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

3 March 2023

English only

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System
Geneva, 6-10 March, and 15-19 May 2023
Item 5 of the Provisional agenda
Intensify the consideration of 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

State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems

Submitted by the State of Palestine

1. The mandate of the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) for 2023 is to "*Intensify the consideration of 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.*"

2. The State of Palestine submits the present proposal for the consideration of the GGE.

Executive summary

3. **Autonomous Weapons Systems** (AWS) are systems that, upon activation by a human user(s), use the processing of sensor data to select and engage a target(s) with force without human intervention.

4. **A nominal human input** after the system's activation does not amount to a human intervention.

5. These systems pose a range of **legal, ethical, humanitarian and security risks**. To deal with these risks, both prohibitions and regulations are required.

6. **Prohibitions** are required on the development and use of autonomous weapons systems that:

(a) Are designed or used to target humans directly;

(b) Cannot be used with meaningful human control.

7. **Meaningful human control** requires that the AWS must meet all of the following requirements. They must be:

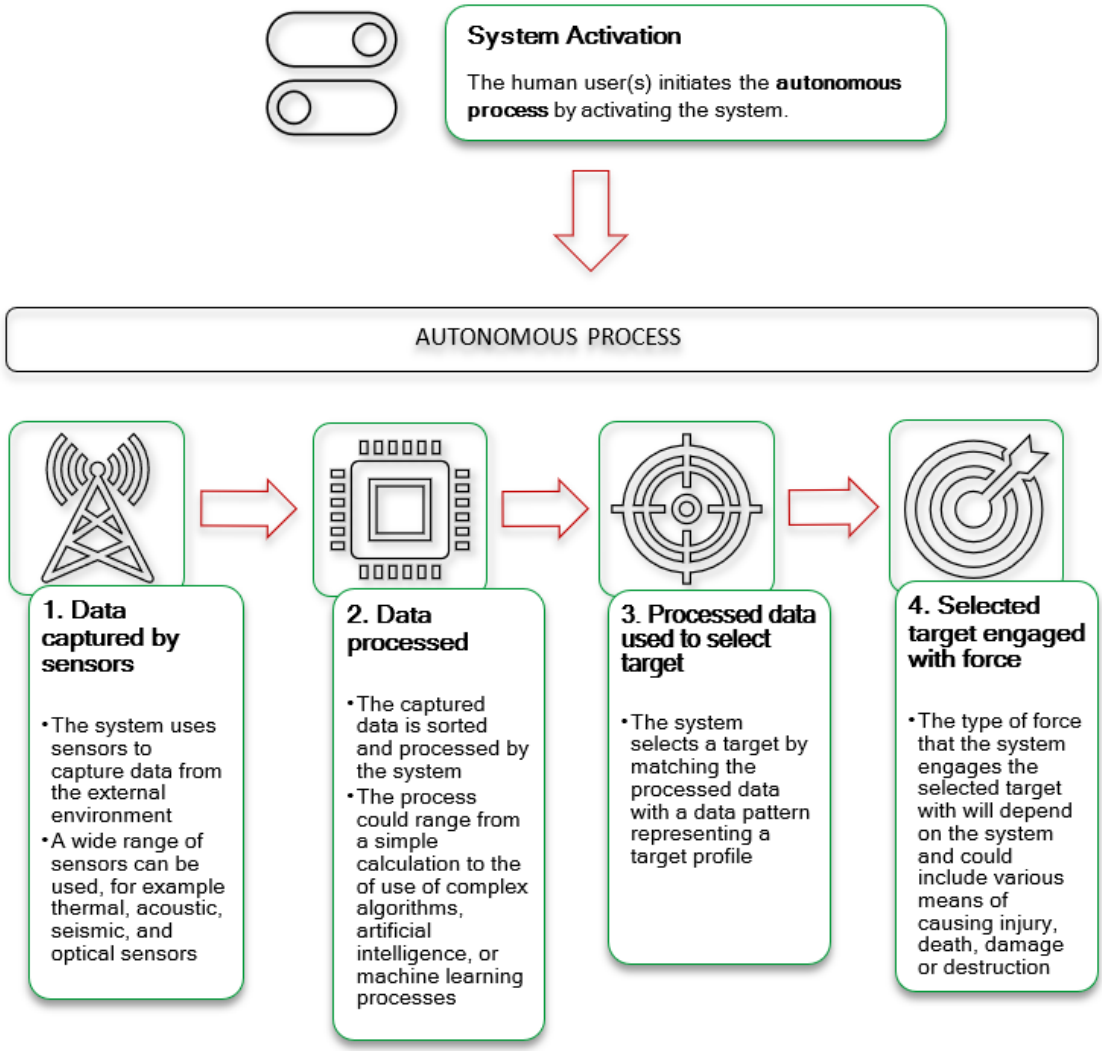
- Predictable
- Reliable
- Understandable and Explainable
- Traceable

8. **Regulations**, including both positive obligations and limits, are required to ensure that AWS can be used with meaningful human control.

