ARTIFICIAL INTELLIGENCE: ISSUES AND CHALLENGES

Written by ***Dr. Abhinandan Bassi***

Assistant Professor of Law, Rajiv Gandhi National University of Law, Punjab, India

ABSTRACT

New technologies are changing human interaction profoundly – including in times of armed conflict. Many States are investing heavily in the development of means and methods of warfare that rely on digital technology.

Cyber tools, increasingly autonomous weapon systems, and artificial intelligence are being used in contemporary armed conflicts. Technological advances can have positive consequences for the protection of civilians in armed conflict: weapons can be used with more precision, military decisions can be better informed, and military aims can be achieved without the use of kinetic force or physical destruction.

At the same time, new means of warfare and the way they are employed can pose new risks to combatants and civilians, and can challenge the interpretation and implementation of IHL.

IHL is applicable to the development and use of new weaponry and new technological developments in war- fare – whether they involve (a) cyber technology; (b) autonomous weapon systems; (c) artificial intelligence and machine learning; or (d) outer space. States that develop or acquire such weapons or means of warfare are responsible for ensuring that they can be used in compliance with IHL

ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) systems are computer programs that carry out tasks – often associated with human intelligence – that require cognition, planning, reasoning or learning. Machine learning systems are AI systems that are "trained" on and "learn" from data, which ultimately define the way they function. Both are complex software tools, or algorithms, that can be applied to many different tasks. However, AI and machine learning systems are distinct from the "simple" algorithms used for tasks that do not require these capacities. The potential implications for armed conflict – and for the ICRC's humanitarian work are broad. There are at least three overlapping areas that are relevant from a humanitarian perspective

Areas of Artificial Intelligence:

The first area is the use of AI and machine learning tools to control military hardware, in particular the growing diversity of unmanned robotic systems – in the air, on land, and at sea. AI may enable greater autonomy in robotic platforms, whether armed or unarmed. For the ICRC, autonomous weapon systems are the immediate concern. AI and machine learning software – particularly for "automatic target recognition" – could become a basis for future autonomous weapon systems, amplifying core concerns about loss of human control and unpredictability. However, not all autonomous weapons incorporate AI.

The second area is the application of AI and machine learning to cyber warfare:

AI-enabled cyber capabilities could automatically search for vulnerabilities to exploit, or simultaneously defend against cyber-attacks while launching counter-attacks, and could therefore increase the speed, number and types of attacks and their consequences. These developments will be relevant to discussions about the potential human cost of cyber warfare. AI and machine learning are also relevant to information operations, in particular the creation and spread of false information (whether intended to deceive or not). AI-enabled systems can generate "fake" information – whether text, audio, photos or video – that is increasingly difficult to distinguish from "real" information and might be used by parties to a conflict to manipulate opinion and influence decisions. These digital risks can pose real dangers for civilians

CYBER OPERATIONS, THEIR POTENTIAL HUMAN COST, AND THE PROTECTION AFFORDED BY IHL

"Cyber warfare" to mean operations against a computer, a computer system or net- work, or another connected device, through a data stream, when used as means or methods of warfare in the context of an armed conflict. Cyber warfare raises questions about precisely how certain provisions of IHL apply to these operations, and whether IHL is adequate or whether, building on existing law, it might require further development.

The use of cyber operations may offer alternatives that other means or methods of warfare do not, but it also carries risks. On the one hand, cyber operations may enable militaries to achieve their objectives without harming civilians or causing permanent physical damage to civilian infrastructure. On the other hand, recent cyber operations – which have been primarily conducted outside the context of armed conflict – show that sophisticated actors have developed the capability to disrupt the provision of essential services to the civilian population.

To develop a realistic assessment of cyber capabilities and their potential human cost in light of their technical characteristics, in November 2018 the ICRC invited experts from all parts of the world to share their knowledge about the technical possibilities, expected use, and potential effects of cyber operations.19 Cyber operations can pose a particular threat for certain elements of civilian infrastructure. One area of concern for the ICRC, given its mandate, is the health-care sector. In this regard, research shows that the health- care sector appears to be particularly vulnerable to direct cyber-attacks and incidental harm from such attacks directed elsewhere. Its vulnerability is a consequence of increased digitization and interconnectivity in health care. For example, medical devices in hospitals are connected to the hospital network, and biomedical devices such as pacemakers and insulin pumps are sometimes remotely connected through the internet. This growth of connectivity increases the sector's digital dependence and "attack surface" and leaves it exposed, especially when these developments are not matched by a corresponding improvement in cyber security. Critical civilian infrastructure – including electrical, water, and sanitation facilities – is another area in which cyber-attacks can cause significant harm to the civilian population. This infrastructure is often operated by industrial control systems. A cyber-attack against an industrial control system requires specific expertise and sophistication, as well as specifically designed cyber tools.

