functions of concern and the interactions of different variables. This would anchor the discussion and set its boundaries. It would also allow discussions to bypass—for the time being—getting bogged down into a technology-centric definitional exercise."

Later that year in our second Observation Report, we noted that focusing on the human side of the autonomy question rather than a technological framing of the issue has several benefits.

- It provides a common language for discussion that is **accessible to a broad range of governments and publics** regardless of their degree of technical knowledge.
- It focuses on the **shared objective of maintaining some form of control** over all weapon systems
- It is **consistent with IHL** regulating the use of weapons in armed conflict, which implicitly entails a certain level of human judgment and explicitly assigns responsibility for decisions made.
- It is a **concept broad enough to integrate consideration of ethics, human-machine interaction and the "dictates of the public conscience"** which are often side-lined in approaches that narrowly consider just technology or just law.

Reaffirming principles on human control/ judgement as well as getting down to work on developing shared understanding of how this applies specifically to the weaponization of increasingly autonomous technologies (how and when human control/judgement is exercised and what makes it meaningful or appropriate) is the urgent next step for the international community. Of course it isn't possible to have the human conversation in a vacuum--we will need to talk about technologies. However, we can use specific technologies to illustrate or test case these human principles, ensure that our values and principles continue to be embedded in decisions about weapon development, design and decisions about their eventual use.

As an additional benefit, this approach will be applicable to yet unimagined technological developments that we might consider to weaponize in the future. At a time when scientific understanding and technology develops at exponential rates, in surprising and non-linear ways, these shared principles will be increasingly necessary.

In conclusion, when approaching the AWS issue with a tech-centric framing there is a temptation to imagine AWS as something other than simply another tool for us to select in order to achieve specific strategic and operational objectives. We must resist this temptation as it clouds both the discussion and perhaps even our judgement. AWS will not be our peers, as they are not human. They will not be our fellow soldiers, no matter how well integrated they are in our military units. They are our tools.

The importance of your work in the CCW cannot be overstated. It will ensure that we do not cede--even unintentionally--our legal or moral responsibilities, nor our humanity, to an object--no matter how technologically sophisticated or capable it is. Putting the human side of the equation first –rather than the technology side--helps us to keep this fundamental distinction at the forefront of our discussions.

Mr Chairman, distinguished colleagues,

The question of autonomy will remain high on the international agenda in the coming years and UNIDIR will continue to be the UN system's thought leader on this topic: providing independent, evidence-based, policy relevant analysis to support Member States as you move forward in your discussions on the weaponization of increasingly autonomous technologies. Building on our work on framing discussions, meaningful human control, ethics and social values, and maritime autonomy, UNIDIR is currently examining how AWS intersect with cyber operations, how the developmental trajectory of learning systems and AI will influence this discussion, about underappreciated risks and uncertainties in the development and deployment of AWS, and considering different arms control approaches to dual-use technologies to ensure peaceful uses are not inhibited.

In this regard, I note that all of project's observation papers and audio files from public events are available on our website, www.unidir.org.

UNIDIR would like to thank Canada, Germany, Ireland, and the Netherlands for their investment in this work at UNIDIR. If your government is interested in supporting UNIDIR's programme of work on this topic, I'd be happy to discuss areas for collaboration with you at any time.

I wish you a productive week of discussions.

Thank you, Mr Chairman.