

Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS)

Joint 'Commentary' on Guiding Principles A, B, C and D

submitted by

Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New-Zealand

The following joint general comments aim to contribute to the work of the GGE LAWS and are issued without prejudice to each State's national positions.

I. Introductory Remarks:

The 2019 Meeting of High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) saw the adoption, by consensus, of eleven guiding principles as affirmed by the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS). In its final report, the 2019 Meeting of High Contracting Parties also stated that the GGE LAWS is to consider those guiding principles which it may further develop and elaborate, and use them, among other elements, 'as a basis for its recommendations in relation to the clarification, consideration, and development of aspects of the normative and operational framework on emerging technologies in the area of lethal autonomous weapons systems'.

In our view, the guiding principles do not in any way constitute the end point of the work of the GGE. We consider them a useful and valuable starting point to build substance towards a normative and operational framework. Four principles in particular are of relevance to building substance towards a normative and operational framework. These are: guiding principles (a), (b), (c), and (d).

Although each of these four principles can be discussed on its own merit, they are clearly interconnected and, together with ethical standards, help to form a coherent approach to understanding and addressing the challenges posed by weapons systems based on emerging technologies in the area of LAWS.

II. Comments on Guiding Principles:

(a) International humanitarian law continues to apply fully to all weapons systems including the potential development and use of lethal autonomous weapons systems.

It is indisputable that all weapons systems must be developed, deployed, and used, in conformity with International Humanitarian Law (IHL). International Law encompasses the key requirements of state responsibility and individual accountability. These obligations entail that States and individuals are responsible and accountable for applying the law and are the ones that must be held accountable for violations.

Application of and compliance with key IHL rules and principles in the conduct of hostilities – such as the principles of distinction, proportionality, and precautions in attack, the prohibition of indiscriminate attacks, as well as the Martens Clause – require context-specific

value-based judgment by a human, which, with respect to emerging technologies in the area of LAWS, must not be substituted by autonomous machines or systems. Human control must therefore be retained in order to allow compliance with IHL.

The key question under guiding principle (a) lies thus in clarifying whether existing IHL is sufficiently specific to address issues arising from the potential use of weapons systems based on emerging technologies in the area of LAWS.

Questions indeed arise around the precise degree and nature of human control over weapons systems based on emerging technologies in the area of LAWS required for ethical acceptability and to comply with IHL rules notably, to limit attacks strictly to military objectives, to assess the civilian harm and military advantage expected from an attack, to refrain from launching a disproportionate attack, to take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event, to minimizing incidental civilian harm, and to uphold the principles of humanity and dictates of public conscience (Martens clause).

In the development of a normative and operational framework the following elements, *inter alia*, should also be considered:

- The necessity of sufficient predictability and reliability of the weapons system (cf. black box concern);
- The necessity of ensuring that the weapons system responds as intended by the developer and user to the operational specificities of pre-planning and dynamic targeting;
- The avoidance of data bias and programming shortfalls in complex systems;
- The necessity of ensuring the weapons system's adaptability to a change in circumstances, including the possibility to cancel or suspend an attack including if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated;
- The necessity of ensuring that weapons reviews are conducted with a full understanding of the weapons' capabilities and limitations, and sufficient confidence about its effects in the expected circumstances of use.

To address these specific issues, an appropriate normative response requires a transparent, process-oriented framework based on a set of criteria for evaluation. Such a framework could help to ensure that weapons systems based on emerging technologies in the area of LAWS are developed, deployed, and used in full conformity with International Law and, in particular, IHL. The criteria upon which this framework would be based are detailed below in our comments on guiding principle (c) on human-machine interaction.

Weapons reviews, including art. 36 reviews, will continue to play an important role in weapons development. Nevertheless, in the evaluation of weapons systems based on emerging technologies in the area of LAWS, key challenges in the regulation and the nature of the systems should be considered, through a regular evaluation process, which should

take into account the criteria detailed under guiding principle (c) below and be applied across the life cycle of a weapons system.

(c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole.

Guiding Principle (c) on human-machine interaction is a key principle to build substance for a future normative and operational framework on weapons systems based on emerging technologies in the area of LAWS. In our view, one of the main tasks of the GGE LAWS will be to elaborate a common understanding of the type and degree of human-machine interaction that will be needed to ensure compliance with International Law and, in particular, IHL.

Human-machine interaction provides an entry point for setting out the building blocks of human control over such weapons. It recognises the necessity of retaining human control over the weapons systems and is a critical element in ensuring that there is no accountability gap in the design, development, deployment and use of weapons systems based on emerging technologies in the area of LAWS. Human control, responsibility and accountability are also intrinsically linked to the important ethical and moral considerations that should form part of the GGE's work. Fundamentally, guiding principle (c) on human-machine interaction should take into account that human control over the critical functions of such a weapons system requires control throughout the life-cycle of the weapon.

Guiding Principle (c) recognises that human-machine interaction may take various forms and be implemented at various stages of the life cycle of a weapon. It also states that a range of contextual (operational context) and technical considerations (characteristics and capabilities of the weapon) should be considered in determining the extent and quality of that interaction.