While attacks against industrial control systems have been less frequent than other types of cyber operations, their frequency is reportedly increasing, and the severity of the threat has evolved more rapidly than anticipated only a few years ago. Beyond the vulnerability of specific sectors, there are at least three technical characteristics of cyber operations that are cause for concern

First, cyber operations carry a risk of overreaction and escalation, simply due to the fact that it may be extremely difficult – if not impossible – for the target of a cyber-attack to detect whether the attacker's aim is to spy or to cause physical damage. As the aim of a cyber operation might be identified only after the target system has been harmed, there is a risk that the target will imagine the worst-case scenario and react much more strongly than it would have done if it had known that the attacker's true intent was limited to espionage, for example.

Second, cyber tools and methods can proliferate in a unique manner, one that is difficult to control. Today, sophisticated cyber-attacks are carried out only by the most advanced and best-resourced actors. But once a cyber tool has been used, stolen or leaked, or becomes available in some other way, actors other than those who developed it may be able to find it, reverse-engineer it, and repurpose it for their own – possibly malicious – ends.

Third, while it is not impossible to determine who created or launched a particular cyber-attack, attributing an attack tends to be difficult. Identifying actors who violate IHL in cyberspace and holding them responsible is likely to remain challenging. The perception that it will be easier to deny responsibility for such attacks may also weaken the taboo against their use – and may make actors less scrupulous about violating international law by using them. While cyber operations have exposed the vulnerability of essential services, they have not, fortunately, caused major human harm so far. However, much is unknown in terms of technological evolution, the capabilities and the tools developed by the most sophisticated actors, and the extent to which the increased use of cyber operations during armed conflicts might be different from the trends observed so far

The limits that IHL sets for cyber warfare:

IHL urges all States to recognize the protection that IHL offers against the potential human cost of cyber operations.

For example, belligerents must respect and protect medical facilities and personnel at all times, which means that cyber-attacks against the health-care sector during armed conflict would – in most cases – violate IHL. Likewise, IHL specifically prohibits attacking, destroying, removing or rendering useless objects indispensable to the survival of the civilian population. More generally, IHL prohibits directing cyber-attacks against civilian infrastructure, as well as indiscriminate and disproportionate cyber-attacks. For instance, even if the infrastructure or parts of it become military objectives (such as a discrete part of a power grid), IHL requires that only those parts be attacked, and that there be no excessive damage to the remaining civilian parts of the grid or to other civilian infrastructure relying on the electricity provided by the grid. IHL also requires parties to conflict to take all feasible pre- cautions to avoid or at least minimize incidental harm to civilians and civilian objects when carrying out a cyber-attack.

Notwithstanding the interconnectivity that characterizes cyberspace, the principles of distinction, proportionality and precautions can and must be respected. A careful examination of the way cyber tools operates shows that they are not necessarily indiscriminate. While some of the cyber tools that we know of were designed to self-propagate and indiscriminately affect widely used computer systems, they did not do these things by chance: the ability to self-propagate usually needs to be specifically included in the design of such tools. Furthermore, attacking specific targets may require custom-made cyber tools, which might make it difficult to carry out such attacks on a large scale or indiscriminately

IHL prohibits directing cyber-attacks against civilian infrastructure, as well as indiscriminate and disproportionate cyber-attacks. For instance, even if the infrastructure or parts of it become military objectives (such as a discrete part of a power grid), IHL requires that only those parts be attacked, and that there be no excessive damage to the remaining civilian parts of the grid or to other civilian infrastructure relying on the electricity provided by the grid. IHL also requires parties to conflict to take all feasible pre- cautions to avoid or at least minimize incidental harm to civilians and civilian objects when carrying out a cyber attack

Notwithstanding the interconnectivity that characterizes cyberspace, the principles of distinction, proportionality and precautions can and must be respected. A careful examination of the way cyber tools operates shows that they are not necessarily indiscriminate. While some

of the cyber tools that we know of were designed to self-propagate and indiscriminately affect widely used computer systems, they did not do these things by chance: the ability to self-propagate usually needs to be specifically included in the design of such tools. Furthermore, attacking specific targets may require custom-made cyber tools, which might make it difficult to carry out such attacks on a large scale or indiscriminately.

