Proceedings of the Bruges Colloquium

New Technologies on the Battlefield: Friend or Foe?

21st Bruges Colloquium
12-16 October 2020

Actes du Colloque de Bruges

Les nouvelles technologies sur le champ de bataille : alliées ou ennemies ?

21ème Colloque de Bruges
12-16 octobre 2020

Brugge
College of Europe
Collège d'Europe
Natolin

COMITE INTERNATIONAL GENEVE
CICR

ICRC Delegation to the EU, NATO and the Kingdom of Belgium

Délégation du CICR auprès de l'UE, de l'OTAN,

et du Royaume de Belgique

Charlotte Giauffret

Sous la direction de
Stéphane Kolanowski

Notice

Collegium is the academic journal of the College of Europe. Copyright remains with the authors. The proceedings of this Colloquium have been written by the speakers or by the Delegation of the International Committee of the Red Cross (ICRC) in Brussels on the basis of audio recordings of the Colloquium. These texts were then reviewed by the speakers. The opinions expressed herein are not necessarily those of College of Europe nor the International Committee of the Red Cross.

Avertissement

Collegium est la revue académique du Collège d'Europe. Les Actes de ce Colloque ont été rédigés par les orateurs ou par la Délégation du Comité international de la Croix-Rouge (CICR) à Bruxelles sur base d'enregistrements audio du Colloque. Ces textes ont alors été revus par les orateurs et n'engagent que ces derniers. Tous droits réservés. Ils ne représentent pas nécessairement les vues ni du Collège d'Europe ni du Comité international de la Croix-Rouge.

I. PROCEEDINGS OF THE BRUGES COLLOQUIUM/ACTES DU COLLOQUE DE BRUGES **p. 7**

Opening Statements

Federica Mogherini, *Rector of the College of Europe* **p. 7**

Gilles Carbonnier, *Vice-President of the ICRC* **p. 8**

Panel 1
Cyber Operations during Armed Conflict
Les cyber-opérations en temps de conflit armé

Dr. Knut Dörmann, *Head of the ICRC Delegation in Brussels* **p. 11**
INTRODUCTION

Prof. Vera Rusinova, *High School of Economics University, Moscow* **p. 16**
THRESHOLDS AND APPLICABILITY OF IHL RULES

Prof. Hongsheng Sheng, *Shanghai University of Political Science and Law, China* **p. 25**
CHALLENGES AND RESPONSES: CYBER OPERATIONS AND THE NOTIONS OF 'ATTACKS' AND 'OBJECTS' UNDER INTERNATIONAL HUMANITARIAN LAW

Prof. Duncan Hollis, *Temple Law School* **p. 29**
IMPROVING TRANSPARENCY: INTERNATIONAL LAW AND STATE CYBER OPERATIONS

DISCUSSION **p. 39**

Panel 2
The Use of Autonomous Weapon Systems: A Challenge to the International Rule of Law?
L'utilisation des systèmes d'armes autonomes : un défi pour l'état de droit international ?

Maya Brehm, *ICRC Geneva* **p. 45**
INTRODUCTION

Colonel Rudolph Stamminger, *French Ministry of Defence* **p. 50**
MILITARY USE OF AWS MUST COMPLY WITH INTERNATIONAL LAW AND, IN PARTICULAR WITH IHL

Thompson Chengeta, *University of Southampton* **p. 58**
ETHICS AND AUTONOMOUS WEAPON SYSTEMS

Netta Goussac, *SIPRI* **p. 63**
WHAT ARE THE ELEMENTS OF HUMAN CONTROL AND HOW CAN THEY INFORM THE SETTING OF LIMITS ON AUTONOMOUS WEAPON SYSTEMS?

DISCUSSION **p. 70**

Panel 3
Artificial Intelligence (AI) and Machine Learning
Intelligence artificielle (IA) et apprentissage automatique

Neil Davison, *ICRC Geneva* **p. 77**
INTRODUCTION

Edward Hunter Christie, *NATO* **p. 79**
AI-SUPPORTED DECISION MAKING IN WARFARE: FAR-REACHING IMPLICATIONS

Dustin Lewis, *Research Director, Harvard Law School Program on International Law and Armed Conflict* **p. 80**
A KEY SET OF IHL QUESTIONS CONCERNING AI-SUPPORTED DECISION-MAKING

Pauline Warnotte, *UNIDIR and University of Namur* **p. 87**
EFFORTS TO GOVERN (MILITARY APPLICATIONS OF) AI

DISCUSSION **p. 94**

Panel 4
Military Space Operations: Constraints under International Law and Potential Humanitarian Consequences
Opérations militaires spatiales : contraintes imposées par le droit international et conséquences humanitaires potentielles

Heather Harrison Dinniss, *Swedish Defence University* **p. 103**
INTRODUCTION

Wen Zhou, *ICRC Geneva* **p. 108**
CONSTRAINTS RELATED TO THE USE OF WEAPONS IN OUTER SPACE UNDER IHL

Mickael Dupenloup, *French Ministry of Defence* **p. 113**
BEYOND THE PEACEFUL USE OF OUTER SPACE: POTENTIAL CONFLICTS?

Jessica West, *Canadian Peace Research Institute, Project Ploughshares* **p. 121**
PROTECTING HUMANS ON EARTH FROM WAR IN SPACE

DISCUSSION **p. 126**

Panel 5

New Technologies and Humanitarian Action

Nouvelles technologies et action humanitaire

Massimo Marelli, *ICRC Geneva* **p. 129**
INTRODUCTION

Edgardo Yu, *Technology Division, World Food Programme (WFP)* **p. 134**
NEW TECHNOLOGIES. WHY IT IS RELEVANT AND USEFUL IN A HUMANITARIAN CONTEXT?

Prof. Carmela Troncoso, *EPFL* **p. 138**
TECHNOLOGY FRAMEWORK: WHAT IS AUTHENTICATION AND IDENTIFICATION, WHAT IS FUNCTIONAL IDENTITY AND FOUNDATIONAL IDENTITY AND WHY DOES IT MATTER?

Wojciech Wiewiórowski, *European Data Protection Supervisor* **p. 141**
DATA PROTECTION PROVISIONS FRAMING THE DIGITAL IDENTITY AND BIOMETRICS (AND THE MAIN PROBLEMATIC AREAS)

Alexandrine Pirlot de Corbion, *Privacy International* **p. 144**
A CIVIL SOCIETY PERSPECTIVE: DIGITAL IDENTITY IN HUMANITARIAN ACTION: WHAT ARE THE IMPLICATIONS FOR PEOPLE AND THEIR RIGHTS?

Catherine Kayser, *Luxembourg Ministry of Foreign Affairs* **p. 150**
A DONOR’S PERSPECTIVE

DISCUSSION **p. 154**

Closing Remarks

Knut Dörmann **p. 163**

II. LIST OF PARTICIPANTS/LISTE DES PARTICIPANTS **p. 167**
III. PROGRAMME (EN) **p. 169**
IV. SPEAKERS’ BIOS/BIOGRAPHIES DES ORATEURS **p. 173**

PROCEEDINGS OF THE BRUGES COLLOQUIUM
ACTES DU COLLOQUE DE BRUGES

Opening Statements

DISCOURS D'OUVERTURE

Federica Mogherini
Rector of the College of Europe

Welcome to this new edition, the 21st edition of the Bruges Colloquium on International Humanitarian Law (IHL), which the College of Europe is proud to host together with the International Committee of the Red Cross.

This year, the Colloquium will be focusing on a very crucial topic, actually on a series of topics, unfortunately online this year. However, I am sure that the quality of the debates and the subsequent exchanges will be so high that, although the exchanges will be online, the outcome will be very relevant for us all.

The Colloquium will be focused on new technologies on the battlefield. I believe that the different aspects that will be discussed, from the challenges that it puts on International Law to the new instruments it puts at the disposal of humanitarian actors, are extremely relevant, not only to experts and academia but also for policy makers. There are issues that need to be addressed that can shape the future of policy making for decades to come. I am sure that the quality of our discussions during the next five days will not only provide food for thought but also elements for decisions and actions to be taken on the ground for all those who are involved in this important field.

I would like to thank, first of all, the International Committee of the Red Cross (ICRC) for the excellent cooperation we have established over the years. For me it is a new challenge and opportunity as I have only been Rector of the College of Europe for one month. I know that this partnership has been extremely fruitful in the past, and I am sure that it will continue to be extremely important and fruitful in the future. I thank the ICRC for this cooperation. I want to thank each of you for joining us online for these five days.

I am very much looking forward to the outcome of these conversations and the suggestions and recommendations that might come out of these important exchanges.

Thank you very much and welcome again to Bruges.

Dr. Gilles Carbonnier
Vice-President of the ICRC

Mme the Rector, dear Federica Mogherini, Distinguished Panellists, Ladies and Gentlemen,

I wish to welcome you on behalf of the International Committee of the Red Cross to this 21st edition of the Bruges Colloquium.

This year, our Colloquium is devoted to a critical issue of increasing relevance: new technologies on the battlefield.

For obvious reasons, we are holding this edition virtually. It is a departure from our traditional Bruges formula where we have some 120 IHL experts, meeting for a day and a half at the College of Europe, in Bruges. This year, the Colloquium consists of a week-long online programme, nicknamed 'the Bruges week', with more than 430 participants from all around the world. This brings its share of challenges, such as dealing with different time zones, but also its share of opportunities: the ability to bring many different perspectives to the table, which will considerably enlighten our debates and discussions.

When we had to select the topic for this 21st edition of the Bruges Colloquium, we quickly agreed to focus on new technologies used on the battlefield, not knowing back then that we would need to rely on some of these very technologies to hold the Colloquium itself.

Ten years ago, we discussed some of these technologies at the Bruges Colloquium. Yet, ten years ago, the world of new technologies was a radically different one. So today it is high time to address the topic again. I am glad to see the very impressive list of speakers who accepted to contribute. Your wealth of expertise will greatly enrich the debate.

Ladies and Gentlemen, now let me turn to the topic at hand. New technologies are changing human interactions profoundly, and these changes extend to the battlefield. Many States are investing heavily in the development of new means and methods of warfare that rely on digital technology. But how is this changing the nature of warfare? What does it mean for

IHL? What does it mean for our capability to ensure that civilians and civilian infrastructure are protected?

Over the course of the week, we will be analysing five main topics from military, technical, ethical, and humanitarian perspectives. The first topic is: cyberoperations. The use of cyber tools as means, and methods of warfare has become a reality in contemporary armed conflicts. While some cyber tools under specific circumstances may contribute to better distinguishing between civilian objects and military objectives, and thus help to reduce or even prevent damage to civilian infrastructure, they also carry significant risks. Cyber operations can dramatically disrupt the provision of vital services. Cyber operations conducted over recent years, primarily outside armed conflicts, have shown that malware can spread instantly around the globe. Of more concerns for us is the threat that cyber operations pose to critical civilian infrastructure such as electricity, water systems, hospitals, or industrial systems, including nuclear facilities. In our view, IHL puts limits to cyber operations during armed conflict, just as IHL limits the use of any other means and methods of warfare, whether old or new. However, in order for IHL to truly protect civilians against the effects of cyber operations, there is a need for greater clarity on how key IHL notions, such as *attacks* or *civilian objects* are interpreted and applied.

Another key development in warfare is autonomous weapon systems, namely systems that can select and attack a target without human intervention. The ICRC is concerned about the loss of human control over the use of force. It is the responsibility of humans to make context-specific judgements to comply with complex principles such as distinction, proportionality, and precaution in attack. A weapon with autonomy in its critical functions that is unsupervised, unpredictable, and unconstrained in time and space would be unlawful under existing IHL rules. It would raise ethical concerns because human agency in decisions to use force is necessary to uphold moral responsibility and human dignity. We will also have the opportunity to discuss this in tomorrow's panel.

The third topic that we will be discussing is artificial intelligence, abbreviated AI. From a humanitarian perspective, the most significant implications of the use of AI in machine learning systems are threefold: first, their use as a basis for increasingly autonomous systems, including weapons; second, the use in cyber and information; third, the impact on decision-making in armed conflict. In terms of decision-making, AI may help facilitate a faster and broader collection of analysis and information, which could contribute to better compliance with IHL and minimise the risk for civilians. However, the same algorithmic analyses or predictions could potentially achieve exactly the opposite, especially given the current limitations of the technology. In any case, the ICRC is convinced of the need for a human-centred approach that ensures that AI is used to augment human decision-making in armed conflict, not to replace it.

The fourth topic is outer space. The military use of space objects has been an integral part of warfare for several decades. The ICRC is concerned by the potentially high human cost if outer space became weaponised. This could directly or incidentally disrupt or destroy civilian and dual-use space objects, for example, satellites transmitting Global positioning System (GPS) signal on which essential civilian services depend. IHL rules apply to any military operations conducted as part of an armed conflict, including in outer space. The ICRC recommends that States acknowledge the potentially significant humanitarian consequences of the use of weapons in outer space and the protection afforded to civilians by IHL. This will be discussed in greater depth on Thursday.

On Friday, we will address the use of new technologies by humanitarian actors themselves, and the implications for the humanitarian sector as a whole, as well as for the people affected by armed conflict.

Ladies and gentlemen, new technologies are a game changer for us all. The 21st Bruges Colloquium is a great opportunity to discuss and anticipate what it holds for our common humanity. How best to seize the opportunities offered by new technologies while addressing the risks and threats they pose in armed conflicts? We are all impatient to explore these issues and listen to the panellists.

I now give the floor to the Head of the ICRC Delegation in Brussels, Dr. Knut Dörmann, and wish you all a stimulating and productive Bruges week!

Thank you very much.

Panel 1
Cyber Operations during Armed Conflict
Les cyber-opérations en temps de conflit armé

INTRODUCTION

Dr. Knut Dörmann
Head of the ICRC Delegation in Brussels

Résumé

Knut Dörmann, Chef de la délégation du Comité international de la Croix-Rouge (CICR) à Bruxelles, était le modérateur de ce premier panel. Après avoir présenté les intervenants, il a introduit le thème de ce panel : la cyber-guerre. Ces vingt dernières années, les cyber-opérations ont pris une importance croissante dans les conflits armés. Bien que peu d'États reconnaissent publiquement en faire usage, ils sont de plus en plus nombreux à développer leurs capacités militaires dans ce domaine. En particulier, les cyber-attaques visant des infrastructures civiles et pouvant altérer, voire stopper, la fourniture de services essentiels constituent un risque majeur pour la vie humaine. Par exemple, des cyber-opérations ont pu servir à couper l'alimentation en électricité ou en eau de pays en conflit, à immobiliser des hôpitaux ou à tenter d'endommager des usines pétrochimiques ou des installations nucléaires. Dans ce contexte, Peter Maurer, Président du CICR, s'est joint en 2020 à un groupe de dirigeants mondiaux appelant tous les gouvernements à travailler ensemble et à affirmer que les cyber-opérations à l'encontre d'établissements médicaux sont illégales et inacceptables en toutes circonstances, y compris en temps de crise ou de conflit armé. Ce panel explore trois grands thèmes cruciaux, et encore en débat, concernant l'application du droit international humanitaire (DIH) aux cyber-opérations dans les conflits armés :

1. *Le seuil et l'applicabilité des règles de DIH : pour le CICR, le DIH régule les cyber-opérations dans les conflits armés. Toutefois, certains États contestent ce point. De plus, la question de savoir quels types de cyber-opérations permettent de qualifier l'existence d'un conflit armé, et sont donc régulés par le DIH, est encore débattue.*
2. *Les concepts « d'attaques » et « d'objets » : en admettant que le DIH s'applique, il reste à savoir si, par exemple, des cyber-opérations visant des ordinateurs ou des systèmes informatiques peuvent être qualifiées « d'attaques ». De plus, si la notion de « données digitales » n'est pas explicitement mentionnée dans les règles et les traités de DIH, les données digitales sont devenues essentielles au fonctionnement des sociétés actuelles. Dès lors, les données digitales bénéficient-elles de la protection accordée aux « biens de caractère civil » ?*

3 *Comment accroitre la transparence et les compétences juridiques : au niveau des opérations, les conseillers juridiques des forces armées doivent savoir comment le DIH s'applique pour pouvoir fournir des conseils juridiques adéquats. Au niveau stratégique, les États doivent être en mesure d'interpréter leurs opérations dans le cyber-espace, en particulier pour évaluer si une cyber-opération va être considérée par un autre État comme déclenchant un conflit armé. Enfin, au niveau diplomatique, la question de l'applicabilité du DIH aux cyber-opérations dans les conflits armés est au cœur de plusieurs processus développés par les Nations Unies. Le développement des compétences dans ce domaine est nécessaire au succès de ces discussions et à la participation de tous les États.*

Ladies and Gentlemen, let me join the Rector of the College of Europe, Federica Mogherini and the ICRC Vice-President, Gilles Carbonnier, in welcoming you all to this 21st edition of the Bruges Colloquium, which is its first virtual edition. I definitely echo Mrs Mogherini's words in saying that the ICRC very much values the relationship it has with the College of Europe. This is true both for the yearly Colloquium and for the students' seminars that we jointly run on both campuses of the College.

It is a true pleasure for me, as the Head of the ICRC Delegation to the EU, NATO and Belgium to chair this first panel on a topic which is particularly dear to me: cyber warfare. It is a topic I worked on as an ICRC Legal Advisor some twenty years ago and have continued to follow closely as the Head of the ICRC Legal Division and Chief Legal Officer.

Let us now turn to the first virtual panel of this year's Colloquium, which is focused on 'cyber operations during armed conflict'.

We have three excellent speakers with us, namely:

- **Dr. Vera Rusinova,** from the High School of Economics University, in Moscow, Russia
- **Prof. Hongsheng Sheng,** from the Shanghai University of Political Science and Law, in Shanghai, China
- **Prof. Duncan Hollis**, from the Temple Law School, Philadelphia, United States of America

You will find their biographies in the 'about the speakers' menu of the Colloquium's website.

Looking at the panellists' expertise on international law and cyber operations, I am certain that we will have an engaging and interesting panel. Moreover, and this makes the panel particularly rich, I am also delighted to have experts from three different regions of the world, not to say three 'cyber hotspots'.

Before giving the floor to our experts, I would like to briefly introduce the topic.

During these past two decades, the importance of cyber operations during armed conflicts has grown continuously. As societies are digitalising, so are armed conflicts. Today, the use of cyber operations during armed conflicts has become a reality. While only a few States have publicly acknowledged resorting to such operations, an increasing number of States are developing military cyber capabilities, meaning their use is likely to increase in the future.

Based on operations seen thus far, cyber technology is used to pursue different objectives, ranging from

- espionage and target identification;
- over cyber operations in support of kinetic operations – you may think of disabling an enemy's military radar station in support of air strikes;
- to cyber operations that are aimed at causing physical effects.

As pointed out by the ICRC Vice-President, cyber operations conducted over the past years – primarily outside armed conflicts – have shown that malware can spread instantly around the globe and affect civilian infrastructure and the provision of essential services. For example, cyber operations have

- cut off electricity supplies and targeted water systems in war-affected countries;
- halted hospitals services in the middle of a global pandemic – most recently a patient died in Germany when a hospital's server was disrupted and the patient had to be diverted to another hospital;
- been used in an attempt to damage a petrochemical plant;
- and especially alarming, there have been cyberattacks against nuclear facilities.

Among the countless malicious cyber operations that are reported daily, cyberattacks against critical civilian infrastructure stand out due to the significant risk they pose to human life. In this context, let me mention that earlier this year the ICRC President Peter Maurer joined a group of global leaders to call on all governments to work together and assert in unequivocal terms that cyber operations against medical facilities are unlawful and unacceptable in time of crisis, in time of conflict, at all times.

In today's panel, we will address some of the key questions on how International Humanitarian Law – in short IHL – applies to, and therefore restricts, cyber operations during armed conflict.

The first question we need to address is the 'threshold' and 'applicability' of IHL rules to cyber operations during armed conflict.

Under this topic, we have at least two larger topics:

- one is whether IHL applies to cyber operations conducted during armed conflict. While for the ICRC there is no question that IHL applies to, and therefore restricts, cyber operations during armed conflict, we are aware of the political discussions among States around this subject and a number of States contesting this;
- moreover, an important and partly unresolved question is: 'what types of cyber operations will trigger an armed conflict and are therefore regulated by IHL?'.

The second issue we will address are the notions of 'attacks' and 'objects' under IHL.

Once we accept that IHL applies to cyber operations during armed conflicts, these issues are at the heart of what States and other experts need to clarify.

For example, most IHL rules on the conduct of hostilities only apply to operations that amount to 'attacks' as defined in IHL. For the past year, there has been significant debate on whether cyber operations that disrupt computers or computer systems qualify as an attack.

Moreover, data is among the most precious 'goods' in cyberspace. Some are calling data the 'gold of the 21st century'. Data – for example civil registries, insurance data, medical data – is essential for the functioning of societies.

IHL rules or treaties do not mention data explicitly. However, under IHL 'civilian objects' are protected against attacks. A key question for the protection of societies from cyber operations during armed conflict is whether digital data enjoys the same protection as civilian objects. Put simply, if paper files are replaced by digital files in the form of data, will this change the protection that IHL affords them?

The third issue on our agenda today is the question of how we can increase transparency and legal capacity on IHL issues in cyberspace.

Admittedly, this is a key question, and important at different levels:

- First, legal advisers of armed forces need to know how IHL rules apply to and regulate the use of cyber operations during armed conflict. Unless they understand, for example whether data qualifies as a civilian object, it will be difficult for them to provide legal advice on operations.

- Second, at the strategic level, States need to know how other States interpret their operations in cyberspace. For example, a State that decides to conduct a hostile cyber operation should know whether the other State will consider this operation as triggering an armed conflict, and therefore the IHL's applicability.
- And third, at the diplomatic level, the question of whether and how IHL applies to cyber operations during armed conflict is receiving significant attention in different United Nations (UN) processes. For these discussions to be successful and to enable all States to participate, capacity building is needed.

All these three points and issues are of utmost importance and I am sure that our three panellists will shed further light and guidance on this issue.

Without further ado, I am now giving the floor to Dr. Vera Rusinova from High School of Economics University, in Moscow, Russia. Dr Rusinova, the floor is yours.

Thank you.

THRESHOLDS AND APPLICABILITY OF IHL RULES
SEUILS ET APPLICABILITÉ DES RÈGLES DE DIH

Prof. Vera Rusinova[1]

High School of Economics University, Moscow

Résumé

Vera Rusinova est docteure, professeure et directrice du département de droit international de la Faculté de droit à la High School of Economics de Moscou. Sa présentation analyse la teneur des différents arguments mis en avant par les États pour affirmer l'applicabilité ou la non-applicabilité du DIH. Elle questionne l'affirmation selon laquelle le DIH devrait s'appliquer à toutes les cyber-opérations dans un conflit armé et appelle à ne pas trop étendre les provisions du DIH, ceci afin de ne pas le vider de sa substance. Elle commence par rappeler qu'il n'existe pas encore de consensus sur l'applicabilité du DIH aux cyber-opérations. Certains États, tels que la Russie ou la Chine, affirment que cela pourrait conduire à une militarisation de l'usage des technologies de l'information et de la communication (TIC). Des États ont aussi insisté sur l'importance du consentement étatique. Cet article propose des éléments de réflexion en trois parties.

Dans une première partie, elle pose un regard critique sur le cadre conceptuel dans lequel s'inscrit cette question du 'si et comment' le DIH s'applique aux cyber-opérations dans les conflits armés. Formulée ainsi, la question mène à une dualité entre le rôle du consentement et de la normativité en droit international, c'est-à-dire entre une approche réaliste et une approche de l'état de droit. Or, ces deux approches se contredisent d'un point de vue théorique. De plus, ce cadre conceptuel légitime une grande variété de positions, y compris les plus radicales. Le problème est que la question du 'si et comment' ne présuppose aucune contrainte légale préexistante. L'usage même de cette formulation suggère finalement que toutes les normes sont basées sur le consentement et que ce consentement peut être retiré à tout moment. Cela menace les principes de base du DIH alors qu'en principe, par leur nature de principes généraux du droit ou de normes coutumières établies, ceux-ci ne peuvent être modifiés unilatéralement.

Dans une deuxième partie, elle se demande s'il faut effectivement souhaiter que les États reconnaissent l'applicabilité du DIH. D'abord, quand l'applicabilité du DIH est reconnue par un État, il est rarement précisé avec plus de détails ce que cela implique au-delà d'une affirmation

[1] Vera Rusinova is Doctor of legal sciences, LL.M (Göttingen), Professor, Head of the School of International Law of the Law Faculty, the National Research University Higher School of Economics; (E-mail: vrusinova@hse.ru).

générale. Ensuite, même si l'argument de la militarisation semble aller à l'encontre de toute l'histoire du développement des normes du jus in bello*, elle suggère qu'une clarification et une délimitation des définitions est nécessaire ; cela afin d'éviter que les États n'appliquent le DIH plutôt que le droit international relatif aux droits de l'homme ou le droit pénal national, au nom de la* lex specialis. *Enfin, bien que largement endossée comme étant progressiste et pro-humanitaire, cette affirmation de l'applicabilité du DIH peut faire oublier la nécessité d'adopter des normes de droit international spécifiques aux cyber-opérations, dans les cas où les règles de DIH ne sont pas adaptées au cyberespace.*

Troisièmement, une reconnaissance collective de l'applicabilité du DIH aux cyber-opérations peut également être insuffisante. Outre l'applicabilité in abstracto, il s'agirait d'examiner plus précisément si les normes de DIH sont pertinentes, adéquates et suffisantes pour faire face aux cyber-opérations de type militaire. Premièrement, il y a un risque que la majorité des cyber-opérations n'atteignent pas le seuil de de « l'attaque ». Or dans un tel cas, les obligations imposées aux parties d'un conflit armé international (CAI) en vertu des articles 51 (1) et 57(1) du premier Protocole additionnel restent très générales et laconiques, se limitant à une obligation de protection générale de la population civile. Ainsi par exemple, la France a reconnu que la plupart des opérations impliquant des cyber-opérations en situation de conflit armé, y compris des opérations offensives menées par la France, n'atteignent pas le seuil de l'attaque et donc restent régies par les principes généraux du DIH. Un deuxième problème survient lorsque les États tentent de contourner les limites du champ d'une « attaque » en DIH en élargissant cette notion à d'autres types de cyber-opérations. Par exemple, aux États-Unis, un conseiller juridique a estimé que toutes les cyber-opérations n'atteignent pas nécessairement le niveau d'une « attaque ». Cela dépend de plusieurs facteurs, donc la présence d'effets cinétiques causés par la cyber-opération. L'auteure suggère que cette méthode peut en pratique entraîner une inapplicabilité des dispositions de DIH consacrées aux « attaques », car ces dispositions s'appliquent aux opérations cinétiques. Enfin, une cyber-attaque peut impliquer différents niveaux d'alliance avec une partie à un conflit armé. Combinée avec la nature spécifique des cyber-opérations, cela peut rendre les règles et notions de « participation directe aux hostilités » plus restrictives.

