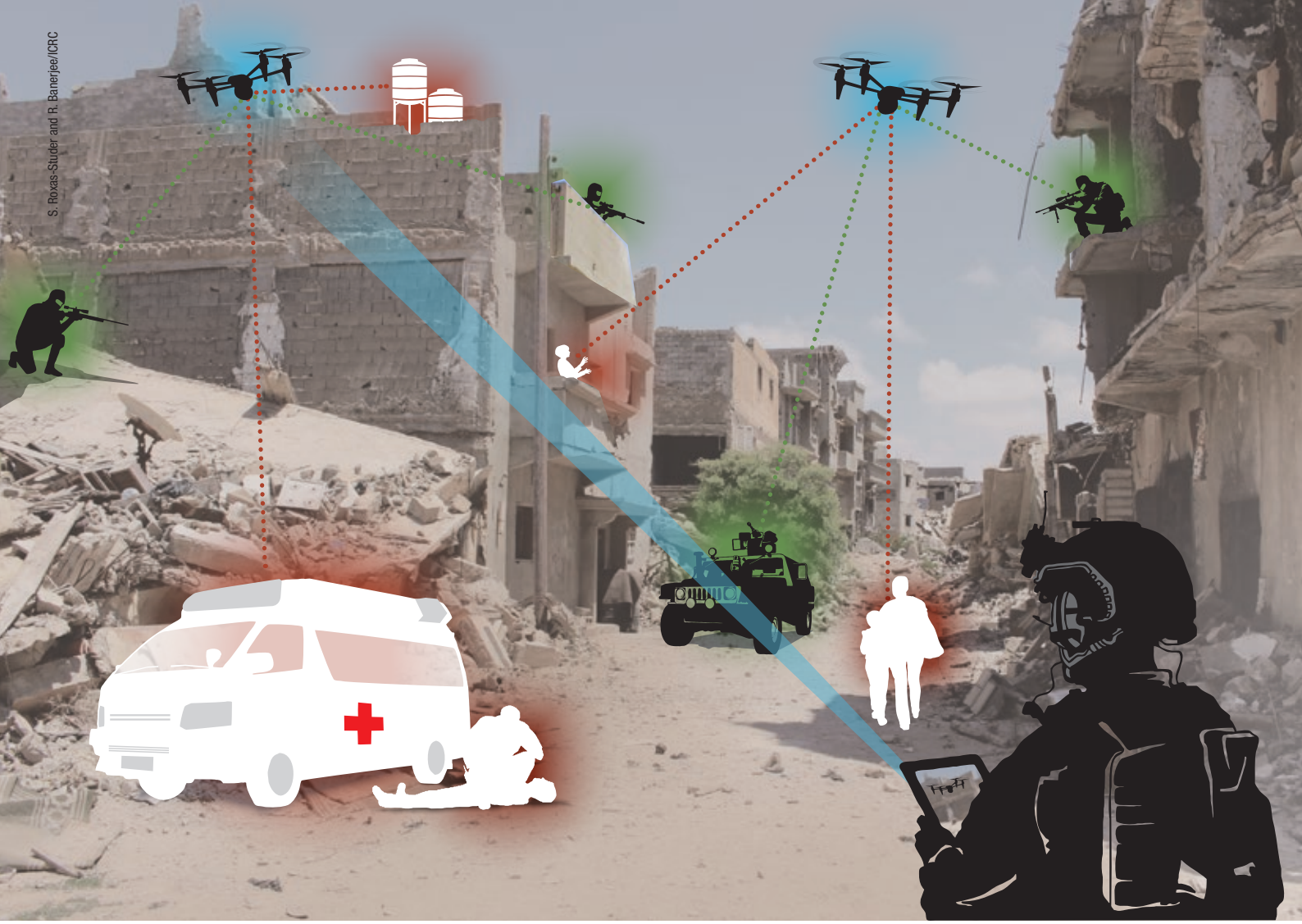
INTERNATIONAL HUMANITARIAN LAW AND THE CHALLENGES OF CONTEMPORARY ARMED CONFLICTS

BUILDING A CULTURE OF COMPLIANCE FOR IHL TO PROTECT HUMANITY IN TODAY'S AND FUTURE CONFLICTS

REPORT

ICRC

TABLE OF CONTENTS



V. APPLYING IHL TO NEW TECHNOLOGIES OF WARFARE

Today, civilians around the world rely on digital technologies in their daily lives: computers and smartphones, but also artificial intelligence, robotics, and outer space infrastructure. At the same time, parties to armed conflicts are making use of such technologies for military purposes. Some years ago, the use of digital technologies of warfare, including the use of artificial intelligence in military decision-making, may have seemed a distant prospect, but this is no longer the case.

The ICRC is concerned by the growing reliance on weapon systems with varying degrees of autonomy, and on systems that use artificial intelligence to inform decisions on who or what to attack and how. There are also worrying trends in contemporary armed conflicts in relation to the use of cyber operations by state and non-state actors to disrupt digital governance infrastructure, essential services, and economies; and in relation to the use of digital communication tools to extend the reach, accelerate, and expand the scale of information operations that fuel violence in violation of IHL. As essential civilian services become more and more dependent on space systems, attention must be paid to the potential human cost of military space operations, and to the legal limits they must respect.

In this chapter, the ICRC presents its legal views on some of these current challenges in applying IHL rules and principles to new technologies of warfare.

1. CYBER OPERATIONS, INFORMATION OPERATIONS AND OTHER DIGITAL THREATS

Societies are becoming increasingly digitalized and interconnected, and many aspects of people's daily lives today are defined or influenced by information and communication technology (ICT). In times of armed conflict, this has an impact on people's needs, and on the risks and threats they might face. For instance, essential services for civilian populations depend on ICT, and civilians rely on digital communication services to contact family members and obtain information about where to shelter or take refuge or obtain the goods and services essential to their survival and well-being. At the same time, state and non-state actors use cyber operations to disable civilian government services or disrupt the provision of essential services, such as electricity, water or medical care. Belligerents have used social media platforms and messaging services to incite violence against civilian populations and military personnel *hors de combat*, and more generally to dehumanize their adversaries. The digitalization of armed conflicts is also increasingly drawing civilians – individuals, hacker groups and tech companies – closer to hostilities, which exposes them as well as other civilians, to the risk of harm.