9. This combination of prohibitions and regulations should be in the form of an **international legally binding instrument**.

I. Characterisation

10. AWS are systems that, upon activation by a human user(s), use the processing of sensor data to select and engage a target(s) with force without human intervention. We shall call this the 'autonomous process'.



11. Note that a system could potentially select and engage multiple targets after being activated and before it is turned off, or its munitions are exhausted.

12. Upon being activated by the human user(s), a process is set in motion that allows AWS strikes to be triggered if / when a pre-programmed target profile is matched by sensor data. The AWS thus proceeds to select and engage a target with force through the autonomous process. The role of the human user(s) in the execution of force therefore takes place at the point of the system's activation.

13. This autonomous process can be incorporated into a broad range of weapons systems. The types of sensors used, the methods of data processing, the nature of the target profile selected, and the scope of force applied, are all variable. If this autonomous process is incorporated into a weapons system, the system will be considered an **AWS**.

14. Note that whereas the above description is limited to AWS that have a "target profile" properly speaking, and that "calculate" and "select" targets, some more rudimentary systems like pressure-activated mines and automatic firing systems set off by motion sensors also fall under the definition of AWS. The latter systems "detect-engage" a target rather than "detect-

select-engage" a target. Note also that other system functions such as navigation, landing and refuelling can also be autonomous- however as they do not involve detection-(selection)-engagement of targets, they do not fall within the scope of this paper.

Nominal Human Input

15. Although the human user(s)' role takes place at the point of the system's activation, a nominal human input may take place after the system's activation and during the autonomous process.

16. A nominal human input is an input performed by a human that **does not materially affect the autonomous process**. A human input will not materially affect the autonomous process if it does not bring to bear any wider information to inform decisions to select and engage a target with force. No further human moral and legal reflection is occurring during that human input. In other words, the human input is "mindless"- its effect being the same as if the autonomous process had occurred without it.

17. If the only human input after the system's activation and during the autonomous process is a nominal human input, this will **not amount to a human intervention** and the system shall **still be considered an AWS**.

18. Examples of a nominal human input within the autonomous process include:

- A human pressing a button that commences the processing of data captured by sensors (i.e. between stages 1-2 of the autonomous process);
- A human clicking to accept target selection based on processed data without consideration of the nature of the target. (i.e., between stages 2-3 of the autonomous process);
- A human triggering a command from the system to use force without verification of the target or scale of force to be applied. (i.e., between stages 3-4 of the autonomous process).

II. Prohibitions

19. Prohibitions are required on the development and use of two types of AWS. These are:

(a) AWS that are designed or used to target humans directly;

(b) AWS that cannot be used with meaningful human control.

20. The risks and challenges associated with each type, in addition to the requirement for their prohibition, are clarified separately below.

(a) AWS that are designed or used to target humans directly

21. A prohibition is required on AWS that are designed or used to target humans directly. Under this prohibition, the data pattern representing a target profile, which is matched with the processed data captured from the external environment using sensors, must not be designed to represent a human.

Risks and Challenges

22. There are serious ethical concerns and legal challenges associated with AWS that are designed or used to target humans directly. These include:

