Table-Top Exercises on the Human Element and Autonomous Weapons Systems

Summary Report

GIACOMO PERSI PAOLI
ALICE SPAZIAN
ALISHA ANAND

UNIDIR UNITED NATIONS INSTITUTE FOR DISARMAMENT RESEARCH

UNIDIR **SECURITY** AND **TECHNOLOGY** PROGRAMME

ACKNOWLEDGEMENTS

Support from UNIDIR core funders provides the foundation for all of the Institute's activities. This project is part of the Security and Technology Programme which is supported by the Governments of Germany, the Netherlands, Norway, and Switzerland and by Microsoft.

UNIDIR would like to thank the 198 experts from 75 countries that took part in this series of exercises organized between September 2020 and June 2021 for their active engagement, inputs and suggestions.

The authors would like to thank Ms Laetitia Zarkan for the technical production of all events and Ms Pauline Warnotte and Ms Sarah Grand-Clement for conducting the exercise in French.

Design and layout by Eric M. Schulz.

ABOUT UNIDIR

The United Nations Institute for Disarmament Research (UNIDIR) is a voluntarily funded, autonomous institute within the United Nations. One of the few policy institutes worldwide focusing on disarmament, UNIDIR generates knowledge and promotes dialogue and action on disarmament and security. Based in Geneva, UNIDIR assists the international community to develop the practical, innovative ideas needed to find solutions to critical security problems.

CITATION

G. Persi Paoli, A. Spazian, A. Anand, *Table-Top Exercises on the Human Element and Autonomous Weapons Systems: Summary Report*, Geneva, Switzerland: UNIDIR, 2021.

NOTE

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The views expressed in the publication are the sole responsibility of the individual authors. They do not necessarily reflect the views or opinions of the United Nations, UNIDIR, its staff members or sponsors.

www.unidir.org
© UNIDIR 2021

TABLE OF CONTENTS

ABOUT THE AUTHORS

Giacomo Persi Paoli is Programme Lead for Security and Technology at UNIDIR. His recent work has focused on arms control, technology horizon scanning, artificial intelligence and cybersecurity. Before joining UNIDIR, Persi Paoli was Associate Director at RAND Europe, where he led the defence and security science, technology and innovation portfolio as well as RAND's Centre for Futures and Foresight Studies. He served for 14 years as a Warfare Officer in the Italian Navy and has been extensively engaged in research on small arms and light weapons in support of United Nations processes. Follow Giacomo on Twitter @GPersiPaoli.

Alice Spazian is a graduate professional for the Security & Technology programme. Her work focuses on artificial intelligence, specifically autonomous weapons systems, as well as on the impact of emerging technologies on international security. She holds a Master's degree in International Security from the Sant'Anna School of Advance Studies jointly with the University of Trento. In her previous positions, Alice worked in the human rights office of the municipality of Bologna and as a researcher in the field of interfaith dialogue.

Alisha Anand is the Research Assistant for Security and Technology at UNIDIR. Before joining UNIDIR, she worked on non-proliferation and export controls with the Disarmament and International Security Affairs Division of the Indian Ministry of External Affairs, the Manohar Parrikar Institute for Defence Studies and Analyses and the Federation of Indian Chambers of Commerce & Industry. She holds a Master's degree in Law and Diplomacy from the Fletcher School, Tufts University, where she specialized in international security and international law. Follow Alisha on Twitter @AlishaAnand912.

ABBREVIATIONS & ACRONYMS

AI	ARTIFICIAL INTELLIGENCE
AWS	AUTONOMOUS WEAPONS SYSTEMS
LAWS	LETHAL AUTONOMOUS WEAPONS SYSTEMS
TTX	TABLE-TOP EXERCISE

1. INTRODUCTION

1.1. BACKGROUND

Since governments began expert meetings in 2014 on lethal autonomous weapons systems (LAWS) in the context of the Convention on Certain Conventional Weapons, maintaining a certain degree of human involvement in the use of emerging technologies in the area of LAWS has been one of the main points of discussion. Many different perspectives were discussed over the years: from maintaining human involvement or control over weapons, or their critical functions, to maintaining control over attacks, the targeting process, and (final) decisions to use lethal force. Although most parties to the debate agree that such involvement should be more significant than the mere possibility of aborting an attack at the final moment, they have yet to reach consensus as to how the human role in the use of lethal force should be defined and implemented.

With a mandate for 2020 and 2021 to produce "consensus 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",[1] Member States participating in the Group of Governmental Experts on LAWS would need to further elaborate upon the general concepts and principles to which they have already agreed, particularly in the area of human–machine interaction and, more specifically, concerning the human element in the use of (lethal) force.