A) IHL LIMITS ON CYBER OPERATIONS

All states agree that international law applies to the use of ICT. States have also explicitly noted in the ICT context that "international humanitarian law applies only in situations of armed conflict", underscoring that IHL principles "by no means legitimize or encourage conflict".[166] This agreement affirms the consensus among legal experts, including the ICRC.[167]

Existing principles and rules of IHL do restrict cyber operations during armed conflicts, but the ICRC is concerned that technological developments and the use of such cyber operations are outpacing normative discussions and developments. In particular, interpretations of IHL that focus on the protection of civilian objects only against physical damage are, in the ICRC's view, inadequate.

166 United Nations General Assembly, *Report of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security* (14 July 2021), para. 71(f); United Nations General Assembly, Resolution adopted on 8 December 2021 (*A/RES/76/19*), para. 2.

167 For further discussion, see ICRC, "International humanitarian law and cyber operations during armed conflicts": Position paper, ICRC, Geneva, 2019, p. 4: https://www.icrc.org/en/document/international-humanitarian-law-and-cyber-operations-during-armed-conflicts.

Cyber operations can disrupt, disable, or physically damage essential civilian services and infrastructure, industrial facilities, communication networks, civilian databases and other civilian sectors of society. They risk injuring or killing people, and jeopardizing assistance to those who need it. As most cyber operations conducted in contemporary armed conflicts disrupt services, disable computers and networks or damage or delete data without causing physical damage, interpreting IHL in light of this reality is critical. Today, many states, whose legal positions on this issue are publicly available, take the view that cyber operations that disable objects, including IT systems or infrastructure, amount to 'attacks' under IHL. Others understand the notion of 'attack' under IHL more narrowly or leave the question open.[168] If the notion of 'attack' under IHL is interpreted as covering only cyber operations that cause physical damage, or effects akin to those caused by kinetic warfare, then most cyber operations against civilian infrastructure would not be constrained by the most detailed IHL rules that originate in the principles of distinction, proportionality and precautions in attack and protect the civilian population and civilian objects.[169] Equally, if data are not regarded as an 'object' within the IHL sense of the term, most cyber operations that damage or delete civilian data would not be prohibited – which would be a reason for serious concern.

Such operations would still be subject to certain limitations under IHL. In particular, military cyber operations must not be directed against specifically protected objects such as medical facilities; and when conducting any military cyber operation, constant care must be taken to spare the civilian population and civilian objects. Directing disruptive cyber operations against civilian objects, including civilian data, or ignoring their incidental effects on civilian populations, would be incompatible with this rule.

Still, if existing rules of IHL are interpreted in ways that undermine the protective function of IHL in the ICT environment, by leaving unaddressed the new kinds of harm resulting from the use of ICTs during armed conflict, additional rules will have to be developed to strengthen the existing legal framework and ensure that it remains adequate for the purpose of setting limits on cyber and other digital operations during of armed conflicts.

B) IHL LIMITS ON INFORMATION OPERATIONS

Information operations have long been conducted in the context of armed conflicts.[170] In recent years, technological developments that allow the instantaneous transmission of information from any distance through digital means, including via social media platforms and messaging apps, have changed the scale, speed, and reach of misleading, inaccurate, hateful or otherwise harmful information. While causal relationships are inherently difficult to demonstrate in this context, information operations are recognized as, among others, having the potential to contribute to or incite violence against people, cause lasting psychological harm, undermine access to essential services, and disrupt the operations of humanitarian actors.[171]

While harmful information is often spread through information operations during armed conflicts, IHL contains several specific rules that impose limits on information-sharing more broadly. For example, civilian and military leadership of a party to an armed conflict must not encourage IHL violations, including through digital platforms.

The ubiquity of smartphone cameras and the widespread practice of publishing photos online have also put renewed strain on detaining authorities in armed conflict to fulfil their obligation to protect all detainees against humiliating and degrading treatment. In particular, prisoners of war and civilian internees must be protected against public curiosity.[172] The public sharing of data, images and videos of persons deprived of liberty will in most cases violate these rules.

168 For an overview of positions taken by states, see Cyber Law Toolkit, 'Attack (International Humanitarian Law): https://cyberlaw.ccdcoe.org/wiki/Attack_(international_humanitarian_law).

169 For further discussion of the ICRC's views on the notion of 'attack' under IHL and the protection of data under IHL, see ICRC, "International humanitarian law and cyber operations during armed conflicts": Position paper, 2019, pp. 7–8.

170 The ICRC understands 'information operation' to mean the use or manipulation of information to influence or mislead the perceptions, motives, attitudes and behaviour of individuals and groups, in order to achieve political and military objectives.

171 See generally, ICRC, *Harmful Information: Misinformation, Disinformation and Hate Speech in Armed Conflict and other Situations of Violence*, ICRC, Geneva, 2021: https://www.icrc.org/en/publication/4556-harmful-information-misinformation-disinformation-and-hate-speech-armed-conflict.

172 See GC III, Art. 13; GC IV, Art. 27.

During armed conflict, the information space may also become fertile ground for the use of false information created through tools powered by artificial intelligence. For instance, deepfake technology can create or modify information, images, audios and videos in ways that make it difficult for people to distinguish them from authentic, original content. IHL rules set limits on certain uses of 'deepfakes'. For example, it is a violation of IHL "to kill, injure or capture an adversary by resort to perfidy", the latter being understood as "inviting the confidence of an adversary to lead him to believe that he is entitled to, or obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence".[173] Resorting to perfidy by harnessing deepfake technology is a violation of IHL. In addition, acts or threats of violence, the primary purpose of which is to spread terror among the civilian population, are prohibited under IHL,[174] including when using deepfakes.

