InterAgency Institute

BEYOND INSTITUTIONAL BOUNDARIES

Report on the

LUIS CAMPANI, RESEARCH AFFILIATE

MAY 2023

On 25 and 26th of April 2023, members of the military, institutions, private sector, field experts and of organized civil society gathered in the south of Luxembourg to discuss the regulation on autonomous weapons systems. The initiative was created by Luxembourg's Directorate of Defense from the Ministry of Foreign and European Affairs. The event was composed of a keynote speech by Paul Scharre on the first day, and five panels on the second, each approaching a different aspect of the debate: politics, defense, regulations, technical and ethics. The event brings to the table the opportunity to fulfill the necessity to include other stakeholders in the conversation, such as the private sector and technicians.

The keynote speech by Paul Scharre started with a strong message: how can an algorithm understand the nuances that go into discrimination between civilians and non-civilians in a war context? While humans can engage with a wide array of behaviors with one another, LAWS are not able to use other means of engaging other than killing its targets. In order to succeed in the regulation of autonomous weapons, we must have clear lines about what is permitted and what is not; States must be able to comply during wartime and we must be able to verify if states are compliant. In this sense, the author highlighted the importance of reciprocity on the prevention of the dissemination of these weapons. Reciprocity in this case is the fear of equal force retaliation by another party, the threat of which acts as a deterrence mechanism for the engagement of this technology.

In the political debate, the most frequent theme was momentum. The understanding that the lengthy procedures that policy making requires – specially when requiring consensus from 126 different states, such is the case of the CCW – are now gaining more traction and attention from the international community, not only with the Luxembourg conference, but also the conference in San José, Costa Rica, that led to a consensual document ratified by 33 countries, the Belén Communiqué, and the following five working papers presented at the CCW in march. These initiatives are a sign that there is a general agreement about the importance of the issue, the necessity of necessity for clear and easy to apply redlines, regulating and constraining the use of autonomous weapons systems, and meeting the need to create accountability for the use of these systems is a key factor in these regulations.

Having the necessity for a legally binding instrument and to include meaningful human control in the technical requirements to these systems been pointed out, in the spirit of Belén, the panel stated the necessity to include States that are not militarily and technologically advanced in the conversation and also stated the necessity of sitting down for negotiations even though we don't have all the answers, in this sense, to seize the opportunity we still have to take prophylactic measures instead of waiting for a catastrophe to start regulations, culminating on the present urgency to construct an ethical framework that establishes clear red lines for the development of such systems.

WWW.INTERAGENCY.INSTITUTE
CONTACT@INTERAGENCY.INSTITUTE

This momentum, if seized, can introduce the discussion on the regulation of LAWS in other forums, such as the European Parliament, that has left the military use of artificial intelligence out of the IA Act.

The military panel debate focused on the interoperability challenges, and why systems with these functions are so enticing to armies worldwide. Regarding the challenges, a common denominator from the panelists was the necessity of a clear and common taxonomy regarding LAWS and a common set of rules for certification and red lines were a common denominator throughout the panel, on the other hand, why is the development of such functions of interest to armies worldwide was also addressed. In the army, autonomous systems are seen as a force multiplier, and the importance of developing such functions as a competitive edge exploiting the global north's comparative advantage regarding technological intensive military equipment and its reduced population. Security dilemma plays a big role in the reasons for the development of autonomous weapons: in the military, nobody wants to get behind in the tech race. Being ahead of such an enterprise means investing heavily in Research & Development of such projects and was considered to be a countermeasure for the development of laws by other countries.

A distinction was made between automated and autonomy in weapon systems. In this case, the first means the system follows preprogrammed rules to a desired outcome, while the second, that it learns from the environment, can lead to unpredictable outcomes. The distinction, though not evidently, masks the fact that in the moment of deployment, the agent that deploys an automated system is not aware how, where, when and who the automated system will engage with. Regarding the regulations, even though no mention of a legally binding document was made, the necessity of clear and applicable red lines was a consensus between the panelists, as was the necessity that all systems with automated functions should comply with international humanitarian law. Meaningful human control was considered a key, even though a definition of the term was not given in this specific panel, both the principle of human in the loop and Artificial Intelligence deciding if someone lives or not, that is, automated engagement, was stated as a red line not to cross.