Dear colleagues, it is a great honour for me to participate in the Bruges Colloquium on International Humanitarian Law. Before I start, I would like to make a *caveat* that while discussing positions of different States, including Russia, I will express my own personal views as a researcher who aims to take a neutral observational position and be critical with respect to three questions articulated in the call of the ICRC a year ago: whether International Humanitarian Law applies to the conduct of cyber operations during armed conflicts, how it applies and whether it is adequate and sufficient[2].

I would like to start my presentation by citing Professor Michael Schmitt who wrote that 'today, no serious international law expert questions the full applicability of International Humanitarian Law to cyber operations'[3]. Well, the Russian Federation, a number of other States, and also some scholars[4] actually do.

The Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security was not able to adopt final reports in 2016–2017 due to the position articulated by Cuba[5] and backed by Russia[6] and China,[7] under which applicability of *jus ad bellum* and *jus in bello* (International Humanitarian Law) may lead to the establishment of the 'equivalence between the malicious use of ICT and the concept of 'armed attack''[8] under Article 51 of the UN Charter, and, thereby militarise the use and the response to information and communication technologies (ICT). The same divergence was found in the positions of the States expressed at the Open-Ended Working Group (OEWG) on Developments in the Field of Information and Telecommunications in the Context of Inter-

[2] ICRC Position Paper, 'International Humanitarian Law and Cyber Operations during Armed Conflicts', November 2019, p. 9, at: <https://www.icrc.org/en/document/international-humanitarian-law-and-cyber-operations-during-armed-conflicts>.

[3] Schmitt M., 'The State of Humanitarian Law in Cyber Conflict', 6 January 2015, in: *Just Security*, at: <https://www.justsecurity.org/18891/state-humanitarian-law-cyber-conflict/>.

[4] See D'Aspremont J., 'Cyber Operations and International Law: An Interventionist Legal Thought', in: *Journal of Conflict and Security Law*, 2016, Vol. 21, Issue 3, pp. 575–593.

[5] Declaration by Miguel Rodríguez, Representative of Cuba, at the Final Session of Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. New York, 23 June 2017, at: <http://misiones.minrex.gob.cu/en/un/statements/71-unga-cuba-final-session-group-governmental-experts-developments-field-information>.

[6] Response of the Special Representative of the President of the Russian Federation for International Cooperation on Information Security Andrey Krutskikh to TASS' Question Concerning the State of International Dialogue in This Sphere, 29 June 2017, at: <http://www.mid.ru/en/foreign_policy/news/-/asset_publisher/cKNonkJE02Bw/content/id/2804288>.

[7] China did not publicly share its position, see: Korzak E., UN GGE on Cybersecurity: The End of an Era? The Diplomat, 31 July 2017, at: <https://thediplomat.com/2017/07/un-gge-on-cybersecurity-have-china-and-russia-just-made-cyberspace-less-safe>.

[8] Declaration by Miguel Rodríguez, Representative of Cuba, at the Final Session of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, New York, 23 June 2017, at: <https://www.justsecurity.org/wp-content/uploads/2017/06/cuban-expert-declaration.pdf>.

national Security meetings in 2019–2020.[9] While the majority of States confirmed the applicability of international law in its entirety to cyberspace,[10] it was contested by a group of States using arguments related to the importance of State consent for the extension of the scope of non-cyber specific norms, indeterminate thresholds of 'armed attack' by cyber-means and the doubtful applicability of International Humanitarian Law to hybrid warfare and to civilian perpetrators of cyber-attacks.[11] The question is: how is this stance to be evaluated? Is it a norm-scepticism, a norm-contestation, or a call to a return to the anarchical 'Wild West', or *vice versa*: can support and affirmation of the applicability of International Humanitarian Law in the cyber context be seen as not being responsible enough, or even as a misuse of law?

To answer these questions, I have divided my presentation into three interconnected parts: the first one deals with the 'whether and how' frame to approach the question of applicability of International Humanitarian Law. The second challenge is the affirmative approach that the *jus in bello* norms are applicable to cyber operations and the last, third one, addresses their applicability in terms of relevancy, adequacy, and sufficiency.

I. 'Whether and How' Frame to Approach the Question of Applicability of IHL

It is since 1998[12] that the topic of cyber-security has appeared and started to gain more and more weight on the international political agenda at universal, regional, and bilateral levels. This agenda had a very clear *legal* segment, and from the very beginning, both governmental and academic discourse concerning the application of international law to cyber operations was put and nurtured in the 'whether and how' ontological frame[13] with the designation of

[9] For instance, Pakistan, Russia, and The Syrian Arab Republic (The Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security, 1st substantive session, 11 September 2019, 2nd substantive session, 11 February 2020, at: <http://webtv.un.org/>) (hereinafter: OEWG, 1st or 2nd subs. session).

[10] Austria, Brazil, Canada, Chile, the Czech Republic, the European Union, Italy, Lichtenstein, New Zealand, Pacific Islands Forum, Sweden, Switzerland, the United Kingdom and others (OEWG, 2st subs. session, 11 February 2020).

[11] Russia raised a question on how the application of international law in cyberspace correlates with voluntary principle (OEWG, 2st subs. session, 11 February 2020).

[12] General Assembly Resolution 53/70, 'Developments in the Field of Information and Telecommunications in the Context of International Security', 4 December 1998.

[13] See Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, 26 June 2015, A/70/174, at: <https://undocs.org/A/70/174>.

forms of possible legal contribution as being confined to interventionist (managerial) and law-making actions[14].

There are two arguments with respect to the application of the ‘whether and how’ epistemic frame to International Humanitarian Law.

Firstly, the ‘whether and how’ frame can easily be explained in terms of the duality between the role of consent (voluntariness) and normativity of international law, between realism and the rule of law approaches. But here the problems arise. Should International Humanitarian Law – both treaty and customary provisions – be regarded as objective law? Such a way of doubting and checking whether it is applicable to cyber operations, which may lead to the same results as kinetic weapons, is counter-normative. Why should the use of cyber operations be treated differently from other types of weapons?[15] However, if some types of cyber operations cannot match existing frames of International Humanitarian Law, why should States affirm its applicability?

Secondly, the question of ‘whether and how’ International Humanitarian Law is applicable to cyber operations on its own legitimises a whole spectrum of possible stances, including two radical ones. On the one side is the position that International Humanitarian Law is applicable even to non-kinetic cyber operations. For instance, France set forth that a ‘cyber operation without physical effects’ may also be qualified as the use of force and suggested using a non-exhaustive list of criteria, i.e. ‘the circumstances prevailing at the time of the operation, such as the origin of the operation and the nature of the instigator (military or not), the extent of the intrusion, the actual or intended effects of the operation or the nature of the intended target’.[16] On the opposite side is the position that a cyber-attack alone is not considered as ‘use of force’ and, thus, cannot give rise to the application of International Humanitarian Law. During the OEWG meeting held on 11 February 2020, Russia took the most stringent position that a cyber-attack alone without the context of an armed conflict does not meet this criterion.[17]

It should be emphasised that the problem is that the frame of ‘whether and how’ is open, as it presupposes no pre-existing conditions. The use of this frame, which invites States to act as if

[14] D’Aspremont J., ‘Cyber Operations and International Law: An Interventionist Legal Thought’, in: *Journal of Conflict and Security Law*, 2016, Vol. 21, Issue 3, p. 575 sqq.

[15] See: International Court of Justice, ‘Legality of the Threat or the Use of Nuclear Weapons’, Advisory Opinion, 8 July 1996, para. 86.

[16] Ministry of Armed Forces, International Law Applied to Operations in Cyberspace, October 2019, p. 7, at: <https://www.defense.gouv.fr/content/download/567648/9770527/file/international+law+applied+to+operations+in+cyberspace.pdf>.

[17] OEWG, 2nd subs. session, 11 February 2020.

there were no legal restraints – to decide 'from scratch'– makes one wonder whether allowing a 'partial gambit' may be legitimate. In the chess game a gambit is used for the operations when the players are sacrificing something with a view to achieving a more advantageous position. Here some States (just a few, of course) may also wish to use the offered rules of the game seriously and say 'no'. What would be sacrificed then are basic principles of International Humanitarian Law, that due to their nature as general principles of law or established customs, cannot be unilaterally altered (some of them because of the *jus cogens* character cannot be altered even in an ordinary process). The 'whether and how' frame is designed in a form that leads to the conclusion (and confusion) that all norms are truly consent-based and that this consent can be withdrawn anytime.

II. Challenging the Affirmative Approach that International Humanitarian Law Is Applicable

However, can we truly celebrate that States affirm the applicability of International Humanitarian Law? Let us examine the innocence of the affirmation.

To begin with, in almost all cases when the application of International Humanitarian Law is confirmed, we do not know in which volume. For instance, 79 States have supported the Paris Call for Trust and Security in Cyberspace that laconically states that 'international humanitarian law' 'is applicable to the use of information and communication technologies by States'[18]. A more or less detailed position has been represented only by a few States, including Australia[19],

[18] Paris Call for Trust and Security in Cyber Space, 11 December 2018, at: <https://pariscall.international/en/call>.

[19] Department of Foreign Affairs and Trade, 'Australia's Cyber Engagement Strategy, Annex A: Supplement to Australia's Position on the Application of International Law to State Conduct in Cyberspace', 2019; 'Australia's Cyber Engagement Strategy, Annex A: Australia's Position on How International Law Applies to State Conduct in Cyberspace', 2017.

Germany[20], the Netherlands[21], the UK[22], the US[23], France[24], Finland[25], and Israel[26] so far.

Turning to the impact of this general affirmation, it is worth questioning whether the fear that it can lead to the militarisation of cyberspace is substantiated. Well, the argument that the applicability of International Humanitarian Law will legitimise militarisation of cyberspace if taken *per se* seems to go against the whole history of the development of *jus in bello* norms. However, this rebuttal is convincing only if it implies a superficial meaning to the argument of militarisation. Another way is to read it as exposing that without a clear determination of borderlines between cyber operations as a 'use of force' or an 'armed attack' in the *jus ad bellum* terms and an 'attack' or a 'military operation' in the *jus in bello* terms, on the one hand, and cyber operations as (ordinary) malicious acts which may take place also during armed conflicts, on the other hand, the shift to International Humanitarian Law can lead to a misuse of a military legal paradigm of international law. So, it would be at the end of the day International Humanitarian Law instead of International Human Rights Law, or national criminal law, which may be well based on numerous international treaties in this respect, as it is not something new when States are sheltering their activities and on the basis of *lex specialis* exclude the application of other regimes.

Finally, the affirmative approach – which is widely endorsed as progressive and pro-humanitarian – can serve to neglect the necessity to adopt cyber-specific norms of international law,

[20] Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Dr. Alexander S. Neu, Andrej Hunko, Wolfgang Gehrcke, weiterer Abgeordneter und der Fraktion DIE LINKE. 'Krieg im 'Cyber-Raum' – offensive und defensive Cyberstrategie des Bundesministeriums der Verteidigung', Drucksache 18/6989, 10.12.2015, S. 4, 5, 7.

[21] Ministry of Foreign Affairs. Letter to the Parliament on the International Legal Order in Cyberspace, 5 July 2019, at: <https://www.government.nl/ministries/ministry-of-foreign-affairs/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace>.

[22] 'Cyber and International Law in the 21st Century', The Attorney General Jeremy Wright QC MP Speech on the UK's Position on Applying International Law to Cyberspace.

[23] Harald Hongju Koh, 'International Law in Cyberspace', Remarks by Harald Hongju Koh, Legal Adviser to the US Department of State, 18 September 2012, in: *Harvard International Law Journal Online*, 2012, vol. 54, pp. 1-12.

[24] Ministère des Armées, France. 'International Law Applied to Operations in Cyberspace', October 2019, at: <https://www.defense.gouv.fr/content/download/567648/9770527/file/international+law+applied+to+operations+in+cyberspace.pdf>.

[25] Finland's National Positions, International Law and Cyberspace, 2020, at: <https://front.un-arm.org/wp-content/uploads/2020/10/finland-views-cyber-and-international-law-oct-2020.pdf>.

[26] Schondorf R., 'Israel's perspective on Key Legal and Practical Issues Concerning the Application of International Law to Cyber Operations', in: *EJILTALK!*, December 9, 2020, at: <https://www.ejiltalk.org/israels-perspective-on-key-legal-and-practical-issues-concerning-the-application-of-international-law-to-cyber-operations/>.

although the International Humanitarian Law regime is full of loose ends and general notions that cannot be seen as self-executing in the cyber context. Hence, the application of International Humanitarian Law can overstretch such norms, their material content and design are not tailored for cyberspace.

III. Applicability of International Humanitarian Law in Terms of Relevancy, Adequacy and Sufficiency

A collective affirmation of the applicability of International Humanitarian Law to cyber operations can also lead to a disappointment, as, besides applicability *in abstracto,* what deserves close scrutiny are the questions of whether International Humanitarian Law norms are relevant, adequate, and sufficient to deal with *military* types of cyber operations. At least three problematic issues can be identified.

First of all, what should be emphasised is the scarcity of International Humanitarian Law provisions applicable to 'military operations' even in international armed conflicts. It creates a problem as the majority of cyber operations will not reach the threshold of the International Humanitarian Law notion of 'an attack', and be qualified as 'military operations'. Under Articles 51 (1) and 57 (1) of the First Additional Protocol the duties of the parties to international armed conflicts are too general and laconic, imposing a general protection and constant care of the civilian population. Let me once again use an example from France. The French Ministry of Defence has developed a broad approach to the 'use of force' and considered cyber operations without physical damage as falling under this notion, but in the end, it had to admit that as 'most operations including offensive cyberwarfare operations carried out by France in an armed conflict situation remain below the attack threshold', 'they remain governed by general principles of IHL'[27].

The second problem arises when States try to circumvent the limitations of the scope of 'an attack' under International Humanitarian Law by stretching this notion to embrace more types of cyber operations. For instance, in his remarks of 10 November 2016, the US Legal Advisor Brian Egan opined that, although 'not all cyber operations rise to the level of an 'attack' as a legal matter under the law of armed conflict', it is still possible to qualify such a cyber operation as an attack, 'considering, among other things, whether a cyber activity results in kinetic or *non-kinetic* effects (emphasis added), and the nature and scope of those effects, as well as the nature of the connection, if any, between the cyber activity and the particular armed conflict in question'[28]. It can be suggested that the use of this method will result in an *objec-*

[27] Ministry of Armed Forces, 'International Law Applied to Operations in Cyberspace', October 2019, p. 13.

[28] Egan B., 'Remarks on International Law and Stability in Cyberspace', 10 November 2016, *Speech at Berkeley Law School*, at: <https://2009-2017.state.gov/s/l/releases/remarks/264303.htm>.

tive inapplicability of International Humanitarian Law provisions dedicated to 'attacks', simply because they are thought and designed to govern kinetic operations.

The third problem connected with the applicability of International Humanitarian Law to cyber operations originates from the fact that perpetrators of cyber-attacks can be in different densities of alliance with the State or a non-governmental party to the armed conflict. Combined with the different nature of cyber operations, this fact can render the rules and concept of 'direct participation of hostilities' in its different incarnations reflected in legal scholarship and jurisprudence[29] under-inclusive. This outcome can result either from requiring the existence of a very strict connection of a person carrying out a cyber operation with the party to the conflict for the purposes of his or her classification as a combatant in international armed conflicts or a member of the armed forces or groups in non-international ones or from the requirement of the infliction of kinetic-like harm, the existence of direct causation, or of a 'belligerent nexus' in a sense that a cyber operation should be 'specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another'[30].

All these arguments encourage us to be critical with respect to the 'whether and how' frame and revisit it once again. The question of whether and how International Humanitarian Law is applicable to cyber operations is answered on the basis of the consequentialist and effects-based logic. From this perspective, it seems to be retrograde and incorrect not to affirm the applicability of the International Humanitarian Law provisions to cyber operations resulting in deaths, injuries, and the destruction of objects. However, should we not think about whether effects-based logic alone is legitimate? Where should we stop deploying the causal link? What is so specific about cyber that we do use the effects-based approach in contrast to the case of economic coercion? Hence, I am suggesting a need to set the thresholds and criteria and not to overstretch them and, thereby, undermine International Humanitarian Law.

[29] See 'Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law', ICRC, 2009, pp. 46-64, at: <https://www.icrc.org/en/doc/assets/files/other/icrc-002-0990.pdf> (*Interpretive Guidance*); another view at the number of key elements of the 'direct participation in hostilities' is represented in the *Targeted Killings* case (Supreme Court of Israel, *The Public Committee against Torture in Israel v. the Government of Israel et al.*, Judgment, 13 December 2006, para. 39, at : <http://elyon1.court.gov.il/Files_ENG/02/690/007/a34/02007690.a34.htm>.

[30] Interpretive Guidance, p. 58.

CHALLENGES AND RESPONSES: CYBER OPERATIONS AND THE NOTIONS OF 'ATTACKS' AND 'OBJECTS' UNDER INTERNATIONAL HUMANITARIAN LAW
QUELS DÉFIS, QUELLES SOLUTIONS ? LES CYBER-OPÉRATIONS ET LES NOTIONS D'« ATTAQUE » ET DE « BIENS » EN VERTU DU DIH

Prof. Hongsheng Sheng
Shanghai University of Political Science and Law, China

Résumé

Hongsheng Sheng est professeur à l'université de droit et de science politique de Shanghai. Dans sa présentation, il analyse comment l'émergence de la cyberguerre redessine les conceptions traditionnelles relatives aux moyens et méthodes utilisés dans la conduite d'un conflit armé. Pour certains experts, les principes fondamentaux du DIH régulent tous les moyens et méthodes de guerre, y compris les cyber-opérations. Pour la Chine, « le DIH ne doit pas s'appliquer à la cyberguerre puisque la cyberguerre est interdite. Affirmer l'applicabilité du DIH au cyberespace revient à introduire des règles d'engagement dans le cyberespace et à reconnaître la légitimité des cyberguerres »[1].

Néanmoins, il est possible de tenter d'appliquer les principes et règles du DIH à la cyberguerre dans l'espace virtuel par analogie avec ce qui est applicable aux conflits armés tangibles. Ainsi par exemple, le concept « d'attaque » traditionnel fait référence à une conduite recourant à la force armée causant des dommages physiques, des pertes en vie humaine et des dommages matériels, parfois des effets psychologiques. Or, les cyber-opérations peuvent-elles être associée à de la force physique ? Les qualifications de victimes ou de combattants, liées au principe de distinction entre civils et militaires deviennent plus complexes. En effet, comment délimiter ceux qui participent aux hostilités dans le cyberespace, entre des soldats, ingénieurs ou d'autres participants en ligne ? La distinction entre objet civil et cible militaire est également rendue floue en raison de la pluralité de fonctions d'une même infrastructure. En cas de doute, du principe d'humanité découle le principe de précaution, qui ferait présumer de la nature civile de l'infrastructure. Quoi qu'il en soit, les principes généraux de droit, tels que le devoir de « diligence » (Manuel de Tallinn, règle 6), continuent de s'appliquer, ainsi que « la clause de Martens », en l'absence d'autres règles. Enfin, bien qu'il existe des tentatives de rédiger des principes et règles communes pouvant servir de base à un traité, tel que le Manuel de Tallinn, cela n'a, à ce stade, aucune force contraignante. De plus, un tel projet serait confronté à de nombreuses difficultés :

[1] Ministry of Foreign Affairs, the People's Republic of China, Intervention by Counselor Wang Lei, Head of Chinese Delegation, on the Application of International Law in Cyberspace at the First Formal Session of UN OEWG (2019/10/18), at: <https://www.fmprc.gov.cn/mfa_eng/wjb_663304/zzjg_663340/jks_665232/kjfywj_665252/t1708831.shtml>, last visit October 10,2020.

intérêts étatiques variés, limites techniques ou encore insuffisance des ressources à la disposition de la communauté internationale. Un tel texte est donc peu probable dans un avenir proche.

It has been nearly 20 years since the first operations were conducted in cyberspace. As a result, the emergence of cyber warfare has affected means and methods employed in armed conflict, if not totally toppled the basic principles and rules of IHL. Thus, challenges have been imposed upon almost all areas of IHL, from subjects of belligerency and the principle of discrimination to the principle of proportionality, and so on and so forth.

Although it was held by some scholars that even in cyber warfare, fundamental principles like humanity should still regulate any possible means and method used in armed conflict of various categories, including cyber operations, the mainstream perspective in this aspect in China is that 'IHL shall NOT apply to cyber warfare since it is unlawful, confirming application of the law of armed conflict in cyberspace amounts to introducing rules of engagement in cyberspace and recognizing the legitimacy of cyber wars'.[2]

Nevertheless, to achieve basic purposes of peace and order, from which international law obtains its own legitimacy, it is permissible, or at least not prohibited, to apply by analogy principles and rules applicable in real, tangible armed conflict to cyber warfare in virtual space, because the ultimate purpose of IHL is just to protect victims of armed conflict by imposing constraints on means and methods of warfare. Furthermore, although operations were carried out in cyberspace, the detrimental effects were in the real world. It was the principle of humanity that justified the application of IHL to all operations resorting to physical force, either in a 'traditional' way or those utilising a variety of new technologies of any sort.

As for further elaborations concerning the theme of the present webinar, some basic notions could be analysed as follows:

1. 'Attacks' in General

It is submitted that the traditional concept normally refers to a type of conduct resorting to armed force, which would cause physical damage, loss of life or damage to property, sometimes with a mental effect. In terms of the notion examined today, it could be rather vague, because a series of questions would be raised, such as: is it still physical force? Is it invisible?

[2] Ministry of Foreign Affairs, the People's Republic of China, Intervention by Counselor Wang Lei, Head of Chinese Delegation, on the Application of International Law in Cyberspace at the First Formal Session of UN OEWG (2019/10/18), at: <https://www.fmprc.gov.cn/mfa_eng/wjb_663304/zzjg_663340/jks_665232/kjfywj_665252/t1708831.shtml>, last visit October 10,2020.

2. ‘Attacks’ in Particular

It could be argued that it is rather risky to over-simplify ‘attacks’ when illustrating their forms and functions in contemporary warfare. They are almost the same as they were in traditional warfare; however, a demarcation line should be drawn between two kinds of attacks: attacks against cyber networks, and attacks against other targets or objects via networks like power stations, water supply systems and banking systems, e.g. what took place in the former Yugoslavia and in Estonia.

3. ‘Combatant’

Another challenge arose from the introduction of new technology into modern warfare like cyber warfare. Perhaps the most prominent one: who can be considered as a combatant? It is quite hard to classify those involved in hostilities in cyber space: soldiers, operators, engineers and even teenage digital game players. Under these circumstances, will the traditional principle of discrimination still be practical and applicable?

4. ‘Victim’

It was held that a question to follow is how to define victims. Normally they refer to civilians and the civilian population. However, as both military and civilian persons depend on some infrastructure and facilities, the ascertainment of victims will be very difficult.

5. ‘Object’ v. ‘Target’

Due to the paradoxical nature of the dual functions of so many facilities, sometimes there is only a blurred boundary between military targets and civilian objects. Under this circumstance, when initiating an attack, it should be limited to the minimum damage, avoiding accidental casualties as much as possible. The bottom line arising from the principle of humanity would lead to the principle of precaution: to impose a presumption of civilian nature, when in doubt.

6. ‘Due Diligence’

It is considered now that what the Tallinn Manual did is merely to re-iterate current international law aimed to apply to cyber operation, so that the general principles of law are applicable to other situations, e.g. domestic situations of violence, turmoil and demonstrations, will still function as usual. Just as the text in the Tallinn Manual indicates, all States bear a responsibility to follow the principle below:

Rule 6 – **Due diligence** (general principle)

A State must exercise due diligence in not allowing its territory, or cyber infrastructure under its governmental control, to be used for cyber operations that affect the rights of, and produce serious adverse consequences for, other States.[3]

7. The Legacy of the Martens Clause

It could be argued that as a matter of fact, there is always a way out when we come across a dead end in terms of new challenges where there are no regulating principles and rules. The Martens Clause will serve as a safety valve to some extent.

'Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity, and the requirements of the public conscience'.[4]

At least, it would be the bottom line to impose constraints on the negative effects by future, yet unknown military technology.

8. Comprehensive Convention Issue

It is concluded that although there are some attempts to draft principles and rules applicable to cyber operations like the Tallinn Manual, yet it is just a work by non-governmental experts only, without legally binding force at present. Nonetheless, there remains the possibility that it could function as a basis for a multilateral treaty to be concluded by States in the future under the auspices of the United Nations. Difficulties lie in the reality that there exist various concerns of State interests, alongside technical issues. Meanwhile insufficient techniques of and lack of resources for drafting an international legal instrument are also facing the international community as a whole, and all these make it difficult to conclude a universally accepted treaty at the moment, and possibly for a long period of time to come.

[3] Michael N. Schmitt, et al (eds), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Cambridge University Press, 2017, p.30.