The ICRC also recalls that in the conduct of military operations, including in information operations that make use of deepfakes, warring parties must take constant care to spare the civilian population, civilians and civilian objects.

C) RISKS AND LEGAL LIMITS WHEN CIVILIANS ARE DRAWN CLOSER TO HOSTILITIES THROUGH THE USE OF DIGITAL TECHNOLOGY

Civilians have long been used to perform tasks supporting the military during armed conflicts. With the digitalization of societies, fundamental shifts have taken place in the types of operations civilians conduct, and the number of civilian actors that take part in such operations. Three trends, in particular, pose risks for civilians. First, an unprecedented number of civilian hackers are conducting cyber operations in the context of armed conflicts, often directing their operations against civilian objects. Second, ICT presents belligerents with new possibilities to encourage civilians to support military operations, for instance by collecting militarily relevant information through their smartphones, thereby exposing civilians to attacks. Third, when civilian tech companies are hired to provide cyber security and other ICT services for the armed forces of parties to armed conflicts – such as connectivity, communication, cloud-computing, or remote sensing – there is a real risk that the assets, infrastructure, and employees of such companies – which are in principle civilian – lose their legal protection against attack.

If individuals and groups, including the employees of tech companies, conduct cyber operations in the context of armed conflicts, they must comply with the limits that IHL sets for such operations. Specifically with regard to civilian hackers operating in the context of armed conflicts, these limits have been summarized into "8 Rules For Civilian Hackers During War", together with four obligations for states to ensure respect for these rules.[175]

IHL is built on the cardinal principle of distinction between who or what is civilian, and who or what is military. Growing civilian involvement in cyber and information operations and the use of civilian ICT infrastructure for military purposes risk undermining this fundamental premise and the protection that it is meant to provide to civilians.

Collecting militarily relevant information through smartphones or other connected devices and providing it to armed forces may, in exceptional cases, amount to "direct participation in hostilities", meaning that a civilian loses their protection against attack if and for such time as this is the case. In the ICRC's view, however, this cannot mean that any civilian using their phone close to military positions or hostilities constitutes a lawful target. For the attacker, it is in most cases impossible to know whether a phone is being used for conduct that may qualify as direct participation in hostilities, or whether a civilian is doing something else, such as warning a friend or contacting a family member. IHL requires that, in case of doubt, a person must be considered civilian and protected as such.[176] Nonetheless, encouraging civilians to collect militarily relevant information risks putting the civilian population at risk.

173 AP I, Art. 37; Convention (IV) respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land (Hague Regulations), 23(B); ICRC, Customary IHL Study, Rule 65.

174 AP I, Art. 51(2); ICRC, Customary IHL Study, Rule 2.

175 ICRC, "8 rules for "civilian hackers" during war, and 4 obligations for states to restrain them": https://www.icrc.org/en/article/8-rules-civilian-hackers-during-war-and-4-obligations-states-restrain-them.

176 AP I, Art. 50(1).

If civilian ICT infrastructure – including infrastructure provided by civilian companies – is used for military purposes, it risks becoming a military objective under IHL and losing its protection against attack. In that case, civilians and civilian objects that are in physical proximity or digitally connected to such targets, or that depend on them, risk being incidentally harmed. To protect civilians and civilian objects from attack or incidental harm, states should, whenever feasible, attempt to segment – that is physically or technically separate – ICT infrastructure (or parts thereof) that are used for military purposes from civilian ones. For example, when deciding whether to store military data on a non-segmented commercial cloud, a segment of a commercial cloud or dedicated military digital infrastructure, military planners and operators should not use the non-segmented commercial cloud.

Even if a belligerent concludes that a civilian or civilian object has lost legal protection against attack because of their involvement in cyber or information operations, the ICRC calls on belligerents to consider carefully whether responding to such threats by kinetic force is actually necessary to achieve a legitimate military purpose or whether other, less destructive (for example, cyber or electro-magnetic) means can be used to achieve their objective.[177]

Long-standing rules of IHL only serve their purpose if applied in ways that ensure adequate protection for civilians, civilian infrastructure, and civilian data in our increasingly digitalized societies. The evolving legal views and practice of states will indicate whether existing law is adequate and sufficient to address the challenges posed by the digitalization of armed conflicts, or whether it needs strengthening to address new dangers posed by this evolution. If new rules are to be developed, they must build on and strengthen the existing legal framework – including IHL.

2. AUTONOMOUS WEAPON SYSTEMS

The deployment of weapon systems with increasingly autonomous modes or functions, particularly small armed drones and loitering munitions,[178] is a fact of contemporary conflicts. The ICRC's assessment is that, generally, these kinds of weapon systems are still remotely piloted or guided. However, with just a software update or a change in military doctrine, they could easily become tomorrow's autonomous weapon systems (AWS), namely weapon systems that select and apply force to targets without human intervention. 'Without human intervention' means that after initial activation by a human, the application of force is triggered in response to information from the environment received through sensors (that measure, for instance, heat, light, movement, shape, velocity, weight or acoustic or electromagnetic signals), and on the basis of a generalized 'target profile' (based on such features as shape, infrared or radar 'signature', or speed and direction of a particular type of military vehicle).[179]

At the same time, there seems to be military interest in loosening the constraints on where, or against what, to use such weapons. This is a trend that may deepen a major concern for the ICRC: potential loss of human control over the use of force in armed conflict. Driven by the need to uphold and strengthen protections for people affected by armed conflict, the ICRC has called on states to urgently establish new international prohibitions and restrictions on AWS that are clear and binding.[180]

177 ICRC, *The Principles of Humanity and Necessity*, ICRC, Geneva, 2023: https://www.icrc.org/sites/default/files/wysiwyg/war-and-law/02_humanity_and_necessity-0.pdf.