1.2. PURPOSE OF THE TABLE-TOP EXERCISES

To support the development of a shared understanding of the nature and type of human–machine interaction in the context of the execution of an attack, UNIDIR organized a series of regional table-top exercises (TTX) focused on the interplay between introducing degrees autonomy in weapon systems and retaining human control in the context of varying scenarios.

This included designing an approach aimed at identifying specific areas of operational concern and action and moving away from more abstract levels of discussion with a view to creating a common knowledge base to support better informed negotiations related to a normative and operational framework for LAWS.

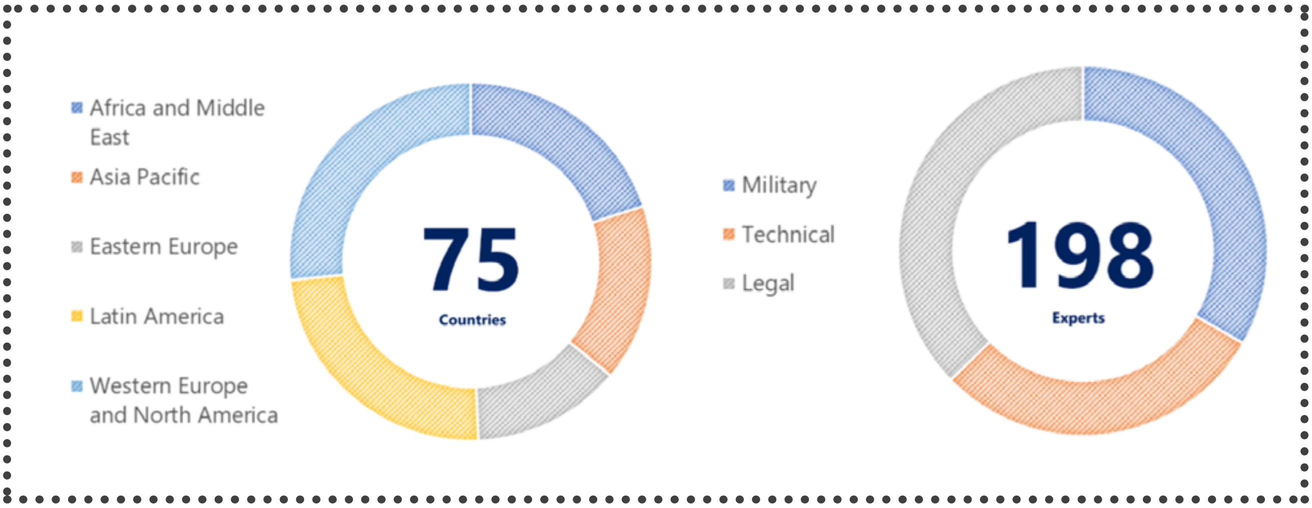
These TTXs brought together technical, military and legal experts to explore the complex interactions between what is **technically feasible**, what is **militarily desirable** and what is **legally permissible** by focusing on two principal research questions:

1 Final Report of the Meeting 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, document CCW/MSP/2019/9, 13 December 2019.

1. Within the complex process leading to the use of force (i.e., the targeting process), what are the technical, military and legal implications of introducing degrees of autonomy in weapon systems?
2. What conditions and circumstances (e.g., type of target, environment, risk to civilians) influence the requirement for human control and involvement in the decision-making process leading to the use of force?

1.3. THE PROJECT IN NUMBERS

Between September 2020 and June 2021, UNIDIR held six TTXs in Africa and the Middle East, the Asia–Pacific, Eastern Europe, Latin America, Western Europe, and North America. **198 individual experts** (some nominated by their governments; others invited by UNIDIR as independent experts) from **75 countries participated in these exercises**. The figures below provide more information on the breakdown by region and by expert type.



