The Ottawa Definition of Landmines as a Start to Defining LAWS

Mark A. Gubrud
Submitted to the Convention on Conventional Weapons Group of Governmental Experts Meeting on lethal autonomous weapons systems • United Nations Geneva • 9-13 April 2018

A working definition of “lethal autonomous weapons systems,” LAWS, suitable for negotiation and treaty language, may be drawn from the way that the Ottawa Convention banning antipersonnel landmines defines those weapons in general:

‘Anti-personnel mine’ means a mine designed to be exploded by the presence, proximity or contact of a person and that will incapacitate, injure or kill one or more persons.

This definition can be parsed into two halves. The second part describes the lethal effects of the weapon on persons. The first part is more interesting. It describes the mine as being “designed to be exploded by the presence, proximity or contact of a person.” Rather than being triggered by their designated and accountable “operators,” mines are triggered by their victims.

In fact, if autonomy is defined broadly as acting without human control, guidance or assistance, mines should be considered LAWS. Legacy mines are extremely simple in comparison to robots and artificial intelligence, but there is no reason that more advanced landmine systems would not incorporate artificial intelligence to more accurately discriminate targets and perhaps tailor responses to varied situations. Surely these would be of interest as potential LAWS.

Building on Precedent

A 2005 recommendation of the Norwegian Ministry of Finance’s Advisory Council on Ethics1 established a precedent that a networked system with sophisticated sensors and weapons, intended to replace legacy landmine systems, would be considered Ottawa-compliant only if an operator is always required to trigger its detonation or other kill mechanism. If throwing a

---

1 Ministry of Finance (Norway), Advisory Council on Ethics, Recommendation concerning whether the weapons systems Spider and Intelligent Munition System (IMS) might be contrary to international law, Oslo, 20 September 2005
https://www.regjeringen.no/no/dokumenter/recommendation-concerning-whether-the-we/id419584/
“battlefield override switch” would enable lethal autonomy, as a fallback for when ‘stuff gets real,’ as they say, then the system would be considered a banned landmine system.

The same precedent must apply in any LAWS ban if it is to be effective in safeguarding peace. Otherwise the “ban” would only establish a norm that the most dangerous forms of autonomy may be turned on precisely when they are most dangerous — in a crisis or hot war.

The CCW LAWS negotiations may wish to exclude landmines, as already addressed by other protocols and treaties, but can they exclude any consideration of advanced, networked and AI-driven systems? In either case, we can start with the Ottawa definition of landmines and generalize it to define LAWS as victim- or target- or condition- triggered weapons, in contrast with weapons triggered by humans:

‘Lethal autonomous weapons system’ means a lethal weapons system triggered by a target or condition, rather than by a human operator.

The principal advantage of this definition is that it avoids unobservable concepts such as target selection and “critical functions” in the targeting cycle, and is framed instead in terms of the system’s observable behavior.

**Triggering**

This statement of the definition achieves brevity by loading a lot into its key concept: triggering. But this is a familiar, simple, intuitive concept.

When we trigger a weapon, we have already aimed it and know what we are aiming at, or at least, if we are acting responsibly, we believe we do. Aiming is not just telling a weapon what to hit, it entails a mechanical interaction, a control loop that includes primary data from senses and sensors – from which the shooter derives her internal representation of an acquired target. Human decision making is not just in the loop; it is the actual point of decision, the point at which everything hangs waiting until the moment the person pulls the trigger. And when we pull the trigger, we are deciding to take immediate action, with minimal delay. Triggering is an act of commitment, and while it remains an obligation, when possible, to rescind an erroneous command and interrupt an erroneous attack, in taking the action of triggering we are affirming the decision to attack and accepting responsibility for the consequences of that action.

Triggering initiates an action which has been prepared, specifically the action of a weapon. Triggering can only occur after targets have been acquired (or found and fixed if you prefer that lexicon), identified, and weapons aimed and armed. After triggering, a mechanical
process may unfold which may include trivial, technical decision points as conditions in a weapon's internal firing sequence are fulfilled. Weapons may even be allowed to control fire and timing within strict limits. However, this level of machine decision will seek only to execute and optimize attacks that have been triggered against particular acquired targets, and never to trigger attacks against newly-acquired targets:

A weapons system is 'triggered by a human operator' if attack is initiated only by an unambiguous signal normally generated by an action of the operator. An interactive confirmation process, which may include a short time delay, may be used to ensure positive control and accountability, but whenever the trigger signal shall be considered to have been confirmed, at that time, and without further delay beyond its internal limitations of speed, the system will commence to execute attacks on particular acquired targets or against specified locations known to the human operator at the time of triggering.