178 A kind of aerial weapon that can hover over, detect and dive onto targets, and detonates on impact.

179 ICRC, Position on autonomous weapon systems, ICRC, Geneva, 2021, p. 2: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

180 ICRC, "Position on autonomous weapon systems", ICRC, Geneva, 2021: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems. See also, Joint call by the United Nations Secretary-General and the President of the ICRC, ICRC, Geneva, 2023: https://www.icrc.org/en/document/joint-call-un-and-icrc-establish-prohibitions-and-restrictions-autonomous-weapons-systems.

A) HUMANS MUST DETERMINE THE LAWFULNESS OF ATTACKS

Despite the growing development of AWS and associated sensor, software and robotics technologies, it is worth recalling that IHL obligations regarding the conduct of hostilities must be fulfilled by human commanders and combatants. These humans must determine the lawfulness of the attacks that they plan, decide upon or execute, and they remain accountable for these assessments. While some proponents of AWS describe the systems as making a 'decision', the decision to launch the weapon, and to carry out the attack, is always made by a human.[181] Accordingly, while certain technical tasks are carried out by machine processes, the determination of the lawfulness of an attack – including whether an object is a military objective – and whether these machine processes will be sufficient for the attack to comply with IHL, is made by a human. It is an exercise in false equivalence to compare a human decision to launch an attack with a machine process that triggers the application of force. A more accurate comparison would be between a human launching an attack using a weapon system that is not autonomous and a human launching an attack using an AWS.

When discussing AWS and compliance with IHL, it is therefore important to emphasize that it is not the weapon system that must comply with IHL, but the humans using it.

Additionally, from an ethical perspective, upholding human agency in critical decisions leading to the use of force is necessary for upholding considerations of humanity, human dignity and moral responsibility.

B) CHALLENGES IN ASSESSING THE LAWFULNESS OF ATTACKS CARRIED OUT USING AWS

Commanders and other users of AWS must make an *ex-ante* assessment of the lawfulness of an attack, and ensure that the weapon system can and will operate only within the confines of what they have assessed as lawful. However, their ability to do so can be limited by the particular way in which AWS function.

The key difficulty is that the user or commander will likely not know important specifics of the attack. That is because, after initial activation or launch by a person, an AWS self-initiates or triggers a strike in response to information from the environment received through sensors and on the basis of a generalized "target profile". The user does not choose, or even know, the specific person or object that will trigger the AWS to strike, or precisely when and/or where the strike will occur. These details will depend on the machine processes (sensors, software, actuators) of the AWS, and the sensory input from the operational environment. This can impede the user's ability to anticipate, and take steps to limit, the effects of an attack – to ensure, for example, that the attack is not indiscriminate.[182]

Commanders and other users thus need to anticipate and assess in advance the lawfulness of all *possible* strikes by the AWS. When doing so, they need to account for all reasonably foreseeable and relevant changes in circumstances, for the entire duration of the weapon system's autonomous operation and the entire area over which this will take place. Such an exercise is likely to be possible only if strict constraints are imposed on the variables applying to the AWS and its operating environment in order to limit the number of potential outcomes.

For instance, in light of the IHL obligation to direct attacks only against military objectives, users must ensure that anything that might trigger an AWS to strike, throughout its area and duration of operation, will satisfy the two-pronged definition of a 'military objective' "in the circumstances ruling at the time" of that specific strike.[183] This will be simplest when considering the potential targeting of objects whose legal classification as military objectives is relatively stable, namely military objectives by nature (the enemy's weapons, vehicles for transporting military equipment, barracks, etc). It will be extremely challenging, however, in the case of civilian objects that could be military objectives by location, purpose or use (e.g. a hill, a hotel temporarily used to accommodate troops, or a bridge about to be crossed by enemy forces). This is

181 ICRC, "Commentary on the 'Guiding Principles' of the CCW GGE on 'Lethal Autonomous Weapons [*sic*] Systems'", ICRC, Geneva, 2020, p. 3: https://documents.unoda.org/wp-content/uploads/2020/07/20200716-ICRC.pdf; see also ICRC, *2019 Challenges Report*, p. 30.

182 ICRC, Customary IHL Study, Rule 12; AP I, Art. 51(4)(c).

183 ICRC, Customary IHL Study, Rule 8; AP I, Art. 52(2): "[M]ilitary objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."

because the effectiveness of the contribution that such objects make to the adversary's military action, and the definite military advantage offered by their destruction, capture or neutralization, may vary significantly and rapidly. For example, a civilian taxi that is temporarily requisitioned to transport soldiers to the front line makes an effective contribution to the enemy's military action – and thus could be characterized as a military objective by use – only for the duration of this use. Its destruction at any other time is unlikely to offer any military advantage.

Furthermore, characterizing an object as a military objective by purpose requires ascertaining the enemy's intentions, the cues for which are nuanced, context-dependent and non-exhaustive, making them ill-suited to standardization in the kind of generalized target profile used by AWS.

Ensuring compliance with IHL would be especially difficult in the case of an attack directed against one or more persons. While the user or commander may have made a general assessment that one or more people in the area constitute a lawful target at the time of launching the AWS,[184] those people's actions and intentions – and hence their classification as a lawful target – can change rapidly and before the AWS strikes. Civilians may lose protection against direct attack only "for such time" as they directly participate in hostilities; determining the beginning and end of specific acts must therefore be done with the utmost care.[185] Similarly, combatants can be wounded or otherwise rendered *hors de combat* at any time – at which point they must not be directly attacked.