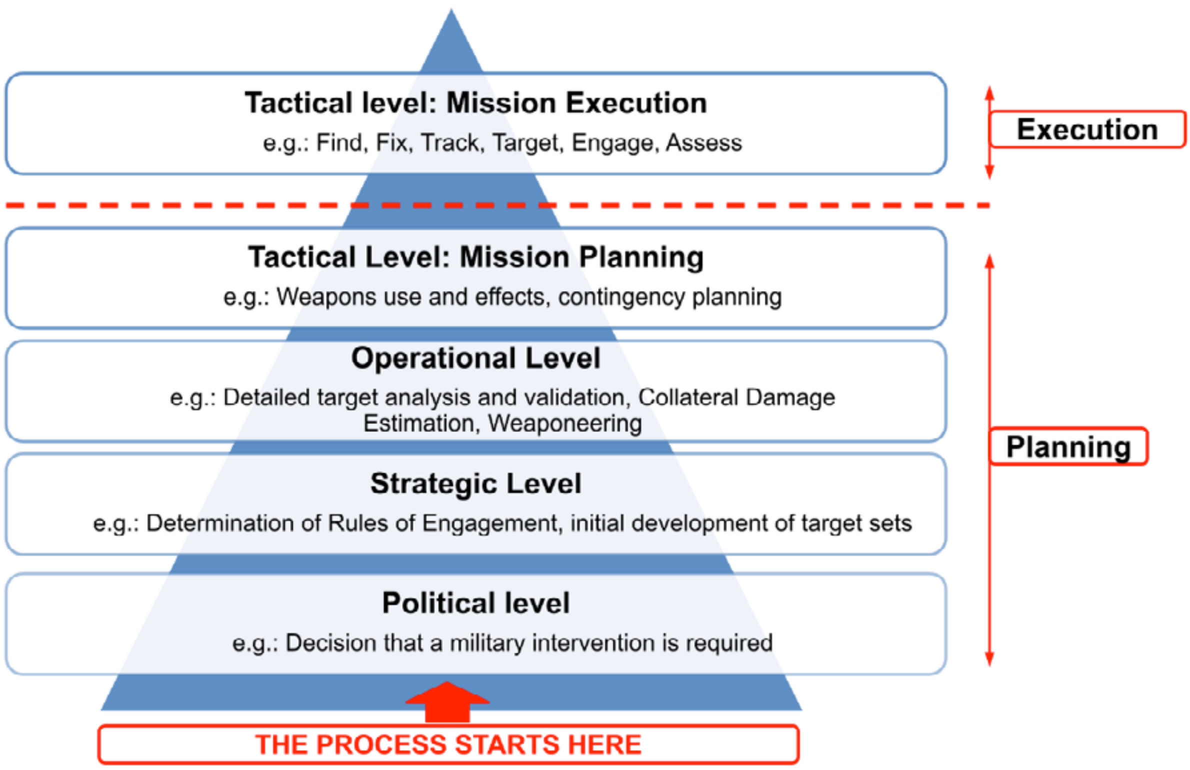
2. METHODOLOGY

2.1. THE ANALYTICAL FRAMEWORK

These exercises were based on the analytical framework known as 'the iceberg' that UNIDIR has developed to describe the human element in decisions about the use of force.[2]

This framework illustrates the complexity of the decision-making process that leads to the use of lethal force in present-day military operations, and explains how the use of force, when conducted in strict compliance with relevant laws and rules through this process, is never random or arbitrary. Even in the case of a more dynamic environment, it remains shaped and constrained by a number of political and operational decisions informed by legal, military and technological factors across all layers of the 'iceberg', not only at the very top of it (i.e., during the actual execution of a mission).

FIGURE 1 *A simplified view of the targeting cycle used during the TTX (adapted from UNIDIR's " iceberg" on the human element in decisions about the use of force)*



2 Merel Ekelhof and Giacomo Persi Paoli, "The Human Element in Decisions about the Use of Force", UNIDIR, 2020, https://unidir.org/publication/human-element-decisions-about-use-force. This document is also available in Spanish and French.

While acknowledging the critical importance of all decision points before the actual execution of a mission (e.g., the establishment of rules of engagement; the selection, vetting and approval of targets; and weaponeering and collateral damage estimation), after conducting a pilot iteration of the TTX it was decided to focus specifically on the tactical steps of mission execution (i.e., the visible part of the iceberg) to provide more time for detailed discussions. An additional set of TTXs focusing on the invisible part of the iceberg is currently under development.

This tactical mission execution phase consists of the following steps:

- **Find**—navigate and maneuver in the battlefield to find the target based on available information, intelligence and data collected in real time.
- **Fix and Track**—once the target is detected, sensors will be used to determine and to maintain positive identification of the target and to monitor the environment.
- **Target**—final checks before the engagement takes place include risk assessment, compliance check for rules of engagement and international law and international humanitarian law.
- **Engage**—the attack is executed, and weapons are released. (An attack can also be suspended or cancelled.)
- **Assess**—the effectiveness of the attack is evaluated and decisions on future action are taken (including re-attack if necessary).

In addition, the exercise presented four different options for human control or involvement in a weapon's execution of each of the above steps:

- **Full direct control**—the system has no autonomy and remains under the full and direct control of the operator for the execution of the given task.
- **Human in-the-loop**—the system implements the given task with autonomy but requires human intervention to validate and implement specific actions.
- **Human on-the-loop**—the system implements the given task in autonomy under the supervision of human operator(s) who can intervene if necessary to correct or abort a specific behaviour or action.
- **Human off-the-loop**—the system implements the given task with full autonomy, without supervision or intervention by human operator(s).