As mentioned in section II(a) above, the process-oriented normative and operational framework should therefore be based on the following criteria:

Contextual considerations :

- a) Whether the weapons system is capable of reading the operational context correctly and whether it demonstrates a sufficient level of situational awareness (i.e. its ability to adequately perceive and react to changing circumstances). These elements should be made sufficiently transparent to the human agent;

Technical considerations:

- b) Whether adequate limits on tasks and types of targets are in place to allow the weapons system to be operated with sufficient degrees of reliability and predictability in the identification, selection and engagement of targets;
- c) Whether adequate environmental limits, including spatial and temporal limits, are in place to ensure that the decisions, made at the planning stage, including legal assessments, are respected throughout the execution stage;

Forms of human-machine interaction:

- d) Whether meaningful human control is exerted and retained over the critical functions of a weapons system - i.e. in the identification, selection and engagement of targets - to ensure the necessary context-specific value judgment required in the application of IHL rules and principles;
- e) Whether the degree of human control allows for human supervision and intervention, where adequate, in order to prevent redefinition of the weapons system's mission without human validation and to interrupt or deactivate the carrying out of autonomous functions if needed.

In order for weapons systems based on emerging technology in the area of LAWS to be operated in conformity with IHL, the following three challenges need to be considered when designing, deploying and using such weapons systems:

- f) Cognitive limitations of the system (lack of common sense and human judgement);
- g) Epistemological limitations (i.e. the system making judgments based on data that are biased, incomplete, or not fully appropriate to the situation);
- h) Algorithmic bias.

(b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system.

AND

(d) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control.

While acknowledging that they are distinct principles, given the commonalities that exist between guiding principles (b) and (d), and in an attempt to avoid repetition, this section of the joint-commentary will address principles (b) and (d) together. Both guiding principles are essential to build substance for the normative and operational framework on emerging technologies in the area of LAWS.

Given that human responsibility and accountability cannot, under any circumstances, be transferred to machines, a normative and operational framework will need to reflect that human responsibility and accountability are maintained throughout the entire life-cycle of any weapons system based on emerging technologies in the area of LAWS.

During discussions at the GGE, the terms ‘responsibility’ and ‘accountability’ have, at times, been used interchangeably. Yet, it is important to recall that they refer to related but distinct concepts. ‘Human Responsibility’ can be considered as encompassing moral and ethical considerations as well as legal obligations and expected conduct. ‘Accountability’ can be considered to relate to legal liability and legal consequences. The concepts are mutually reinforcing with clear and distinct lines of responsibility improving the accountability and attribution process.

International Law, particularly International Humanitarian Law, International Human Rights Law, and International Criminal Law, including the rules of attribution and responsibility applicable in a given case, apply fully to any weapons system. Thus, in our view, both human responsibility and accountability apply throughout the design, deployment and use of any weapons system.

In terms of scope, it is also important to stress that in International Law, responsibility and accountability apply at State and individual levels. A human chain of command and control must always be ensured during the deployment and use stages of the life cycle of such weapons systems.

The starting point in interpreting guiding principles (b) and (d) is that such systems must not be designed, deployed or used without a clear line of responsibility and full accountability. This highlights the necessity of maintaining human control of the systems to ensure responsibility and accountability, and underscores the importance of developing a common understanding of guiding principle (c).

A key issue reflected in guiding principles (b) and (d) is the recognition that an increasing level of autonomy in weapons systems may pose challenges in the attribution of conduct to individuals and holding them to account. Issues including, but not limited to, mens rea, recklessness, negligence, or misconduct in the deployment and use of weapons systems based on emerging technologies in the area of LAWS may be particularly difficult to assess.

The ultimate goal in including guiding principles (b) and (d) in a normative and operational framework is to prevent any ambiguities or inconsistencies in the attribution of responsibility and accountability that may arise from the design, deployment or use of such systems since any ambiguity would increase the risk of impunity and undermine confidence in the efficacy of the framework. It is essential that responsibility for the use and for the consequences of the use of a weapons system can be clearly assigned.

With advances in artificial intelligence and machine learning, predictability is another complicating factor that merits further attention. For instance, it may be necessary to take precautions to ensure that a weapons system is not capable of changing certain mission parameters without human validation. It must also be ensured that commanders and

operators are informed about any new characteristics, functions and parameters of weapons systems and are trained accordingly before the deployment or use of such systems in the field.

A State that deploys or uses weapons systems based on emerging technologies in the area of LAWS must and will be accountable for the consequences of its use. Including these principles in the normative and operational framework aims to ensure that relevant actors, particularly those in the chain of command, have sufficient understanding of weapons systems based on emerging technologies in the area of LAWS under their control.

III. Concluding Remarks:

Considering all the issues developed above, we are of the view that a normative and operational framework should ensure that human control is exerted and retained over critical functions of any weapons system based on emerging technologies in the area of LAWS.

The nature and degree of human control may vary during the life cycle of a weapons system. There is also no accountability without human control in all phases such as design, development, deployment and use of any weapons systems.

Human control, responsibility and accountability are furthermore intrinsically linked to the important ethical and moral considerations that should form part of the GGE's work and apply across the development, deployment and use stages of a weapons system. In developing a normative and operational framework, it will be necessary to consider the implications of guiding principles (b) and (d) with respect to developers and manufacturers as well, as they bear responsibility in the design and programming stages of the weapon. This is particularly relevant for issues related to data bias, which can impact targeting, and malicious or careless programming.