- **Ethical Concerns**

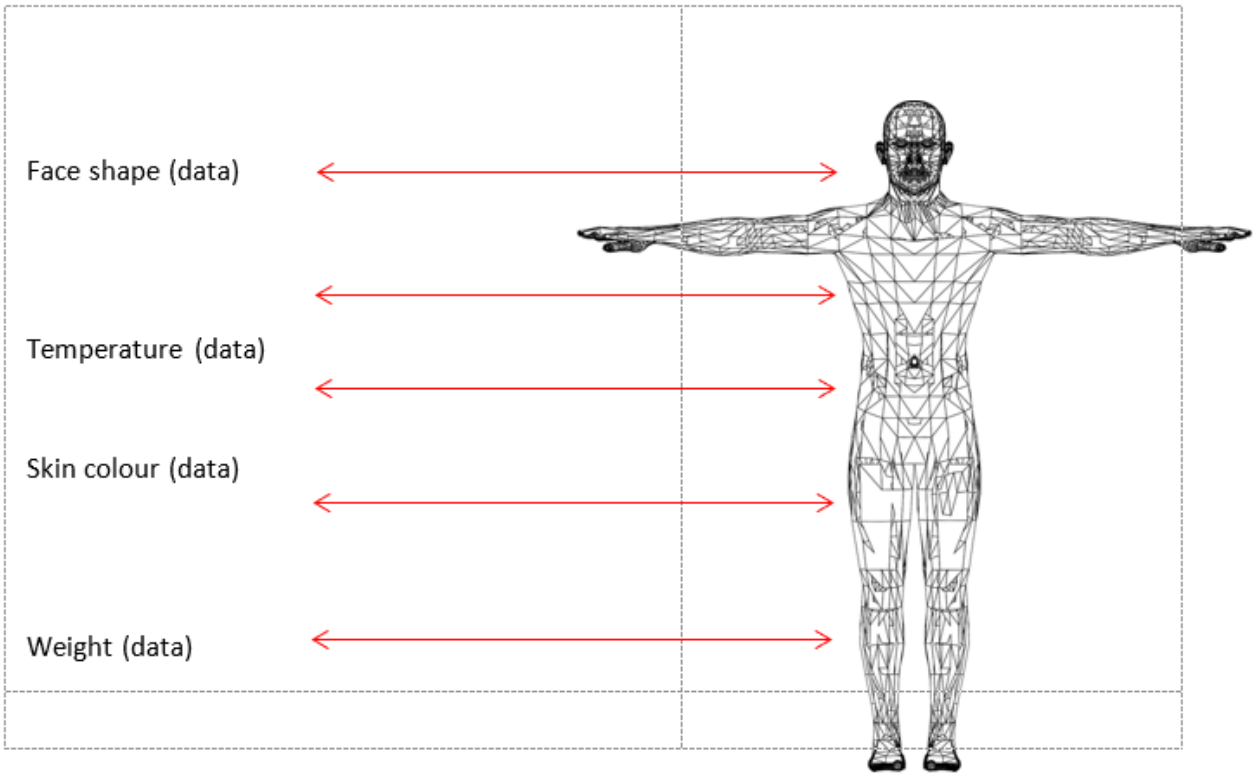
23. The process of encoding a data pattern to represent a human would necessitate the reduction of a human to data points. This could include, for example, data patterns representing human features and attributes, such as human faces, shapes, temperature, skin

colour, gait or movement speed. These data patterns would be used as proxy indicators for humans and matched with data captured from the external environment by sensors for the purpose of selecting human targets to attack.

24. This process of reducing humans to digital data for the application of force against them is dehumanising. Understanding and respecting the inherent dignity of a human is a fundamental ethical concern. This entails due regard for the intangible qualities of humans, including compassion, judgment, our relationships to one another and to our cultural and natural environment. These qualities of humans must not be reduced to data, nor can they be. The process of reducing humans to data for the purpose of executing force against them would be profoundly unethical and must be prohibited.

Digital Dehumanisation

The process of encoding a data pattern to represent a human would necessitate the reduction of a human to data points. This data would be used to select and engage a human with force.



25. Delegating the facility to take a human life to machines should be recognized as a challenge to the principles of humanity because machines cannot respect human dignity or show compassion. The profound ethical concerns expressed by faith leaders, scientists, States, private sector tech companies, military veterans, the public and the UNSG prove that these emerging weapons systems are problematic under the dictates of public conscience. In global surveys, the most cited reason for opposition to autonomous weapons systems is that they would "*cross a moral line as machines should not be allowed to kill*" (Ipsos, February 2021). Here it is also important to highlight that the principles of humanity and dictates of public conscience are not only ethical principles but also customary international humanitarian law under the Martens Clause.

26. This process would also likely entrench bias and discrimination through flawed profiling of human characteristics, particularly if seeking to target some people rather than others. Through reducing humans to data, problematic judgements might then be made concerning the characteristics of a human that are inadequate or inappropriate as a basis for targeting them. In addition, features like skin colour, which is complex and variable, should not be used as a basis for using force against people. Features like gait and movement speed

are not adequate proxies for legal grounds permitting targeting of persons due to the wide variability of how people express themselves and the complexities of social interactions. Reliance on such features further risks entrenchment of bias discrimination which raises additional ethical concerns and lends further support to a prohibition of systems that are designed to target humans directly.