2.2. THE SCENARIOS

The four scenarios used in the exercise were carefully designed to frame a more practical discussion than normally achievable at a more theoretical level. The TTXs were not intended to be representative of all the possible operational and tactical contexts as the purpose was not to identify a catalogue of permissible or non-permissible use cases for LAWS.

The scenarios represented four different tactical situations which shared some common characteristics, such as the wider context (i.e., international armed conflict between two States) and domain (i.e., land domain). In addition, the scenarios proposed different combinations of other critical factors such as:

- **Target**
  - Fixed or mobile
  - Inhabited or uninhabited
  - Location known or unknown
- **Collateral damage and risk to civilians**
  - Low risk or high risk of civilian casualties
  - Low risk or high risk of damage to civilian or dual-use infrastructure, some of which were included in no-strike lists (e.g., civilian housing)

Each scenario was built using a different combination of the parameters above while maintaining the same overall framework of context and domain. This allowed experts to focus their discussions and increased the comparability of tactical situations, while reinforcing the non-transferability of observations and arguments to other contexts.

The figure below summarizes the four scenarios used in the TTX.

FIGURE 2 *Overview of scenarios used in the TTX*

2.3. THE EXERCISE

For each of these scenarios, experts were asked to assess the most suitable level of autonomy (no autonomy, human in-the-loop, on-the-loop or off-the-loop) for each of the steps in the tactical execution phase of the targeting mission.

Based on the type of expert, the notion of 'most suitable' was interpreted either based on the technological maturity, today and in the near future, of the artificial intelligence (AI) and autonomous systems (i.e., technical feasibility), or the perceived military advantage that AI and autonomy could bring in the specific context (i.e., military desirability), or the perceived legality of introducing degrees of AI and autonomy in the given context (i.e., legal permissibility). The latter two elements were informed by the technical feasibility to ensure that assumptions were based on realistic and achievable system performance.

In order to allow experts (particularly non-military experts) to familiarize themselves with the different steps, this assessment was initially conducted at the conceptual level (i.e., in a scenario-agnostic context) and then repeated for each scenario. This approach elicited general concerns and considerations before putting them to the test in the four scenarios.

Finally, experts were asked to reflect on and assess the relative relevance and influence of a range of factors in their decisions for each scenario:

- **type of target** (fixed or mobile, manned or unmanned, pre-planned/on call/not planned, etc.);
- **environment** (e.g., urban or open, mountain, desert or forest, etc.);

- **domain** (e.g., air, land, maritime);
- **type of mission and mission parameters** (e.g., time of attack, desired effect);
- **assessment of risks** to civilians or own forces;
- **technical characteristics** of the system (e.g., understandability, predictability, reliability); and
- **other** factors (which experts were asked to specify in their inputs).

2.4. SCOPE AND ASSUMPTIONS

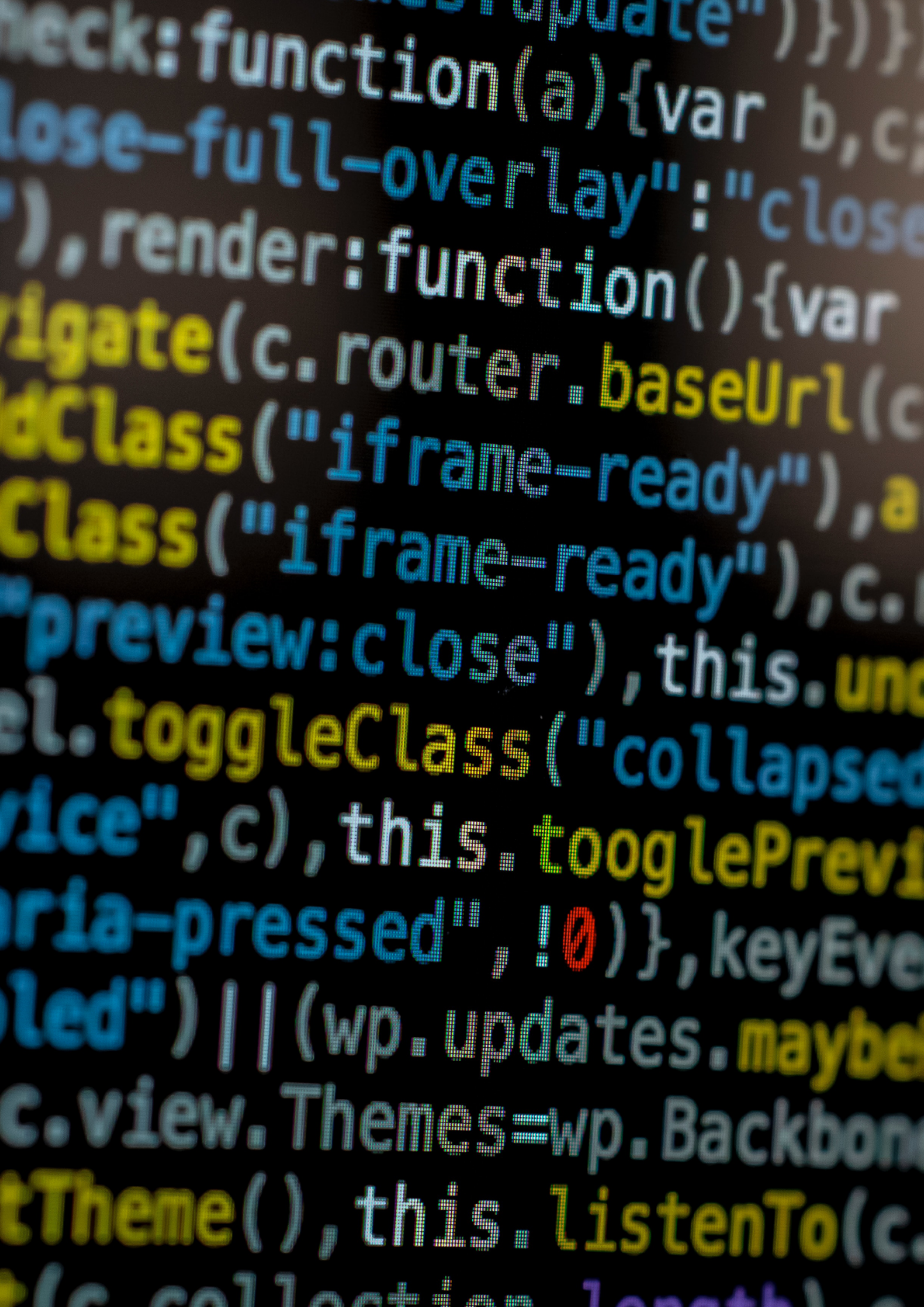
A number of assumptions and constraints on scope were applied to better focus the discussions and to make an efficient use of the limited time available:

