Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System
Geneva, 7-11 March, and 25-29 July 2022
Item 5 of the agenda
Consideration of proposals and elaboration, by consensus, of possible measures, including taking into account the example of existing protocols within the Convention, and other options related to the normative and operational framework on emerging technologies in the area of lethal autonomous weapon systems, building upon the recommendations and conclusions of the Group of Governmental Experts related to emerging technologies in the area of lethal autonomous weapon systems, and bringing in expertise on legal, military, and technological aspects

Group of Governmental Experts (GGE) document on the application of International Humanitarian Law to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)

Submitted by the United Kingdom

I. Introduction

1.   CCW High Contracting Parties have long agreed the need for substantial progress in the area of Lethal Autonomous Weapons Systems. Central to this is the need to increase understanding and agreement of the ways in which systems with autonomy can be developed and used ethically, responsibly and in compliance with International Humanitarian Law (IHL).

2.   The Group of Government Experts on emerging technologies in the area of Lethal Autonomous Weapons Systems is mandated by the Sixth Review Conference of the CCW to "consider proposals and elaborate, by consensus, possible measures, including taking into account the example of existing protocols within the Convention, and other options related to the normative and operational framework on emerging technologies in the area of lethal autonomous weapon systems, building upon the recommendations and conclusions of the Group of Governmental Experts related to emerging technologies in the area of lethal autonomous weapon systems, and bringing in expertise on legal, military, and technological aspects." The Review Conference further agreed that "international law, in particular the United Nations Charter and international humanitarian law, as well as relevant ethical perspectives should guide the continued work of the Group."

3.   In accordance with the mandate of the GGE, this paper proposes the elaboration of a document, or manual, that would constitute an authoritative and comprehensive statement of the application of International Humanitarian Law and agreed best practice.

4.   The object of this proposal is informed by similar authoritative texts which apply a substantive and contextual application of legal principles and norms of practice, in their respective areas of applications, such as The Montreux Document on Private Military and Security Companies, The Tallinn Manual on the International Law Applicable to Cyber

Please recycle

Operations, and the San Remo Manual on Armed Conflict at Sea. We believe that not only is this a realistic, important and deliverable output for the GGE, it would assist with putting high level principles into practice. It would have a tangible impact on the development and use of future weapon systems and would be an effective means to address concerns associated with the integration of autonomous functions within weapon systems, while preserving space for the legitimate development of AI.

This proposal is without prejudice to other proposals including any proposed additional principles or good practices or other proposals to be considered by the Group of Governmental Experts.

II. The Document - Possible Structure and Contents

5.   It is proposed that the GGE commission a document that sets out guidelines, advice and best practices on how states should approach the development and use of emerging technologies in the area of LAWS at each stage of its lifecycle. This would include an assessment on any characteristics which would be necessary for compliance with IHL and any characteristics which would be incompatible with IHL.

6.   The document might include examples that illustrate what good practice is or might look like. This would provide a clear framework for the operationalisation of the guiding principles and actionable guidance for policy, technical, and military stakeholders. It would thereby encourage the adoption of national regulations designed to strengthen respect for international law.

7.   A suggested outline for the document is included in the annex to this paper. It does not attempt to suggest conclusions or add detail beyond each area, which would be for High Contracting Parties to discuss and agree together, in accordance with the GGE's mandate. The proposal should therefore be viewed as a handrail to help guide further discussions in the most fruitful way possible, by energising parties and engineering a path towards the creation of an authoritative and comprehensive statement of the application of the law and agreed best practice. With that in mind, the headings, questions and potential examples within it should not be viewed as exhaustive or definitive – we submit this to the GGE in the hope that together we add additional aspects as well as the granular detail behind them.

8.   Best practice and commentary may also address questions of ethics in an applied and context appropriate manner. This prevents the 'ethics issue' becoming an intangible catch-all which defies inclusion in either consideration of the legal framework or as a matter of practice. Rather it frames, as far as possible, ethical concerns at each stage of the weapon's lifecycle and informs best practice on the basis of the concern. An ethical concern in relation to research methodology and an ethical concern in relation to command accountability are two entirely different concepts, both of which can benefit from a conceptual application through development of best practice guidelines.

II. Way Forward

9.   This document should be owned and completed by the GGE – it should be a reflection of the agreements of those parties who sit within it. Completion of it will be a significant task; some potential ways the GGE may wish to progress are listed below. Again, these are not exhaustive.

- The creation of a sub-group, or series of sub-groups, within the GGE to work on specific areas within the desired document. Membership would be made up of parties to the GGE on a voluntary basis and should involve a diverse range of GGE parties. A group per heading may work, for example. Any new group would report to the GGE who would then be required to agree the findings and recommendations of the group before information is included in the document.