It was pointed out that developers should know the rules and laws are limiting the development of their systems, and the necessity of speeding up the definition of the system.
The regulatory panel was consensual in identifying the imminent necessity of a legally binding instrument as an end goal for negotiations. It mentioned the importance of beginning negotiations and setting standards, even though we don't have all the answers. LAWS were specified as a functionality in weapons systems, that is, a clear definition was given: LAWS are considered weapons that have functionalities to select and apply force to targets without human intervention. It was said that leadership in the regulation of LAWS is coming from the Global South, as can be seen in the CCW and the Belén Communiqué.

While in the international community there is a common recognition of the problems and risks and we are witnessing progress in States positions, the likelihood of some States opposing starker stance against LAWS induce States that have a democratic understanding of the issue to offer diluted positions to appease States that might oppose them, making it even more difficult to overcome the roadblocks. The panel was consensual about the problematic of the ethical dimension, leaving life and death for a machine to decide is something to be avoided, and it was said that by withdrawing the human-in-the-loop and replacing a machine as the ultimate decision-maker would be undermining the experience of military officers.

Meaningful human control was defined as a scenario where the operator can predict the effects of the attack, understand what's going to happen and explain what the system is actually doing. It was suggested that autonomous engagement in weapons that target humans directly should be prohibited, therefore, the prohibition of automatic engagement in anti-personnel weapons. It was stated that the regulations should follow a two-tier approach consisting in prohibition and regulations, therefore it doesn't mean that everything within that scope of autonomous weapons will be prohibited, but the goal of the regulations is setting clear red lines of what is or is not possible to make within a defined framework that capture all humanitarian concerns and meet the functional approach criteria in order not be outdated. Even though technological increment is a cross-sectorial trend, specially in the military context it is imperative to understand and mitigate the risks created by the development of these functions.
Potential escalation of conflict due to unpredictable ways LAWS might react in operation, both particular systems and in system interoperability. Moreover, the issue of lethal autonomous weapons systems goes beyond the battlefield, where there is a wide array of nuances that require human judgment. With the development of such systems the possibility arises that States without regard to international humanitarian law are able to develop them, or non-State actors to gain access, and the current debate on the CCW is not yet able to capture this wider aspect of externalities. If the technology is put in practice before the legislation, it sets a precedent for what meaningful human control will look like to govern the use of these systems.

With the unparalleled pace of technological development, the technical panel comes to answer fundamental questions to the debate on LAWS, how can a norm be implemented in a machine learning model? How can we translate the moral concerns into requirements? Is it possible to confirm that a system complies with the regulations? There were identified an array of threat categories for AI models, such as: privacy & data, supply chain, adversarial machine learning. Also, there are operational risks: cyberattacks and electronic interference. A big part of the risk assessment discussion was given to cyberthreats. For that, it was said that you need to consider the whole pipeline of the Software Development Life Cycle (SDLC), and not only the system itself, but all other support infrastructure for the system. It was stated that due to its autonomous nature, LAWS posses a low attack surface, there are less entry points to attack while it is operational, therefore, an identified externality were advanced persistent threats, that is, an attack the system at the start of its SDLC, introducing malware before the system becomes operational.

That can be exploited via Open Source software, for example. It was also said that if there is a lack of communication between a human and the machine, there is less chance of detecting malware while the system is operational, for this reason, vulnerabilities in autonomous weapons systems are emphasized because a human can't detect if something goes wrong. On the other hand, at the tactical level, it is much easier to contain drones with kinetic action than with cyberattacks. It was also stated that there are going to be tradeoffs between security and operational advantage, such as, not using encrypted communication.