[4] See: Martens Clause, Military Wiki Fandom, at: <https://military.wikia.org/wiki/Martens_Clause#cite_note-RT-3>. Last visit 22 November 2020.

IMPROVING TRANSPARENCY: INTERNATIONAL LAW AND STATE CYBER OPERATIONS

POUR PLUS DE TRANSPARENCE : LE DROIT INTERNATIONAL ET LES CYBER-OPÉRATIONS MENÉES PAR LES ÉTATS

Prof. Duncan Hollis

Temple Law School

Résumé

Duncan Hollis est professeur à la Temple Law School et rapporteur d'un projet concernant la transparence des États sur les cyber-opérations au Comité juridique interaméricain de l'Organisation des États américains (OEA). Dans sa présentation, il présente le projet et ses principales conclusions.

Initié en 2017, le projet a quatre objectifs principaux : 1. identifier des zones de convergence d'interprétation entre États ; 2. identifier les divergences ; 3. réduire le risque d'escalade ou de conflit dû aux différentes interprétations ; 4. Exemplifier le rôle de l'OEA et ses États Membres dans ces négociations et faire entendre leur voix sur la scène internationale. Le projet comprenait un questionnaire envoyé aux États Membres et une demi-journée de réunion avec des représentants des ministères des Affaires étrangères de 16 États Membres.

Le questionnaire a été envoyé à 44 États et a reçu 9 réponses, dont un renvoi à de précédentes déclarations, ce qui en soi souligne les limites et les défis posés à la transparence des positions des États sur le sujet. Dans l'ensemble les États ayant répondu – Bolivie, Chili, Costa Rica, Équateur, Guatemala, Guyane, Pérou, États-Unis, Brésil – partagent un même attachement à l'état de droit. Ils reconnaissent également que les États ont des obligations positives et négatives, régulant leurs pratiques dans le cyberespace.

Toutefois, dans le détail, les réponses révèlent un réel manque de compétences, tant techniques que juridiques, ainsi qu'une divergence de positions sur de nombreux sujets. Ainsi, la notion de responsabilité et les règles d'attribution dans le cyberespace présentent des difficultés pour les Etats, notamment pour distinguer « contrôle effectif » et « contrôle global ». Les questions portant sur les différences entre conflit armé et conflit cyber n'ont pas non plus abouti à des réponses claires. Sur la notion d'attaque, le Chili, le Pérou et les États-Unis se sont accordés sur le fait qu'une cyber-opération qui ne cause pas de morts, de blessés ou de dommages physiques directs ne peut être qualifiée d'attaque, tandis que pour d'autres États, une perte de fonctionnalité peut entrainer la qualification d'une cyber-opération d'« attaque » *sous certaines circons-*

tances. Enfin concernant les données civiles, aucun des États ayant répondu ne s'est aligné sur la position du CICR selon laquelle les données civiles essentielles peuvent être en soi qualifiées de biens civils. Les États se rapprochant de cette position, par exemple le Chili et la Guyane, ont préféré mettre en avant les effets plutôt que l'appellation pour décider du principe de distinction.

Au-delà des compétences techniques, le manque de transparence sur ces sujets est également lié à des questions politiques : manque de formations des administrations et des gouvernements ou volonté de ne pas se positionner dans un débat impliquant les grandes puissances. Certains États font aussi le choix de rester silencieux sur cette question pour maintenir une flexibilité opérationnelle et ne pas créer ce qui peut être perçu comme des contraintes à une potentielle action future.

Enfin, le Comité juridique a recommandé à l'OEA d'adopter une résolution affirmant l'applicabilité du droit international au cyberespace. Le Comité poursuit ses efforts pour plus de transparence et de clarification de l'application du droit aux cyber-opérations étatiques.

The editorial team transcribed the contribution below based on the recording of the presentation of Professor Duncan Hollis.

1. Introduction

First of all, I would like to thank the ICRC for having invited me. This is my first Bruges Colloquium and I am very pleased to be here, even if only virtually.

We are in a world now where we are seeing extensive cyber operations being associated with States. Since 2005 thirty three States have been publicly associated with one or more cyber operations. In 2019, there were at least 76 State or State-sponsored cyber operations.[1] Moreover, Covid-19 has catalysed a new range of cyber operations targeting hospitals, targeting the World Health Organisation, and most recently targeting vaccine research.

As a result, States have increasingly recognised the need for rules on cyberspace, most often looking at international law as the vehicle for doing so. There is widespread agreement that international law applies to States' cyber operations. See, for example, recognition in:

- United Nations General Assembly (UNGA) Resolution 73/266 of 22 December 2019 on Advancing responsible State behaviour in cyberspace in the context of international security;

[1] <https://www.cfr.org/cyber-operations/>.

- Association of Southeast Asian Nations-United States (ASEAN-US) Leaders' Statement on Cybersecurity Cooperation (2018)[2];
- European Union (EU) Statement (2018)[3];
- UN Group of Governmental Experts, First Committee, UNGA (2015).[4]

The challenge has been moving beyond the general availability of international law in regulating State behaviour in cyberspace to its particulars. We still know very little about *how* international law applies. Most States have remained silent on how they view the application of international law. They just have not given an opinion at all. For the minority of States that have opined, moreover, we see key differences.

- First of all, there are existential disagreements, where the very existence of a legal regime, like IHL (or due diligence), is challenged by several States;
- Second, when we do have agreement on the application of international law (e.g. the duty of non-intervention) we see large interpretive disagreements, such as the need for 'coercion' to trigger the duty of non-intervention, not to mention disagreement on what behaviour constitutes coercion with respect to cyber operations.
- Third, we have the challenge of attribution. And here I do not mean the challenge of just technical attribution – which computers or which users are responsible for deploying cyber exploits but – what are the international legal evidentiary standards and the amount of control needed over proxies to trigger State responsibility for identified internationally wrongful behaviour with respect to cyber operations?

[2] Available at: <https://asean.org/storage/2018/11/ASEAN-US-Leaders-Statement-on-Cybersecurity-Cooperation-Final.pdf>.

[3] EU Statement – United Nations 1st Committee, Thematic Discussion on Other Disarmament Measures and International Security (Oct. 26, 2018) ('EU Statement'), at: <https://eeas.europa.eu/delegations/un-new-york/52894/eu-statement-%E2%80%93-united-nations-1st-committee-thematic-discussion-other-disarmament-measures-and_en>.

[4] See also UN Secretary-General, 'Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security', para. 24, UN Doc. A/70/174 (July 22, 2015); see also UN Secretary-General, 'Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security', para. 19, UN Doc. A/68/98 (June 24, 2013).

2. The Organization of American States (OAS) Juridical Committee's Improving Transparency Project

Over the last few years, several States have tried to offer 'official views' on how international law applies in the cyber context to bridge these gaps: US (2012, 2015, 2016) [5], UK (2018)[6], France (2019)[7], Estonia (2019)[8], the Netherlands (2019)[9], Australia (2019)[10], Iran (2020)[11]. These statements are so far dominated by Western, and specifically European, voices. There are other non-State actors' views: the ICRC[12] has been quite prominent in that respect as have the

[5] See e.g. Brian Egan, 'Remarks on International Law and Stability in Cyberspace' (Nov. 10, 2016), in: *Digest of U.S. Practice in Int'l Law* 815 (2016); 'U.S. Submission to Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security' (Oct. 2016), in: *Digest of U.S. Practice in Int'l Law* 823 (2016) ('2016 US GGE Submission'); 'U.S. Submission to Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security' (Oct. 2014), in: *Digest of U.S. Practice in Int'l Law* 732 (2014) ('2014 US GGE Submission'); Harold Koh, 'International Law in Cyberspace' (Sept. 18, 2012), in: *Digest of U.S. Practice in Int'l Law* 593 (2012). In 2020, the General Counsel of the US Department of Defense offered views on several key questions of international law's application to cyberspace. It is not yet clear, however, whether his views reflect those of the whole United States or only the US Department of Defense. See Paul C. Ney, 'DOD General Counsel Remarks at U.S. Cyber Command Legal Conference' (March 2, 2020), at: <https://www.defense.gov/Newsroom/Speeches/Speech/Article/2099378/dod-general-counsel-remarks-at-us-cyber-command-legal-conference/>.

[6] Jeremy Wright, QC, MP, 'Cyber and International Law in the 21st Century' (May 23, 2018), at: <https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century> ('U.K. Views').

[7] Ministère des Armées, « Droit international appliqué aux opérations dans le cyberespace » (Sept. 9, 2019), at : <https://www.defense.gouv.fr/salle-de-presse/communiques/communiques-du-ministere-des-armees/communique_la-france-s-engage-a-promouvoir-un-cyberespace-stable-fonde-sur-la-confiance-et-le-respect-du-droit-international> ('French Ministry of Defense Views').

[8] Kersti Kaljulaid, President of Estonia, 'President of the Republic at the opening of CyCon 2019' (May 29, 2019), at: < https://www.president.ee/en/official-duties/speeches/15241-president-of-the-republic-at-the-opening-of-cycon-2019/index.html> ('Estonian Views').

[9] 'Letter to the parliament on the international legal order in cyberspace', July 5, 2019, Appendix 1, at: <https://www.government.nl/ministries/ministry-of-foreign-affairs/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace> ('The Netherlands Views').

[10] Australian Mission to the United Nations, 'Australian Paper—Open Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security' (Sept. 2019), at: <https://s3.amazonaws.com/unoda-web/wp-content/uploads/2019/09/fin-australian-oewg-national-paper-Sept-2019.pdf> ('Australian Views');

[11] Available at: <https://unoda-web.s3.amazonaws.com/wp-content/uploads/2020/02/second-submission-by-iran-feb-2020.pdf>.

[12] See, e.g., ICRC, 'Position Paper on International Humanitarian Law and Cyber Operations during Armed Conflicts' (Nov. 2019) ; see also ICRC, 'Report on International Humanitarian Law and the Challenges of Contemporary Armed Conflict' (Nov. 2019); ICRC, 'International Humanitarian Law and the Challenges of Contemporary Armed Conflicts', (Oct. 2015), pp. 39-44.

Tallinn Manuals[13], although these are not authoritative. In addition, ongoing efforts at the UN Group of Governmental Experts (GGE)[14] support having more national statements about how international law applies.

Yet, the reality is that the application of international law is not exclusively a UN project, nor does it depend only on European States (or only on those States that have the most capacity in space) speaking out. It does in some ways depend on all States. As the EU has emphasised, we need all States to have the opportunity, and be encouraged, to offer their national views to delineate and describe their opinions on international law's applications to cyber operations.

If State silence is the dominant position, however, we need to ask: how can we move beyond that? How can we get States to offer their views on these issues?

That is the premise of the project I have led at the Organisation of American States (OAS), as part of the Inter-American Juridical Committee. Although I am a professor at Temple Law, today I am also speaking on behalf of the OAS. The Juridical Committee is one of its principal organs. We are eleven lawyers, elected by the OAS General Assembly to deal with and advise on issues of international law.

Beginning in 2017, the Committee put on its agenda the challenge of transparency and State operations in cyberspace. I was selected as the project's Rapporteur. The project has four goals.

1. First, to identify areas of convergence: where do we see agreement on States' understandings of which rules of international law apply and how they do so?
2. Second, to identify the divergences, to set the baseline for further dialogue, for reconciliation, or maybe even for efforts to reach converging interpretations.
3. Third, to reduce the risk of inadvertent escalation or conflict due to differing States' understandings of international law's application. For example, if States have different thresholds for identifying an attack whether under the *jus in bello* or the *jus ad bellum*.

[13] See Michael N. Schmitt (ED.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2017) ('Tallinn 2.0'); see also ICRC, 'Report on International Humanitarian Law and the Challenges of Contemporary Armed Conflict' (Nov. 2019); ICRC, 'International Humanitarian Law and the Challenges of Contemporary Armed Conflicts', (Oct. 2015) pp. 39-44.

[14] Group of Governmental Expert (GGE): 'In GA resolution 73/266, the Secretary-General was requested to establish a Group of Governmental Experts on Advancing responsible State behaviour in cyberspace in the context of international security', at: <https://www.un.org/disarmament/group-of-governmental-experts/>.-

4. Fourth, to instantiate the OAS in these global conversations about international law's application and to elevate its Member States' voices.

The project has proceeded through two threads. First, the OAS asked its Member States, via a questionnaire circulated by the Department of International Law, what they think about the application of international law in cyberspace. And then later, in June 2020, I hosted a half-day meeting under the Chatham House rule with representatives from 16 Member States' Foreign Ministries.

With respect to the questionnaire, we received nine responses. Seven of them were direct and substantive responses: Bolivia, Chile, Costa Rica, Ecuador, Guatemala, Guyana, Peru. The United States forwarded references to earlier statements while Brazil referenced its forthcoming work in the context of the GGE.

Given this is the Bruges Colloquium, I thought I would focus my remarks on the questionnaire responses most relevant to those affiliated or associated with IHL and the ICRC.

3. What States' Responses Revealed

First, it was interesting that all the States that did respond shared a dedication to the rule of law. There was an endorsement that there are positive and negative legal obligations governing State behaviour in cyberspace. That said, there is also a real and concrete lack of capacity, both technically and legally.

I had ten questions in my questionnaire. If I may, let me talk about four that are most relevant to this group.

State Responsibility

First, there were two questions about State responsibility: that is about proxies, about when a State would be responsible for the acts of a non-State actor (NSA) in cyberspace, and whether those standards vary depending on whether the operation is occurring inside or outside an armed conflict. It is worth recalling the different control standards we have seen articulated by different international judicial bodies, with the International Court of Justice (ICJ) famously endorsing a standard of 'effective control' to trigger State responsibility.[15] and the International Criminal Tribunal for the former Yugoslavia (ICTY) *Tadić* case having a looser

[15] 'Military and Paramilitary Activities in and Against Nicaragua' (*Nicaragua v. U.S.*) [1986] ICJ Rep. 14, para. 115 (June 27).

'overall control' standard which was later endorsed by the International Criminal Court (ICC).[16] Part of the question was: does that 'overall control' standard get endorsed by States in the context of cyber operations involving IHL?

Here, there was a general concern by States about the difficulty of attribution in cyberspace. States are still cautious with this, even if a number of them, including the United States and a number of companies, like Microsoft and Google, are getting more comfortable attributing behaviour in cyberspace.

Peru indicated that if States have the capacity to control a National Security Agency (NSA) who commits a cyber-attack, that could give rise to the attack being attributed to the State.

Bolivia stressed the corollary that States should not bear responsibility for an NSA's acts outside of the State's control or beyond its technological or judicial capacity.

A number of States – Chile, Guyana and Peru – cited Article 8 of the Articles of State Responsibility[17] to affirm that where the State has control, it can be held internationally responsible.

However, it remains unclear what 'in control of' means. Chile was the only State to choose a side in the 'effective' versus 'overall' control question. It said that the narrower effective control standard should apply and it should apply across all situations, both inside armed conflicts and outside.

Differences between Armed Conflicts and Cyber Conflicts

- Peru did not detail the level of control required but shared the idea that whatever the standard of State responsibility, it would apply across all State operations; there would not be a different standard in an armed conflict and outside.
- Guatemala suggested that this is a topic that bears further discussion, that we need to have further dialogue on this question: 'international forums must continue their discussions on the uniquely different aspects that a conflict in cyberspace would entail, particularly regarding such issues as attribution and territorial considerations'.

[16] *Prosecutor v. Dusko Tadić aka 'Dule'* (Judgment), ICTY-94-1-A (15 July 1999), para. 131-145, 162; see also *Prosecutor v. Lubanga*, Case No. ICC-01/04-01/06, Trial Chamber, Judgement (Int'l Crim. Court, March 14, 2012) para. 541.

[17] Article 8 of the Articles on State Responsibility: 'The conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct'.

- Other States focused on the question as to whether there are differing standards of State responsibility for international and non-international armed conflicts, which was not what the question had sought to address.

Cyber Operations and the Notion of 'Attack'

Next, we had a question on when would a cyber operation constitute an 'attack' under IHL. Can a cyber operation constitute an attack for IHL purposes if it does not cause death, injury or direct physical harm? To put it another way, can we have a cyber-attack for IHL purposes with just a loss of functionality? I know that the ICRC has itself emphasised how important it is to have this question answered; how widely or narrowly the notion of 'attack' is defined in cyberspace has tremendous implications for IHL's application.

- On this point, Chile, Peru, and the United States agree that a cyber operation could not qualify as an attack if it fails to cause death, injury or direct physical harm. The implication is that if a cyber operation produces non-kinetic or reversible effects (say as ransomware does as it is deployed in your system, it prevents you from using your system but it does not require you to replace any physical components), then that might not constitute an attack even if it might, for example, deprive vaccine facilities of much needed access to clinical trial results.
- Guatemala and Ecuador opined that in some cases, functionality loss alone, rather than death, injury, or destruction of property, could qualify as an attack.
- Bolivia and Guyana's responses were more equivocal. They both emphasised that you need an ability to cause loss of life, injuries, death or destruction to be an attack, but at the same time they said that if an operation disabled a State's basic infrastructure (e.g. water or electricity), that would be an attack, without addressing what would happen if that disabling operation was only functional in that it did not cause any kinetic harm (e.g., to prevent people from accessing water or electricity, just if the lights were turned off, so to speak).

These responses suggest a need for continuing dialogue, particularly on how proximate the death or destruction must be to have functionality losses qualify as an 'attack' for IHL purposes.

Civilian Data

Finally, we asked Member States a series of questions about data: is an operation that targets civilian data protected by IHL? If a cyber operation only targets data, can that qualify as an 'attack'? Is data an object to which we can apply the principle of distinction? Is there civilian data and military data? While the Tallinn Manual says that data cannot alone qualify as an object for IHL purposes, the ICRC has been more expansive, using the term 'civilian essential data' to suggest there would be IHL protections[18].

- None of the responding States appeared willing to adopt the ICRC's view just yet. No State endorsed civilian data alone as subject to the principle of distinction in armed conflict. Several emphasised the principle of distinction without implying whether or not data could be an object.
- Chile acknowledged that data cannot qualify as an object under Additional Protocol I because data is intangible. At the same time, Chile suggested the potential knock-on effects of targeting data could trigger the principle of distinction (i.e., an operation targeting data could affect civilian populations in ways that do cause death, destruction or physical harm indirectly, if not directly). As such, Chile suggested that States should take the military and civilian distinction into account when they are targeting their cyber operations, unless the data is being used just for military purposes.
- Guyana similarly focused on effects rather than on the label of data as an 'object' (or not) as a way to decide on the applicability of the principle of distinction to a cyber operation.
- Peru added more details in terms of distinguishing data used within a military setting by distinguishing which data might be a lawful versus an unlawful target, i.e. a military objective or not. Peru suggested that data about troop communications in the field, synchronisation of a military arsenal, or location of enemy aircraft could be the object of a cyber operations, but other data systems used in a conflicts – such as data for a field hospital or medical information – would be off limits, without necessarily explaining whether it was off limits because other IHL rules provide protections to medical facilities or if the data itself warrants such protection.

4. Challenges to Further Transparency on IHL's Application in Cyberspace

The eight States offering their views to the OAS provided us with interesting and important information on where they are willing to opine. That said, asking thirty-four Member States for their views and only getting eight substantive responses affirms that States remain reluctant to talk openly about whether and how international law applies.

This point was really confirmed at the June meeting I held with representatives of sixteen OAS Member States. There, it became clear that there are capacity issues, including technical, legal

[18] See for instance the ICRC Report, 'International Humanitarian Law and the challenges of contemporary armed conflicts', 2019, p.28, at: <https://shop.icrc.org/international-humanitarian-law-and-the-challenges-of-contemporary-armed-conflicts-recommitting-to-protection-in-armed-conflict-on-the-70th-anniversary-of-the-geneva-conventions-pdf-en>.

and internal issues. Technically, some States are still building up their resources to engage in cyber operations or just getting to the very minimum levels necessary to defend against them. Legally, there is a lack of governmental expertise (or resources) on cyber-related issues. And, just as importantly, there was also in some States a lack of capacity to understand how international law manifests itself in the cyber context. Some OAS Member States were unfamiliar with many of the issues that have been so prominent in the ICRC's analysis for almost two decades now. There are also political issues with States looking to avoid taking views that may entangle them in political tensions or disagreements with other States.

These capacity issues suggest that we need more training for Foreign Affairs ministries and Defence departments' lawyers to understand the relevant issues. Without this, it is not surprising that States have not had the internal coordination that is required to formulate a view. Some States are, by their own admission, quite behind with the question of 'what is the appropriate control standard for cyber operations', 'whether civilian data can be the object of a military operation'. Also, some States, being quite candid, indicated that they want to avoid entanglements in great power politics, that taking a view on these issues means they may be dragged into a larger debate in which they do not wish to be involved. Other States' silence may be due to a desire to preserve operational flexibility: these States do not want to draw a line now in an international statement that might constrain their own future cyber operations.

5. The Future?

The Juridical Committee has recommended that the OAS itself adopts a resolution affirming international law's application in cyberspace, including IHL. Meanwhile, the OAS Juridical Committee plans to continue working on the application of international law in cyberspace:

- On 7 March 2021, it will host another virtual discussion with OAS Member State representatives on the International Association of Journals and Conferences' (IAJC) work
- January 2021: Mariana Salazar Albornoz takes up the topic as the new Rapporteur with the prospect of further questionnaires and/or legal capacity building for Member States' legal representatives.

I am hopeful that this effort of transparency, just to clarify how States understand these issues will help clarify international law's applications to cyber operations, both inside the IHL context and beyond. In doing so, we can also look more at whether the law can provide concrete and effective ways to govern States' cyber operations and contribute to the rule of law.

Thank you for your attention.

DISCUSSION

This first virtual Q&A session allowed participants to raise and discuss in depth the following topics.

1. Applicability of IHL to Cyber Operations during an Armed Conflict

General Applicability

A first question raised by the moderator was aimed at clarifying the applicability of IHL to cyber operations during an armed conflict. He noted the lack of clarity in States' views as to whether cyber operations alone would amount to an armed conflict. He wondered how States who showed reluctance to acknowledge IHL applicability during an armed conflict would look at situations where cyber operations are used in addition to kinetic operations. The moderator gave the example of the attacks that are carried out against hospitals by cyber-means as part of an ongoing armed conflict. A cyber operation may target and destroy the electricity generating functions of hospitals and therefore generate the death of the hospital's patients. He noted that IHL rules protecting medical facilities are not presupposing that harm is caused by the sometimes controversially discussed notion of an attack. According to one of the panellists, in many of these situations, cyber operations which meet the criteria of the threshold being analogous to the kinetic operations and being military by nature, providing that these operations can be attributed to a party to the conflict, indeed fall under the application of IHL. One should also wonder whether the notion of proximity is verified, i.e. if the causal link between a cyber operation and the results are analogous to kinetic attacks. This would be the case in the example cited in the question. However, the panellist stressed that this proximity would not be easily established in all cyber operations adversely affecting one of the parties to the conflict.

States' Positions

A major issue concerning the applicability of IHL to cyber operations in armed conflict which was raised in the discussion relates not necessarily to the theoretical applicability but, rather, to the actual positions of States. For instance, following a question on the position of the Russian Federation, a panellist explained that it might be difficult to assess Russia's position. Russia, like other States, has expressed its views on several topics but its position is not necessarily publicly shared on all topics. Based on the official positions, the one of the Russian Federation is rather strict with respect to the possibility for a cyber operation to qualify as an attack and to reach the threshold so that the application of IHL begins. For the Russian Federation, the applicability of IHL depends on whether a cyber-attack which takes part or is conducted outside the kinetic armed conflict can give rise to an international armed conflict (IAC) or a non-international armed conflict (NIAC). The Russian Federation has concerns especially on cyber operations which are carried out by civilian perpetrators. From a legal perspective, such a situation can be assessed based on the rules of attribution in public international law, in order to assess whether these operations can be attributed to a State or to a body participating in a conflict.

Applicability to Specific Situations, e.g. Cyber Operation Targeting Financial Institutions

A participant asked a question on the applicability of IHL to cyber operations specifically targeting financial institutions, for instance if some groups that are party or are linked to a party to an armed conflict take control of a cryptocurrency. The moderator recalled that first, the basic prerequisite for IHL to be applicable is that the operation occurs in the context of an armed conflict. Then, a panellist suggested that this raises the question of the link between the groups and the party to the armed conflict. If the link is not close enough, this means that the law applicable in times of peace, e.g. criminal or domestic law, could regulate some operations or activities. Lastly, a panellist explained that if there was already an armed conflict, the rules of IHL would apply, including the principle of distinction. The question thus becomes related to the dual use, i.e. whether financial institutions are used by civilians who are within the zone of the conflict primarily or essentially for civilian purposes, or are also used by the military. Assuming that it is not the case, this situation is reflected by well-known examples of kinetic operations from the past where forces in the Balkan wars refrained from targeting banks or other financial institutions, for the precise reason that they were concerned that doing so would breach the principle of distinction.

2. Legal Framework Applicable to a Cyber Operation Falling below the Threshold of an Attack

A participant asked what it would mean if cyber operations were used before engaging in armed conflict, for instance to disrupt utilities in order to create public disorder and foment civil unrest to degrade the adversary's ability to defend itself: would that fall under IHL? Building on the question, the moderator suggested that this question could include two different scenarios, the first one disrupting utilities before any hostilities and kinetic operations have started, and the second doing this amid ongoing hostilities.

Before an Armed Conflict

A panellist explained that no State has claimed that cyber operations which are not analogous to kinetic attacks can give rise to an armed conflict, neither IAC nor NIAC. Indeed, if IHL would apply to such situations, this would mean that some IHL provisions also apply in peacetime. Therefore, the provisions dedicated to and governing hostilities and military operations would not be applicable to this situation. An unrest or public disorder would be below the threshold of applicability of IHL either because of the lack of intensity or because of the lack of required level of organisation for being considered a party to a (non-international) armed conflict. However, other legal provisions would be applicable, although they may also need further clarification. Firstly, the principle of non-intervention would be applicable, despite several gaps and disagreements on how it should be interpreted. Secondly, the principle of sovereignty may also apply. But again, there are also discussions among States on the interpretation. For instance, several States, initially the United States, officially supported by the United Kingdom, have claimed that sovereignty is a fundamental principle but not a rule in public international law. In any event, the situation above should be assessed in the light of the principle of non-intervention, with the criteria which were also developed by the International Court of Justice and depending on the intensity of the actions which constituted the cyber operation.