When experts were asked to reflect on future developments in 5–10 years, technological developments were considered to be evolutionary/incremental based on the current state of technological maturity.

The exercise focused exclusively on physical autonomous weapons systems (AWS) capable of being deployed that achieve kinetic effect with various degrees of human involvement. As such, physical autonomous systems used for intelligence, surveillance and reconnaissance tasks, or autonomous cyber weapons, were out of scope.

AI-enabled decision-support systems that may be used for planning purposes were out of scope.

The exercise focused on decisions to deploy LAWS. Political decisions to acquire or develop such technology, as well as considerations relevant to the development, review, testing and acceptance of such systems, were out of scope.


function(a){var b,
full-overlay":
),render:function(){var
(c.router.baseUrl(
("iframe-ready"),
("iframe-ready"),c.
preview:close"),this.
toggleClass(
",c),this.
pressed",!0)},
")||(wp.updates.
c.view.Themes=wp.
(),this.listenTo(

3. FINDINGS

This section provides an overview of key findings of the series of TTXs as well as some open questions that emerged from the discussions with experts.

Several key points emerged from the technical, military and legal discussions and will be discussed in sections 3.1, 3.2 and 3.3 respectively. The key overall takeaways are summarized as follows:

KEY TAKEAWAYS

Decisions on the type and level of human control in any given operation will be informed by:

» the technical characteristics of the system;

» the doctrine, concept of operation, specific rules of engagement and further restrictions/guidelines that were developed for the system based on its review; testing, evaluation, verification and validation; and other earlier processes;

» the characteristics of the operation, the environment, the target and the adversary;

» the military advantage/disadvantage of autonomous functions relative to the objective and compared to alternative weapons systems available; and

» the legal considerations applicable to the specific situation, context and mission

3.1. SUMMARY OF TECHNOLOGICAL CONSIDERATIONS

Having a shared and well-informed foundational understanding of what the technology is and is not, what it can and cannot do, and the parameters that influence its performance is an important factor to advance the discussions on LAWS. This section summarizes the main points that emerged from the interventions of the technical experts.

- The ability of a system to perform actions without human intervention sits on a scale that ranges from automatic, to automated, to autonomous. Factors that determine where on the scale a specific system or functionality lies include the number of environmental variables the system has to take into account to inform its actions, the amount and variety of data it has to process, and the complexity of the 'decision' it has to take. The higher the number of variables, the larger the amount of data to be considered, and the more complex the decisions to be made, the greater the system's capacity for 'autonomous' behaviours.

Automatic vs Automated vs Autonomous: an illustrative example

A practical example may help better capture the difference between different degrees of a system's ability to perform actions without human intervention.