With a wide array of newly developed AI, we have distributed AI, how to verify such distributed models? What are the costs? Will it be feasible? A lot was said about Certifications and Test, Evaluation, Verification and Validation (TEV&V), that is, in the operationalization of the TEV&V, there should be an approach to the full SDLC. Regarding the verification, ensuring that a software is compliant with the functional (how input should be processed into output) and non-functional (how a system should be, in terms of security, reliability, etc.) requirements that are formalized in the specification. The theme was intertwined with risk assessment and management tools to evaluate AI model performance, such as the joint initiative carried out by the International Organization for Standardization and the International Electrotechnical Commission.

These techniques can have a dual use and be integrated to verify and audit software to the world of autonomous weapons systems. It was also stated that the general cybersecurity regulations should be applied to AI and all the weapon systems.

Regarding the recommendations, it was mentioned that even though learning and adapting as functional requirements for LAWS, learning should not be done statically to avoid non-deterministic behavior and boundaries should be put in place for the adaptive functionality of the software in order to facilitate the verification of such software.

Also, the necessity to create constraints and limitations as to: where the data comes from, components one is using in software, to verify the different components that make up for the autonomous system and all other systems attached to it. Furthermore, in critical systems in which AI is used, to have a set of data that is able to reboot it. In the governance side of the technical debate, it was appealed for more countries should publish guidelines for LAWS, and support different forums that are already looking into AI regulation and risk assessment, support the academic community in operationalizing the ethical and juridical implications in technical specifications, and to bridge the gap for responsible AI.

The ethics panel came to discuss the broader implications of the decision to conceive autonomous lethal functionalities to these systems. On one hand, it was stated that AI may never be capable of addressing these issues. On the other hand, it was also said that, given that these are not moral and legal agents, but it becomes possible to use it according to certain values, and that mankind should not delegate moral and legal responsibilities to them, that would be an abrogation of our own moral and legal duties. A panelist framed the discussion as a decision to engage lethality to machines we already know to be flawed.

Moral relativism was cited as an obstacle to encoding morality into the specifications of these weapon systems. How do you choose which moral code to implement into these systems? In response to that, came a statement that human dignity and principles of humanity can be used as a baseline concept, as they are the basis also for IHL, in addition, a statement was made that an IHL compliant weapon is not a sufficient condition to ensure adequate functionment of the weapon system. It was also proposed that the Martens Clause should be the ethical basis for those sorts of laws, having IHL as an ad hoc baseline until proper legislation was established.

The necessity to include the concept of meaningful human control in the regulation of autonomous weapons systems was defended. It was stated to have a strong legal and moral basis for regulation and as a take away from the panel on regulations, It was observed a lot of common ground as to what meaningful human control should look like. It was said that the high moral standards claimed by democracies are not reflected by the positioning of some of those counties, and a call to action was made to more militarily-developed nations, if they believe their technological developments can ensure the tests that meaningful human control requires, then join these discussions and make sure you use the responsibility that they have at the global stage.

Took place a statement that, given the recent trends in military technology, such as the scalability of drones and the proliferation of loitering munitions are currently paving the way for LAWS, and that the countries that are most afraid that their counterparts are able to develop this technology are developing an arms race for this technology instead of advocating more intensely in favor of regulations. The real concern of the increasing tensions a global arms race of increased rivalry can lead, make States coming together to negotiate a treaty that limits the use and properties of autonomous weapons an ethical imperative. A participant asked himself: who is benefiting from this? His answer was that the dominant military powers and the arms industry would profit from LAWS, and that as long as they are prominent and dominating the way that we are going to use this technology in warfare, in contrast, the people that are most likely the receiving end of it will be people from the Global South.

Deepening the point that decisions over the use of force cannot be taken by a machine or algorithm, digital dehumanization was addressed. It was stated that it is not ok to objectify humans just as a data point, understand the value of life, and reflect on the reasons why life should be taken.

If we automate the targeting process, we would leave the recognizing, understanding and reflecting, out of the picture, and would be a violation of human dignity by the lack of human agency. In this sense, when it comes to human dignity, it was said that It is particularly undignifying about being killed by a machine because you cannot understand why the machine chose to kill you, you would not be targeted as an individual, but as a data point, among the hundred or thousand that this machine looks to kill. Humans, in this processesses, would become mere target proxies, as these systems have no representation of what it is killing, just data points, the target profiles that have been coded into the machine, translating ones and zeros into life and death decisions. This distance, both geographical and mental, from the act of killing, violates the dignity of the target. It was added that this distance, plus the increased availability of such weapons systems, will lower the threshold for the use of violence.