During an Armed Conflict

The second scenario, that the question entails, relates to the same type of cyber operation disrupting utilities to create public disorder, but after the outbreak of hostilities. The panellist suggested that it raises the question of which governments participate in the conflict, which parties, both governmental and non-governmental, are actually supported by the States. Assessing the situation would also require thinking about whether the cyber operations have effects which are analogous to kinetic attacks. However, if it is not the case that the question implies, e.g. just information campaigns or fake news, one would have to assess whether the acts committed occur with a link to the hostilities or a conflict, e.g. are committed by the governmental side of an ongoing NIAC, or without a link. In the latter case the acts would not be covered by IHL, in the former case it has to be assessed whether they are falling under the notions of ‘attack’ or ‘military operations’ or not. Based on what is suggested in the question the situation seems unlikely to reach either of the thresholds.

No Legal Void

Following up on these conclusions, another panellist insisted on a previous point: there is no situation in which international law does not regulate State behaviour. If a State is involved and engaged in the operations that are described in the question, whether it is an IAC or a NIAC, the question is which international law rules would apply. Despite the remaining debates on what the principles of non-intervention or sovereignty mean, these principles are not rejected by any State. Any situation in which a behaviour is attributable to a State would end up with a legal discussion, no matter what.

3. Adequacy and Interpretations of IHL Provisions

Adequacy of IHL

On a question on the adequacy of IHL provisions to cyber operations in armed conflict, overall, the panellists emphasised that the adequacy is not necessarily systematic or perfect, but reiterated that this certainly does not mean that there is a legal vacuum in the matter. A panellist suggested that if cyber operations are similar to kinetic operations, IHL provisions may certainly be found adequate. Nonetheless, the panellist also suggested that scholars should not assume or pretend that IHL rules provide adequate answers in all cyber operations. Scholars should recognise that there can be cyber operations of different natures, they can target different objects and can be of different proximity, with sometimes more or less causal links, and they can also target different sectors. However, another panellist rephrased the question in terms of 'default rules'. Indeed, the challenge is: what do we do by default? Could we have a better set of rules of law tailor-made to cyber space? Although this panellist is convinced that a cyber-specific set of rules could theoretically exist, this is not the case at this stage. Therefore, the question becomes: how should existing rules apply? Some situations at the very least would possibly be covered by IHL, e.g. the electricity system of a hospital that has become dysfunctional and causes deaths. In other situations, in which IHL may not apply, international law or other legal frameworks remain applicable.

Notions of 'Attacks' and 'Immaterial Effects'

Regarding the notion of 'immaterial effects' a participant asked to what extent they could be considered as attacks under IHL, for instance if a cyber operation destroys data or entails a total loss of functionality of an infrastructure.

For one of the panellists, data, like information, is neutral, i.e. it is not civilian or military by nature, it depends on the purpose for which the data is to be used. Information has a very peculiar character which is that it may be shared by anyone. That is the difference between information and intelligence. Intelligence can be considered as a refined type of information, which is possessed for a purpose, for a meaningful function: a military operation. Therefore, one cannot say that a cyber operation on data is a lawful attack under IHL at all times. It depends on the purpose, what are the functions of the data, functions which could have been exploited.

Another panellist emphasised that there are ongoing discussions among States on this issue. Several States rely on Additional Protocol I or IHL more generally and look for a kinetic or tangible element to define an attack or an object. However, an operation targeting data is not kinetic or tangible at first sight, so that some States conclude that such an operation falls outside the terms described under IHL. At the same time, States are aware that scholars and organisations such as the ICRC are looking for analogies to what was protected under the traditional principle of distinction. The notion of data entails a grey area. This is novel in some of the ways the data can be used and the effects an operation affecting data can have. Anyone would agree that targeting somebody's shed in the backyard in an armed conflict would trigger the rules of IHL. However, the debate on data is ongoing, while there are certain essential civilian data whose targeting could have a greater effect on human lives than the targeting of somebody's shed in their backyard. For instance, taking out all hospital records for various patients, all their drug interactions, all their prescriptions, or targeting social security data would have severe consequences. The panellist recalled that the principle of humanity remains the overarching principle that motivates IHL. There are situations, particularly the rise and the ability to have cyber operations, that can disable a facility in its entirety. They can stop a hospital from working, with real knock-on effects that can cause death: someone who is unable to go to hospital would drive and then die on his/her journey to another hospital because the first hospital was shut down. According to the panellist, this is a material effect of an operation which at first sight looks immaterial. It is a foreseeable consequence that if a cyber operation turns off all the systems in a hospital, it can cause a real risk of mortality. In addition, the panellist stressed that a mere loss of functionality can have such great impact on civilians that it is crucial to ask about IHL's ability.

If it is believed that IHL does not cover a situation, legal experts should seek further interpretation or legal clarifications for the sake of the protection of civilians and the great risks these operations entail. The moderator concluded that when the effects are similar to kinetic attacks, which can be measured in death, injury and destruction, it is likely that IHL applies. Improving and sharing knowledge on the different forms of cyber operations and the impact they can have on civil society may lead to more acceptance or thinking that perhaps the initial restrictive interpretations were not sufficiently protective from an IHL perspective.

4. Steps Forward to Enhance the Protection of Civilians

The last question related to the protection of civilians: how can the developing understanding of IHL applicable to cyber operations best incorporate the importance of civilian protection? How is harm to be defined, identified and mitigated against? It was answered that cyber operations have been discussed for the last 20 years in certain legal circles. What cyber operations can and cannot do has changed over these last 20 years. Initially, talking about cyber operations almost amounted to talking about science fiction in legal scholarship. Now, the ability to cause physical destruction through cyber operation and the ability to have lethal consequences are real. Therefore, the first stage is to make sure that all States are aware of what cyber operations are capable of. A dozen States now have civil and military capabilities.

It is critical for lawyers, governments and others to understand the factual possibilities, what is possible, what not, how difficult it is to conduct certain operations and what pre-conditions must exist to conduct certain operations. If lawyers do not understand the operational environment, it becomes very difficult to apply, for instance, the principle of distinction. A second step is to explain the concrete ways in which, directly and indirectly, a cyber operation can implicate civilians: what losing access to the online environment would mean, what happens to people if the internet is shut down, whether it is inside an armed conflict or outside it, but also what the loss of access to medical information or to critical infrastructures entails, whether it is electricity or something similar. There is a real need for capacity building both on the technical and on the legal sides.

In conclusion, the discussion highlighted different perspectives, which is also paying tribute to the political discussions that are taking place currently in various fora. A crucial point to bear in mind is that the use of cyber operations can and have already created a wide range of disastrous consequences for civilians. More is known on this topic now than 20 years ago. It is time now to find the right avenues and that these issues be sufficiently understood so that the policy and legal developments required to protect civilians take place.

Panel 2
The Use of Autonomous Weapon Systems: A Challenge to the International Rule of Law?
L'utilisation des systèmes d'armes autonomes : un défi pour l'état de droit international ?

INTRODUCTION

Maya Brehm
ICRC Geneva

Résumé

Maya Brehm est conseillère juridique au CICR et modératrice de ce panel. Les armes autonomes, parfois appelés « robots tueurs », font l'objet de nombreux débats, en particulier dans le cadre du Groupe d'experts gouvernementaux (GGE) sur les systèmes d'armes létaux autonomes (SALA) relevant de la Convention sur certaines armes classiques (CCAC)[1]. Les États parties à cette Convention reconnaissent l'applicabilité du DIH aux armes autonomes, la nécessité d'une interaction entre l'humain et la machine et le besoin de clarifier la nature et l'étendue de cette implication ou ce contrôle humains. Pour le CICR, ce sont des armes « ayant la capacité de sélectionner (chercher, détecter, identifier, pister) et attaquer (intercepter, faire usage de la force contre, neutraliser, endommager ou détruire) des cibles sans intervention humaine[2] ». Une fois lancée ou activée, une arme autonome utilise la force de sa propre initiative sur la base du codage d'une cible et du traitement de données qu'elle détecte dans son environnement. Il est difficile pour l'utilisateur de prévoir les frappes et donc les conséquences de l'utilisation de ces armes. Cela pose un réel risque de dommages à l'encontre des civils qui ne prennent pas part aux hostilités ou des biens à caractère civil, par exemple les habitations, les écoles ou les hôpitaux.

[1] La Convention sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discrimination du 10 octobre 1980, telle qu'elle a été modifiée le 21 décembre 2001

[2] ICRC, 'Statement to the Convention on Certain Conventional Weapons' (CCW), Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 13 April 2015, at: <https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS>https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS. For more detail, see: ICRC, 'Characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention', CCW GGE on LAWS, 9 - 13 April 2018, at: <https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/statements/10April_ICRC.pdf>.

Ce panel propose d'explorer trois types de questions soulevées par les armes autonomes. D'abord, des questions quant à l'interprétation et à l'application du DIH. Le DIH limite l'utilisation des armes autonomes et interdit certaines armes autonomes. Toutefois, certaines questions demeurent quant aux contraintes et conditions juridiques dérivant du DIH, concernant l'utilisation (et la conception) de ces armes. Deuxièmement, les armes autonomes soulèvent des questions éthiques, notamment l'absence d'élément humain dans les décisions de recours à la force, la dilution de la responsabilité morale et la perte de la dignité humaine, d'autant plus avec les armes autonomes pouvant porter atteinte à la vie humaine. Troisièmement, les préoccupations humanitaires, juridiques et éthiques du CICR sont largement partagées et soulèvent la question de savoir si le droit existant apporte une réponse suffisante à l'autonomie croissante des systèmes d'armes. Les États parties à la CCAC ont des points de vue différents sur la nécessité et la forme d'une possible réponse multilatérale, même s'il existe une convergence croissante sur le type et le degré requis de contrôle humain.

Le CICR a organisé plusieurs réunions d'experts sur les armes autonomes et a identifié les questions urgentes nécessitant une attention politique dans plusieurs rapports, déclarations et commentaires. Sur la base de ce travail, le CICR est convaincu qu'il est urgent de fixer des limites internationales strictes à l'autonomie des systèmes d'armes – que ce soit sous la forme de nouvelles règles juridiques, de normes politiques ou de meilleures pratiques.

Autonomous weapons systems (AWS) – also known as 'killer robots'[3] – command considerable media attention. They are also a topic of animated political and legal debate, including in the framework of the Convention on Certain Conventional Weapons (CCW)[4] Group of Governmental Experts (GGE) on emerging technologies in the area of lethal autonomous weapons systems (LAWS).[5]

States party to the CCW agree that International Humanitarian Law (IHL) applies to AWS and that '[h]uman-machine interaction' is key to ensuring that their use complies with IHL.[6]

[3] See, notably: 'Campaign to Stop Killer Robots', at: <https://www.stopkillerrobots.org/>.

[4] 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 of 10 October 1980, as amended on 21 December 2001 (CCW).

[5] United Nations Office at Geneva, 'Background on Lethal Autonomous Weapons Systems in the CCW', at: <https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument>.

[6] 'Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Final Report', CCW/MSP/2019/9, December 2019, Annex III, letters (a) and (c).

States Parties also recognise that '[f]urther work is required to determine the type and extent of human involvement or control' necessary for compliance with IHL and to respond to ethical concerns raised by AWS.[7] In line with its humanitarian mandate, the ICRC seeks to support and inform States' deliberations on AWS.

The ICRC understands an AWS to be a weapons system 'that can select (i.e. search for or detect, identify, track) and attack (i.e. intercept, use force against, neutralise, damage or destroy) targets without human intervention.'[8] After launch or activation by a human operator, an AWS – in contrast to other weapons – self-initiates force application on the basis of an encoded target profile and the processing of data captured by sensors in its environment. Consequently, an AWS user does not know what, *concretely*, an AWS will strike, nor when and where, *specifically*, strikes will take place. This makes it challenging for the user to predict the consequences of AWS use.

The AWS user's difficulty to predict who or what will be harmed in a specific attack – a difficulty that is more or less pronounced depending, notably, on the constraints under which the system operates – can pose a real risk of harm to persons and objects that, under IHL, may not be attacked and must be protected against the effects of hostilities, notably civilians who are not taking part in the fighting, as well as civilian objects, such as homes, schools or hospitals.[9]

It also raises pressing questions about the interpretation and application of IHL to the use of AWS. As a means of warfare, AWS must be capable of being used and must in fact be used in compliance with IHL rules governing the conduct of hostilities.[10] IHL thus limits the use of AWS

[7] 'Commonalities in National Commentaries on guiding Principles', paper submitted by Jānis Kārkliņš, Chairperson of the 2020 meetings of the GGE on LAWS, September 2020, at: <https://documents.unoda.org/wp-content/uploads/2020/09/Commonalities-paper-on-operationalization-of-11-Guiding-Principles.pdf>.

[8] ICRC, 'Statement to the Convention on Certain Conventional Weapons' (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 13 April 2015, at: <https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS>https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS. For more detail, see: ICRC, 'Characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention', CCW GGE on LAWS 9 - 13 April 2018, at: <https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/statements/10April_ICRC.pdf>.

[9] This includes both the risk that persons and objects protected under IHL could be harmed as a result of an AWS striking in their vicinity and the risk that they could fall within the target profile of an AWS.

[10] Notably, the rules on distinction (ICRC Customary IHL Study, (Rule 1, at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule1>), proportionality (Rule 14, at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule14>), and precautions (Rule 15, at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_cou_ch_rule15>).

and prohibits some of them.[11] However, many open questions remain about what constraints and requirements derive from existing IHL (and other applicable law) for the use (and design) of AWS. Some of these are discussed in more depth by Colonel Rudolf Stamminger in his contribution to this volume, which also sets out the French legal and operational perspective.

In addition, the ICRC has ethical concerns about AWS. These relate in particular to the erosion of human agency in decisions to use force, the diffusion of moral responsibility and the loss of human dignity. Ethical concerns are most acute with AWS that present risk for human life and may preclude the development and use of AWS designed or used to target humans directly ('antipersonnel' applications).[12] In his contribution, Dr Thompson Chengeta raises ethical concerns about AWS and the policy debate on their regulation and suggests adopting an inclusive ethical approach to the question.

The ICRC's humanitarian, legal, and ethical concerns about AWS are widely shared and raise the question whether existing law provides a sufficient response to increasing autonomy in weapons systems. Whereas States party to the CCW hold different views on the need for a multilateral response to AWS and its form, there is increasing convergence on elements of substance that are critical to ensuring the required type and degree of human control. The contribution by Ms Netta Goussac, co-author of a study published jointly by the ICRC and Stockholm International Peace Research Institute (SIPRI)[13] provides more detail on elements of human control and how these can inform the elaboration of limits on autonomy in weapons systems.

Over the past years, the ICRC has convened several expert meetings on AWS[14] and identified pressing questions in need of political attention in numerous reports, statements and com-

[11] Notably, the prohibition on weapons that are by nature indiscriminate. See ICRC Customary IHL Study, Rule 71, at: <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule71>.

[12] ICRC, 'Ethics and autonomous weapon systems: An ethical basis for human control?', 3 April 2018, at: <https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control>.

[13] ICRC and SIPRI, 'Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control', June 2020, at: <https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy_0.pdf>.

[14] ICRC, 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects', March 2014 – report of an expert meeting, at: <https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014>; ICRC, 'Autonomous weapon systems: Implications of increasing autonomy in the critical functions of weapons', March 2016 – report of an expert meeting, at: <https://www.icrc.org/en-publication/4283-autonomous-weapons-systems>; ICRC, 'Ethics and autonomous weapon systems: An ethical basis for human control?', August 2017 – report of an expert meeting, at: <https://www.icrc.org/en/document-ethics-and-autonomous-weapon-systems-ethical-basis-human-control>; ICRC, 'Autonomy, artificial intelligence and robotics: Technical aspects of human control', June 2018 – report of an expert meeting, at: <https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control>.

mentaries.[15] Based on this work, the ICRC is convinced that strict, internationally agreed limits are urgently needed on autonomy in weapons systems – whether in the form of new legal rules, policy standards or best practice.

[15] See e.g. ICRC, 'International humanitarian law and the challenges of contemporary armed conflicts: Recommitting to protection in armed conflict on the 70th anniversary of the Geneva Conventions', Report, 33rd International Conference of the Red Cross and Red Crescent, Geneva, October 2019, at: <https://www.icrc.org/en/document/icrc-report-ihl-and-challenges-contemporary-armed-conflicts>; 'ICRC commentary on the 'Guiding Principles' of the CCW GGE on "Lethal Autonomous Weapons Systems"', July 2020, at: <https://documents.unoda.org/wp-content/uploads/2020/07/20200716-ICRC.pdf>.

MILITARY USE OF AWS MUST COMPLY WITH INTERNATIONAL LAW AND, IN PARTICULAR WITH IHL
L'USAGE MILITAIRE DES SYSTÈMES D'ARMES AUTONOMES DOIVENT S'INSCRIRE DANS LE RESPECT DU DROIT INTERNATIONAL, EN PARTICULIER DU DIH

Colonel Rudolph Stamminger
French Ministry of Defence

Summary

Colonel Rudolph Stamminger is a Legal Adviser with the French Ministry for the Armed Forces. In his contribution he explains that the military use of AWS must comply with IHL, although other bodies of law also apply to questions related to AWS. According to the Permanent Representative of France to the Conference on Disarmament in Geneva, States agree that weapons systems must comply with IHL. However, there are disagreements on how IHL should apply. Discussions within the Group of Governmental Experts (GGE) related to the Convention on Certain Conventional Weapons (CCW) revealed how complex and diverse the positions on Lethal Autonomous Weapon Systems (LAWS) are, both from a technical legal perspective and from a political and strategic perspective.

The first two sections detail why LAWS are an important and complex topic, including for France. France introduced the issue of LAWS within the CCW in 2013. It actively participates in the GGE, especially on the 11 guiding principles on emerging technologies in the area of LAWS, which the GGE adopted in 2019. In the Declaration by the Alliance for Multilateralism on LAWS, France called on States to pay particular attention to the challenges associated with future AWS. The Ministry for the Armed Forces enacted the principle of developing an artificial intelligence defence system that complies with international law, in particular IHL and fundamental rights. The French State does not have LAWS at this stage and developing them is not even desirable. LAWS are systems evading human supervision or subordination to a chain of command by assigning objectives to them or by modifying, without human approval, their initial programming (rules of operation, use, commitment) or the framework of their mission. Autonomy must be distinguished from tele-operated systems, e.g. armed drones, or automated systems, aircraft autopilot. The military has been using the latter for a long time, contrary to AWS.

The third section shows that complying with IHL is a priority. IHL does not favour, constraint or prohibit specific types of technologies. It applies to all weapons, including new ones, as confirmed by the International Court of Justice in the Advisory Opinion on the Legality of the

Threat or Use of Nuclear Weapons of 8 July 1996. IHL is also fully endorsed by the Armed Forces. Recently, the French Ministry for Armed Forces has drawn up a directive specifying the modalities for implementing a legal review of new weapons, means and methods of warfare, provided for in Article 36 of Additional Protocol I to the Geneva Conventions. The rigour and precision of the process suggests that a purely AWS would not pass the first stage of the French legal review. In addition, France does not intend to develop LAWS because they would go against its own values and would have no operational advantages, partly because they would not be subject to the human chain of command. Lastly, France published a non-paper on 'the human-machine interaction' at the GGE in August 2018, which stressed: the need for an informed human command, the subordination to the human command, the communication links between the system and the command, and human control over the decision to use lethal force. The use of force remains the responsibility of the human command. However, it is worth noting that the nature and quality of human supervision depends on the type and spatio-temporal dimension of the mission, the nature of the targets, the nature of the data on which the selection of targets is based, the environment, etc. Moreover, international law does not provide for a specific type of human control.

Nonetheless, Colonel Stamminger explains in the fourth section that developing weapon systems with some degree of autonomy regarding certain functions is a crucial strategic objective for the military. The military needs to take into account new developments in order to maintain a strategic operational advantage, especially in comparison with other actors and possibly terrorist groups. France therefore intends to develop its capacities in the area of LAWS and artificial intelligence along the lines of three key principles: 1) complying with existing international law, including IHL; 2) having sufficient human control; 3) safeguarding the responsibility of human command. Because of these developments, the compliance with international law and especially IHL will remain under significant vigilance by the Ministry of Armed Forces and within the GGE designated in the framework of the CCW. Lastly, the Minister of the Armed Forces set up an Ethical Ministerial Committee in January 2020 to contribute to the debate on the use of new technologies.

Mesdames, Messieurs,

Je suis très heureux de participer pour la première fois à cette manifestation prestigieuse qui me conduit au cœur des enjeux qu'affrontent les forces armées françaises dans un contexte d'une intensité particulière, à un triple titre : des engagements extérieurs qui sont à un haut point depuis le second conflit mondial ; un système multilatéral mondial qui traverse une passe difficile ; une évolution technologique accélérée dont la portée stratégique, le questionnement éthique et l'encadrement juridique posent des questions redoutables.

Ce sujet dépasse le champ du seul Droit international humanitaire (DIH). D'autres droits comme le droit européen, le droit de la responsabilité, les régimes encadrant l'usage qui peut être fait de l'informatique ou encore la protection des droits de l'homme et des libertés fondamentales sont nécessairement concernés par la question des systèmes d'armes létaux autonomes (SALA). Cependant dans le cadre de cette intervention nous resterons concentrés sur le DIH.

Le développement croissant des technologies dans le domaine des systèmes d'armes létaux autonomes interroge les grands principes du DIH. Il soulève également des questions en matière d'engagement de la responsabilité de l'opérateur ou du commandement. Dans un tel contexte, l'application du DIH, sous-tendu par des considérations d'humanité, doit faire l'objet d'une attention particulière.

Cependant, et afin de rassurer sur ce sujet complexe et parfois mal compris, je reprendrai les mots du représentant permanent de la France auprès de la Conférence du désarmement à Genève : l'application du DIH « ne fait pas débat, aucun État ne remettant en cause les principes du DIH ni l'obligation pour tout système d'armes de le respecter ».

Face à ces enjeux, le développement et l'utilisation des systèmes d'armes létaux autonomes font l'objet de discussions dans le cadre de la CCAC (Convention sur certaines armes classiques). Les discussions du Groupe d'experts gouvernementaux ont ainsi pu révéler à quel point le débat sur l'encadrement des SALA était d'une grande complexité.

- Sur le plan technique d'une part, parce qu'il est encore difficile de parvenir à une définition universellement acceptée de ce que serait un SALA et parce qu'il serait contre-productif de vouloir limiter ou restreindre les recherches dans le domaine de l'autonomie.
- Sur le plan politique et stratégique d'autre part, parce que tout encadrement international des SALA, pour être crédible et effectif, devrait engager l'ensemble des États susceptibles de développer un jour de tels systèmes.

Les interventions de ce panel illustrent la variété et la complexité des questions juridiques et éthiques auxquelles nous sommes confrontées.

Il est important d'avoir un discours dénué de sensationnalisme, rappelant le cadre d'action des armées françaises (notamment sur le respect du DIH et la responsabilité du commandement). L'enjeu consiste donc à aborder la question des SALA de manière objective pour distinguer le mythe de la réalité.

Avant toute réflexion sur le respect du droit international (DI) et du DIH en particulier, je souhaiterais revenir sur ce dossier majeur, complexe et multifacettes (I) ainsi que sur la définition française des SALA, qui pour la France n'existent pas à ce jour et dont le développement n'est pas souhaitable (II).

Nous verrons ensuite que l'application du Droit international humanitaire est un enjeu de premier ordre (III) et son respect, à ce titre, fait l'objet d'une grande vigilance de la part du Ministère des armées (MINARM) (IV).

1. Les SALA : un sujet majeur, complexe et multi-facettes

- Il s'agit d'un dossier majeur :
  - Pour l'État, c'est la France qui a introduit la problématique des SALA au sein de la Convention sur certaines armes classiques (CCAC) en 2013. Depuis, elle participe activement aux travaux du Groupe d'experts gouvernementaux (GGE), dernièrement par sa contribution à l'opérationnalisation des 11 principes directeurs relatifs aux technologies émergentes dans le domaine des SALA, principes qui ont été adoptés par le GGE en 2019. De même, dans la Déclaration de l'Alliance pour le multilatéralisme sur les systèmes d'armes létales autonomes de septembre 2019, la France a appelé les États à accorder une attention particulière aux défis associés aux futurs systèmes d'armement comportant des fonctions autonomes.
  - Pour le MINARM, avec notamment l'édiction du principe de développement d'une intelligence artificielle (IA) de défense respectueuse du droit international, en particulier du DIH et des droits fondamentaux, principe défendu avec constance par la Ministre dans ses interventions. Si la question des technologies dans le domaine des SALA ne peut être déconnectée de celle de l'intelligence artificielle, c'est bien sur cette première question que porte cette intervention.
  - Et enfin pour la Directions des affaires juridiques (DAJ) qui prend une part active dans la définition de la position française sur ce domaine.
- Il s'agit d'une question d'une grande complexité technologique dont la réponse doit intégrer le droit et la morale. L'évolution de ces technologies est probablement sans précédents dans l'histoire récente et constitue une révolution technologique.

  Cependant, cette situation s'est déjà présentée dans l'histoire des technologies et des armes. Le droit et la morale ont souvent été confrontés aux mêmes défis (p. ex. dans l'histoire de l'aviation militaire).

2. Les SALA n'existent pas et ne sont ni des systèmes d'armes que la France pourrait développer, eu égard à ses engagements internationaux, ni des systèmes d'armes souhaitables pour les forces armées

- Il nous faut tout d'abord définir les limites du sujet abordé :

  Clarification terminologique : autonomie vs système télé-opéré (drones armés actuels) et automatisé (pilote automatique des aéronefs).