Moreover, anyone using an AWS against an enemy fighter in armed conflict would need to preserve a reasonable possibility for that person to surrender. Employing an AWS that prevents the user from recognizing an adversary's clear communication of intent to surrender – however that is expressed – and ceasing an attack, would violate the prohibition against conducting hostilities on the basis of leaving no survivors (denial of quarter).[186]

In short, it is hard to envisage realistic combat situations where the use of autonomous weapons against humans would not pose a significant risk of IHL violations. Therefore, if there is no clear prohibition against anti-personnel AWS, it will create an unacceptably high risk of such AWS being deployed without sufficient safeguards to respect IHL.

In addition to the legal challenges, the ethical concerns raised by AWS have been emphasized by many states, the UN Secretary-General, civil society and leading figures in the technology industry and scientific community. These concerns are centered on the interrelated loss of human agency, moral responsibility and human dignity in life-and-death decisions.

Humans have moral agency and responsibilities that guide their decisions and actions whereas inanimate objects (e.g. weapons, machines and software) do not. When human agency and human determination are absent, it can be said that there has been neither morally responsible decision-making nor recognition of the human dignity of those targeted or affected. Removing human agency in decisions about life and death also removes the possibility for restraint, a human quality that might guide people not to use force even if it would be lawful.

These concerns are most acute with AWS designed or used to target persons directly (as opposed to AWS that target objects). They reduce life-and-death determinations to sensor data and machine processing based on generalized target profiles, the consequence of which is that people are treated as mere targets instead of human beings. It would effectively amount to "death by algorithm" – the final frontier in the automation of killing.[187]

184 Whether as a combatant, fighter or civilian directly participating in hostilities, and not *hors de combat*.

185 ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, ICRC, 2009 (hereafter *Interpretive Guidance*, 2009): https://shop.icrc.org/interpretive-guidance-on-the-notion-of-direct-participation-in-hostilities-under-international-humanitarian-law-print-en.html.

186 AP I, Art. 40; ICRC, Customary IHL Study, Rule 46.

187 ICRC, "Position on autonomous weapon systems", ICRC, Geneva, 2021: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems (p. 8). See also ICRC, "Ethics and autonomous weapon systems: An ethical basis for human control?", ICRC, Geneva, 2018: https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control.

C) THE NEED FOR NEW INTERNATIONAL LAW RULES ON AWS

In light of the serious risks of harm to those affected by armed conflict, challenges for compliance with IHL and ethical concerns raised by them, the ICRC has, since 2021, been calling for new, binding international law rules on the development and use of AWS.[188] These rules should clarify and formalize specific prohibitions and restrictions concerning the design and use of AWS. Any such limits would be additional and complementary to existing IHL rules, including weapons treaties, and would not displace them. They would strengthen and build on existing legal protections in order to respond to the specific risks and ethical concerns raised by AWS.

In particular, new rules must:

- prohibit unpredictable autonomous weapons that do not allow a human user to understand, explain and predict the system's functioning and effects. Users of AWS must be able, with a reasonable degree of certainty, to predict the effects of the weapon, in order to determine whether it can be directed at a specific military objective and take steps to limit those predicted effects, as required by IHL. This entails the ability to understand the functioning of the AWS: the nature and functioning of its sensors, the definition of its target profile and the potential effects in the circumstances of use, including any risk of error or malfunction. Autonomous weapons that are likely to produce effects that are unpredictable include those controlled by machine-learning software, along with certain swarm technologies;
- prohibit autonomous weapons designed or used to target humans directly. This is required because of the significant risk of IHL violations and the unacceptability of anti-personnel autonomous weapons from an ethical perspective, as outlined above.

Even in the case of an AWS that is sufficiently predictable, and designed and used only against objects, the user's reduced ability to know all the specifics of an attack, including the ultimate target and any incidental harm, will still create residual challenges for their context-specific application of IHL's rules on the conduct of hostilities. To reduce the risk of violations, new rules must also strictly constrain the design and use of AWS, including through a combination of:

- restricting targets of the AWS to only those that are military objectives by nature;
- limiting the duration and geographic scope of the AWS' operation;
- limiting the scale of use, including the number of engagements that the AWS can undertake;
- limiting the situations of use, namely constraining them to situations where civilians or civilian objects are not present;
- ensuring, to the maximum extent feasible, the ability for a human user:
  - to effectively supervise; and
  - in a timely manner, to intervene and, where appropriate, deactivate operation of the AWS.

Where this is not feasible, the AWS must be equipped with an effective mechanism for self-destruction or self-neutralization. Against the backdrop of rapid and expanding development and use of AWS, the establishment of these prohibitions and restrictions on AWS, in clear and binding international law, is an urgent humanitarian priority. The ICRC, together with the UN Secretary-General, calls on states to take bold and principled political action to conclude negotiations of such rules by 2026.[189] The ICRC has submitted its views, for consideration by states and the UN Secretary-General, on how these rules could be drafted in a legally binding instrument.[190]

188 ICRC, "Position on autonomous weapon systems", ICRC, Geneva, 2021: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

189 Joint call by the United Nations Secretary-General and the President of the ICRC, 2023: https://www.icrc.org/en/document/joint-call-un-and-icrc-establish-prohibitions-and-restrictions-autonomous-weapons-systems.

190 ICRC submission on AWS to the UN secretary-general, 2024: https://www.icrc.org/en/document/autonomous-weapons-icrc-submits-recommendations-un-secretary-general.

3. ARTIFICIAL INTELLIGENCE IN MILITARY PLANNING AND DECISION-MAKING

Armed forces are investing heavily in artificial intelligence (AI). While AI technology can be incorporated in AWS (see section V. 2), one of the most widespread, and increasingly prominent, military applications of AI is in 'decision-support systems' (AI-decision-support systems). These are computerized tools that bring together data sources – such as satellite imagery, sensor data, social media feeds or mobile phone signals – and present analyses, recommendations or predictions based on them to decision makers.