- **Challenges to International Humanitarian Law (IHL)**

27. In the context of armed conflict, an AWS that is designed or used to target humans directly would also pose serious challenges to compliance with IHL, including the requirement of distinction, the prohibition of indiscriminate attacks and the protection of the civil population against the effects of hostilities.

28. The principle of distinction requires humans using force to distinguish between people who participate in hostilities and may therefore be attacked, and all others, including civilians, combatants hors de combat, medical personnel, etc. As such, who may be targeted calls for a context-specific, case-by-case assessment, requiring judgement and control to be exercised at the time at which force is executed. Through relying on a data pattern that has been designed to represent humans, at a time and place removed from when and where the force is executed, an AWS designed or used to target humans directly would undermine the requirement of exercising judgement and control to distinguish between civilians, enemy combatants, and enemies hors de combat on a contextual and case-by-case basis.

29. As noted above, additional serious problems relating to prejudice and bias and higher rates of error would arise if systems used target profiles to identify ‘targetable people’ based on certain characteristics such as race, gender or age. For example, a system may associate all military-aged males with combatancy or direct participation in hostilities, creating the risk that certain persons trigger an AWS strike although they are not legally permissible targets. Furthermore, systems that are particularly complex or opaque, that cannot be readily understood or explained, would make it difficult or perhaps impossible to offer a meaningful explanation of why certain people were targeted in certain circumstances, and thus to ascribe responsibility.

30. If systems were enabled to target all people in a particular location, they would risk having indiscriminate effects, contrary to established rules of IHL. It might be argued that use could be restricted to areas from which civilians are excluded. Yet such an approach would be insufficient since medical personnel and combatants hors de combat, among other people, may also not be attacked. It is also likely to be unreliable in practice and would shift the burden of avoiding harm onto the civilian population, thus eroding the presumption of protected status, and undermining the general principle of the protection of the civilian population against the effects of hostilities.

- **Challenges to International Human Rights Law (IHRL)**

31. Under IHRL, the use of force against a human is only permitted if it has a sufficient legal basis. In addition to the requirements for relevant court proceedings with due process, the use of force requires the principles of necessity and proportionality to be met. These require complex human judgements, and sensitive decision-making processes to ensure compliance with the law. The process of autonomy in weapons systems risks execution of force absent the required legal processes or judgements to be fulfilled.

32. AWS that target humans would reinforce or exacerbate existing structures of inequality. Bias and discrimination can be introduced into the autonomous process, including through data collection, training of algorithmic models, their evaluation, their use and their archiving or disposal, and harms can be amplified through flawed connections between identity markers, including gender, race, age, and ability. These issues raise acute concerns around the prohibition of discrimination under IHRL.

33. AWS designed or used to target humans directly would also fuel bulk collection of data, including individuals’ biodata. Apart from legal concerns around the use of force, this would enhance the capacity of organizations and governments to undertake surveillance, interception and data collection and would raise concerns around the right to privacy. Under IHRL, interference with the right to privacy can take place only if States’ measures respect

legality, legitimacy and proportionality. The potential bulk collection of data could raise challenges to these rights.

34. Concerns also arise regarding infringements of the right to remedy under IHRL. The increased opacity in decision-making processes through the use of AWS poses serious challenges to the ability to ensure accountability for unlawful acts. This is particularly problematic in the case where a victim has been killed by an AWS attack.

Requirement

35. A prohibition is required on AWS that are designed or used to target humans directly. Under this prohibition, the data pattern representing a target profile that is matched with the processed data captured from the external environment using sensors must not be designed to represent a human.

(b) AWS that cannot be used with meaningful human control

36. Upon activation, an AWS proceeds to select and engage a target with force through the autonomous process. This raises serious ethical, legal, humanitarian and security risks. To safeguard against these risks, AWS that cannot be used with meaningful human control must be prohibited.

Risks and Challenges

37. Upon activating an AWS, the human user(s) does not determine when, where or against what, force is applied. Without sufficient controls, this process is likely to create unpredictable outcomes and undermine the context-based, dynamic, multidimensional, and situation-dependent requirements needed in the use of force to comply with international law.