  Les forces armées s'appuient de longue date sur des technologies automatisées pour des séquences bien définies où les tâches n'ont pas une sensibilité nécessitant une validation humaine (tri automatique de données pour présentation à un opérateur, guidage automatique pour aller d'un point A à un point B...). En revanche, le recours à des technologies autonomes n'est pas encore effectif.
- La France considère que les SALA *stricto sensu,* c'est-à-dire des systèmes échappant à toute supervision humaine ou de subordination à une chaine de commandement en s'assignant eux-mêmes des objectifs ou en modifiant, sans validation humaine, leur programmation initiale (règles de fonctionnement, d'emploi, d'engagement) ou le cadre de leur mission, n'existent pas à ce jour.

3. L'application du Droit international humanitaire est un enjeu de premier ordre

- Le droit international humanitaire fournit un cadre juridique neutre du point de vue des technologies car les principes fondamentaux qui le sous-tendent en sont indépendants. Le DIH régit la conduite des hostilités et protège les personnes qui n'y participent pas. Le DIH ne favorise pas, ne restreint pas ou ne prohibe pas une technologie en particulier.
- Le DIH est d'autant plus pertinent qu'il est réputé applicable même aux armes nouvelles. Dans un avis consultatif du 8 juillet 1996, la Cour internationale de justice, statuant sur la licéité de la menace ou de l'emploi d'armes nucléaires, a estimé que la nouveauté d'une arme ne pouvait être invoquée en soutien d'une quelconque dérogation aux principes et règles établis du droit humanitaire applicable dans les conflits armés : « une telle conclusion méconnaîtrait la nature intrinsèquement humanitaire des principes juridiques en jeu, qui imprègnent tout le droit des conflits armés et s'appliquent à toutes les formes de guerre et à toutes les armes, celles du passé, comme celles du présent et de l'avenir ».
- Le DIH n'est pas obsolète. Au sein des forces armées françaises qui sont animées d'un « légalisme profond », nul ne conteste la nécessité de faire respecter le DIH.
- Ainsi, du fait de l'examen de licéité des nouveaux armements, dont la France a récemment renforcé les modalités, le développement de SALA semble improbable.

En effet, la Direction des affaires juridiques a élaboré, en concertation avec l'État-major des armées (EMA) et la Direction générale de l'armement, une instruction qui précise les modalités de mise en œuvre de l'examen de la licéité des nouvelles armes, des nouveaux moyens et méthodes de guerre, prévu par l'article 36 du Protocole additionnel I aux Conventions de Genève. Cette instruction est en vigueur depuis le 31 octobre 2019.

Le champ d'application de l'article 36 du premier Protocole additionnel est interprété largement dans les premières phases de développement ou d'acquisition d'une nouvelle arme ou d'un moyen de guerre. Concrètement, cela signifie qu'avant tout engagement contractuel ou financier, la Direction générale de l'armement (DGA) et l'État-major des armées intègrent une analyse de licéité qui vise à déterminer si le système ne présente pas de fonctions susceptibles d'être contraires principes du Droit international humanitaire ou aux interdictions conventionnelles spécifiques. Dans ce cadre, la DGA et l'EMA se laissent la possibilité de conduire un nouvel examen en cas de changement des circonstances de droit ou de fait.

La DAJ est associée à ce processus et intervient en cas de doutes sérieux. La rigueur et la précision dont fait preuve le Ministère des armées lors de cet examen de licéité m'amène à conclure qu'un système purement autonome ne passerait pas le premier stade de l'examen de licéité, tel qu'il est conçu par la France.

- Par ailleurs, la France n'a pas l'intention de développer des SALA car ces systèmes d'armes seraient contraires à ses valeurs et dépourvus d'intérêt opérationnel, en partie parce qu'ils échapperaient au commandement humain.

  Tous les systèmes d'armes doivent rester subordonnés à un commandement humain, y compris dans l'usage d'un système présentant de forts degrés d'autonomie. Cette subordination vaut en particulier pour la décision d'emploi de la force et la délimitation du cadre de la mission qui doivent appartenir à l'humain.

  Ainsi, les règles de fonctionnement, d'emploi et d'engagement ainsi que les missions de tout système d'arme, notamment les systèmes présentant des fonctions autonomes, doivent être impérativement validées par l'humain et intégrées dans un cadre précis et défini, la mission devant par essence être limitée dans le temps et dans l'espace à une zone de confrontation déterminée.

  À ce jour, pour une armée respectueuse du droit international, il n'y a pas d'intérêt opérationnel identifié à se doter de systèmes disposant de capacités d'auto-apprentissage, qui les rendraient capables de se reprogrammer en cours de mission au point de pouvoir sortir du cadre de celle-ci, sans en référer à la chaîne de commandement, ou bien qui leur permettraient de s'assigner eux-mêmes des objectifs sans validation humaine. En effet, le besoin majeur pour les forces armées est de contrôler l'effet des armes employées.

- La France a initié des réflexions sur la notion de supervision humaine, notamment en publiant un non-papier sur « l'interaction humain-machine » lors du GGE d'août 2018. Ce document définit des principes qui doivent régir cette interaction : (i) le commandement humain doit avoir une compréhension du fonctionnement du système qui lui permette d'en apprécier le comportement prévisible ; (ii) les systèmes doivent rester subordonnés au commandement humain qui définit le cadre de leur mission, leurs règles d'emploi et d'engagement ; (iii) des liens de communication, même intermittents, doivent exister entre le système et le commandement lorsqu'il est nécessaire de faire évoluer le cadre de la mission ; (iv) le commandement humain doit conserver le contrôle sur la décision de recourir à la force létale.

  Comme il a été souligné plus haut, le besoin majeur pour les forces armées est de contrôler l'effet des armes employées. À cet égard, le commandement doit conserver la capacité de prendre les décisions s'agissant du recours à la force létale, y compris dans le cadre de l'utilisation de systèmes présentant des degrés d'autonomie ou faisant appel à diverses composantes de l'intelligence artificielle.

  L'emploi de la force reste une responsabilité intrinsèque du commandement humain, en particulier en cas de violation du Droit international humanitaire.

  Il convient ici de rappeler que la nature et la qualité de la supervision humaine nécessaire ne peuvent pas être déterminées *in abstracto* pour tous les systèmes et pour tous les contextes d'emploi imaginables. Cela dépendra de la nature et de la dimension spatio-temporelle de la mission, de la nature des cibles et des données sur lesquelles la sélection des cibles est fondée, de l'environnement d'emploi, etc. En outre, le droit international n'exige pas une nature spécifique de contrôle humain.

4. Développer des systèmes intégrant de l'autonomie dans certaines fonctions est en revanche un enjeu essentiel pour les armées, qui nécessitera une grande vigilance de la DAJ afin d'assurer le maintien du respect du droit international

- Le développement de ces technologies présente des avantages cruciaux pour les armées. En effet, il nous faut tenir compte de l'évolution des conflits afin d'assurer à nos forces armées l'avantage opérationnel sur nos futurs adversaires, étatiques ou non-étatiques, et de ne pas se retrouver dans une position désavantageuse. Il s'agit de conserver notre supériorité militaire et notre liberté de manœuvre dans des contextes toujours plus complexes. En effet, d'autres acteurs, y compris des groupes terroristes, pourraient quant à eux choisir de développer des systèmes d'armes employant des fonctions autonomes, voire des SALA. Le risque de dissémination de ces nouvelles technologies est également à prendre en considération : de tels systèmes seront susceptibles d'être utilisés contre les forces armées françaises par nos ennemis sur les théâtres d'opération.

- La France entend développer ses capacités en matière de SALA et plus particulièrement dans le domaine de l'IA militaire dans le respect de trois principes : (i) le respect du droit international existant, notamment du DIH ; (ii) le maintien d'un contrôle humain suffisant ; (iii) la responsabilité du commandement humain, qui doit définir et valider les règles de fonctionnement, d'emploi et d'engagement des systèmes d'armes.
- Ainsi, la France appelle à clarifier et élaborer un cadre normatif et opérationnel des SALA par le groupe d'experts gouvernementaux (GGE) désigné dans le cadre de la Convention des Nations Unies sur certaines armes classiques (CCAC).
- Dans ce contexte, un travail de conviction et de pédagogie paraît indispensable, en particulier face à la volonté de nombre d'acteurs d'élargir la réflexion à des systèmes existants et de parvenir à une interdiction préventive des systèmes d'armes létaux autonomes, au besoin en exportant le débat en dehors de la CCAC.
- Le GGE sur les technologies émergentes dans le domaine des SALA a donné des résultats encourageants. En particulier, la session 2019 du GGE a permis de confirmer le consensus autour de onze principes directeurs qui constituent une base solide pour la poursuite des travaux du GGE.
- Ces principes affirment notamment : (i) que le Droit international humanitaire s'applique à ces systèmes ; (ii) que la décision d'en faire usage doit toujours relever d'une responsabilité humaine ; (iii) que les États doivent examiner au stade de leur conception la licéité des armes nouvelles qu'ils développent ou acquièrent.

  En cela, ces principes sont très proches sur le fond des positions de la France.

  Le GGE 2020 a conforté le consensus sur ces *principes directeurs*.
- Dans la perspective des travaux du GGE et des interrogations qui peuvent subsister sur les SALA (acceptabilité morale de l'usage de la force sans intervention de l'homme, distanciation de l'humain par rapport à la chaine d'engagement, réflexions sur la chaine de responsabilité et de commandement, par exemple), la ministre des Armées a souhaité que soit créé un comité d'éthique ministériel pour alimenter notamment la réflexion sur l'emploi des nouvelles technologies. Deux grands sujets à forts enjeux ont été proposés pour initier les travaux du comité d'éthique : le « soldat augmenté » et les systèmes d'armes létaux autonomes.

  Ce comité d'éthique, a été instauré en janvier 2020 et est composé de 18 membres issus des armées mais aussi de personnalités qualifiées extérieures, nommées par la ministre des Armées, pour un mandat de trois ans, renouvelable une fois, entretiendra un dialogue ouvert avec de nombreux interlocuteurs.

ETHICS AND AUTONOMOUS WEAPON SYSTEMS
ÉTHIQUE ET SYSTÈMES D'ARMES AUTONOMES

Thompson Chengeta
University of Southampton

Résumé

Thompson Chengeta est un juriste et chercheur dans le domaine de la violence liée aux drones et de l'éthique de l'intelligence artificielle à l'Université de Southampton. Sa contribution avance trois arguments principaux. Premièrement, il regrette que certains États puissants veuillent exclure ou atténuer le rôle de l'éthique dans la formulation de la politique sur les armes autonomes. Deuxièmement, il est également regrettable que la plupart des références se concentrent sur l'éthique et la philosophie occidentales, en excluant d'autres régions. Enfin, même s'il peut y avoir des interprétations divergentes, il est possible de dépasser ces difficultés, en les envisageant sous l'angle de l'éthique relationnelle.

La première partie explore les dilemmes éthiques que soulèvent les armes autonomes. D'abord, les armes autonomes soulèvent la question de savoir s'il est moralement et éthiquement acceptable que des machines décident de la vie et de la mort d'êtres humains : des armes devraient-elles être autonomes dans l'exercice de fonctions critiques ? Pour le Secrétaire général des Nations Unies António Guterres, les armes autonomes dans leurs fonctions critiques sont moralement et politiquement inacceptables et devraient être interdites par le droit international. Plusieurs universitaires et États partagent cette position, en particulier ceux qui adoptent une éthique déontologique, s'appuyant sur les notions de responsabilité morale et de dignité humaine. D'autres universitaires et États s'opposent à cette position, sur la base d'une éthique conséquentialiste ou téléologique. Une autre question est de savoir si les considérations éthiques sont pertinentes dans la formulation d'une réponse politique. Au-delà des questions juridiques, les dilemmes soulevés par les armes autonomes impliquent des questions d'adhésion à nos valeurs. Dans son récent ouvrage[1], Thompson Chengeta se demande comment les processus liés à la Convention sur certaines armes classiques (CCAC) pourraient aboutir à un cadre réglementaire commun, alors même que les inquiétudes et objections à l'utilisation des armes autonomes sont ancrées dans des valeurs subjectives. Au-delà des approches éthiques déontologiques et téléologiques, Thompson Chengeta raisonne avec une approche relationnelle de l'éthique. L'éthique relation-

[1] See Chengeta, Thompson. 'Autonomous Armed Drones and the Challenges to Multilateral Consensus on Value-Based Regulation', in: C. Enemark (ed.), *Ethics of Drone Strikes: Restraining Remote-Control Killing*, Edinburgh University Press, Edinburgh, 2021, pp. 170–189. JSTOR, at: <www.jstor.org/stable/10.3366/j.ctv1c29rjh.14>. Accessed 18 Jan. 2021.

nelle se concentre sur comment vivre ensemble, indépendamment des questions de race, religion, nationalité, genre, sexe, sexualité, handicap, statut social ou tout autre forme de discrimination.

La deuxième partie adopte cette approche relationnelle pour réinterroger les fondements mêmes des références éthiques communes, en insistant sur les notions de diversité et d'inclusion. Les Nations Unies et d'autres organisations internationales telles que le CICR ont noté qu'il était nécessaire de prendre en compte l'éthique et les valeurs pour formuler une réponse politique aux armes autonomes. Toutefois, Thompson Chengeta souligne la convergence autour de perspectives éthiques occidentales et l'exclusion des approches éthiques africaines, ou venant du monde arabe, ou de celles d'autres groupes marginalisés qui aussi risquent d'être affectés par les armes autonomes. Certaines organisations internationales, telles que l'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO), ont soulevé l'importance de la diversité pour formuler les cadres politiques et de gouvernance liés à l'intelligence artificielle. Pourtant, les approches actuelles, en particulier au sein d'instances onusiennes, font peu de cas de l'impact potentiel des armes autonomes sur des groupes qui ont été historiquement opprimés ou dominés. Au sein du débat onusien actuel, le jus in bello *est considéré comme un cadre stable à l'aune duquel les armes autonomes doivent être mesurées. Cependant, ces déclarations et suppositions ne questionnent ni l'adéquation des normes du jus in bello, ni sa capacité à couvrir les diverses valeurs des différents peuples. Cela semble problématique si l'on considère que les États les plus faibles n'étaient pas à la table des négociations lorsque les pouvoirs impériaux décidèrent des fondations du* jus in bello *en Suisse en 1864 et que ces mêmes pouvoirs impériaux planifièrent la partition de l'Afrique et l'assujettissement de ces peuples à Berlin en 1884. Les pouvoirs impériaux n'avaient certainement pas l'intention de codifier les valeurs des peuples qu'ils entendaient dominer et opprimer[2]. Il semble donc présomptueux de supposer que les valeurs codifiées en 1864 sont suffisamment diversifiées pour être encore utilisées aujourd'hui pour mesurer l'acceptabilité de nouvelles armes.*

1. The Relevance of Ethics to the AWS Debate, Different Approaches to Ethics and Challenges to Multilateral Consensus on Relevant Values/ Ethics

One of the key ethical questions that has been asked is whether it is morally and ethically acceptable for machines to make decisions as to who lives or dies? In other words, should weapon systems have autonomy in the critical functions of selecting, targeting and releasing force against human targets? The United Nations Secretary General, Antonio Guterres, says

[2] Voir également : Amanda Alexander, 'A Short History of International Humanitarian Law', in: *European Journal of International Law*, Volume 26, Issue 1, February 2015, pp. 109–138, à: <https://doi.org/10.1093/ejil/chv002>.

that AWS that have autonomy in critical functions are morally repugnant, politically unacceptable and should be banned by international law. This view is supported by various scholars and States, particularly those who subscribe to deontological ethics, condemning certain AWS for lacking moral responsibility and, therefore, making their use inconsistent with human dignity. Yet, these views have been opposed by some scholars and States, particularly those who approach the issue from a consequentialist standpoint or teleological ethics.

Another burning question has been whether ethics are relevant to the AWS debate and the formulation of a policy response to this issue. I cannot say it better than what the ICRC has observed: the challenges raised by AWS go beyond considerations of compatibility with our laws to encompass issues of acceptability to our values. The ICRC has also noted that ethics may form the basis for human control over use of force, a concept that I believe will be discussed by my co-panellists.

Again, the relevance of ethics, particularly values such as human dignity, has been questioned by those who see human dignity as a conversation stopper, a term that is incapable of precise meaning for purposes of regulation or policy formulation. In my recent book chapter titled 'AWS and the challenges to multilateral consensus on a value-based regulation'[3], I pose and discuss the following question: given that some of the critical concerns on, and objections to, the use of AWS are anchored on our values – values that are arguably subjective – what are the chances of a regulatory framework emerging from the UN CCW, a multilateral process that operates on consensus? I considered this question particularly important given that States in the UN GGE on AWS continuously highlight that the UN CCW is the appropriate forum within which AWS must be discussed.

Now, while scholars and States have more fully referred to deontological and teleological ethics relevant to AWS, in my current research I focus on relational ethics, a contemporary approach to ethics that situates ethical action explicitly in how we relate to each other as peoples of this Earth. The premise is, if the idea of ethics is about how we should live, it is essentially about how we should live together regardless of race, religion, nationality, gender, sex, sexuality, disability, social status or any other basis of discrimination. In consideration of relational ethics, I more fully emphasise the issue of diversity and inclusion in consideration of relevant ethics and values to the AWS debate and formulation of policy.

[3] See Chengeta, Thompson. 'Autonomous Armed Drones and the Challenges to Multilateral Consensus on Value-Based Regulation', in: C. Enemark (ed.), *Ethics of Drone Strikes: Restraining Remote-Control Killing*, Edinburgh University Press, Edinburgh, 2021, pp. 170–189. JSTOR, at: <www.jstor.org/stable/10.3366/j.ctv1c29rjh.14>. Accessed 18 Jan. 2021.

2. Diversity and Inclusion in Consideration of Relevant Ethics and Values to the AWS Debate

The UN and many other organisations such as the ICRC have noted that in formulating a policy response to AWS, there is a need to take into account ethics and 'our values'. In my current research titled 'Re-examining the *jus ad bellum – jus in bello* dichotomy from an African ethics perspective: Towards a comprehensive response to autonomous weapon systems', I discuss a number of limitations in the current ethical approaches to AWS.

I have noted and observed that there has been a focus on Western ethical perspectives to the exclusion of African ethics, ethics from the Muslim world, and values of other marginalised groups that are likely to be adversely affected by AWS. There is sufficient evidence to support this assertion, but that is not the thrust of my discussion today. Suffice to say that exclusion of ethics and values from marginalised groups and less powerful nations exacerbates epistemic injustice and global social injustice that currently characterise armament and disarmament processes across the globe.

While international organisations such as United Nations Educational, Scientific and Cultural Organization (UNESCO) have noted the importance of diversity in the framing of policy and governance frameworks for AI technologies such as AWS, in the current approaches to AWS, particularly, in the UN *fora,* there is no deliberate expression or discussion on the potential impact of AWS on groups that have been historically oppressed or dominated. For those who mention the values of marginalised groups, they present them as if they are ideas that are meant only to plug into a set of stabilised global arrangements and values that should not be disturbed. In my opinion, if global arrangements such as those provided for in *jus in bello*, are insufficient or enable or sustain the continued oppression of certain nations and other marginalised groups, they should be open to alterations and additions.

In the UN debate on AWS, there are general declarations or assumptions that existing *jus in bello* should be the yardstick upon which the acceptability or otherwise of AWS must be premised or predicated. In other words, *jus in bello* is regarded as the stable framework against which AWS must be measured. Those declarations and assumptions neither address the question of adequacy of existing *jus in bello* standards nor their breadth in covering diverse values of different peoples across the globe. That assumption may be problematic if one considers that most of the weaker States were not at the negotiation table in Switzerland in 1864 when imperial powers laid the foundations of *jus in bello*. In fact, in 1884, the same imperial powers met in Berlin to plan the partition of Africa and subjugation of its people. Surely, when laying the foundations of *jus in bello*, the imperial powers did not necessarily intend to codify the values of peoples they intended to dominate and oppress. One scholar once noted that in addition to the history of *jus in bello,* which is often glossed as a law that sought to humanise war, 'there is another story about *jus in bello,* which describes it not only as a history of compassion and civilisation but, rather, as a history of oppression and imperialism [...] a history in which military or Western needs have consistently trumped humane values, exposing civilians to the violence of war and legitimising their suffering.'[4] It is, therefore, a huge assumption, if not arrogance, to assume that the values codified in 1864 – which have essentially remained the same regardless of further treaties adopted – are sufficiently diverse to still be used today to measure the acceptability or otherwise of new weapons.

Whenever I highlight the inequality and exclusion of values and ethics of marginalised groups and weaker nations, and that the disparity of power in State relations as evidenced in disarmament processes is a source of abuse and oppression of certain peoples, I am often told, that, according to Thompson, inequality is an inevitable part of society and therefore, any attempt to tackle inequality is doomed to failure. Indeed, I agree that the world we live in is extremely complex and presents major challenges for anyone who seeks to change the *status quo*. Yet, while we should, of course, avoid naive idealism, we should equally avoid extreme cynicism and defeatism, particularly, in seeking freedom from oppression at the instance of misuse of AI technologies such as AWS.

To conclude, let me mention three main take-aways from my short discussion today. First, it is bad that some powerful States want to exclude or undermine the role that ethics should and ought to play in formulating policy on AWS. Second, it is also bad that most of the reference to ethics in the current discussions on AWS focus on Western ethics and philosophy while excluding those from other regions. Third and finally, even though there may be diverging views and interpretations of relevant ethics, it is possible to move past such difficulties, and one way is to look at it from a relational ethics standpoint.

[4] See Amanda Alexander, 'A Short History of International Humanitarian Law', in: *European Journal of International Law*, Volume 26, Issue 1, February 2015, pp. 109–138, at: <https://doi.org/10.1093/ejil/chv002>.

WHAT ARE THE ELEMENTS OF HUMAN CONTROL AND HOW CAN THEY INFORM THE SETTING OF LIMITS ON AUTONOMOUS WEAPON SYSTEMS?
QUELLES SONT LES CARACTÉRISTIQUES DU CONTRÔLE HUMAIN ET COMMENT CES CARACTÉRISTIQUES CONTRIBUENT-ELLES À POSER DES LIMITES AUX ARMES AUTONOMES ?

Netta Goussac
SIPRI

Résumé

Netta Goussac est chercheuse senior associée dans les domaines de l'armement et du désarmement du SIPRI et conseillère spéciale chez Lexbridge. Cette présentation portait sur ce qui caractérise le contrôle humain et les limites que ce concept impose aux armes autonomes. La présentation s'appuie sur les conclusions et recommandations d'un rapport conjoint SIPRI-CICR sur les limites de l'autonomie publié en juin 2020[1]*. Elle propose notamment des éléments de réflexion et de précisions concernant le concept de contrôle humain.*

Un consensus semble émerger des discussions entre États et autres entités concernées au sein du Groupe d'experts gouvernementaux (GGE) sur les systèmes d'armes létales autonomes (SALA) autour de l'idée que l'autonomie d'un système d'armes ne peut être illimitée et que les humains doivent conserver la responsabilité de l'utilisation de ces systèmes. En 2019, le GGE a conclu que, bien qu'il y ait un accord sur l'importance de l'élément humain, des précisions supplémentaires sont nécessaires sur le type et le degré d'interaction homme-machine, notamment les éléments de contrôle et de jugement. Cela témoigne de l'avancée du débat mais indique à la fois un besoin de réflexion supplémentaire sur ces sujets.

Le concept de contrôle humain soulève un problème commun à tous les systèmes d'armes, quelle que soit l'étendue de la définition de cette catégorie : la notion de contrôle humain est le pendant du passage d'un contrôle direct de l'arme à l'autonomie.

Il y a trois raisons d'utiliser la notion de contrôle humain comme cadre conceptuel pour les armes autonomes. Une première raison est d'ordre juridique. Bien que le droit international s'applique au développement et à l'utilisation des armes autonomes, les modalités d'application du droit posent plusieurs difficultés. Ces difficultés concernent d'abord la difficulté de porter des jugements basés sur des valeurs prescrites par le DIH, tout en se s'appuyant sur de la programmation

[1] SIPRI-ICRC, 'Limits on autonomy in weapon systems, Identifying Practical Elements of Human Control', June 2020. Available at: <https://www.icrc.org/en/document/limits-autonomous-weapons>.

ou des données sensibles, et de prendre des décisions en se basant sur le contexte ou encore d'anticiper les effets des armes automatiques. Ensuite, une autre raison d'utiliser le concept de contrôle humain comme cadre conceptuel est d'ordre éthique. L'autonomie de ces systèmes rend difficile de maintenir la responsabilité morale humaine dans la prise de décision concernant l'usage de la force. Enfin, une dernière raison est d'ordre opérationnel. Bien que l'autonomie offre certains avantages militaires, elle implique aussi un degré d'imprévisibilité concernant les conséquences des opérations militaires, y compris en termes de sécurité et d'efficacité.

Les armes autonomes posent deux menaces principales. D'abord, que ce soit d'un point de vue juridique, éthique ou opérationnel, l'utilisateur a besoin d'un niveau raisonnable de certitude par rapport aux effets des armes autonomes dans un environnement donné. Cela implique une compréhension à la fois de l'environnement et du système utilisé. De plus, que ce soit d'un point de vue juridique ou éthique, l'utilisateur doit faire preuve de discernement et d'intention dans l'usage de la force lors d'une attaque.