In taking the triggering action, the operator accepts responsibility for initiating attacks on the particular acquired targets or specified locations known to the operator at that time. This action will normally be pursuant to the authority of a human commander, who may be the same person.

When acting on the basis of identifying particular acquired targets with previously specified targets, groups or classes, such identification must be the responsibility of the human commander who authorizes attacks.

The system's execution of attacks triggered by a human operator may entail a programmed process which may include machine decision to sequence events in weapons fire and, within strict limits, to control aiming and timing of fire. Such machine decisions may be used to optimize execution of attacks the operator has triggered against particular acquired targets or specified locations, but may never include the triggering of attacks against any other targets or locations, including targets newly-acquired or re-acquired during the execution of triggered attacks.

If a weapons system is not triggered by a human operator, it will typically be triggered by targets or conditions. Such weapons systems are 'autonomous.'

The target identification and attack decision processes of autonomous weapons, their "critical functions" in the language of the ICRC, are not directly observable. However, their triggering by humans is observable and can be documented. Compliance with a commitment to accountable human control and triggering by human operators only can be
documented with cryptographic authentication for verification while protecting confidential information. This is the possible basis of a verified ban of LAWS.

**Targeting in the Fog of War**

The need for qualification of targets as “particular” and “acquired” adds a complication, but is needed in order to nail down the ambiguity of “target selection” and what it means for something to be a target.

An “acquired target” is one whose presence has been detected and is being observed in actual data deriving from the real world via senses and sensors. A target is “particular” if it corresponds to a particular person or object in the real world. Particular targets can be members of groups or classes of targets, but groups or classes are not particular targets.

*A ‘particular acquired target’ is a representation of a particular target object or person whose presence in the real world, in real time, is believed to be revealed by signals in sensor data and/or human senses.*

We may identify an acquired target with a previously specified target, something we think may be out there, that we are on the lookout for or actively hunting. Specified targets may specified in more or less detail, and their specification might include something about their behavior as well as targeting instructions and rules of engagement.

However, when a target is acquired, its identification with a specified target is a non-trivial decision. It is important to remember that all targets, as known by targeting systems, are only representations and interpretations. Identification of an acquired target with a specified target — drawn from some portfolio of enemy combatants, say — is always subject to possible error. Once an acquired target has been identified, its identification can be maintained on the basis of tracking only as long as acquisition is not lost and there is no possibility of confusing targets.

**Framing, Defining and Banning LAWS**

Under the approach advocated here, lethal weapons systems would have to be designed to only attack particular acquired targets or specified locations known to the commander authorizing and the operator aiming and triggering the weapon, or else they would be considered LAWS.

This would capture many things we may not want to ban, or to ban with a particular instrument. However, we don’t want things we haven’t thought of to be allowed by default.
Therefore we should agree that all LAWS are prohibited except for some types which we shall name and enumerate as either specifically allowed or not subject to regulation under this treaty. Some of these are obvious, such as defense shields against incoming uninhabited munitions. Others will be more contentious, such as some existing and evolving types of homing and loitering missiles.

Meaningful human control, including triggering by human operators, and accountable human control for verification, are the basis for this approach, but rather than attempting to specify exactly how humans should interact with machines—what information should be displayed and so on—we simply require that the particular acquired targets that the system will attack are already determined when attacks are triggered, and that in triggering attacks the human operator accountably accepts responsibility for authorizing those attacks on those particular acquired targets.

**Lethality**

Finally, it may be necessary to define the terms “lethal” and “lethality” for this context, since it is not acceptable to allow autonomous weapons to attack materiel targets, at least with destructive or damaging, kinetic force, which will inevitably also endanger humans. The CCW may still want to avoid stirring “cyber” warfare into this pot. But it cannot free robots to fight other robots with live weapons, as if robotic warfare would not become general conflict. Therefore the definition of “lethal” must here be something like:

*Lethal weapons systems apply physical force against objects or persons, with effects which may include impediment, harm, damage, destruction or death.*

It is most important to remember that this is only an attempt at a working definition of LAWS which can be the basis for an agreement to ban or regulate some or all of them. The approach does not seek to ban everything that falls under the definition of LAWS. But to avoid a narrow ban that could be circumvented just by changing one detail of a future weapon system so that it fell outside the definition, it bans all LAWS by presumption and enumerates exceptions for weapons systems, such as close-in defenses against uninhabited munitions, that may be considered desirable, or for existing systems that we agree, by negotiation, to grandfather in.

**Author contact:**

Mark Avrum Gubrud  
Adjunct Assistant Professor,  
Curriculum in Peace, War & Defense  
University of North Carolina, Chapel Hill USA  
gubrud@ad.unc.edu