In contemporary conflicts, one AI-decision-support system might analyse drone footage and apply image-classification technology to identify and classify potential targets. Its output might feed another system running simulations to recommend the 'optimal' weapon available to attack the target. These could also link to a system using predictive analytics to forecast how the adversary might respond to the attack. Such AI-decision-support systems can have a significant impact on human decisions about who or what to attack and where, when and how.[191]

Increased situational awareness and faster decision-making cycles are often cited as the advantages to be had potentially from the use of AI-decision-support systems. However, the ICRC has previously cautioned that the human cost of these technologies will depend on the way they are designed and used.[192] Importantly, the use of AI-decision-support systems can never ameliorate targeting methodologies and other policies that do not otherwise comply with IHL; applying AI-decision-support systems within such frameworks will serve only to replicate and likely exacerbate unlawful or otherwise harmful effects faster and on a larger scale.

A) UNDER IHL, HUMANS MUST MAKE LEGAL DETERMINATIONS

As outlined in the previous section, the ICRC takes the view that IHL requires individuals to make legal determinations, such as whether the expected incidental harm from an attack will be excessive in relation to the concrete and direct military advantage anticipated.

This is not to say that, in making these legal assessments, commanders and combatants cannot, or even that they should not, use tools – including AI-decision-support systems. In fact states have already adopted a broad range of military decision-making tools, at all levels, to assist members of their armed forces during the planning, ordering and conduct of attacks. In some states, for instance, the operational process of conducting an estimation of incidental civilian casualties is computerized, for feeding into an assessment of whether an attack will be proportionate under IHL. The important point is that these computer outputs can inform, but must not displace the need for legal determinations. In the ICRC's view, this means that in designing and using any AI-decision-support systems, militaries and other armed actors must account for the ways in which these AI tools function and the tendencies of human users interacting with them.

Integrating AI into decision-support systems can increase the rate of unforeseen errors, and perpetuate and propagate problematic biases, particularly against individuals or groups of a certain age, gender or ethnicity, or persons with disabilities. Trends indicate that these challenges will increase with more complex forms of AI, such as machine learning, which can make it more difficult, even impossible, for the user to understand

191 ICRC, *Artificial Intelligence and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts: Current Developments and Potential Implications*, ICRC, Geneva, 2024, pp. 8–9: https://shop.icrc.org/expert-consultation-report-artificial-intelligence-and-related-technologies-in-military-decision-making-on-the-use-of-force-in-armed-conflicts-current-developments-and-potential-implications-pdf-en.html; See also, ICRC, *Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making*, ICRC, Geneva, 2024, p. 20: https://shop.icrc.org/decisions-decisions-decisions-computation-and-artificial-intelligence-in-military-decision-making-pdf-en.html.

192 ICRC, "Position paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach", *IRRC*, Vol. 102, No. 913, April 2020: https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913.

how and *why* the system generates its output from a given input.[193] Moreover, when a number of different decision-support systems build on and contribute to decisions in a single process, an error in one can become compounding or cascade across a planning and decision-making process.

When humans interact with machine systems, they exhibit 'automation bias', meaning a propensity to trust machine outputs over other sources of information. This is most particularly the case in situations of stress or pressure, such as in armed conflicts.[194]

Taken together, these factors can hamper a user's ability to scrutinize the information available. The practical consequence might be that someone may plan, decide upon or launch an attack based on an AI-decision-support system's output, rather than actually assess the attack's lawfulness – thereby effectively serving as a human rubber stamp.

B) AI IS NOT SUITED TO ALL TASKS

Applying AI – particularly machine learning – to problems for which it is not well-suited can negatively impact human decision-making.

Generally, AI will perform better when given clear, well-defined goals and access to data of good quality. The contextual, qualitative assessments required by IHL are unlikely to produce clear goals for an AI-decision-support system; they are notoriously difficult and generally cannot be reduced to mathematical formulae or numerical values. Furthermore, armed conflicts are characterized by uncertainty and volatility, compounded by adversaries seeking to deceive one another, all of which makes it hard to source representative, transferable data.

An AI-decision-support system would be ill-suited to inferring something open-ended, such as the purpose behind a person's action (e.g. determining the 'belligerent nexus' in the context of direct participation in hostilities),[195] or an enemy's intention (e.g. assessing whether an object constitutes a military objective by purpose).[196] Similarly, predictions about enemy behaviour are likely to be unreliable. Application of AI-decision-support systems would be more appropriate when the possible outcomes are finite, and for which there are more and better test and simulation data. These AI-decision-support systems can be used, for instance to optimize own-force logistics or in transportation planning or choosing between available weapons.

In short, to ensure that an AI-decision-support system supports rather than hinders decision-making in armed conflict – and assists in ensuring respect for IHL – parties to conflict must carefully assess its suitability for the specific task and context. AI-decision-support systems may have to be ruled out altogether in some areas. One clear example would be that such tools must never be incorporated in nuclear-weapon command-and-control systems.[197]

193 ICRC, "Position paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach", *IRRC*, Vol. 102, No. 913, April 2020: https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913; See also, *Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making*, ICRC, Geneva, 2024, p. 31, 54: https://shop.icrc.org/decisions-decisions-decisions-computation-and-artificial-intelligence-in-military-decision-making-pdf-en.html.

194 ICRC, *Artificial Intelligence and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts: Current Developments and Potential Implications*, ICRC, Geneva, 2024, p. 17: https://shop.icrc.org/expert-consultation-report-artificial-intelligence-and-related-technologies-in-military-decision-making-on-the-use-of-force-in-armed-conflicts-current-developments-and-potential-implications-pdf-en.html.

195 *Interpretive Guidance*, 2009, p. 59.