Le rapport identifie trois caractéristiques de la notion de contrôle humain :

1. *les paramètres d'utilisation du système, par exemple : limitation du type de cibles et de tâches, de la durée de l'opération ou de la liberté de mouvement ;*
2. *l'environnement d'utilisation, par exemple : environnements où les civils et objets civils ne sont pas présents, zones d'exclusion, barrières physiques, avertissements ;*
3. *l'interaction entre humain et machine, par exemple : outils de supervision tels que reprise de contrôle direct, possibilité d'opposer un veto, d'annuler des fonctions ou d'interrompre une tâche.*

Ces trois éléments de contrôle humain – par rapport aux paramètres, à l'environnement et à l'interaction humain-machine – constituent un cadre pratique qui permet d'identifier les limites nécessaires à l'utilisation d'armes autonomes. Les trois éléments de contrôle seront à mettre en œuvre quel que soit le scénario de leur utilisation, avec des combinaisons spécifiques à chaque contexte, en fonction de l'environnement et des caractéristiques de chaque système d'armes. Si un type de contrôle n'est pas adapté à un contexte, alors il devra être compensé par un degré plus élevé d'un autre élément de contrôle. Par exemple, s'il s'avère difficile d'exercer un contrôle sur l'environnement, alors le contrôle exercé sur les autres paramètres devient d'autant plus important.

The editorial team has transcribed the contribution below based on the recording of the presentation of Ms Netta Goussac.

Thank you, Maya and thank you also to the ICRC and the College of Europe for inviting me to take part in the Colloquium today. This is my first participation at this event, I am very sad not to have the opportunity to be in the beautiful city of Bruges, but I am grateful for the opportunity to be able to connect with you all from Canberra, Australia.

Today, the topic of my presentation is: 'what are the elements of human control and how can these elements inform the setting of limits on autonomous weapon systems (AWS)?' Indeed, what I will be doing is presenting the findings and the recommendations from the Stockholm International Peace Research Institute SIPRI-ICRC report on the limits of autonomy[2] that was published in June this year. In the time available, I will be able to only share a very small subset of this rich report. Today, I would like to cover three key reasons why focusing on human control as a conceptual framework is warranted with respect to autonomous weapons, three key elements of human control identified by SIPRI and the ICRC which are explained in the report and then three ideas for how those elements of control could inform about limits on AWS for the future.

Before I start on these three issues, there appears to be an emerging consensus evidenced in the discussions among States and other participants in the Group of Governmental Experts (GGE) on lethal autonomous weapon systems (LAWS) that autonomy in a weapon system cannot be unlimited, that humans must retain or exercise responsibility for the use of these weapon systems. In 2019, the GGE concluded that although there is agreement on the importance of the human element, further clarification is needed on the type and degree of human-machine interaction required, including elements of control and judgement. This showed a marker for how far States and civil society and other actors have progressed in the conversation, but also indicated a need for further reflection on what human-machine interaction means and on what elements of control could exist or could be applied.

One of the reasons why this concept of human control is attractive is because it is not technocentric, but it addresses an issue that is common to all types of weapon systems, no matter how broadly that category is defined. What these systems have in common or what the various definitions of AWS have in common is that there is no human intervention in the functions of selecting and applying force to targets as part of the weapon system. This kind of shift from direct control to automation or autonomy is what is common to AWS and the focus on human

[2] SIPRI-ICRC, 'Limits on autonomy in weapon systems, Identifying Practical Elements of Human Control', June 2020. Available at: <https://www.icrc.org/en/document/limits-autonomous-weapons>.

control is the other side of that coin. This led the ICRC and SIPRI to collaborate on a project that identifies more in detail the elements of human control and to publish the report.

When I talk about AWS, I have in mind the broad working definition that is mentioned in the report and which has been adopted by the ICRC. I do not use this definition in order to set policy boundaries, but rather to allow a common basis for discussion and to be able to draw from the broadest range of experiences how weapon systems are in fact being used today, with varying degrees of autonomy.

1. Three Reasons to Focus on Human Control as a Conceptual Framework

Three reasons to focus on human control as the conceptual framework have been identified. The previous presentations respectively addressed the legal perspective and the ethical perspective on this issue.

To briefly recapitulate, from a legal perspective, of course international law applies to the development and use of autonomous weapons, but there are some challenges that arise from the unique characteristics of this type of weapon systems. These unique characteristics may in some circumstances make the application of IHL more difficult and may even raise questions of how IHL should be interpreted. The report outlines three specific challenges based on discussions with experts:

1. firstly, the ‘number challenge’, which is the difficulty of making value-based judgement that are required by IHL when relying on sensitive data and programming;
2. secondly, a ‘context challenge’, which is a challenge for a user of an AWS to make context-based decisions in light of the circumstances ruling at the time of the attack while, in fact, when using AWS, these kinds of decisions must be made rather earlier and more distantly from a resulting application of force;
3. finally, the third challenge is that the unpredictability introduced by AWS can hinder the user from anticipating and limiting the effects of a weapon as required by IHL.

Regarding the ethical perspective, I could complement the previous presentation by pointing to one of the other ethical concerns: the question of human agency and the importance of upholding moral responsibility in decision making on the use of force being seriously challenged by the autonomy inherent to those weapon systems, particularly because of a cognitive distance in time, in space, in understanding between the human decision to use an AWS and the effective use of force by an AWS.

Finally, an important third reason to focus on human control as a conceptual framework is the operational reason, the operational perspective. Autonomy in weapon systems offers certain military advantages in terms of speed of action, the ability to operate with communication and remote control, but on the other hand it can also create unpredictability in the consequences and challenges for ensuring safety and efficiency in military operations. This provides yet another reason to focus on human control.

As a bottom line for these three perspectives, two common threats can be identified. First, whether from a legal, ethical or operational perspective, the user needs to have a reasonable level of certainty about the effect of an AWS in a specific environment of use. That includes both a sufficient understanding of the environment and of the system that is being used in order to make the kind of judgements that are required from a legal, ethical and operational perspective. The second common threat that can be discerned is that users must exercise judgement and intent or agency in the use of force in specific attacks, both to ensure compliance with IHL and to uphold the moral responsibility and ensure accountability for the consequences of the use of these weapons.

2. Three Elements of Human Control

Based on these common threats and discussions with experts, the ICRC and SIPRI identified three elements of human control. The first is control on the weapon system's parameters of use, that is measures that restrict the type of targets and the tasks that the AWS is used for, measures that place temporal and spatial limits on its operations or constrain its effects or that would allow for the deactivation of the fail-safe mechanism. That is the first group of measures that could be elements of control.

The second group is control on the environment of use. Measures that control or structure the environment in which the AWS is used, and which overlap with the weapon system's parameters. For example, using an AWS only in environments where civilians or civilian objects are not present or excluding their presence for the duration of operation through strict temporal or spatial constraints on the use of autonomous weapons, or exclusion zones, physical barriers, warnings. These are controls on the environment.

A third group of elements of human control would be control on the interaction between human and machine. By that, I am including measures that allow the user to supervise the weapon system, to intervene in its operations when necessary, including, if needed, through direct active control, vetoing or overriding its functions, aborting a task, etc. To many of us this sounds like familiar measures, already applied when using weapon systems with any degree of autonomy or automation. It is indeed from this broad group of weapon systems that these ideas are from.

What do we do with these ideas and how can they help the discussions of governmental experts? I want to step outside of the policy discussion on what form any kind of international action might take, whether it is rule, policies, guidelines, best practices, and really to focus on the substance of how this kind of elements of human control can help us identify substantive areas, substantive limits that could be applied to weapon systems no matter the form.

3. Three Ideas to Inform Limits on AWS

Here I just want to mention a couple of things. One is that these kinds of control measures, control of the parameters of use, control of the environment, control of the human-machine interaction, are a practical framework that can help us identify the necessary limits. The reason for that is that all three types of control measures will have to be applied to some extent in every scenario when an AWS is used, in some kind of combination. The particular combination of measures, the particular types of measures that would be needed, may vary according to the specific context, including the environment of use and the characteristics of the specific weapon systems. But somehow, they all must combine, in order to ensure that the user has reasonable certainty about the effects of the weapon system and that the user exercises judgement and intent. When one type of control is inadequate for whatever reason, if it is insufficient or if it is challenging to implement, then other types of control measures rise in prominence. When, for example, the environment is difficult to control or to structure, then control over the weapon system's parameters become ever more important.

From here, we can think of three types of limits that draw from these elements of human control. I put this out to the audience for consideration and discussions.

The first kind of limits that one can imagine is limits or restrictions on the types of targets against which an AWS may be used: constraints, for example, on its freedom of movement or its duration of operation, requirement for deactivation or fail-safe mechanism. These types of limits may be linked to the elements of control of parameters.

The second category of limits that can be discussed and should be discussed is the development of limits in avoiding or at least in minimising the risks of harm to civilians and to civilian objects which are not to be attacked and are to be protected from the effects of attacks. By this I mean, for example, limits that may see the use of AWS precluded in certain areas or otherwise excluding civilians from a weaponised area of operations.

The final set of limits that should be discussed would be limits addressing the need for this human-machine interaction, human supervision of the AWS, the ability to intervene in its functioning to deactivate it. One way to explain this kind of limits is what is called ‘human on the loop’ control in the GGE or human supervisory control.

I put these ideas about possible limits in the connection to human control to the group for discussion and also more broadly within the ICRC and SIPRI to try to contribute to the ongoing discussions amongst States and other actors in the Group of Governmental Experts and give some substance and some ideas to what to do with this ever-reaching discussion that has been building since 2017.

DISCUSSION

The Q&A session allowed the following topics to be raised and discussed in depth:

1. Human Control and Limits to AWS

The moderator recalled that international policy discussions recognise the need for further clarification of human control, human judgement or perhaps more generally the involvement of a human person in the use of autonomous weapons. Building on previous propositions aiming at limiting the weapons' autonomy, the moderator asked whether, in the panellists' views, these types of limits facilitate compliance with the law by the commander in charge of a particular attack, whether any of these raise operational or legal challenges, and whether these kinds of limits take into account the diverse ethical concerns in autonomous weapons.

A first panellist emphasised that for some States such as France, full autonomy does not exist. In terms of chain of command, a system with no limits in time or space in a conflict is not conceivable. Rules are needed even when considering autonomous systems. This echoes the word 'heteronomous', i.e. a system which is autonomous but with some rules of external control. Any system has rules of engagement, joint operation areas, orders, directives, etc. The chain of command would not accept a system, at least in France, that would choose a target without knowing the law of armed conflict and the rules of engagement, who the enemy is, who the enemy is not, what the target is. A weapon which is fully autonomous would certainly raise operational and legal challenges. However, if the framework exists and the system is circumscribed within a specific area and is told to choose specific targets, there are still some challenges, but it does not seem to raise unsolvable issues.

Another panellist answered that the debate on parameters brought to light an immediate concern about autonomy in the critical functions of weapon systems, i.e. the targeting of human beings without human control or human input. From a normative aspect, there should not be autonomy in the critical functions, for both legal and ethical reasons. Ethics is one of the major reasons for not allowing autonomy in the critical functions. The panellist recalled that for instance, the issue of limiting human casualties is already a binding principle of IHL regardless of the type of weapons. He expressed his fear that raising some of these issues closes the debate about the real concern. In 2018, the issue was not necessarily to deal with targeting or the existing parameters of these systems, the issue of the type of targets was specifically about the right to life, i.e. the targeting of humans. The focus should be on the idea that machines should not make decisions in critical functions without human input, rather than trying to identify the parameters. The parameters do not matter if the weapon is autonomous in its critical functions. Without human control or human input to choose the target, then that issue still exists even if the weapon is being activated for only a few seconds. Temporary limits and parameters may not necessarily be helpful as far as the discussion of what the standard of human control ought to be.

Another panellist concluded with some additional comments. Firstly, although the discussion focuses on legal, ethical and operational perspectives, this is not an exclusive list of approaches that can be taken on the question of AWS. There are broader political, and strategic, concerns at stake, as well as concerns of peace, security and human rights, that simply fell outside the scope of this conversation, but which are nonetheless critical. Secondly, there is a whole range of views about how IHL should be interpreted and applied. There is also a divergence of practice among States about war fighting generally. All these questions and ideas have a bearing on how AWS are viewed. Remarks about whether it is even lawful to rely on these technologies in carrying out some of the critical functions of selecting and applying force to the targets is debated among experts. That kind of nuance is important to keep in mind.

2. Design of the AWS

Several participants raised questions relating to the design and programming of AWS. More specifically, a participant suggested that human control may be needed over the ability of the machine to modify its programming and asked for details on the risk for an autonomous weapon to be hacked. Another participant wondered if the AWS technology could ensure that targeting without human intervention is compatible with IHL.

Risks Associated with Machine's Self-programming and Hacking

A panellist said that both questions – on the ability to modify programming or on programming a weapon system to comply with IHL – are questions of design. On a general level, this is a question of risk. Although for the military there is no problem that cannot be resolved, the question is how much risk militaries or parties to a conflict are willing to take in order to achieve their aims. In IHL, there is always a level of discretion for the parties involved on the amount of acceptable risk: this is a question of how much risk parties are willing to take. The question of the ability of a machine to modify its programming seems to introduce a level of risk that is simply unmanageable from a legal perspective. This may require additional technological expertise, but the idea would be that a system, after it is activated, is reacting to its environment and continuing to develop a functioning use of machine learning. From a legal perspective, it seems very challenging to understand how to limit the effect of that kind of system in compliance with IHL. This is a question that should come up not necessarily at the stage of deployment but much earlier, at the stage of design and development, i.e. the legal review stage. This is the stage where there is an assessment made about the lawfulness of weapon systems as they are being considered for adoption or being developed. It is already at that stage that the question of human control arises. Human control is not only a question of the operational use of the weapon system, it is a concept that reaches all the way from the design and development of autonomous weapons to the use and the aftermath when it comes to questions of responsibility and accountability. At the very early stage of legal reviews, it is already difficult to determine how the user of an autonomous weapon system that is capable of modifying its functioning after it was activated will be able to limit its effects as required by IHL and to achieve the level of certainty that is required about its effects in a specific attack.

Another panellist added that the risk for a machine to be hacked is a crucial technical challenge which arises at the development stage. France in particular wants to make sure that the system is natively built not to be hacked and with a permanent possibility to liaise with the weapon system. These requirements can also be an issue because the possibility to liaise may imply that the system is more vulnerable to hacking. It is a very important question, which however requires more technical expertise.

Programming AWS Compliance with IHL

A panellist recognised that the idea of programming an AWS to comply with IHL seems very enticing. It has been said by several experts and States that this could remove certain human fallibilities affecting IHL compliance and therefore improve the ability to protect civilians from the effects of armed conflict. There is a humanitarian perspective behind that. At the same time, it raises very specific questions. Firstly, 'what IHL is' is not necessarily agreed at all times. IHL is not a static concept. Secondly, it is not clear how to translate that into the technical indicators and values that are used when programming software. For instance, how do you translate the rule of proportionality, which requires a value judgement about the loss of civilian lives, civilians' injuries and expected or anticipated military advantage? However, what could be programmed in a weapon system is: constraints on the type of targets, constraints on when force may be applied or under what circumstances, constraints based on limits of operations. These kinds of technical constraints which assist the users to ensure that they themselves comply with their obligations under IHL. But that is a slightly different way of conceptualising what is meant by programming an AWS.

A panellist was of the opinion that if a weapon system goes through Article 36 on the Legal Review of New Weapons, Means and Methods of Warfare, the assessment becomes an ethical problem rather than a question of law. That is why the Ethics Committee was set up in France.

The aim is to be sure that the Ministry of Army, the President or other governmental authorities will be able to answer additional questions. The panellist acknowledged that this would be challenging. However, if the weapon system goes through the Article 36 review and is therefore IHL-compliant, there is no legal issue. From an operational military perspective, the aim is always to operate from as far away as possible in order not to be shot. In this sense, AWS is just another technological evolution, although there are additional ethical concerns. This is why France has an Ethics Committee. The panellist agreed with the previous panellist that for instance the principle of proportionality is a real issue in targeting. Nonetheless, it does not mean that the development of AWS has to be stopped, because there may be a way to solve existing issues. France does not want to stop the development of AWS, although for now this remains very challenging from a legal perspective.

3. Ethics, Inclusiveness and Diversity

A panellist addressed a question by a participant who asked for an example of IHL norm that does not correspond to African values. The panellist first reflected upon the question itself. He suggested that this question should arise after thorough, extensive and critical research on African values and values of certain marginalised groups. For the question to arise, there should have been a research finding that nothing warrants the idea that perhaps current *jus in bello* may not be sufficiently diverse to be used as a yardstick to measure the acceptability of new weapons. Otherwise, the question itself would confirm the panellist's suggestion that there is need for critical reflection on the values of other peoples from other regions, beyond the assumption that existing norms encompass everything else in the world.

On the content of the question itself, the panellist recalled that the current emphasis within the United Nations Office for Disarmament Affaires (UNODA), for example, is that the *jus in bello*, IHL, should be the yardstick upon which it is decided whether certain AWS, particularly those which are autonomous in their critical functions by design, are acceptable or not acceptable. That approach in itself is already limited because it is assuming that the problems or the concerns that all the people around the globe may have are well contained or represented in IHL. The oppression and domination because of the subjugation of African peoples to other peoples has been, for example. possible on account of the possession of superior weapons by Western powers. That is a historical fact. Some people have already referred to the current development of AI technologies of AWS as neo-colonialism or re-colonialisation, through what they call 'algorithmic colonialisation of certain people in domination across the globe'. LAWS would make it very easy for certain nations to continue infringing on the territories of others. One could argue that this is a matter of *jus ad bellum*, prohibition of the use of force, and those are not the values which we consider on deciding whether or not a given weapon is acceptable. But this argument amounts to using a specific framework that was set when some peoples were not present, then to assess what is acceptable, depending on these criteria. But one should look at who set the criteria. Therefore, consideration of re-evaluation of how certain particular weapons or certain particular technologies should be accepted or not, should at least be part of this approach.

4. Conclusions on the International Framework

The moderator informed the participants that there was another set of questions that regrettably, the panel would not have time to consider but which were worth bringing to the attention of the audience. Several participants had pointed to the peace and security dimensions of this issue, questions about whether the exclusive focus or dominance of IHL in this debate is likely to miss the point, at least in part. As a last concluding thought, she asked panellists whether an international convention to regulate autonomous weapons or additional international legal limits would be needed. From the perspective of the ICRC, IHL regulates autonomous weapons and already sets limits. It therefore also prohibits some autonomous weapons. The ICRC calls urgently for internationally agreed limits, but it is yet to be said in what form these can be legally or politically binding rules. Much is at stake and much is expected of stakeholders in the discussions in the CCW.

A panellist concluded that the general concern is not necessarily that autonomy should never be used at all. However, a particular line which should not be crossed is that weapon systems would make decision or be given tasks that have historically been taken by human beings both in law and in ethics. These include the execution of critical functions or choosing who is the legitimate target. Article 49 of Additional Protocol I to the Geneva Conventions on what it means to carry out an attack gives a sense of how a target may acquire the status of a legitimate target. This must continue to be the exclusive prerogative of human beings, there is no reason to give that task to machines. This is an important line that humanity must not cross. Lastly, all considerations of how we should appropriately respond to this issue, are not only a UN concern. In all meetings, it is important to consider very broadly affected people and their interests in order to see how to best respond to the issues raised by autonomous weapon systems.

Another panellist emphasised that France very positively participates in the GGE on LAWS and continues to work on the topic. The GGE on emerging technologies in the area of LAWS has produced encouraging results. It allowed the experts to confirm the consensus around 11 guiding principles. These principles are a solid foundation, which France values very much. These principles state, *inter alia*, that IHL applies to such systems, that the decision to use them must always fall within the scope of human responsibility and that States must consider at the design stage the lawfulness of the new weapons they develop or acquire. In this respect, these principles are very close to the French position and France will continue to defend them in this forum.

Lastly, a third panellist emphasised that these discussions demonstrate that there are continuing questions among members of the public, officials and experts, about the acceptability, the desirability and the compliance of these kinds of systems, or the ability to comply with IHL when using them. These issues are worthy of discussion. What to do is a political question. Having legal clarity is desirable but it is not the only way to respond to these concerns. The complexity involved and the challenges in achieving consensus should not dissuade us from continuing the conversation because it is crucial to examine the exact concerns, whether there are sufficient protections for civilians in the law and whether there are other kinds of measures, legal or otherwise, that need to be taken in order to keep the level of protection which is seen as essential.

Panel 3
Artificial Intelligence (AI) and Machine Learning
Intelligence artificielle (IA) et apprentissage automatique

INTRODUCTION

Neil Davison
ICRC Geneva

Résumé

Neil Davison est conseiller scientifique et politique au CICR. Son introduction à ce panel s'inspirait du résumé d'un rapport du CICR sur « une approche centrée sur l'humain de l'intelligence artificielle (IA) et de l'apprentissage automatique (machine learning) dans les conflits armés »[1].

L'IA a de nombreuses conséquences, qui ne sont pas encore pleinement comprises. Le CICR s'intéresse en particulier à deux grands domaines d'application de l'IA et de l'apprentissage automatique : (1) dans la conduite de la guerre ou autres situations de violence ; (2) dans l'action humanitaire, pour assister et protéger les victimes de conflits armés. Le rapport auquel l'article fait référence présente la perspective du CICR sur l'utilisation de l'IA et de l'apprentissage automatique dans les conflits armés, ses conséquences humanitaires potentielles ainsi que les obligations juridiques et considérations éthiques devant en réguler à la fois le développement et l'utilisation. L'IA et l'apprentissage automatique pourraient affecter le rôle de l'humain dans les conflits armés, en particulier en ce qui concerne les armes autonomes, les nouvelles formes de cyberguerres et plus généralement la prise de décision.

Pour le CICR, il est nécessaire d'adopter une approche centrée sur l'humain lorsque ces technologies sont utilisées dans les conflits armés. Il sera essentiel de préserver le contrôle et le jugement humains dans les applications de l'IA et de l'apprentissage automatique pour les opérations et les décisions qui peuvent avoir des conséquences graves sur la vie des personnes, et lorsque opérations et décisions sont régies par le droit international humanitaire (DIH). L'IA et l'apprentissage automatique sont des outils qui doivent être utilisés pour servir les acteurs humains et améliorer la prise de décision, et non les remplacer.

[1] ICRC, 'Artificial intelligence and machine learning in armed conflict: A human-centred approach', 6 June 2019. Available at: <https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach>.

Neil Davison is Scientific and Policy Adviser at the ICRC. His introduction of the topic was based on an ICRC article entitled 'Artificial intelligence and machine learning in armed conflict: A human-centred approach' published on the ICRC website on 6 June 2019[2]. The article summarises the ICRC report on the topic and is reproduced below.

Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach

The ICRC, like many organisations across different sectors and regions, is grappling with the implications of artificial intelligence (AI) and machine learning for its work. Since these are software tools, or algorithms, that could be applied to many different tasks, the potential implications may be far-reaching and are yet to be fully understood.

There are two broad – and distinct – areas of application of AI and machine learning in which the ICRC has a particular interest: its use in the conduct of warfare or in other situations of violence; and its use in humanitarian action to assist and protect the victims of armed conflict.

This paper sets out the ICRC's perspective on the use of AI and machine learning in armed conflict, the potential humanitarian consequences, and associated legal obligations and ethical considerations that should govern its development and use.

AI and machine-learning systems could have profound implications for the role of humans in armed conflict, especially in relation to increasing autonomy of weapon systems and other unmanned systems; new forms of cyber and information warfare; and, more broadly, the nature of decision making.

In the view of the ICRC, there is a need for a genuinely human-centred approach to any use of these technologies in armed conflict. It will be essential to preserve human control and judgement in applications of AI and machine learning for tasks and in decisions that may have serious consequences for people's lives, especially where they pose risks to life, and where the tasks or decisions are governed by the rules of International Humanitarian Law.

AI and machine-learning systems remain tools that must be used to serve humans, and augment human decision makers, not replace them.

[2] Ibid.

AI-SUPPORTED DECISION MAKING IN WARFARE: FAR-REACHING IMPLICATIONS

L'IA EN APPUI DE LA PRISE DE DÉCISION EN TEMPS DE CONFLIT ARMÉ : DES IMPLICATIONS CONSIDÉRABLES

Edward Hunter Christie

NATO

The transcript of this presentation has not been made available for publication.

A KEY SET OF IHL QUESTIONS CONCERNING AI-SUPPORTED DECISION-MAKING

QUESTIONS CLÉS SUR LE DIH CONCERNANT LA PRISE DE DÉCISION ASSISTÉE PAR L'IA

Dustin Lewis

Research Director, Harvard Law School Program on International Law and Armed Conflict

Résumé

Dustin A. Lewis est directeur de recherche du programme de droit international et des conflits armés à la Harvard Law School. Il présente plusieurs questions clés que la prise de décision assistée par l'intelligence artificielle (IA) pose au Droit international humanitaire (DIH).

Dans une première partie, il propose plusieurs définitions et exemples. Il n'existe pas de définition internationalement reconnue de l'IA. La science de l'IA est liée à la traduction de comportement intelligent en langage informatique. Certains experts distinguent deux sous-catégories générales d'IA. Une première catégorie est celle des « handcrafted-expert-knowledge systems » pour lesquels des humains programment des « connaissances » dans un système. C'est le cas par exemple du programme d'échecs à IA Deep Blue, qui a battu le champion du monde d'échecs en 1997. Une deuxième catégorie – celle dont il est question dans cette présentation et qui concentre l'attention de certaines armées aujourd'hui – est l'apprentissage automatique (« machine learning »). Dans ce cas, les « connaissances » se construisent à partir d'assemblages de données et d'algorithmes, et en particulier à partir d'un algorithme qui fonctionne sur un programme d'apprentissage et produit un modèle d'intelligence artificielle. Il existe plusieurs familles d'algorithmes d'apprentissage automatique, qui se caractérisent par le fait que les données sont étiquetées ou non et par la manière dont le système reçoit ses données d'entrée. Concernant les « deep-learning neuronal networks » (réseaux neuronaux à apprentissage profond), il peut être difficile – voire impossible – pour les personnes physiques (1) de prévoir raisonnablement le comportement et les effets du système ; (2) de comprendre, superviser et gérer de manière fiable les performances et les effets du système pendant son fonctionnement ; (3) de retracer et comprendre le fonctionnement et les effets du système après coup.