A ban on autonomous functions in anti-personnel weapons and the adoption of a two-tier approach for the remaining systems was proposed as a means to address human dignity in this debate. It was also cited the UNESCO draft text of the Recommendation on the Ethics of Artificial Intelligence has elaborated on principles which could easily applied to the military sector as well, such as the articles 14th and 36th, that say respectively that: during any phase of the life cycle of AI systems no human being or human community should be harmed or subordinated and that life and death decisions should not be ceded to AI systems.

Came up also, during the panel, the issue of accountability. As told, accountability is already an issue in drone warfare and the war on terrorism, the increase of availability of autonomous weapons would only enhance the accountability problem. A panelist reiterated that people, not machines, have to be held responsible, and that a human has both to understand and an option to intervene. It was proposed that this interactive obligeance should be a functional requirement in the specification of these softwares. In conformity to statements made in the technical panel, it was said that all phases of the SDLC should be regulated, including the choice of data, in order to avoid bias (as they reflect pre existing societal biases), and in cases that the database is not a good fit.

Also, the necessity to ensure the necessary communication links for intervention.Importance of understanding that since this issue revolves around Artificial Intelligence, we are not only in the world of military, but present in civil society, and in that sense, who do we want to be as a species and then discussing which AI systems we need for that. An initiative between Airbus and the Fraunhofer Institute to design autonomous functions with human control was cited, and the panelist added that there must be an international regulation that sets normative standards and that decision should not be left to industry. Regarding the taxonomic problems of the debate, it was said that there is no simple solution for creating consensus on definitions. Panelists reinforced that there should be limits in the specification as to functionalities such as aiming and engaging autonomously.

The proactive stance of the Luxembourgish government has helped further the debate on the regulation of lethal autonomous weapons systems. The initial key-note speech has given in-depth considerations for what regulations should look like, and the major moral dilemmas regarding the debate. In the political panel, we can note the interest of institutions present in proceeding with the regulations.

Nevertheless, topics such as meaningful human control and a legally binding instrument were addressed in the panel, the topics could have been taken into a deeper level, helping in consolidating the taxonomy of meaningful human control and pointing ways to deal with the current challenge States are facing regarding the adoption of such a document. From the military panel it could be a way of counterbalancing asymmetrical warfare for countries with reduced population, and the perception of the advantages of staying ahead in the Artificial Intelligence race. In the panel, a lack of knowledge over the taxonomy was demonstrated, terms such as Meaningful Human Control, and even Autonomous Weapons Systems, were thoroughly debated in this panel, despite the already existing definitions, as panelists from the regulatory panel pointed out, and a big conceptual debate, that didn't seem to be familiar to panelists.

The regulatory panel made an assessment of the 10 years of debate on the topic, while pointing out risks imminent to technological innovations in the military sector, they have made points for the need of a human-in-the-loop-approach. Also, recommendations were made, such as the two-tier (regulations and restrictions) approach and the ban of autonomous functions in anti-personnel weapons. The panelists were very clear about the overlapping consensus about the importance of the issue, and pointed out difficulties regarding consolidating a legally binding document such as diluted State positioning in the CCW.

While the tech panel has raised the debate on what VT&T should look like for these systems (with a big emphasis in dual use of AI protocols), it has not been able to give insights of how the norms (from IHL, for instance) should be translated into requirements for these systems. Simultaneously, the risk assessment made on the tech panel about the use of open source software in this context seems to overlook the risk of paramilitary groups and non-state actors having access or being able to develop their own Autonomous Weapons Systems.

The moral panel has raised the issue of digital dehumanization, accountability and dove into meaningful human control, helping solidify the conceptual debate in the event. Overall, the event has demonstrated leadership from the Luxembourgish government, reinforced the momentum created by the conference in Costa Rica, enriched the debate on LAWS, and led to clues as to areas in which different sectors struggle to communicate with each other.