In fact, many of the cyber-attacks that have been observed appear to have been rather discriminate from a technical perspective. This does not mean they were lawful or would have been lawful if carried out in a conflict; on the contrary, in the ICRC's view, a number of the cyber-attacks that have been reported in public sources would be prohibited during armed conflict. However, their technical characteristics show that cyber operations can be very precisely designed to have an effect only on specific targets, which makes them cap- able of being used in compliance with IHL principles and rules. IHL rules protecting civilian objects can, however, provide the full scope of legal protection only if States recognize that cyber operations that impair the functionality of civilian infrastructure are subject to the rules governing attacks under IHL.

Moreover, data have become an essential component of the digital domain and a cornerstone of life in many societies. However, different views exist on whether civilian data should be considered as civilian objects and therefore be protected under IHL principles and rules governing the conduct of hostilities. In the ICRC's view, the conclusion that deleting or tampering with essential civilian data would not be prohibited by IHL in today's ever more data-reliant world seems difficult to reconcile with the object and purpose of this body of law.. Finally, parties to armed conflicts must take all feasible precautions to protect civilians and civilian objects under their control against the effects of attacks. This is one of the few IHL obligations that States are required to implement in peacetime. Affirming that IHL applies to cyber warfare should not be misunderstood as encouragement to militarize cyberspace or as legitimizing cyber warfare. Any use of force by States, whether cyber or kinetic in nature, will always be governed by the UN Charter and relevant rules of customary international law. IHL affords the civilian population an additional layer of protection against the effects of hostilities. In the coming years, the ICRC will continue to follow the evolution of cyber operations and their potential human cost, in particular during armed conflicts. It will explore avenues to reduce that cost and work towards building consensus on the interpretation of existing IHL

rules and, if necessary, on the development of complementary rules that afford effective protection to civilians

The use of digital technology during armed conflicts for purposes other than as means and methods of warfare:

In recent conflicts, certain uses of digital technology other than as means and methods of warfare have led to an increase in activities that adversely affect civilian populations. For example, misinformation and disinformation campaigns, and online propaganda, have fused on social media, leading in some contexts to increased tensions and violence against and between communities. Unprecedented levels of surveillance of the civilian population have caused anxiety and increasing numbers of arrests, in some instances possibly based on disinformation. Disinformation and surveillance are not unique or new to armed conflicts; however, the greater scope and force-multiplying effect provided by digital technology can exacerbate – and add to – the existing vulnerabilities of persons affected by armed conflicts. Developments in artificial intelligence and machine learning are also relevant in this regard. IHL does not necessarily prohibit such activities, but it does prohibit acts or threats of violence the primary purpose of which is to spread terror among the civilian population. Moreover, parties to armed conflict must not encourage violations of IHL. Other bodies of law, including international human rights law, might also be relevant when assessing surveillance and disinformation

IHL does not necessarily prohibit such activities, but it does prohibit acts or threats of violence the primary purpose of which is to spread terror among the civilian population. Moreover, parties to armed conflict must not encourage violations of IHL. Other bodies of law, including international human rights law, might also be relevant when assessing surveillance and disinformation.

The global digital transformation is changing not only warfare but also the nature of humanitarian action. Digital technologies can be leveraged to support humanitarian programmes, for instance by capturing and using data to inform and adjust responses or by facilitating two-way communication between humanitarian staff and populations affected by conficts.24 For example, the ICRC analyses “big data” to anticipate, under- stand, and respond to humanitarian crises, and uses internet-based tools to interact with beneficiaries as well as

with parties to armed conflicts. The ICRC also uses digital tools to restore family links and, if possible, to facilitate communication between detainees and their loved ones; the ICRC does all this also to help parties to implement their IHL obligations. These new possibilities entail new responsibilities: humanitarian organizations need to strengthen their digital literacy and data-protection measures, in accordance with the "do no harm" principle.25 The ICRC encourages further research, discussion, and concrete steps by all actors to enable humanitarian actors to safely adapt their operations to digital changes

AUTONOMOUS WEAPON SYSTEMS

The autonomous weapon systems as: Any weapon system with autonomy in its critical functions. That is, a weapon system that can select and attack targets without human intervention. Autonomy in critical functions – already found in some existing weapons to a limited extent, such as air defence systems, active protection systems, and some loitering weapons – is a feature that could be incorporated in any weapon system.