Ces systèmes peuvent soulever des problèmes particulièrement aigus au regard du DIH.

- *Premier exemple, sur le ciblage : l'apprentissage automatique « supervisé » signifie qu'un superviseur – un humain ou un système logiciel – a associé chacune des entrées de données avec une sortie. Un belligérant peut chercher à utiliser un système d'apprentissage automatique supervisé pour classer les images satellites des transports militaires, des dépôts*

d'armes, des installations radar, etc. d'un adversaire. Ensuite, des humains pourraient utiliser ces classifications dans le cadre d'une décision de ciblage, par exemple une décision visant à satisfaire l'obligation en DIH de faire la distinction entre les objectifs militaires et les biens civils et de diriger les attaques uniquement contre les objectifs militaires.

- *Deuxième exemple, sur la détention : les algorithmes « non supervisés » peuvent extraire des caractéristiques des données sans avoir besoin d'un résultat idéal attendu par les personnes en charge du système. Un belligérant pourrait utiliser un tel système pour catégoriser un ensemble de textes, de vidéos et d'images provenant de médias sociaux, afin de repérer des schémas non identifiés auparavant et d'aider à détecter si un adversaire est susceptible de lancer une opération militaire. Des humains pourraient utiliser ces catégorisations générées par l'apprentissage automatique dans le cadre d'une décision liée au DIH concernant les adversaires qui pourraient être détenus afin de déjouer une attaque.*

Dans une deuxième partie, Dustin A. Lewis explore plusieurs questions de DIH concernant l'utilisation potentielle de techniques ou outils liés à l'IA dans des prises de décision évaluatives ou de jugements normatifs encadrés par le DIH. De nombreux principes, règles et normes requièrent de telles évaluations, parmi lesquels :

- *l'interdiction de détruire ou de saisir des propriétés ennemies, sauf les cas où ces destructions ou ces saisies seraient impérieusement commandées par les nécessités de la guerre*[1] *;*
- *la proportionnalité dans l'attaque*[2] *;*
- *l'obligation d'évaluer pendant les hostilités si un combattant blessé « s'abstient de tout acte d'hostilité » de telle sorte qu'il est protégé en tant que « blessé et malade » au sens du DIH*[3] *; et*
- *n'ordonner l'internement ou la mise en résidence forcée des personnes protégées que si la sécurité de la partie au pouvoir de laquelle ces personnes se trouvent le rend absolument nécessaire.*[4]

De telles évaluations ont des conséquences importantes et sont extrêmement difficiles à mener à bien. À cet égard, deux aspects pourraient être étudiés par les acteurs internationaux élaborant des positions sur la question :

- *D'abord il s'agit de savoir si ce sont des personnes physiques qui doivent prendre ces décisions et jugements normatifs (ou de valeur), ainsi que d'autres décisions d'évaluation man-*

[1] Convention de la Haye IV (1907), art. 23(g) ; voir aussi Convention de Genève IV (1949), art.53.

[2] Protocole additionnel I (1977), art. 51, par. 5b), art. 57, par. 2a)iii) et art. 57., par. 2b).

[3] Protocole additionnel I (1977), art. 8a).

[4] Convention de Genève IV, art. 42, 1er para.

datées par le droit – et, dans l'affirmative, quelles prescriptions et proscriptions relatives à des techniques ou outils particuliers liés à l'IA peuvent en découler.

- *Un deuxième élément concerne l'établissement et la validation des informations sur lesquelles ces décisions et jugements de licéité sont fondés, et éventuellement dans quelles circonstances et sous quelles conditions.*

1. Introduction

The organisers have asked me to speak about key International Humanitarian Law questions concerning decision making supported by artificial intelligence. In this intervention, I will set out definitions and examples, then turn to one set of IHL questions in this area, namely, the potential use of AI-related techniques or tools in relation to the making of legally mandated evaluative decisions and normative judgments and to configuring the information on which those assessments are based.

2. Definitions and Examples

There is no internationally agreed definition of 'artificial intelligence' in this area.[5] It may be useful to consider these issues from the broad frame of AI science. According to my understanding, AI science may be said to pertain in part to the development of computation- based understanding of intelligent behaviour, typically through two interrelated steps.[6] One of those steps concerns the determination of cognitive structures and processes and the corresponding design of ways to represent and reason effectively. The other step relates to developing theories, models, data, equations, or algorithms that embody that understanding.

Some commentators have explained various approaches to artificial intelligence by distinguishing two general subcategories of AI systems.[7] Both subcategories have been around for decades.

[5] For materials reflecting the vast range of States' views on definitional aspects, see Dustin A. Lewis (ed.), *A Compilation of Materials Apparently Reflective of States' Views on International Legal Issues pertaining to the Use of Algorithmic and Data-reliant Socio-technical Systems in Armed Conflict* (Harvard L. School Program on Int'l L. & Armed Conflict 2020), at: <https://pilac.law.harvard.edu/a-compilation-of-materials-apparently-reflective-of-states-views-on-international-legal-issues-pertaining-to-the-use-of-algorithmic-and-data-reliant-socio-technical-systems-in-armed-conflict>.

[6] See Dustin A. Lewis, 'Legal reviews of weapons, means and methods of warfare involving artificial intelligence: 16 elements to consider' [21 Mar. 2019] ICRC Humanitarian Law and Policy Blog, at: <https://blogs.icrc.org/law-and-policy/2019/03/21/legal-reviews-weapons-means-methods-warfare-artificial-intelligence-16-elements-consider/>.

[7] See e.g. Greg Allen, 'Understanding AI Technology' [Apr. 2020], U.S. DoD Joint Artificial Intelligence Center, at: <https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf>.

The first is handcrafted-expert-knowledge systems. These systems use rules-based software to codify subject matter knowledge of human experts into a series of programmed rules. With handcrafted knowledge systems, humans program the 'knowledge' into the system. As an example, think of the AI chess system Deep Blue that defeated the world champion in chess in 1997. That system was created through a collaboration between computer programmers and human chess grandmasters to write an algorithm in computer code that considered many potential moves and countermoves and reflected rules for chess play. I will not focus on these systems here.

The second subcategory, which attracts much greater attention from some militaries today, is machine learning.[8] With machine-learning systems, the 'knowledge' is learned from assemblages of data and algorithms, particularly from an algorithm that runs on a training dataset and produces an AI model. There are different families of machine-learning algorithms. And these families can be distinguished by whether or not the training data is labelled or unlabelled and how the system receives its data inputs. For some machine-learning systems, not least deep-learning neural networks, it may be difficult – and perhaps impossible – for natural persons: (1) to reasonably predict the system's behaviour and effects; (2) to reliably understand, supervise, and administer the system's performance and effects during operations; or (3) to sufficiently trace and understand the system's performance and effects after the fact.[9] Such systems may present especially significant issues from an IHL perspective.[10]

Consider two potential examples of applications of machine-learning methods related to armed conflict. One concerns targeting, and the other relates to detention.

[8] See, e.g., Kelley M. Sayler, 'Artificial Intelligence and National Security' [21 Nov. 2019], Congressional Research Service Report No. R45178, at: <https://fas.org/sgp/crs/natsec/R45178.pdf>; International Committee of the Red Cross, 'Autonomy, artificial intelligence and robotics: Technical aspects of human control' [Aug. 2019], at: <https://www.icrc.org/en/download/file/102852/autonomy_artificial_intelligence_and_robotics.pdf>.

[9] See e.g. Arthur Holland Michel, 'The Black Box, Unlocked: Predictability and Understandability in Military AI' [2020] U.N. Inst. for Disarmament Research, at: <https://unidir.org/publication/black-box-unlocked>; Jenna Burrell, 'How the machine "thinks": Understanding opacity in machine learning algorithms' [Jan.–June 2016] Big Data & Society 1. On legal aspects of automatic target recognition systems involving 'deep learning' methods, see Joshua G. Hughes, 'The Law of Armed Conflict Issues Created by Programming Automatic Target Recognition Systems Using Deep Learning Methods' [2018] 21 YBIHL 99.

[10] See e.g. Vincent Boulanin, Neil Davison, Netta Goussac and Moa Peldán Carlsson, *Limits on Autonomy in Weapon Systems* [June 2020] Stockholm International Peace Research Institute and International Committee of the Red Cross, at: <https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy_0.pdf>.

First, 'supervised' machine learning means that some supervisor – such as a human or a software system – has accurately labelled each of the data inputs with its correct associated output. As an example, if the goal of the AI system is to correctly classify the objects in different images as either 'ambulance' or 'tank', the labelled training data would have image examples paired with the correct respective classification label of 'ambulance' or 'tank'. A party to an armed conflict might seek to use a supervised machine-learning system to classify satellite images of an adversary's military transports, weapons depots, radar installations, and the like.[11] One or more humans employing this machine-learning system might use those classifications as part of a targeting decision, such as a decision aimed at satisfying part of the IHL obligation to distinguish between military objectives and civilian objects and to direct attacks only against military objectives.

Second, 'unsupervised' algorithms are those that can extract features from the data without the need for an ideal expected result – ideal, that is, according to the people in charge of the system. For example, a party to an armed conflict might seek to use an unsupervised machine-learning system to categorise a collection of social media-based texts, videos, and images to spot previously unidentified patterns to help detect whether an adversary may launch a military operation. One or more humans might use those machine learning-generated categorisations as part of an IHL-related decision concerning which perceived adversary may be detained to thwart an anticipated attack.[12]

[11] See e.g. Nathan Strout, 'Inside the Army's futuristic test of its battlefield artificial intelligence in the desert' [25 Sept. 2020], C4ISRNET, at: <https://www.c4isrnet.com/artificial-intelligence/2020/09/25/the-army-just-conducted-a-massive-test-of-its-battlefield-artificial-intelligence-in-the-desert/>.

[12] Authorities in Israel have reportedly used algorithms as part of attempts to obviate anticipated attacks by Palestinians through a process that involves the filtering of social-media data, resulting in over 200 arrests. See 'Israel claims 200 attacks predicted, prevented with data tech' [12 June 2018] CBS News, at: <https://www.cbsnews.com/news/israel-data-algorithms-predict-terrorism-palestinians-privacy-civil-liberties/>. (Reporting with respect to that system does not indicate whether or not the system included unsupervised algorithms as such). More generally, see Dustin A. Lewis, 'AI and Machine Learning Symposium: Why Detention, Humanitarian Services, Maritime Systems, and Legal Advice Merit Greater Attention' [28 Apr. 2020], Opinio Juris, at: <http://opiniojuris.org/2020/04/28/ai-and-machine-learning-symposium-ai-in-armed-conflict-why-detention-humanitarian-services-maritime-systems-and-legal-advice-merit-greater-attention/>; Tess Bridgeman, 'The viability of data-reliant predictive systems in armed conflict detention' [8 Apr. 2019], ICRC Humanitarian L. and Policy Blog, at: <https://blogs.icrc.org/law-and-policy/2019/04/08/viability-data-reliant-predictive-systems-armed-conflict-detention/>; Ashley Deeks, 'Detaining by algorithm' [25 Mar. 2019], ICRC Humanitarian L. and Policy Blog, at: <https://blogs.icrc.org/law-and-policy/2019/03/25/detaining-by-algorithm/>; Ashley S. Deeks, 'Predicting Enemies' [2018] 104, Virginia LR 1529.

3. Framing a Set of Key IHL Questions

With respect to IHL, from my perspective, a set of key questions in this area concerns the diverse array of evaluative decisions and normative (or value) judgments mandated in international law applicable in relation to armed conflict.[13] Numerous principles, rules, and standards require such assessments, including provisions relating to:

- the prohibition of the destruction or seizure of the enemy's property unless such destruction or seizure is 'imperatively demanded by the necessities of war';[14]
- the prohibition on attacks that may be expected to cause 'incidental' loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof that would be 'excessive' in relation to the 'concrete and direct military advantage' anticipated;[15]
- the obligation to assess during hostilities whether an injured fighter 'refrains from any act of hostility' so as to become protected in terms of being 'wounded and sick' in IHL terms;[16] and
- ordering the internment or placing in assigned residence of protected persons only if 'the security' of the detaining power 'makes it absolutely necessary'.[17]

[13] See e.g. Switzerland, 'Towards a "compliance-based" approach to LAWS [Lethal Autonomous Weapons Systems] [30 March 2016], Informal Working Paper, p. 3, para. 16, at: <https://www.unog.ch/80256EDD006B8954/(httpAssets)/D2D66A9C427958D6C1257F8700415473/$file/2016_LAWS+MX_CountryPaper+Switzerland.pdf>.

[14] Art. 23(g), Hague Regulations IV (1907); see also Art. 53, GC IV (1949).

[15] Arts. 51(5)(b), 57(2)(a)(iii), and 57(2)(b), AP I (1977).

[16] Art. 8(a), AP I (1977).

[17] Art. 42, first para., GC IV. See also Lewis, AI and Machine Learning Symposium, above note 14 ('[T]o raise two doctrinal examples, IHL envisages assessments as to whether a protected person may be interned or placed in assigned residence either in an international armed conflict because "the security of the Detaining Power makes it absolutely necessary' or in a situation of occupation because the Occupying Power deemed it that it was "for imperative reasons of security." Transposing those evaluative decisions and normative judgments partially or fully into algorithmically generated assessments by way of data-shaped probabilities in concrete cases presents a far-from-frictionless exercise, to say the least. Moreover, the rules for detention relating to non-international armed conflict are, in many respects, even less clear than their international-armed-conflict counterparts. That would appear to leave larger space for those employing AI technologies to rely upon potentially problematic domestic criminal-law examples without enough international law to guide them').

These and other[18] highly consequential assessments may be extremely difficult to make.[19]

Two potential AI-related components related to this set of IHL questions may be distinguished. A first one concerns whether natural persons must make those decisions and judgments. In that connection, it may be warranted for international stakeholders to establish positions on whether or not these and other legally mandated evaluative decisions and normative (or value) judgments *may be reposed only in one or more natural persons* – and, if so, what prescriptions and proscriptions relative to particular AI-related techniques or tools may arise therefrom.

A second component concerns establishing and validating the information on which these legally mandated decisions and judgments are based. In particular, it may be warranted for international players to establish positions on whether – and, if so, under what circumstances and subject to what conditions – reliance may be placed on AI-related techniques and tools to partly or wholly establish or validate that information.

[18] Additional examples (among many others) include provisions relative to evaluative decisions and normative (or value) judgments concerning: the presumption of civilian status in case of 'doubt' (Arts. 50(3) and 52(3), AP I (1977)); the betrayal of 'confidence' in relation to the prohibition of perfidy (Art. 23(b), Hague Regulations IV (1907) and Art. 37(1), AP I (1977); see also Art. 8(2)(b)(xi), ICC Statute (1998)); whether an object – by its nature, location, purpose or use – makes 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 (Art. 52(2), AP I (1977), Art. 2(4), Protocol II to the CCW (1980), Art. 1(3), Protocol III to the CCW (1980), Art. 2(6), Protocol II to the CCW (1996 amend.), and Art. 1(f), Second Protocol to the Cultural Property Convention (1999)); and whether or not a civilian takes a 'direct' part in hostilities (Art. 51(3), AP I (1977) and Art. 13(3), AP II (1977)).

[19] See, e.g., ICRC, *GC I Commentary* (2016 update), para. 1348 ('Under combat conditions, in the very moment that a person is injured, *it may be extremely difficult* to determine with any degree of certainty whether that person is wounded in the legal sense, and in particular whether he or she is refraining from any hostile act.') (emphasis added).

EFFORTS TO GOVERN (MILITARY APPLICATIONS OF) AI
COMMENT RÉGLEMENTER L'IA, EN PARTICULIER SES APPLICATIONS MILITAIRES

Pauline Warnotte

UNIDIR and University of Namur

Résumé

Pauline Warnotte est chercheuse au sein du programme Sécurité et technologie de l'Institut des Nations unies pour la recherche sur le désarmement (UNIDIR) et enseigne à l'université de Namur. Dans cette contribution, elle aborde, à titre personnel, les tentatives de réglementation des applications militaires de la prise de décision assistée par intelligence artificielle (IA) dans les conflits armés.

La première partie présente une typologie de l'IA dans les conflits armés et discute de sa pertinence pour l'éthique et la gouvernance de l'IA militaire. La typologie de l'IA relative au processus de prise de décision militaire comprend trois catégories principales : les systèmes offensifs ; les systèmes non offensifs ; les systèmes d'IA appuyant la prise de décision dans la planification générale d'une opération militaire. Chacun de ces types d'intelligence artificielle militaire soulève des préoccupations éthiques et des questions juridiques spécifiques, découlant de plusieurs facteurs, notamment : le type de mission (létale ou non létale), l'environnement (peuplé ou non), le processus d'interaction homme-machine, l'opérateur humain éventuel, ses connaissances et sa compréhension du système, ainsi que d'éventuels biais conduisant à une confiance excessive dans le système. Des questions de même nature se posent dans l'utilisation non militaire de l'IA. Ces développements ont donné lieu à un nombre exponentiel de normes, codes de conduite, lignes directrices, procédures et autres principes élaborés notamment par le secteur privé, les institutions nationales et les organisations internationales. Malgré le manque d'homogénéité, il existe certains éléments saillants : la référence au respect du droit international ; la notion de responsabilité ou d'obligation de rendre des comptes ; la nécessité pour l'IA de répondre à des critères qualitatifs ou à des normes, tels que la robustesse, la sécurité, la transparence, la conformité, l'explicabilité et la prévisibilité ; et la nécessité d'effectuer des essais et tests, des vérifications et des réexamens périodiques du système et des principes eux-mêmes.

La deuxième partie présente quelques initiatives visant à réguler plusieurs aspects du développement et de l'utilisation de l'IA et de l'apprentissage automatique appuyant la prise de décision militaire. En 2019, 11 principes directeurs ont été adoptés au sein du Groupe d'experts gouvernementaux sur les armes létales autonomes. Ils réaffirment la pertinence et l'applicabilité du

droit international et consacrent la notion d'interaction homme-machine ainsi que la question du « risque » qui devrait être abordée dès le début de la conception. L'auteure donne ensuite plusieurs exemples de mesures de gouvernance visant à mettre en œuvre ces principes. Des États ou entreprises privées déclarent qu'ils n'utiliseront pas l'IA à des fins militaires – tout en continuant à la développer à d'autres fins de défense. Des mesures visent à assurer la conformité éthique et juridique d'un système d'IA militaire dès sa conception. D'autres visent à garantir la qualité des systèmes d'IA en insistant sur la fiabilité, tant morale que technique. Dans ce cas, la mise en œuvre correcte de ces caractéristiques ou même la pertinence d'un seuil de confiance « suffisant » qui serait exprimé en pourcentage posent question. Enfin, certaines directives précisent dans quelles situations le processus décisionnel peut être délégué à un algorithme et dans quelles situations il doit être mené ou validé par un être humain.

Malgré le nombre considérable de directives tant éthiques que techniques et leur apparente exhaustivité, la question de leur pertinence et de leur efficacité demeure. L'innovation en matière d'IA n'évolue pas dans un vide juridique. Outre l'obligation de respecter le DIH dans la conduite des hostilités, les États sont également tenus, en vertu du droit international, d'examiner la licéité des nouvelles armes, ainsi que des moyens et méthodes de guerre qu'ils souhaitent développer ou acquérir. Toutefois, selon l'auteure, les mesures de bonne gouvernance peuvent favoriser le respect du droit international pour au moins deux raisons. Premièrement, l'élaboration et la mise en œuvre de mécanismes de gouvernance par les acteurs privés permettent une proximité maximale avec les questions scientifiques, techniques ou éthiques auxquelles ils sont confrontés. Deuxièmement – en dépit de certaines faiblesses inhérentes à la nature volontaire des lignes directrices provenant du secteur privé – la contribution possible de ces acteurs à la gouvernance de l'IA pourrait permettre une meilleure adhésion aux lignes directrices développées et/ou adoptées par les acteurs eux-mêmes.

L'auteure conclut que les règles fondées sur le droit international et les autres mesures prises par les gouvernements et/ou les acteurs privés sont complémentaires et se renforcent mutuellement pour contribuer à l'élaboration d'un cadre global pour la gouvernance de l'IA militaire. Cette approche multiforme aurait le mérite de dépasser le « dilemme de Collingridge » qui est souvent invoqué à propos des nouvelles technologies.

Enfin, elle identifie quatre considérations clés pour que ces principes de bonne gouvernance puissent être considérés comme pertinents :

1. *leur élaboration et leur adoption devraient impliquer toutes les parties prenantes, en exigeant leur participation et en visant leur adhésion ;*
2. *ils doivent couvrir l'ensemble du cycle de vie d'un système ;*

3. *les personnes concernées doivent être sensibilisées au contenu de ces principes, y adhérer et être formées pour les mettre en œuvre ; et*
4. *une évaluation périodique de ces principes devrait être effectuée afin de s'assurer qu'ils restent pertinents, notamment au regard d'une évolution possible du cadre juridique.*

I first would like to thank the ICRC and the College of Europe for the pleasure and the honour of speaking today at this edition of the Bruges Colloquium, among such distinguished panellists.

For this last intervention of the panel, I will briefly address, in my personal capacity, the efforts to govern military application of AI-supported decision making in armed conflict.

I will structure my presentation around three points:

- first, I will briefly discuss the typology of AI in armed conflict and its relevance for ethics and the governance of military AI.
- second, I will highlight salient points of some initiatives that have been put in place to date to try governing several aspects of the development and the use of AI.
- and third, I will discuss the relevance of ethical AI principles as a possible way of moving forward in the development of a relevant and viable governance framework for the use of AI-supported decision making in military operations.

Due to the focus of this presentation, I will not deal today with other questions related to strategic and political considerations and the broader spectrum of arms control measures.

I would also like to note that Artificial intelligence comprises a vast number of fields, including machine learning, natural language processing, robotics, computer vision, and knowledge representation and reasoning. For the purposes of this presentation, I refer to AI-driven and machine learning-driven technologies specifically.

Military AI Topology, Ethical Concerns, Legal Questions

We can consider that the AI typology regarding the military decision-making process, – which, as Dustin expertly described – encompass both handcrafted-expert-knowledge systems and machine learning, includes three main categories:

- Offensive systems, kinetic or otherwise adversarial, such as a jamming system.
- Non-offensive systems, that could be for instance identification and facial recognition systems.

- AI systems supporting decision making in the broader planification of a military operation. For example, this could be a system that would support the determination of the most appropriate rules of engagement to achieve the intended objective.

The scope of application of military artificial intelligence (MAI) is therefore much broader than the mere question of autonomous weapon systems (AWS), lethal or not. Furthermore, each of these type of MAI raises specific ethical concerns and legal questions requiring tailored answers. I will try to avoid using the term 'problem': they are not problems but questions and/or concerns arising from several factors including:

- the type of mission (lethal or non-lethal),
- the environment (populated or not),
- the human-machine interaction process,
- but also the human operator themselves (if any) including his or her knowledge and understanding of the system and a possible automation bias leading to over-rely on the system itself.

Similar questions are also encountered in the non-military use of AI. The use of AI and machine learning in decision-making process is now broadly implemented and the AI ecosystem encompasses the full spectrum of human activities, from healthcare to education, agriculture, administrative decisions, and the finance sector.

It is therefore unsurprising that we witnessed the blooming of an exponential number of standards, codes of conduct, guidelines, procedures and other principles developed notably by the private sector, national agencies and international organisation. This is a euphemism to speak about a lack of homogeneity in the AI regulation or governance approach. Nevertheless, I would like to point some salient if not common elements:

- the reference to the compliance with (international) law;
- the notion of responsibility or accountability in the design and development;
- the necessity for AI to meet qualitative criteria or standards, although addressed under various concepts, such as robustness, security, transparency, compliance, explainability and predictability, and
- the need to conduct testing, verification, and periodical review of both the system and the principles themselves.

Some Initiatives Aiming at the Governance of (Military) AI/ML Supported Decision Making

I will therefore now turn to my second point and briefly mention some initiatives aiming at the governance of AI or machine learning-supported decision making.

As was mentioned yesterday, in 2019, a set of 11 guiding principles was adopted at the GGE on LAWS level. Those principles mainly reaffirm the relevance and applicability of existing international law. Other important aspects are the notion of human-machine interaction and the emphasis on the question of 'risk', which should be addressed from the early design stage, implying therefore the necessity to implement those considerations.

Governance measures can target the design, development and use of an object but also a choice or behaviour. Some States and private companies such as Google, based on legal and ethical considerations, have declared that they will not weaponise AI – while still developing it for other defence purposes.

Measures can also aim at ensuring that legal and ethical considerations are discussed at the earliest stage of the development of military AI. One example is the French-German Future Combat Air System Forum for the responsible use of new technologies launched by Airbus earlier this year. These measures feature both the definition of ethical guidelines 'based on international law' and their 'technical implementation', aiming at achieving ethical and legal compliance by design. To this aim, a commission of experts, which is one of the good practices that we also see blooming, has been set up. The experts come from numerous fields, including computing science, law and ethics.

Some governance measures have also been created to guarantee the qualitative characteristic of developed AI systems. Examples are the 2019 Beijing Principles, the EU Commission's Ethics Guidelines for Trustworthy AI, the Model AI Governance Framework in Singapore and the UNESCO draft recommendations on the ethics of artificial intelligence. Common characteristics include trustworthiness and reliability. However, one can ask how to ensure the correct implementation of those characteristics, including testing and verification. On its side, Amazon decided for a recommended very high confidence threshold of 99% in case of law enforcement use of facial recognition technology. This however raises other questions. Is 99% a sufficient threshold for this kind of situation? How is this confidence threshold calculated? And what happens in case of failure?

As a final example, regarding the use of AI, some guidelines specifically indicate, based on an impact assessment, in which situations the decision-making process can be delegated to an algorithm and in which situation it has to be made or validated by a human being. This is notably the case in the Canadian Directive on Automated Decision Making.