196 ICRC, *Commentary on the Additional Protocols*, 1987, para. 2022.

197 ICRC, Statement to the 78th session of the UN General Assembly, First Committee General Debate, 11 October 2023.

C) POTENTIAL FOR AI-DECISION-SUPPORT SYSTEMS TO SUPPORT COMPLIANCE WITH IHL AND MITIGATION OF CIVILIAN HARM

At the same time, careful use of AI-based systems may facilitate quicker and more comprehensive information analysis, which can support decisions in a way that enhances IHL compliance and minimizes risks for civilians. In the context of urban warfare in particular, the ICRC has recommended that open-source repositories online should be used to gather information about the presence of civilians and civilian objects.[198] AI tools can likely assist in collecting and synthesizing such sources. The use of AI-decision-support systems to support weaponeering may also inform the choice of means and methods of attack that can best avoid, or at least minimize, incidental civilian harm.[199]

Importantly, IHL imposes obligations to take constant care to spare the civilian population and to take all feasible precautions in attack. Therefore, in developing and using AI-decision-support systems, armed forces should be considering not only how such tools can assist them to achieve military objectives with less civilian harm, but also how they might be designed and used specifically to protect civilians. This could include tools to recognize and track civilian populations and alert forces to their presence, or to recognize distinctive emblems or signals that indicate protected status.

As stated above, the efficacy of any such tools will depend on access to data of good quality. It appears that militaries are increasingly building and maintaining datasets to support target identification, but it is not clear whether they are making a corresponding investment in gathering data to support the identification of people and objects that are *not* lawful targets. States and other actors developing and deploying AI-decision-support systems must address this gap. The ICRC recommends prioritizing research and investment in tools and data that can facilitate better compliance with IHL and increase protection for civilians.

When drawing on an AI-decision-support system's output for targeting decisions, combatants and commanders must assess information from all sources reasonably available. Relying solely on the output of one AI-decision-support system is unlikely to meet this standard, especially during pre-planned targeting processes when more time is available to assess different sources of information. Commanders and users of AI-decision-support systems should therefore cross-check the outputs of these tools against all other available intelligence.

D) PRESERVING TIME AND SPACE FOR HUMAN DELIBERATION

One of the main military benefits of AI-decision-support systems that is touted, and is behind their development and use, is their ability to accelerate planning and decision-making processes, giving an advantage over the adversary. Increasing the speed of military operations can, however, create additional risks, for both civilians and combatants, including by increasing risks of miscalculation and escalation.

To alleviate these risks, planners and commanders have long employed practices such as 'tactical patience': deliberately pausing to allow a situation to unfold in order to increase situational awareness and develop more options. Parties to an armed conflict should consider how to maintain such practices even while employing AI-decision-support systems. This will likely require slowing down points of the planning and decision-making processes on purpose, in order to preserve time for deliberating about decisions on the conduct of hostilities.[200]

AI tools are having a significant influence on military planning and decision-making processes. They have the potential to facilitate decisions that minimize risks for people affected by armed conflict. States and non-state actors should consider how to develop and use such systems to support compliance with IHL. That being said, technical limitations, lack of good quality data and human behavioural tendencies when interacting with machines mean that AI-decision-support systems will not be suitable for all tasks and contexts. Their

198 ICRC, *Reducing Civilian Harm in Urban Warfare: A Handbook for Armed Groups*, ICRC, Geneva, 2023, p. 15: https://shop.icrc.org/reducing-civilian-harm-in-urban-warfare-a-handbook-for-armed-groups-pdf-en.html.

199 As per obligation in AP I, Art. 57(2)(a)(ii).

200 ICRC, *2019 Challenges Report*, p. 32.

use can also create additional risks for civilians and other protected persons, especially in connection with decisions on targeting. These risks must be carefully considered and addressed when developing, reviewing the legality of, and using these tools.

4. REDUCING THE HUMAN COST OF MILITARY OPERATIONS IN OUTER SPACE

The military application of technology enabled by space systems is an integral part of modern military operations. Outer space is becoming increasingly contested as a number of states view space as an operational domain, put in place dedicated space-defence strategies and commands, and are engaged in developing, testing and deploying kinetic or non-kinetic 'counterspace' capabilities.

At the same time, essential civilian services are coming to depend more and more on space systems. Today, space systems – particularly navigation, communications and remote-sensing satellites – play an indispensable role in the functioning of critical civilian infrastructure, especially in the energy and communications sectors. These sectors enable the provision of the essential services on which civilians depend, such as food production and supply, water, electricity, health care, sanitation and waste management, and humanitarian operations.[201]

The expanding role of space systems in military operations during armed conflicts increases the likelihood of their being targeted, putting the functioning of essential civilian services on earth, which rely on such systems, at risk.

A) EXISTING LIMITS UNDER INTERNATIONAL LAW ON MILITARY OPERATIONS IN, OR IN RELATION TO, OUTER SPACE

Military operations in, or in relation to, outer space[202] – whether through kinetic or non-kinetic means – do not occur in a legal vacuum. They are constrained by existing international law. Relevant international law includes, in particular, the Charter of the United Nations, space law treaties, the law of neutrality and IHL.[203]

First and foremost, treaty and customary rules prohibit or restrict the choice of weapons, means and methods of warfare that could be placed and/or used in, or in relation to, outer space. The placement in orbit of objects carrying nuclear weapons or other weapons of mass destruction, the installation of such weapons on celestial bodies, or the stationing of such weapons in outer space in any other manner is prohibited. The testing of weapons of any kind and the conduct of military manoeuvres on celestial bodies are forbidden. In addition, the prohibition against indiscriminate weapons, weapons of a nature to cause superfluous injury or unnecessary suffering,[204] and against a number of other specific types of weapons[205] apply to outer space. And the prohibition against military or other hostile use of environmental modification techniques contrary to the Environmental Modification Convention apply equally on Earth and in outer space.[206] These rules become

201 See further, Gilles Doucet and Stuart Eves, *Protecting Essential Civilian Services on Earth from Disruption by Military Space Operations*, ICRC, Geneva, 2024, pp. 39–56: https://shop.icrc.org/protecting-essential-civilian-services-on-earth-from-disruption-by-military-space-operations-pdf-en.html.