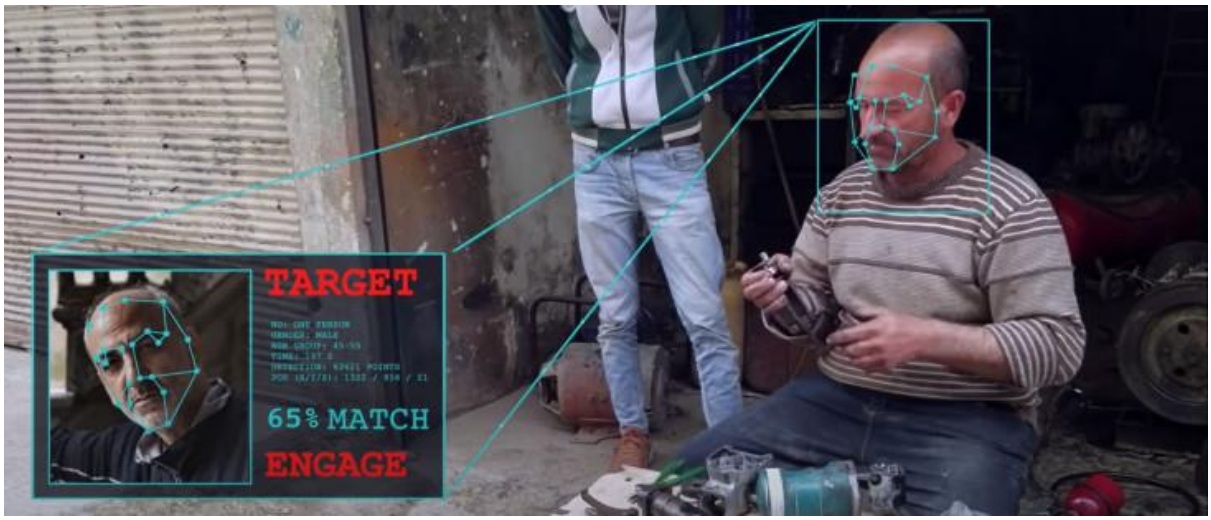
The most important aspect of autonomy in weapon systems – from a humanitarian, **legal and ethical perspective** – is that the weapon system self-initiates, or triggers, an attack in response to its environment, based on a generalized target profile. To varying degrees, the user of the weapon will know neither the specific target nor the exact timing and location of the attack that will result. Autonomous weapon systems are, therefore, clearly distinguishable from other weapon systems, where the specific timing, location and target are chosen by the user at the point of launch or activation.

Depending on the constraints under which a system operates, the user's uncertainty about the exact timing, location and circumstances of the attack(s) may put civilians at risk from the unpredictable consequences of the attack(s). It also raises legal questions, since combatants must make context specific judgements to comply with IHL. And it raises ethical concerns as well, because human agency in decisions to use force is necessary in order to uphold moral responsibility and human dignity

A person activates an autonomous weapon, but they do not know specifically who or what it will strike, nor precisely where and/or when that strike will occur. This is because an

autonomous weapon is triggered by sensors and software, which match what the sensors detect in the environment against a 'target profile'.

For example, this could be the shape of a military vehicle or the movement of a person. It is the vehicle or the victim that triggers the strike, not the user.