When a military drone loses the communication link with its ground control station, it automatically triggers a countdown for a predetermined period of time. When the countdown reaches zero, if the communication link has not been re-established, then the drone automatically triggers an automated navigation mode to relocate to a predetermined area and loiter for another period of time, after which, for example, the drone will trigger another automated process (e.g., return to base, controlled crash/landing in another area designated for that purpose before the mission begins). This scenario illustrates how simple inputs (e.g., the loss of communication) can trigger automatic outputs that are purely logic-based (e.g., if no communication then start timeout procedure, if the timeout runs out then relocate to area X) and automated processes (e.g., fly from A to B, land in area X, return to base).

If instead the drone was equipped with autonomous capabilities, then, for example, it may be delegated to choose the best area to conduct a controlled crash or attempt a landing based on a number of parameters (e.g., at least X km away from the area of operation, at least X km away from buildings, people or natural obstacles, following a course that minimizes risks of adversarial interference, etc.). This level of delegated decision-making requires the system to interact with its environment in a much more complex way and to replicate, to a certain degree, decision-making processes that would be followed by human operators.

- Today's research and development in the area of technologies to enable such autonomous functions mostly focuses on AI that can conduct specific tasks, under specific circumstances and in specific contexts (sometimes referred to as 'artificial narrow intelligence'). The idea of a single AI capable of transferring learning from one task or context to the other (also known as 'artificial general intelligence'), remains an aspirational concept. As such, one of the most common views raised by technical experts across sessions was that state-of-the-art AI today and for the foreseeable future, when employed in real-world environments, remains highly brittle and limited in its ability to transfer to novel domains or tasks, or even to marginally unanticipated inputs.
- Applying these considerations to weapon systems, many experts expressed that the introduction of autonomy will be gradual, incremental and limited to those tasks or functions where the AI meets the predictability and reliability requirements set by end users (i.e., military forces) and are able to operate within applicable regulatory frameworks. This of course does not exclude the possibility that some reckless actors, including non-State actors, may decide to integrate brittle AI into critical roles as soon as they can, prioritizing the technology's potential military advantage over system safety considerations.

The Parallel with the Automotive Sector

The gradual introduction of autonomy in physical systems is the approach followed by the automotive sector in the development and roll-out of autonomous driving. In this context, an internationally accepted and widely used taxonomy exists with detailed definitions for six levels of driving automation, ranging from no driving automation (level 0) to full driving automation (level 5), in the context of motor vehicles and their operation on roadways.[3] It is important to note that, as of today, fully autonomous cars are undergoing testing in several parts of the world, but none are yet available to the general public.[4]

In the autonomous driving context, the taxonomy identifies three primary actors: the (human) user, the driving automation system, and other vehicle systems and components. This is particularly relevant for autonomous weapon systems where the 'intelligent' component will ultimately have to be integrated within a wider system of systems.

Finally, another factor taken into consideration by the autonomous driving taxonomy is the environment in which the autonomous system will be used, with a key distinction between systems capable of performing only in specific conditions and environments as opposed to autonomous driving systems capable of operating in all circumstances, environments and conditions. The environmental factor is an important point of difference between autonomous vehicles and autonomous weapon systems; the former are meant to be used in an environment that will be designed and adapted to facilitate their operations (e.g., road signals, sensors, signal amplifiers) while the latter are meant to be used in a denied environment where exactly the opposite will likely be encountered (e.g., jamming, spoofing, camouflage).

- Data are a critical factor that have significant impact on all stages of an autonomous system development, training, testing and deployment.[5] With the application of AI in the military domain still in its infancy, no common international standards exist on the issue of data—which data to use, how to collect it, how to secure it, if/how to share it for safety purposes, etc. This is particularly relevant for testing, evaluation, verification and validation, which is a key step to calibrate the trust between human and system, as well as a potential area for capacity- and confidence-building among States. However, according to some experts, this remains a step that is heavily under-resourced in the military domain compared to research and development budgets.

3 SAE International. 2021. "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles". J3016_202104. https://www.sae.org/standards/content/j3016_202104/.

4 Todd Litman. 2021. "Autonomous Vehicle Implementation Predictions". Victoria Transport Policy Institute. https://www.vtpi.org/avip.pdf.

5 For a detailed overview of the critical importance of data, see Arthur Holland Michel. 2021. "Known Unknowns: Data Issues and Military Autonomous Systems", UNIDIR. https://doi.org/10.37559/SecTec/21/AI1.

3.2. SUMMARY OF MILITARY CONSIDERATIONS

Technology should not be considered separately from the military requirement that would lead its development and employment. This section summarizes the main points that emerged from the interventions and considerations of the military experts that took part in the exercise.

