Group of Governmental Experts of the High Contracting Parties to the 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

Geneva, 13–17 November 2017
Item 6 of the provisional agenda
Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

Food-for-thought Paper

Submitted by the Chairperson

Introduction

1.	The Group of Governmental Experts (GGE) is mandated by the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) to examine issues related to emerging technologies in the area of lethal autonomous weapon systems in the context of the objectives and purposes of the Convention. The CCW GGE is today the only inter-governmental forum in the United Nations system formally seized of the matter, bringing together relevant inter-disciplinary insights and expertise. For the CCW process to be of value and continued relevance it is important that there is an assurance of continued financial support to the meetings of the GGE and that all High Contracting Parties and other stakeholders, including industry, contribute to the discussion.

2.	Given the curtailment of the time devoted to the GGE for 2017 from 10 to 5 days, the Chair believes that the GGE can make a good start this year by conducting a thorough review of the current state of the developments in the technology domain and their instantiation in specific military systems. The GGE could also continue the useful discussion of the past three years on legal and ethical issues of relevance to LAWS, including emerging developments in national law as well as the corporate domain to regulate the use of civilian autonomous systems such as self-driving vehicles. A thorough three-fold review of technology, military effects and legal/ethical considerations would help the GGE move on to the next stage of discussions, which could focus on framing devices such as definitions and other concepts with the potential of narrowing the line of sight to policy pathways.

3.	With a view to stimulating discussion and assisting delegations in preparing for the GGE, the Chair offers the following questions grouped under three broad categories. These questions are meant only to help start a substantive review of each of the three categories. They are neither exhaustive nor fully representative of all views on the subject; in particular certain policy-related questions that belong to the next stage of discussions have been kept aside for now.

Please recycle

Technology

- What are the technologies that are contributing to or could contribute to lethal autonomy in weapon systems? Could they be broadly characterized as artificial intelligence/autonomous systems (AI/AS)?

- What civilian uses are they being put to or could be put to?

- Are autonomous systems best visualized as physical robots (discrete) or virtual machines (spread out information processing systems)? Could there be alternate technology pathways (neuromorphic, human enhancement for example) that the LAWS discussion has not captured?

- What are the various dimensions of autonomy seen in the current suite of AI/AS technologies? Where are we on these dimensions (energy autonomy, computational autonomy etc.) in comparison to full autonomy (even if idealized)?

- Has there been a shift in recent years (Google's cat-face detector and DeepMind's DQN algorithm based Atari 2600 2D game player, chatbot Eugene Goostman passing the Turing test, Alpha GO etc.) or is there likely to be a shift from specific applications (narrow or weak AI) to General Intelligence (AGI) leading to an eventual "intelligence explosion"/"singularity"/"super intelligence"(ASI)?

- How are existing systems verified (was it built right) and validated (was the right system built)? Are existing and planned autonomous systems scrutable (what do you know and how do you know it)? Can machines describe their learning?

- How are human and social safety issues (for example hacking and privacy) being tackled by industry? Can autonomous machines be made foolproof against hacking?

- Can there be software/hardware locks on machine behaviour and could a learning machine be prevented from bypassing/changing them?

- Does the transformative character of AI and its possible ubiquity limit the LAWS discussion in any manner or is AI like other dual-use technologies in the past?

Military effects

- What are the specific areas where autonomous technologies are being deployed or are likely to be deployed? Logistics? Intelligence, Surveillance and Reconnaissance (ISR)? Border and area defence? Could the manner of their deployment lead to LAWS through mission creep or otherwise?

- Are there specific domains (such as the maritime surface and subsurface domains) where potential deployment of LAWS is more likely and why?

- Which would LAWS impact more, offensive or defensive capabilities? Could the potential deployment of LAWS lower the threshold of use of force? Could it enhance asymmetric deployment of force or covert use of force? Are there arms race and stability considerations of relevance to the CCW?

- Could potential LAWS proliferate/learn to act in conjunction with terrorists and other unlawful non-state actors?

- Are there specific military applications (Explosive Remnants of War disposal, demining etc.), which could be desirable even in the current CCW context?

- Could potential LAWS be accommodated under existing chains of military command and control? Are there specific doctrinal issues unique to LAWS?

- How do potential LAWS fit with existing notions of transparency in armaments with regard to deployment, holdings, trade or transfers?

Legal/Ethical issues

- Where does legal accountability and liability reside for existing or planned autonomous systems? With the planner-developer, the legal owner, the user and/or the machine?

- What are the main features of national or regional laws planned or already in place for the regulation of autonomous systems such as driver-less cars, entertainment or companion robots and chatbots?[1]

- Are there existing codes of conduct or industry standards that can have relevance for the LAWS discussion?[2] What is of relevance to the CCW in these reference documents?

- Are there regulatory insights from other domains such as genetics?

- Could international humanitarian law developed for human and State-controlled behaviour continue to apply *mutatis mutandis* to potentially autonomous machines and through which mediatory mechanisms?

- How do potential LAWS fit with existing national weapons review prior to development, testing and deployment?

- Are there legal gaps in regard to potential autonomous systems in other international policy domains such as trade (Rules of Origin etc.) and intellectual property?

- Can machines be truly intelligent in the sense of humans (phenomenally conscious, intentional, creative, empathic, evolutionary, free agents with embodied intelligence)?

- What system of ethics or specific ethical values could apply to machines? How can different views of consciousness and human-material world interaction in different ethical/theological systems be reconciled in a universally applicable set of injunctions for learning machines?

- Ethics/morality related concerns have focused so far on machines taking life. What about the human-machine pair acting collaboratively or human enhancement?

- How can these ethical injunctions be coded? Can a machine become a moral being?

---

[1] For example, Republic of Korea's "Intelligent robots development and distribution act" and the European Union's Parliament Committee on Legal Affairs' report (2015/2103(INL).

[2] Such as the Institute of Electrical and Electronics Engineers (IEEE)'s P7000-7002 under the Global Initiative for Ethical Considerations in Artificial Intelligence & Autonomous Systems and the United Kingdom's EPSRC-AHRC Principles of Robotics.