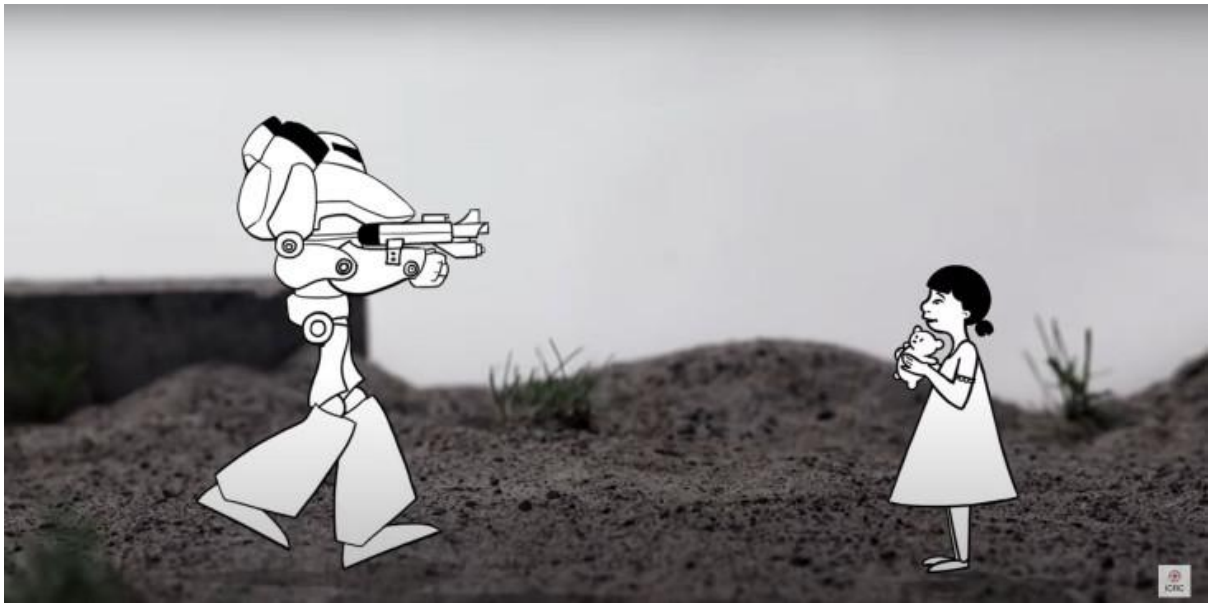


Our concern with this process is the loss of human judgement in the use of force. It makes it difficult to control the effects of these weapons.

The way autonomous weapons function i.e., where the user does not choose the specific target or the precise time or location of a strike makes this difficult. Under what conditions could users of an autonomous weapon be reasonably certain that it will only be triggered by things that are indeed lawful targets at that time and will not result in disproportionate harm to civilians?

Autonomous weapons also raise challenges from the perspective of legal responsibility. When there are violations of international humanitarian law, holding perpetrators to account is crucial to bring justice for victims and to deter future violations. Normally investigations will look to the person who fired the weapon, and the commanding officer who gave the order to attack.

There are many questions about whether alleged perpetrators of war crimes could be held responsible under existing legal regimes.



ETHICAL CONCERNS

Most fundamentally, there are widespread and serious concerns over ceding life-and-death decisions to sensors and software. Humans have a moral agency that guides their decisions and actions, even in conflicts where decisions to kill are somewhat normalized. Autonomous weapons reduce – or even risk removing – human agency in decisions to kill, injure and destroy. This is a dehumanizing process that undermines our values and our shared humanity. All autonomous weapons that endanger human beings raise these ethical concerns, but they are particularly acute with weapons designed or used to target human beings directly.

CONCLUSION

There is increasing interest in relying on AI, particularly machine learning, to control autonomous weapons. Machine learning software is 'trained' on data to create its own model of a particular task and strategies to complete that task. The software writes itself in a way. Often this model will be a 'black box' – in other words extremely difficult for humans to predict, understand, explain and test how, and on what basis, a machine-learning system will reach a particular assessment or output. As is well known from various applications, for example in policing, machine learning systems also raise concerns about encoded bias, including in terms of race, gender and sex.

With all autonomous weapons the policy should be made to make the responsibility of the product manufacturer and also the person who is in the control of a machine in the case of any causality if it happens.

REFERENCES

1. DoD, Summary of the 2018 Department of Defense Artificial Intelligence Strategy, 2019.

2. DoD, Defense Innovation Board, AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense, 31 October 2019.

3. DoD, "DOD Adopts Ethical Principles for Artificial Intelligence", news release, 24 February 2020, available at: www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

4. French Ministry of Defence, "Florence Parly Wants High-Performance, Robust and Properly Controlled Artificial Intelligence", Actualities, 10 April 2019, available at: www.defense.gouv.fr/english/actualites/ articles/florence-parly-souhaite-une-intelligence-artificielle-performante-robuste-et-maitrisee.

5. ICRC, ICRC Strategy 2019–2022, Geneva, 2018, p. 15, available at: www.icrc.org/en/publication/4354-icrcstrategy-2019-2022.

6. ICRC, above note 8, p. 22. 44 Google, Perspectives on Issues in AI Governance, January 2019, pp. 23–24, available at: http://ai.google/ perspectives-on-issues-in-AI-governance.