- There are several perceived advantages of introducing autonomy in the perfromance of specific tasks during the execution of a mission. These include, speed, accuracy, more efficient resource management, and ability to operate in communication-denied environments. On this latter point, views were split—some military experts highlighted that the ability of autonomous systems to operate in situations where communication links cannot be sustained is an important advantage, while others underlined how a lack of communication would undermine the operator's or commander's ability to maintain situational awareness of the system's behaviour and to intervene if needed.
- Which tasks would benefit from the introduction of autonomy would depend on the context and missions parameters—for example, type of target, space and time limitations, complexity of the environment, risk to civilians and to own or allied forces, availability of alternative means, adversarial action, and rules of engagement. Experts expressed that there is no incentive for military forces to deploy a weapon system they do not understand or that has unpredictable behavior that cannot be constrained or limited, particularly when considering critical functions.
- Compliance with international humanitarian law will be a core determinant in whether military forces choose to use an autonomous system, and how it is used.
- The availability of a new technological solution does not mean that such solution will be automatically developed into a full military capability and adopted by military forces. A military capability is much more than a piece of technology: it includes everything from doctrines, to organizational structures, training and education, facilities, processes, logistics, etc. This point was stressed throughout the exercise by military experts, highlighting the important role that such elements would play as part of a wider risk management approach for systems with autonomous capabilities, particularly AI-enabled autonomy. There is no reason to believe that a responsible military would bypass any of these elements when employing weapons that include autonomy.

3.3. SUMMARY OF LEGAL CONSIDERATIONS

Legal considerations permeate the entire decision-making process that leads to the use of force. The main points that emerged from the interventions of the legal experts that took part in the exercise included the following:

- Given that the (legal) use of force in military operations is never random or arbitrary and it is subject to many layers of vetting and approval, the legal assessment of an attack starts well before the execution of a mission. However, the dynamic nature of the battlefield requires that checks are made to ensure that the attack remains compliant with international humanitarian law at the time it is delivered.
- The assessment of the legal permissibility of introducing autonomy in specific steps of the targeting cycle would be informed by how much autonomy is introduced in preceeding and following steps.
- As elaborated in one of the guiding principles developed by the Group of Governmental Experts, the use of autonomy in weapon systems does not make human decision makers less legally liable for the outcomes of an attack. However, a number of experts pointed out that there may be novel challenges in appropriately assigning human responsibility and criminal liability for harms resulting from autonomous system errors.
- Full direct control of all steps in the targeting cycle is not the only way in which humans can retain ‘legal ownership’ of the actions and results of an autonomous system. Depending on the circumstances and on the specific functions performed by autonomous systems, control by veto (i.e., human-on-the-loop) or control by validation (i.e., human-in-the-loop) were also considered potentially acceptable by expert participants depending on the wider context and circumstances.
- However, this consideration is built on the assumption that some form of communication between human operators and the weapon system can be maintained at all times. When this is not the case, specific measures should be implemented at the technical and military levels to avoid any action by the system that could result in a violation of international humanitarian law and other applicable legal constraints.

3.4. SUMMARY OF SCENARIO-BASED CONSIDERATIONS

The exercise included a scoring component to capture expert views on the different scenarios as described in section 2.3. The purpose of this element was to trigger discussions on concrete (and visually clear) results rather than collect statistically significant data to enable their quantitative analysis. However, some high-level descriptive trends can be extrapolated from the results and be summarized as follows:

- The scenarios did not include information about own forces' order of battle, adversarial capabilities or wider information about the operational situation and context around the specific tactical problem that experts were asked to discuss. These were considered key factors by some experts in informing their assessment of the introduction of autonomous functions in weapon systems.
- None of the scenarios emerged as a clear situation where all degrees of autonomy should be completely prohibited or permitted across all steps of the targeting cycle.
- This indicates that technological maturity alone cannot drive decisions on the use of autonomy in weapon systems. The relevant operational and tactical contexts will also need to be considered.
- Consensus did not emerge. That being said, when provided with a range of options for control, most experts converged towards options that would allow humans to retain a form of involvement. Few participants opted for either full direct control (i.e., no autonomy) or human off-the-loop (i.e., full autonomy) across the steps of the targeting cycle.

Reflecting on the specific steps of the tactical execution of an attack, it is possible to identify the following high-level findings:

- The degree of autonomy deemed appropriate for any given step may depend on what degree of autonomy is assigned to previous or successive steps, and will always depend on context, environment and other factors beyond the specific tactical situation presented in the scenarios.
- On average the steps of Find, Fix and Track emerged as those where autonomy might be more feasible, desirable and permitted. Some variations emerged across experts on the extent of human involvement to be retained—technical and military experts converged more towards human on-the-loop or off-the-loop, while legal experts preferred human in-the-loop for those scenarios where risks to civilians was considered higher.
- The Target step was, on average, considered to require a more conservative approach (i.e. more human involvement) in those circumstances where risk of collateral damage or civilian casualties was higher. While legal experts opted more for full direct control, technical and military experts were open to consider a human in-the-loop configuration as a valid alternative. Where risks of collateral damage or civilian casualties were lower, all experts were open to a less strict control configuration.