- Discussions solely at GGE meetings. This would severely restrict the speed at which such a document could be completed.

- Voluntary contributions by parties on each section, to be submitted ahead of GGE meetings and then discussed at each meeting. The outcome of the GGE meeting would be to determine which information could be agreed and should form part of the document. Again, this would restrict the speed at which such a document could be completed.

- The creation of an independent sub-group to report into the GGE on the real-world impacts of the use of autonomy within weapons systems. The make-up of this should include experts from legal, technical and academic areas as well as industry. Its numbers should be capped (for example, at 15-20) with people appointed by the GGE. Those within the group would not act on behalf of the GGE (and therefore be independent of their state, if applicable). The GGE would not be bound by its findings but could use them to reach conclusions of its own.

Annex A - a possible outline

1.   The following is one possible outline for an IHL Manual as proposed by the UK, by way of illustration of the overall concept. It is based on eight headings, many of which reflect the concept, development and use cycle ('lifecycle') of a weapons system (adapted from the Sunrise diagram from the 2018 report of the GGE). Within each of these headings is a series of sub-headings and questions, which when answered or expanded upon would increase understanding of the ways in which – at each stage of a system's life – compliance with IHL can best be assured. Human-machine interaction, which is relevant to all aspects, is included throughout.

2.   We recognise that agreement on each element may not be possible – where this is the case, such a document could include a collation of the potentially different methods of good practice, as long as those methods are designed to enable overall compliance with IHL.

The purpose of the document

3.   It is expected that this section would contain a statement as to the aims of the document, how it has been developed and the expected use that states would make of the contents.

International Humanitarian Law

4.   A capping section to introduce IHL and its overall applicability. This could involve a clear summary of applicable law and explanation of accountability covering:

- Basic Rules of International Humanitarian Law: The object of this section is to re-affirm the relevance and application of the core principles and rules of IHL. These principles and rules have general and continuous application in relation to the behaviour of parties to armed conflict. No advanced method of warfare or autonomous weapons system permits the derogation or relaxation of the rules of IHL and it is imperative that states and parties to armed conflict ensure that their conduct conforms to these requirements in all circumstances and no matter what means or method of warfare is adopted.

- Distinction: Parties to an armed conflict must at all times distinguish between the civilian population and those hors de combat and combatants, and between civilian objects and military objectives and shall direct their operations only against military objectives. The exercise of distinction requires the ability to observe, recognise and exercise situational judgement. The deployment of an autonomous system in a manner that does not have regard to these requirements is unlawful.

- Proportionality: Parties to an armed conflict are prohibited from launching an attack that is expected to cause incidental harm to civilians that exceeds the direct military advantage anticipated from the attack. This principle of proportionality requires the individual taking a decision to attack to appreciate of the context and object of the attack before it is possible to assess the legality or the illegality of the action. The application of the principle requires qualitative, subjective and strategic appreciation of the military advantage and the expected impact of the attack.

- Necessity: The principle of military necessity permits measures, including measures and weapons which engage autonomous functions, which are necessary to accomplish a legitimate military purpose and are not otherwise prohibited by international humanitarian law.

- Humanity: Parties to an armed conflict are prohibited from the infliction of all suffering, injury or destruction which not necessary for achieving the legitimate purpose of a conflict. Compliance with the principle of humanity

requires judgement and understanding as to the nature of human suffering and any use of an autonomous weapons system which fails to have regard to this requirement for judgement and understanding in a manner that is able to balance this against the legitimate purpose of the conflict is unlawful.

- Martens Clause: The right of parties to conflict to choose means and methods of warfare is not unlimited, IHL instruments state that in cases not specifically covered by the provisions of its instruments, the civilian population and the combatants shall at all times remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.

- Feasible precautions: In the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All feasible precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians and damage to civilian objects. This obligation falls on persons who plan or decide upon an attack, autonomous systems may be engaged in realising this obligation, but the obligation cannot be divested onto the system.

5.   The rest of the document would also make reference to the requirements of IHL but within the more specific context of each of the sections. This introductory piece should demonstrate that the core, binding legal framework applies to LAWS and does not allow derogation on the basis of the type of weapons system.

Political control: national policies and direction

6.   This would detail good practice for decisions on the overall requirement for autonomy in systems. It revolves around the decisions made by politicians, policy makers and military leadership. Potential areas for further information could include:

- How is political and policy direction determined?

- How is public opinion considered?

- When overall political direction is provided, how should states develop their policies and strategies?

- How should policies be tested or challenged; when and by whom?

Research and Development

7.   This section would focus on capability planning and how the activities conducted to research, develop and acquire capability can enable (or prevent) compliance with IHL. Potential areas for further clarification could include:

- Requirement setting: how are requirements properly set?