38. This unpredictability poses fundamental challenges to the application of IHL. IHL prohibits weapons that are by nature indiscriminate, that is, weapons that cannot be directed at a specific military objective or whose effects cannot be limited. If the human user of an autonomous weapons system cannot reasonably anticipate what would trigger an application of force, or cannot control or limit the system's effects, its use would undermine the prohibition of weapons that are indiscriminate by nature.

39. To comply with IHL principles, including the principles of distinction, precaution, and proportionality, a degree of control is needed over the use of force. This includes an ability to determine with reasonable certainty the target type, geographical location and spatial expanse, duration and extent of the force applied. If the user cannot dictate the effects of the weapon system or the system does not function reliably as intended, the user cannot ensure compliance with these principles. For example, with regard to the proportionality principle, AWS without meaningful human control lack human judgement to engage in case-by-case balancing tests of the infinite number of ever-changing scenarios on the battlefield, and they cannot be preprogramed to deal with them. A lack of control would also present challenges for compliance with IHRL, especially the right not to be arbitrarily deprived of life.

40. In using an AWS, the human user(s) may lack understanding of how the process is functioning. The functioning of an AWS could be opaque, notably if it relies on artificial intelligence and machine learning techniques, or because it changes during use in a way that affects the use of force (e.g., machine learning enables changes to targeting parameters over time). If an AWS' functioning is opaque, then humans responsible for the application of IHL rules – both persons entrusted with the legal review of an AWS and persons responsible for compliance with IHL during its use – could not reasonably determine its lawfulness under IHL. If the human user is unable to understand how the system functions or explain why a particular person or object was struck, this would undermine the ability to hold perpetrators of IHL violations to account.

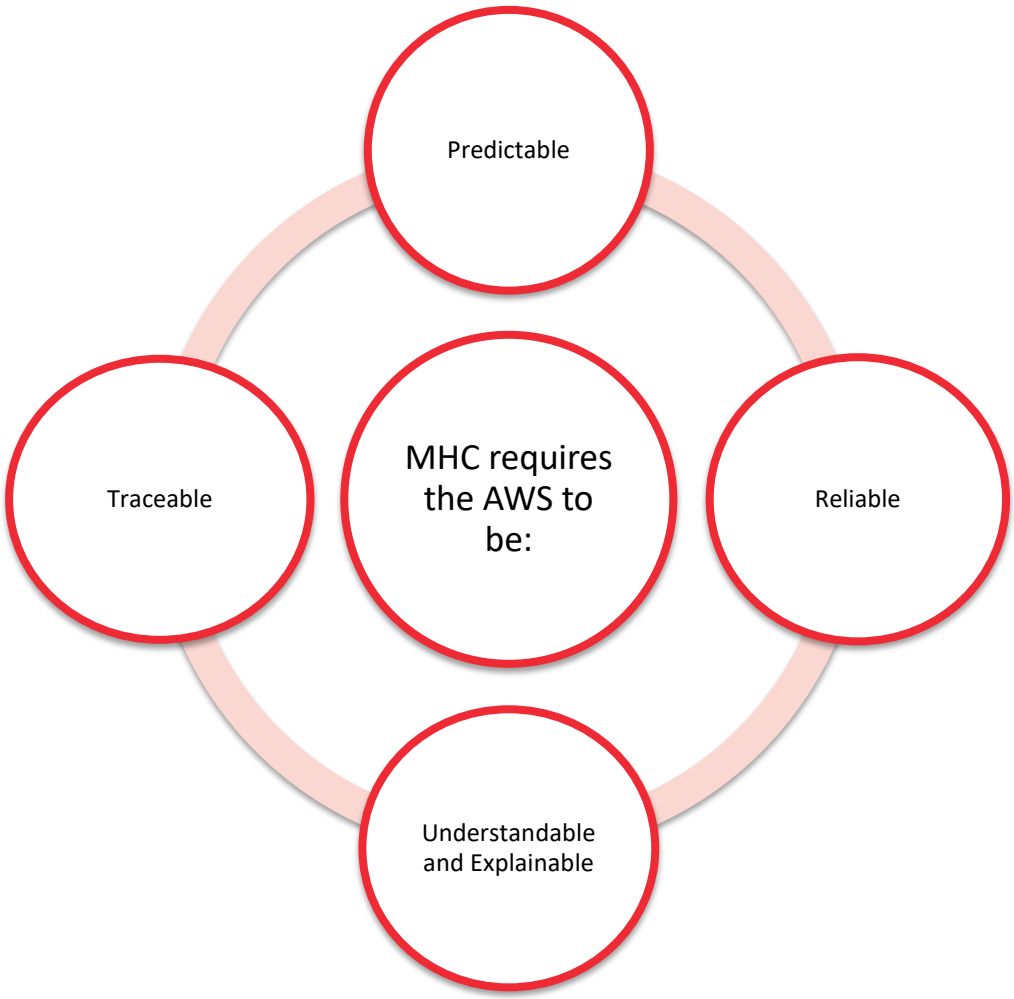
41. AWS lacking meaningful human control thus create a responsibility gap under international law. Responsibility for compliance with international law lies with humans and not systems. However, if an AWS is used in contravention of international law, attributing