- The Engage step was, on average, the step requiring stricter control by human operators in most scenarios, except those where location of the target was known and the risk of collateral damage or civilian casualties was low or non-existent. In that case, views were almost equally distributed across all options available.
- The Assess step elicited the most diverse profile of preferences. In general, preferences were distributed across all four available options for control, with the technical and legal groups being more cautious and clustering around full direct control or human in-the-loop control, while military experts were more distributed across the three options "in-the-loop", "on-the-loop" and "off-the-loop control".

3.5 OPEN QUESTIONS

In addition to providing insights into the subject of the human element in autonomous weapon systems, expert engagement in these TTXs allowed the identification of important open questions that would require further reflection and investigation.

Technical questions

- When evaluating systems with autonomous functions, should the decision to deploy such systems be based on a comparison with human performance in a comparable situation, rather than with the absolute characteristics of the system? How can one compare human performance against machine performance in a scientific, measurable manner?
- How can one ensure that a system has been developed using appropriate datasets for real-world deployment? And how does one certify that a system's development reflects all the possible contexts that the system might encounter in real life?
- Related to the above, to what extent can synthetic data replicate the complexity and diversity of real operational data? Should synthetic data be used for testing, evaluation, verification and validation purposes?
- How can one manage human control in systems with continuous/adaptive learning capabilities?

Operational questions

- How does the inclusion of intelligent agents in a decision-making process affect the chain of command?
- In deciding whether to use an AWS or a human-controlled weapon, how does one weigh the risks stemming from potential AWS malfunctions against the risks stemming from human error in an equivalent conventional weapon?
- Does the fact that a system is autonomous change the threshold of tolerance of how much risk military commanders are willing to take?

Legal questions

- Is there a difference in discussing required human control versus allowed intelligent functionality?
- Can evaluative decisions and normative judgments mandated by international humanitarian law be assigned only to natural persons or also to artificial agents?
- Is the use of socio-technical proxies (e.g., gender, age) for IHL-relevant characteristics permissible? And, if so, under what circumstances?
- What technical information is necessary to conduct an appropriate legal review of an AWS?
- How can legal requirements be quantified/translated/implemented into technical requirements to increase the likelihood that the use of an AWS is compliant with international humanitarian law?

4. CONCLUSIONS

This series of table-top exercises provided an opportunity to investigate the interplay between human involvement, autonomy and use of lethal force. Based on the discussions with experts, we might conclude that:

- Systems that perform all steps of the targeting cycle outside of any form of human control or supervision **are not** technically feasible, militarily desirable or legally permissible.
- On the other hand, those systems which introduce degrees of autonomy in the execution of selected tasks **might be** technically feasible, militarily desirable or legally permissible, depending on the context, conditions and circumstances and provided that appropriate operational and legal frameworks are applied throughout the life cycle of the weapon system and across all steps of the targeting cycle.

A secondary outcome of the exercise was the horizontal transfer of knowledge—both within and among expert communities—which will support informed participation of Member States to the important discussions within the GGE on LAWS.

APPENDIX

APPENDIX I - LIST OF COUNTRIES REPRESENTED IN THE TABLE-TOP EXERCISE BY AT LEAST ONE EXPERT (NOMINATED OR INVITED)

REGION	COUNTRIES
Africa and Middle East	Burkina Faso, Burundi, Democratic Republic of the Congo, Jordan, Kenya, Kuwait, Lebanon, Morocco, Nigeria, Palestine, Qatar, Rwanda, Saudi Arabia, Uganda, United Arab Emirates
Asia–Pacific	Australia, India, Japan, Lao People's Democratic Republic, Mongolia, New Zealand, Philippines, Republic of Korea, Sri Lanka, Viet Nam
Eastern Europe	Bulgaria, Czech Republic, Estonia, Georgia, Republic of Moldova, Romania, Russian Federation, Slovakia, Slovenia, Ukraine
Latin America	Argentina, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Uruguay, Venezuela
Western Europe and North America	Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States

This report summarises the findings of UNIDIR's series of regional Table-Top Exercises conducted between September 2020 and June 2021. The project brought together 198 experts from 75 countries to discuss the technical, military and legal implications of introducing autonomy in various steps of the targeting cycle. By summarizing the main findings of this series of exercises, this report aims at creating a common knowledge base to support informed negotiations related to a normative and operational framework for LAWS.

UNIDIR UNITED NATIONS INSTITUTE FOR DISARMAMENT RESEARCH

@unidirgeneva @UNIDIR un_disarmresearch