- How do project and programme teams enable full understanding of the goals?

- How are they empowered to challenge?

- How are necessary constraints identified and implemented?

- How is accountability at the point of use being considered?

- How is compliance with requirements documents tested and ensured?

- Human understanding: what levels are acceptable and how could it be measured or assessed? Could an end user understand the concept for use? What level of understanding is required by each individual within the authority chain of weapon deployment?

- Industry and civil society involvement: how are industry and civil society brought in to understand goals, risks, policy and legal parameters?

- Waypoints: how might these be used as checks along the entire lifecycle to ensure that requirements are being met?

Testing And Evaluation

8.   This section would set out what might be involved in the robust assessments of a system, and what practices help to provide confidence in its performance within the intended context of use. Acknowledging the through-life process, activities within this section would focus on compliance with the requirements documents, and legal, regulation and certification requirements. Potential areas for further clarification could include:

- Testing frameworks: What kind of testing is needed? How robust must measures be? What is an acceptable fail rate? How might this vary and why?

- Requirement documents: How is compliance with requirements documents tested and ensured? Who should do this to ensure independence?

- What happens if a system fails a test?

- How can bias and unintended harms, whether through their original rollout or as they learn, change and are redeployed, be identified and eliminated?

- How is risk best calculated?

- How are systems best regulated or certified?

- What is good practice in identifying and implementing appropriate constraints?

- How is complexity best managed? What levels of understanding are acceptable? Would we need to consider T&E at a system of systems level as well as at the individual system level?

- How is accountability at the point of use being considered?

- How might Machine Learning "in the wild" impact on T&E

- How might adjustments that are made by systems integrators and maintainers during use impact on T&E validity?

Training and planning for use

9.   Linked to the above section, this would set out how states might best ensure a system is understood by those who would operate it. Potential areas for further clarification could include:

- How is training developed? How do varying levels of autonomy change this at both individual and collective levels?

- How does training prepare the force for the use of AI enabled systems under the most stressing scenarios (this might include situational complexity, uncertainty, high workload, high tempo etc)?

- How does training address the procedural, C2 and safety precautions that support compliance with IHL?

- How is complexity managed? What levels of understanding are acceptable?

- How does this enable adequate understanding of a system and allow calibration of appropriate trust? How is 'adequate' determined?

- How are Rules of Engagement best drafted?

- How are unforeseen circumstances mapped?

- How are necessary constraints identified and implemented?

- How might different interactions (based on geography, time) be accounted for?

- How is accountability agreed and tested – how does this vary depending on the system, its use or use within different scenarios?

Article 36 Weapon Reviews

- In the context of autonomous systems what information and level of understanding is necessary to inform an effective weapons review?

- Does the inclusion of AI functionality make a difference in relation to the scope of the review or the resulting authorisation?

- Does machine learning necessitate re-review and authorisation? If so how is this built into the review process and operationalised to ensure that the system does not exceed authorities?

- How should the approval parameters applied to the system as a result of a weapons review be best effected through the authorisations process and through into Rules of Engagement (RoE)?

Deployment and use

10. This section would set out agreements and good practices for how operators and commanders can ensure compliance with IHL when using systems with autonomy in operations. Potential areas for further clarification could include:

- How might the implementation of the system as part of a wider system of systems impact on compliance with IHL?

- How might incorporation of autonomous weapons systems as an element within command and control structures impact on compliance with IHL?

- How might potential use of the system be incorporated into mission-specific RoE?

- How might varying levels of autonomy affect target identification, the application of RoE and the assessment of proportionality?

- How accountability and responsibility are determined, agreed and set out – how this changed from previous conclusion earlier in the cycle?

- Are operational limits or parameters needed – how are these best determined and defined?

- When might commanders wish to abort the use of such systems? What considerations should be taken into account? How should this be this handled in practice?

- How are decisions as to the impact of the specificities of combat in different locations (geographic, temporal) considered?

- How is situational understanding passed between human and system?

- Where must human control be present or absent, or what degrees of control are necessary

Analysis

11. This section would focus on what should happen after a system has been used – was it used and operated as intended and was accountability possible. Potential areas for further information could include:

- How is specific battle use assessed and analysed? How is damage assessed?

- How are lessons identified and learnt?

- How is feedback provided?

Human-Machine Interaction

12.   This would be an overarching theme throughout the whole document. The document would need to address the level of human involvement is necessary to achieve the IHL ends. How does this involvement vary across different systems and in different environments in order to meet the best outcomes in terms of applying the principles of IHL? Equally what are the implications of autonomous systems for chains of accountability and the need for humans to be accountable for IHL outcomes?

———————————