responsibility may be complicated as the user does not determine exactly when, where or against what, force is applied. The development of an AWS may be distributed across several actors, including different teams of individuals or different producers and for different purposes. Tracing responsibility back to one of these individuals and identifying who among the various programmers, designers, data labellers and others are responsible for a certain act or omission could be a significant challenge. A complicating factor is that military commanders could adapt the parameters of an AWS during deployment. This difficulty in determining responsibility would make accountability for breaches of international law more complex.

42. The use of AWS also poses risks relating to their reliability of operation. Reliability encompasses the principles of both safety and security. Safety refers to the proper internal functioning of a system and the avoidance of unintended harm, while security addresses external threats. If an AWS' sensors malfunction, or the processing of sensor data is incorrect, the AWS may be unsafe to use and could cause unlawful harm. If an AWS is vulnerable to being hacked, or interrupted by an external variable, it could be insecure and result in dangerous outcomes contrary to the user's intent.

Requirement

43. To safeguard against such risks, AWS that cannot be used with meaningful human control must be prohibited. To be able to use the system with meaningful human control, the autonomous weapons system must be sufficiently **predictable, reliable, understandable and explainable, and traceable.**



- **Predictable**

44. An AWS must be sufficiently predictable. This means that the human user(s) must be able to reasonably anticipate how the system will function (so that it its effects are in accordance with the user's intent).

- **Reliable**

45. An AWS must be sufficiently reliable. This means that it must be capable of performing as expected to, consistently, safely, and securely.

- **Understandable and Explainable**

46. An AWS must be sufficiently understandable and explainable. This means that the human user(s) must have a sufficient understanding of how the system functions and what circumstances will trigger an application of force, to comply with legal and other requirements. Retrospectively, human users must be able to provide an explanation that satisfies legal and other requirements as to why a particular object or person was struck by the AWS.

- **Traceable**

47. An AWS must be sufficiently traceable. This means that the AWS must be designed and used in a way that ensures responsibility of human individuals for wrongful outcomes.

48. If an AWS does not satisfy of all these components, then it cannot be used with meaningful human control, and its use should be prohibited.

III. Regulations

49. Regulations, including positive obligations and limits, are required to ensure that AWS can be used with meaningful human control i.e. with sufficient **predictability, reliability, understandability and explainability, and traceability.** These regulations are required at both the development and use stages, to ensure each component of MHC is satisfied. These include:

Predictability: *The human user must be able to reasonably anticipate how the system will function (so that it its effects are in accordance with the user's intent).*

Regulations for development of AWS	Limits shall be placed on the type of target against which an AWS can be used. AWS should generally be constrained to pursue only targets that are military objectives "by nature" (see Article 52.2 of Additional Protocol I to the Geneva Conventions), in other words targets whose legal qualification as a military objective is relatively stable and unlikely to change between the activation of an AWS and the ensuing strike. Examples include munitions, military radar, or military naval vessels. This will reduce the risk that circumstances may change during an attack such as an object becoming protected under IHL in between the activation of an AWS and the eventual application of force. Limits shall be placed on the duration of an AWS operation. Generally, AWS should only be capable of operating for short periods of time. Limits shall be placed on the spatial parameters of an AWS operation. Generally, AWS should only be able to operate within a limited space. Limits shall be required on the scale or number of strikes an AWS may conduct. Limits shall be required on data processing methods. The AWS shall incorporate features that allow for human supervision and provide a human with the possibility to intervene after the activation of the system, where necessary to interrupt and deactivate the system.