202 For the purpose of this chapter, military operations in, or in relation to, outer space include military operations in, to, from and through outer space and those against space systems, whether they are space components or ground components or any communication link between them.

203 For a detailed discussion on existing limits under international law, including IHL, on military operations in or in relation to outer space during armed conflicts, see ICRC, "Constraints under international law on military operations in outer space during armed conflicts": https://www.icrc.org/en/document/constraints-under-international-law-military-space-operations.

204 ICRC, Customary IHL Study, Rules 70 and 71.

205 ICRC, Customary IHL Study, Rules 72–84; see also all the treaties regulating specific means and methods of warfare, as listed in the ICRC's IHL treaties database: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/vwTreatiesByTopics.xsp#view:__id1:__id2:__id260:repeat1:1:labelAnchor.

206 Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques, 1976, Arts I and II.

particularly relevant when states decide to study, develop, acquire or adopt any new weapon, or means or method of warfare, that could be used in, or in relation to, space.[207]

Beyond the prohibitions against specific types of weapons, or limitations on them, IHL imposes more general constraints on military operations conducted in the context of an armed conflict, including those that are carried out in outer space or the effects of which extend to outer space. These rules include, notably, the principle of distinction, the prohibition against indiscriminate and disproportionate attacks and the obligation to take all feasible precautions in attack.[208]

Furthermore, international law, IHL in particular, also affords specific protection to certain persons and objects in armed conflict, including objects indispensable to the survival of the civilian population,[209] medical personnel, units and transports,[210] humanitarian relief personnel and objects,[211] cultural property,[212] the natural environment,[213] works and installations containing dangerous forces such as dams, dykes and nuclear power plants,[214] and astronauts.[215] These enhanced protections must be upheld at all times, including when carrying out military operations that may be expected to affect space systems critical to the protection, safety or functioning of these persons and objects.

Finally, belligerents must take all feasible precautions to protect civilians and civilian objects against the effects of military operations in, or in relation to, outer space, which is an obligation that states must already have fulfilled in peacetime.[216] Measures that could be considered include physically or technically segmenting space systems (or parts thereof) used for military purposes from those put to civilian use, and working towards identifying space systems serving specifically protected objects, such as hospitals and objects indispensable to the survival of the civilian population. If a space object is put exclusively to civilian use, the state of registry should register it as such, clearly indicating its protected status under IHL.[217]

B) WORKING TOGETHER TO PREVENT AND ADDRESS THE RISK OF CIVILIAN HARM DUE TO MILITARY SPACE OPERATIONS

In line with its humanitarian mandate and mission, the ICRC is concerned primarily with the potential human cost on Earth of the use of weapons and other military operations in, or in relation to, outer space. Given the indispensable role of space systems in the provision of essential civilian services, humanitarian considerations should be a cornerstone of any multilateral discussion or normative development with regard to space security.

To this end, the ICRC has made preliminary recommendations to the international community on the possible further development of legally binding and/or non-binding instruments, focusing on measures to minimize the risk of civilian harm posed by threats to space systems. These recommendations aim at, first, ensuring protection for space systems necessary for the provision of essential civilian services and for specifically protected persons and objects under international law; second, mitigating the risk of space debris by refraining from developing, testing and using kinetic counterspace capabilities and other harmful operations with

207 Notably, states party to AP I are required to review the legality of such a new weapon, means or method of warfare, in order to ensure that its employment would comply with IHL and other relevant rules of international law; see AP I, Art. 36.
208 ICRC, Customary IHL Study, Rules 1, 7, 11–14 and 15–21; AP I, Arts 48, 51 and 57.
209 ICRC, Customary IHL Study, Rule 54; AP I, Art. 54; AP II, Art. 14.
210 See, for example, GC I, Art. 19; GC II, Art. 12; GC IV, Art. 18; AP I, Art. 12; AP II, Art. 11; ICRC, Customary IHL Study, Rules 25, 28 and 29.
211 AP I, Arts 70(4) and 71(2); AP II, Art. 18(2); ICRC, Customary IHL Study, Rules 31 and 32.
212 See, for instance, AP I, Art. 53; AP II, Art. 16; ICRC, Customary IHL Study, Rules 38 and 39.
213 AP I, Art. 35(3); ICRC, Customary IHL Study, Rules 43–45.
214 Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, 1996, Art. V.
215 AP I, Art. 56; AP II, Art. 15; ICRC, Customary IHL Study, Rule 42.
216 AP I, Art. 58; ICRC, Customary IHL Study, Rules 22–24.
217 Convention on Registration of Objects Launched into Outer Space, 1974, Art. IV(1)(e).

similar effects; and third, enhancing international cooperation to increase the resilience of space-based services that humanitarian relief and emergency response rely on and ensuring uninterrupted access to them.[218]

More broadly, the ICRC urges states to carefully consider the human and societal cost if they decide to develop military space capabilities or use them during armed conflicts. In light of the risks of significant civilian harm, states may decide to set general prohibitions or specific limits with regard to weapons, hostilities or other military operations in, or in relation to, outer space for a range of reasons; one of these reasons must be the humanitarian impact of such operations. If new legally binding rules or voluntary norms in this regard are developed, they must be consistent with, build on and strengthen the existing legal framework, including IHL.

218 A detailed elaboration of these recommended measures can be found in ICRC, "Preliminary recommendations on possible norms, rules and principles of responsible behaviours relating to threats by States to space systems", 27 January 2023: https://www.icrc.org/en/document/preliminary-recommendations-on-reducing-space-threats, and in the statement made by the ICRC at the open-ended intersessional informal consultative meeting on further practical measures for the prevention of an arms race in outer space on 29 February 2024: https://www.icrc.org/en/un-outer-space-ihl-statement.