Predictability: *The human user must be able to reasonably anticipate how the system will function (so that it its effects are in accordance with the user's intent).*	
Regulations for use of AWS	Limits shall be placed on the operational context in which a human will be allowed to use an AWS. This will enable users to have the necessary real-time situational awareness to predict the consequences of an AWS strike, such that the likelihood of the system selecting and engaging a wrong target is minimized. For example, even if the target is a military objective, it may not be targetable at a particular time due to the risk of disproportionate harm to civilians nearby. AWS should not be used in dynamic, congested or complex civilian environments such as cities or towns; instead, they should only be used in places where civilians and civilian objects are not present, such as on the high seas far from shipping lanes or fishing areas

Reliability: *To ensure that the system performs as expected to, consistently, safely and securely.*	
Regulations for development of AWS	The AWS shall be designed such that its proper internal functioning, as expected, is guaranteed and the possibility of causing unintended harms is minimized. The AWS shall be designed so as to minimise the risk of external threats, including by embedding safeguards against hacking or data spoofing.
Regulations for use of AWS	Requirements for the user to verify the correct functioning of the AWS before using it. The potential impact of contextual factors on the functioning of the autonomous process shall be considered before activating the weapons system. For example, the weather (sunny/windy/rainy) could lead to variations in the functioning of certain sensors of an AWS.

Understandability and Explainability: *To ensure the human user has a sufficient understanding of how the system functions and what circumstances will trigger an application of force, to comply with legal and other requirements.*	
Regulations for development of AWS	Limitations shall be placed on the complexity of data processing methods and AI / machine learning-based data processing shall be avoided. Indeed, such processes may be too complex, unexplainable and unpredictable for the user of a weapon to understand how it functions or what its limits are, in order to effectively predict and control the consequences of its use. The AWS shall be designed in such a way that it is not able to change mission parameters without sufficiently informed / meaningful / mindful human validation. The AWS shall be developed in conjunction with detailed manuals explaining how the system functions.
Regulations for use of AWS	Clear procedures shall be put in place to ensure that human users receive adequate training about the functioning of the system, including what the system might identify as a target, the limitations of the sensors being used by the system, the data processing methods at work, their known limitations and uncertainties about their functioning, and the circumstances that will trigger an application of force. Here it must be highlighted that, no matter how sophisticated sensors may be, sensor readings do not in themselves provide an effective proxy for that which triggers an AWS strike and are not equivalent to a human assessment of what an object is, where it is, what it is being used for and who is nearby. In a simple

Understandability and Explainability: *To ensure the human user has a sufficient understanding of how the system functions and what circumstances will trigger an application of force, to comply with legal and other requirements.*	
	example, a pressure sensor on an anti-vehicle mine is not a realistic proxy for a military vehicle (as it can also be triggered by a civilian vehicle).

Traceability: *To ensure that the AWS are designed and used in a way that ensures responsibility of human individuals for wrongful outcomes.*	
Regulations for development of AWS	Technical features of AWS that facilitate traceability, such as digital logs, shall be embedded. Mechanisms that could facilitate the task of tracing specific conduct back to one or more agents involved in the decision-making process shall be audited. The AWS shall be designed in such a way as to permit the attribution of responsibility for the consequences of its development and use to individuals and States under international law.
Regulations for use of AWS	Parameters of AWS use shall be set in such a way that responsibility for the consequences of use can be attributed to individuals and States under international law.

IV. Legally binding instrument

50. An international legally binding instrument on AWS is required with a combination of both prohibitions and regulations, as set out above. Existing legal rules are insufficient to address the serious risks and challenges posed by AWS.

51. A legally binding instrument would provide a durable framework offering the benefit of legal certainty and stability for the development and use of AWS now and in the future.

52. In addition, new legally binding rules would strengthen existing requirements for weapon reviews under Article 36 of Additional Protocol I of the Geneva Conventions. These reviews provide an obligation to review weapons, means and methods of warfare against States' legal obligations. The establishment of international legal obligations relating to the development and use of AWS would promote standardization of States' review processes around the world and help ensure that they are conducted transparently and consistently.

53. A non-binding instrument would risk promoting the widespread development and use of AWS, absent clear legal and ethical boundaries of acceptability. Without universalization of clear legal standards, there would be significant discrepancies in how States apply any non-binding practices, resulting in AWS being considered acceptable in one State, but unacceptable in another. Non-binding principles and practices should not therefore be agreed prior to establishing an international legally binding instrument.