Despite the huge number of ethical and other technical guidelines and their apparent comprehensiveness, one can however ask the question of their relevance and their effectiveness.

Value and Merits of Guidelines and Principles and their Possible Development in the Military Domain

Innovation regarding AI does not operate in a legal vacuum. Indeed, next to the obligation to adhere to IHL in the conduct of hostilities, States are also required under international law to legally review new weapons, means and methods of warfare that they intend to develop or acquire. One can also assume that States that would themselves have adopted domestic rules and policies on AI or machine learning to support decision making, derived from their international obligations, would comply with those requirements.

But governance measures are also relevant to achieve or, at least, to foster compliance with international law for at least two reasons.

First, the development and implementation of governance mechanisms allows States to stay as close as possible to scientific, technical or ethical questions. This is particularly true and important for specific fields where final users do not always have the required knowledge to understand how systems are built and operate. The adoption of a relevant set of guidelines to be applied to AI in the military domain should, in my opinion, ideally involve both the scientific community having to implement those guidelines and the governmental authorities having recourse to the service of non-governmental entities.

Second – and even though this is contested in some studies – the possible contribution of private players to governance of AI allows governments to secure better adherence to guidelines developed and/or adopted by those bodies themselves. It creates a better opportunity to check their relevance compared to a traditional rule-making process.

It is however also true that guidelines derived from the private sector show some inherent weaknesses, notably the need to establish an enabling environment in order for those guidelines to achieve effectivity, their voluntary nature questioning how their compliance can be ensured and the possible difficulties to verify their implementation and application.

Nevertheless, in my opinion, they can be seen as a possible median way that would allow to navigate between what is technically achievable, legally authorised and socially acceptable.

Conclusion

Rules based on international law and other measures driven by governments and/or private players are not only complementary, they are also mutually reinforcing and help build a comprehensive framework for military AI governance. Amongst the merits of this multiform approach, I would like to underline its ability to move beyond the so-called 'Collingridge dilemma' that is often invoked in relation to new technologies, considering that there is a current lack of sufficient information to address the questions while once these systems are finally operational, it might be too late to regulate them.

However, to be relevant, certain key considerations related to the adoption process and scope of those governance mechanisms should be present:

1. their development and adoption should engage all relevant stakeholders, requiring their involvement but also aiming at their adherence;
2. principles should cover the whole life cycle of a system, going from the design level to the development phase, including the validation, verification, testing and evaluation of the system, pre and post deployment;
3. the relevant persons should be aware to the content of those principles, adhere to them and be trained to implement them; and
4. a periodical evaluation of those principles should be conducted to ensure their continued relevance, notably regarding the possible evolution of the legal framework.

Thank you very much.

DISCUSSION

In session three the following topics were discussed in depth.

1. Obligations and Responsibility in AI-Supported Decision Making

The moderator highlighted that all the presentations relate to the underlying obligation and responsibility of the people using AI machine learning to support decision making in armed conflict. Others have used the term 'human-centred' or 'human compatible' AI to think about ways of approaching these issues from the perspective of human obligations. As a starting point, the moderator invited the panellists to discuss whether they saw the question in terms of technical approaches of how AI machines are used in decision making, in terms of limits of the types of decisions it is used for or in terms of human or other implications.

In a first panellist's opinion, there are potentially four levels of analysis, which are not necessarily mutually exclusive. The first level of analysis would be that of overarching principles which could be made to apply across, not just AI, but all sub-functions in a system, including other types of technical equipment, physical or not. Such principles could be for instance: reliability, traceability or governability. These are generally desirable for many types of technological equipment, not just military ones. A second layer of analysis could be the criticality-of-use case of the system under discussion. Is it about logistics, intelligence gathering, combat operations? One may want to slightly or strongly differentiate according to that criterium. The third level would be the functions within a given system which should be subject to stricter or lighter human supervision. As examples of functions, in an aerial system for instance there are: navigation, communicating back to the command post, tracking potential targets. The fourth level which may be considered is the nature and the reliability of the control mechanisms to be envisaged. This might seem counterintuitive because the debate focuses on these issues as human-controlled v. machine-controlled, but there may be a third way. According to the panellist, serious considerations should be given to the use of automated software solutions to tackle unintended AI behaviour. It is attractive because it could be much faster than a human supervisor, meaning that one could stop the potential damage faster. This is by no means a new concept. In the civilian sector, there are already many industrial applications, e.g. in power plants or with control software for safety purposes, or even in the financial markets.

Another panellist suggested stepping back to envision the development of legal regulations concerning war technologies in the coming years, with machine learning playing an increasingly important role for some militaries. This raises questions of what it means to have legal compliance and accept responsibility in this area. There is a huge opportunity to try to systematically connect substantive legal roles with the realisation of these assumptions underlying international legal responsibility while being very clear about what States and other international legal actors are or are not willing to commit to in this area, especially given the current prevalence of approaches on military secrecy. Those linkages between the rules and those assumptions could help operators detect and instantiate the general responsibility concepts and the legal requirements pertaining to particular uses, including of AI or machine learning. Those concepts and attributes might be able to be used to identify the technical properties required for a particular use of these technologies, whether in the detention context. in the humanitarian context or in targeting for the conduct to be considered legally responsible. The panellist suggested that previous and current approaches which focus, sometimes very narrowly, on the legal compliance dimension solely in the eyes of the individual, might merit reconsideration: a broader frame of legal responsibility should be taken into account. This could translate into adding new dimensions to the traditional legal assessment, which provides that technologies are capable of being, and are in fact subject to, an assessment concerning legality in respect to each contemplated use or actual use. One new dimension of the legal compliance could be to assess the capability of the technology to be subject to an evaluation regarding international legal responsibility, in order to discern whether or not a breach exists and to whom the conduct is attributable. Such evaluation would therefore aim to determine whether the technology is capable of being subject to an application of international legal responsibility in case of violation. This is an opportunity to think through more broadly what can be demanded of the legal system of protection in war. The moderator noticed that this topic is making legal experts return to old issues and explore new legal questions.

Lastly, the third panellist agreed that there is a need to return to the definition of a legal obligation and the notion of responsibility. According to the panellist, a human-centred approach and an approach which is compatible with being human does not lie outside of this framework. The 11 Guiding Principle which were adopted by the GGE on LAWS reiterated that the human being is responsible when using artificial intelligence. The human being is responsible throughout the entire life cycle of the system, from the stage of design until the system is actually implemented. A holistic approach would ensure that humans are at the centre of decision making, in a manner which commensurate with each one’s role in these decisions. This approach is not new. It is a reinterpretation of legal notions within the framework of existing international law, which ensures that despite technological developments, international law remains the basis and remains relevant, so that the technological developments comply with the existing legal framework.

2. Opportunities and Risks of AI-Supported Decision Making

A participant asked whether beyond the concerns AI could be an opportunity to improve physical military capability and decision making to reduce error or unintended consequences, e.g. preventing the targeting of civilians.

A panellist emphasised that this is a recurrent topic. The idea with AI is to make a better use of or to add to what already exists, in order to improve the processes and tools in military operations. One such way of improving processes, which also fulfils an obligation under IHL, is to ensure that the number of individuals involved is kept to a minimum. The use of artificial intelligence can be beneficial from this perspective, which must be a constant consideration in hostilities. Nonetheless, AI is not a panacea. Most importantly, voices are being raised on issues that are not necessarily linked to legal issues, although the legal aspect comes into play, but which are aspects linked to ethics and to what is meant by the responsibility and the role of humans more generally in military operations. A question that arises in the current discussions is: even if these systems are completely legal, are they acceptable in terms of the current recognised societal norms? The panellist also referred to what was said in the panel of the previous day regarding non-European and more diverse ethical standards. According to the panellist, biases are also due to the language and access to other ethical norms, as well as the available academic papers online which are mostly in English and from Western sources. However, when discussing ethics within the GGE on LAWS, it is crucial to recall that international law and especially IHL remains the bottom line and a universal common ground. The panellist therefore opined that IHL should be the basis, even in ethics considerations.

For another panellist, under certain circumstances and subject to certain conditions, certain AI-related techniques and tools might be relied upon to help configure the information necessary to make decisions that IHL mandates, e.g. targeting military objectives or determining who should be detained. However, as compared with other technologies, there is a concern with respect to particular families of machine learning and algorithms and the potential lack of ability for the human operators across the life and death cycle in relation to targeting. One should consider if the operators know enough about the system to be able to know beforehand what it will reliably do in a reasonably foreseeable sense, whatever that standard is. So far there is no generalisable standard applicable to all these potential applications. This could be discussed further and clarified. A second dimension is monitoring and administration during the operations, where the applicable standards are not clear. From a legal perspective, there will hopefully be reflection on ethical and normative commitments on that dimension. Thirdly, some of these machine-learning algorithms' families might raise concerns about the ability to trace and to reconstruct what has triggered the recommendation to attack certain area or certain military installations v. others. This is often framed in terms of explainability, interpretability, understandability. This is an area where there are not only legal but also ethical and political concerns. There is therefore a potential for certain applications to give rise to what could be considered outcomes of less civilian harm as one of the main aims of IHL, but a lot of work is needed to figure out exactly what the dependencies are in these systems, whether or not the operators understand them and whether or not they would even be capable of external juridical scrutiny. For instance, would an agency trying to assess whether a crime was committed – which is an obligation under Additional Protocol I and for grave breaches under the four Geneva Conventions – be able to reliably assess after the facts what actually happened and whether or not a grave breach has occurred?

A panellist noted the question reflects a relatively common argument by those who are more technology enthusiasts as regards robotics in general, and military applications of robotic systems more specifically. In the panellist's opinion, it is a very strong argument: outside of the realm of AI, there is a much longer term trajectory towards developing weapon systems that tend to be more precise. They are not necessarily always so, it depends on how they are used, but there is a trend towards precision-guided weaponry. Back in the 1960s, the emerging technology was intercontinental anti-ballistic missiles and even more powerful hydrogen bombs. According to the panellist, current technological developments are very far away from these types of potentially massive indiscriminate attacks on population centres with hydrogen bombs. There is a trend towards more precision, and more precision is consistent with IHL and helps to better comply with IHL. Although it will not be perfect and war crimes will still occur, more precision means more accuracy, better distinction and less indiscriminate actions. The panellist therefore agreed with the assumption in the question that AI included in machine and deep learning techniques can help the military to be more accurate.

The panellist added that although second generation AI deep learning has an explainability problem, incidentally so do human brains. Deep learning is based on the notion of artificial neuronal networks. Once it has been trained enough, the machine sets itself up with hundreds or thousands or millions of parameters to become an efficient protection machine, which is good at pattern recognition. Occasionally, it will get it wrong. However, from a very pragmatic perspective, if after repeated training and testing and training and testing again, a pattern recognition piece of software does better than humans in terms of both precision and recall, i.e. in terms of both avoiding false positives and avoiding false negatives, it is rather tempting to use it. It will not be 100% accurate, but humans are not 100% reliable neither. Therefore, someone who has a utilitarian framework's reference may want to accept the deployment of machines that use, not exclusively but also use some elements that are based on deep learning, provided they have very high performance and in the end, on average, one would expect more precision and less violations as a result. There are other arguments as well. For example, algorithms would not develop emotions, like hatred, resentment, bitterness, etc. For the panellist, it is worth delving into these questions in more details from all of the various perspectives.

On the comparison between humans and machines, the moderator concluded that beyond the distinction, future developments may lead one to think in terms of different configurations: humans with one type of machines and no humans with another type of machines. However ultimately, there would be humans deploying and using these technologies. The way they decide to use them can perhaps reinforce existing good practices but could also amplify bad practices where they exist, which is a very complex issue.

3. The Danger of Biased Algorithms

Another participant raised a question on how the panellists would assess the danger of biased algorithms, e.g. gender or ethnicity biases, and how to solve that problem.

A first panellist answered that although it is a very relevant and important question, from a defence perspective, most of the cases where this is highlighted are civilian types of applications within AI societies, e.g. how companies interact with their customers. For instance, a bank or an insurance company might have biases in one of its algorithms and end up inadvertently – or maybe even on purpose – mistreating certain segments of the population groups. This is a huge concern, and there are ways of addressing this. When it comes to military applications *per se*, there is the risk of having bias either in the dataset, in the ways that the data was collected, or maybe even in the design: the military must therefore be aware of those problems. However, according to the panellist, this would not necessarily raise the same questions as in the civilian domain, especially regarding combat-oriented applications. In other words, the type of bias depends on the types of applications. Ethnicity and gender biases would typically be a risk in a population to which the application is used. But for military applications, the panellist believes that it is hard to see how that could happen, except for an instance with civilians facing military activities in which deployed troops would be using an AI application to interact with the public. For purely military applications, other biases might be more relevant. An example would be an algorithm which is trained to object detection, i.e. to recognise objects in a certain landscape, to recognise friend v. foe and friend and foe v. civilians: one needs to make sure that the examples used to train it, training datasets, cover the whole range of the possible.

Another panellist said that bias might be a type of problems that cannot be solved. Bias will exist irrespective of the application involving a machine learning system. Some writings suggest that regarding the review of weapons, means and methods of warfare involving AI- related techniques and tools, it should be incumbent upon the reviewer, often from the outset, to discern what biases are capable of arising, precisely because there is no way to set up an 'unbiased system' that is just based on the nature of the technologies with these assemblages of algorithms and data. Concerns about the applications of the existing technologies would include for instance an image classification system. Military secrecy entails that the labels or the training data would be left unknown and would almost never be disclosed, unlike in a civilian context. There is a transparency and secrecy concern in this respect. There are ongoing discussions on the necessary number of labelled images, e.g. would five thousand labelled images be sufficient to be able to nominate someone for targeting, which somebody else would need to actually validate and decide to nominate? This type of information is left unknown for those outside the government, even though experts outside the government and other organisations should have a say in this legal system of protection on which they are working. It is a bit of a black box. Other scholars have even talked about 'the double black box' in this context. In terms of biases too, there is a potential for weighting the different values of the inputs, i.e. proxies for legally relevant characteristics. Ethnicity and gender would be excluded as proxies for legally relevant characteristics in armed conflict. However, in some context some parties might say that if some individuals are members of certain tribal groups or certain age groups or even of a certain gender, e.g., male, that those might actually be considered, internally – within the party, not subject to external scrutiny – proxies for what they would consider legally relevant characteristics and would weigh those accordingly to make an assessment as to whether or not somebody made them prioritised or nominated for targeting. Bias is an extremely difficult area, one that can be addressed but not solved in that it brings back the question concerning the scrutinised ability of these systems internally but also externally. The moderator added that it raises the question that also came up in the discussion of the previous day on autonomous weapons about technical indicators as proxies for whether someone or something can be attacked or otherwise.

Lastly, a panellist commented that the governance principles are crucial. Discussions of trustworthy AI presuppose an algorithm which is without bias. Indeed, how could AI be trustworthy AI if the algorithm is biased by design? The guidelines on these principles are key for the implementation of the principles. It is not up to a legal adviser to say whether this can technically be achieved or not. However, the panellist agreed that this should be addressed, at least, at the latest during the legal review of the system. One of the main challenges is to detect these biases before the system is used. There is a very important element of validation or verification and later training and evaluation of the system during the design phase of the system, during the development phase of the system, later when assessing the value of the system and when finally, if it has ticked all the boxes, when deploying the system. Detecting biases in a system is very hard especially because human beings could be the ones who implemented the bias in the first place. Therefore, developing an algorithm to detect bias in another algorithm may not be very relevant in this framework. As the black box has been mentioned several times, the panellist referred to a recent study published by The United Nations Institute for Disarmament Research (UNIDIR) specifically on the black box's understandability and predictability in military AI[1].

4. Attribution, Responsibility and Accountability

Two questions from the audience fell under the broader issue of accountability, from different perspectives.

Involvement of the Private Sector

A question brought up the issue of the private sector which was debated in the discussions, but also the question of accountability and responsibility. The participant noted that governments may hand over AI decision making to private vendors. These vendors then use data analytics to code policy choices and how it works is often unknown to both the government agencies and the public. The participant asked how this could be mitigated. A panellist explained that the private sector involvement is a very important issue in many domestic contexts given that lots of the AI applications, a lot of the technical ingredients including some of the datasets and the computing power, are often administered by private entities and then adopted by the government. From a legal perspective there are at least a few aspects, including rules around attribution of conduct in governments which assume that once governments undertake conduct even of other entities, that is attributable to them. There may be some various peculiar legal rules, however there is also some measure of doubt about whether or not they are sufficiently comprehensive. There is another concern about whether or not governments sufficiently understand and are capable of exercising sufficient knowledge and intent over the use of these techniques and tools once they have been adopted.

Responsibility and Accountability of the User

Another participant raised another accountability concern: the use of AI systems may allow those in power or those using the systems to target people and then say it was not their fault. On the potential to escape responsibility and not being able to hold somebody accountable for an application of AI involved in targeting, a panellist suggested that this is the subset of the questions that arise or have been discussed in detail at the GGE on LAWS but have not yet been resolved satisfactorily. It raises the broader question of what should be demanded of these legal systems, including whether or not parties to arm conflicts ought to commit to

[1] Arthur Holland Michel, 'The Black Box, Unlocked, Predictability and understandability in military AI', UNIDIR, 22 September 2020. Available at: <https://unidir.org/publication/black-box-unlocked>

ensuring that any employment of these technologies may be ascribed to the individual human agents that compose the party. From the panellist's perspective, without that level of commitment, there is a potential responsibility concern: it will be difficult to attribute responsibility for an action, which means that the parties will not be able to have their conduct scrutinised from a legal perspective.

A panellist stressed that in view of many defence organisations, an AI system is not a human and cannot be held legally accountable. Only humans can be held legally accountable. Governments would agree that it always has to be clear who is accountable and that the 'who' is a human, not the machine. It would be unacceptable for both military and civilian applications to simply shrug their shoulders and say that it is the machine that did it. From an ethical perspective, and basic legal principled perspective, accountability is always human and that always has to be absolutely clear. According to the panellist, that is the position of most governments. There is always the notion that there is a chain of command and the commander is responsible for defining the parameters and sending equipment into the battle, knowing what the equipment might do and what consequences it could cause.

Another panellist emphasised that there is a link with the first question on obligations and responsibility. There should be a development of those policies by both the government and the private sector. The question encompasses many aspects: one is about the knowledge, but also the responsibility and probably the liability in developing the system, as well as the responsibility of the user and probably also the liability of the user of these systems. The right hand cannot ignore what the left hand is doing in these systems. From an administrative perspective, someone working in government is responsible for the objects that are created by the people that the government representative hires or that the government representative procures to a certain extent. This is linked with the so-called accountability gap that has been invoked with respect to certain autonomous systems. The panellist argued that more educational efforts should be made to describe all the types of accountability and responsibility that will play into these processes. There are many types of responsibility: State responsibility, responsibility of the individual, criminal liability, civilian court liability, liability for recklessness behaviour, all those could probably come into play; and there is even a way to invoke the responsibility of the industry. In addition, at least for the Council of Europe and the European Court of Human Rights (ECtHR), there is an additional procedural layer on this framework beyond the material dimension: the victim of, for instance, the violation of the right to life has the possibility to file a complaint with the ECtHR. The claim has to be assessed and there needs to be at least a procedure to be launched, which should also be enshrined into the legislation of the State. Therefore, there are many ways to apprehend this responsibility and accountability approach, which needs to be investigated further.

5. Compliance with IHL of AI-Supported Decision Making

A participant from the audience asked what would be needed to create an AI system that would help the user comply with IHL. A panellist answered that it would depend on the application. The panellist suggested that there could be an expert system, a piece of software that helps the user navigate certain cases and the principles of IHL. Such a system may already exist. If it does not, it could be designed. Another panellist said that it would be potentially very useful to be able to use the AI applications and tools in limited respect to be able to increase the quality and amount of information available to decision makers to make these very difficult life and death decisions in armed conflict. This information should be phrased in a way that is easily understood by humans and is clear about what its limitations are. That would potentially be of great value to increase the informational basis on which these very difficult judgements and decisions are made. The last panellist suggested that the key elements are informational intelligence and time, so that the system can process the fact-based information and then leave what would be called the human-based judgement to the human decision makers. This would lead to an improvement of the adherence to IHL at least regarding targeting operations.

In conclusion, the moderator emphasised that it is encouraging to hear a lot of overlap on the key issues and what needs to be looked at. It is clear from these discussions that AI-supported decision raises important issues of legal compliance and responsibility, and more fundamentally as regards the balance of human and machine in warfare. From an ICRC perspective, a human-centred approach is necessary, an approach that would ensure sufficient human control and judgement in decisions that have serious consequences for life, and that would be based on specific IHL rules. The ICRC recognises that there are some challenges, given the tension and the push to accelerate military decision making. Even if the intention is to improve decision making, that acceleration could potentially squeeze out the necessary meaningful human intervention. There is a lot of work to do to establish what are the terms of responsible use, legal compliance, ethical acceptability, human-centred means and practice from both the legal and ethical perspective. This discussion will certainly feed the agenda in the months and years to come.

Panel 4
Military Space Operations: Constraints under International Law and Potential Humanitarian Consequences
Opérations militaires spatiales : contraintes imposées par le droit international et conséquences humanitaires potentielles

INTRODUCTION

Heather Harrison Dinniss
Swedish Defence University

Résumé

Heather Harrison Dinniss est maître de conférences au Centre de droit international de l'Université suédoise de la défense. En tant que modératrice, elle a introduit le thème de ce panel portant sur les opérations spatiales militaires, les contraintes que le droit international impose à ces opérations et les potentielles conséquences humanitaires de ces opérations. Le secteur spatial est utilisé quotidiennement par les États, les forces armées et les civils. Les forces armées en particulier ont un intérêt certain à utiliser l'espace : observation, systèmes d'alerte rapide, suivi météorologique, collecte de renseignements et surveillance. Pendant la guerre du Golfe de 1991 par exemple, les outils d'imagerie et de localisation satellitaires ont apporté un avantage militaire significatif à la coalition dirigée par les États-Unis. Surtout, cette contribution propose des éléments de définition des « opérations militaires spatiales ». Les opérations militaires spatiales incluent des opérations conventionnelles qui s'appuient sur ou dépendent des outils spatiaux (suivi météorologique, reconnaissance, communication et système de commande, etc.). Toutefois, les opérations militaires spatiales vont au-delà de la définition traditionnelle d'une opération militaire en DIH. L'OTAN donne une définition plus large : « une séquence ou des actions coordonnées ayant un objectif défini, de nature militaire et ayant un lien matériel avec l'espace ». Heather Harrison Dinniss propose quatre manières d'établir ce lien :

- *dans l'espace, par exemple, les opérations en orbite de proximité et de rendez-vous spatial. Certaines manœuvres des systèmes satellites russes ou chinois près de systèmes étasuniens entrent dans ce cadre.*
- *depuis l'espace, par exemple, les intercepteurs de missiles balistiques basés dans l'espace ; ils sont actuellement encore au stade de la recherche.*

- *vers l’espace, par exemple les missiles antisatellites en « ascension directe » par énergie cinétique. En 2019 l’Inde est devenu le quatrième pays à lancer un tel missile, après la Russie, les États-Unis et la Chine. Ces opérations entrainent la création de nombreux débris dans l’espace, qui menacent ensuite les autres systèmes spatiaux.*
- *à travers l’espace, par exemple les missiles balistiques transitant par l’espace. Dans ce cas, la qualification d’opération militaire dans l’espace est en débat car ces missiles n’effectuent en général qu’une fraction d’orbite et pas une orbite complète. Cependant, dans la mesure où ces missiles fonctionnent dans l’espace, il existe tout de même un lien.*

Ce panel explore les règles de droit s’appliquant à ces opérations et les processus multilatéraux et bilatéraux visant à assurer la sécurité spatiale, tels que la Conférence du désarmement, le Comité des utilisations pacifiques de l’espace extra-atmosphérique (CUPEEA), les accords d’Artemis, ou les efforts des spécialistes pour clarifier les règles applicables, tels que le manuel de Woomera et le manuel de McGill (MILAMOS).

Welcome to you all from wherever in the world you are joining us. It falls to me to introduce the topic of today’s session, which is military space operations and the constraints that international law places on those operations, as well as the potential humanitarian consequences of military space operations. The importance of space to our modern everyday life is immense. The idea of a ‘day without space’ has been played out in multiple fora, but to give you an idea I have taken a snippet from one of them:[1]

> You wake up and turn on the TV. Your usual shows are not airing. You flip on the radio and learn that the Paris and Tokyo stock markets have closed. Back on TV, CNN is trying to use Skype in an attempt to cover what is happening around the world following a solar superstorm. *[Note that it could equally be massive debris damage]*
>
> In a US bunker, the military has lost contact with armed drones flying over hostile areas in the Middle East. Loss of global communication satellites makes it difficult to send commands and surveillance data to soldiers, ships and aircraft, rendering them vulnerable to attack.
>
> Throughout the day, more challenges arise. First responders don’t have access to their location systems. Delays in ground and air traffic begin to develop. Systems that depend

[1] Pål Brekke ‘A Day without Satellites’ presentation at ION GNSS+, 17 September 2019, cited in: Tracey Cozzens ‘A Day without Satellites’ Taking Position, *GPS World*, October 2019, 8. (Editorial note added), at: <https://editions.mydigitalpublication.com/publication/?m=59713&i=625808&view=articleBrowser&article_id=3501731&ver=html5>

> on GPS time stamps – ATMs, power grids, computer-data and cell-phone networks – begin to fail, and the cloud becomes unstable. The internet soon collapses.

One can start to see the importance of space security not just for States and armed forces, but for every one of us and for civilian life on Earth as a whole.

Military uses of space have been there since the beginnings of space exploration. Early military uses included reconnaissance and observation as well as early warning systems to track any launch of intercontinental ballistic missiles (ICBM) (from the US side). Uses then expanded to weather monitoring, intelligence gathering and surveillance. GPS, the US global navigation satellite system, was originally planned for military use but was rapidly overtaken by civilian use of the system, causing problems with security, reliance and predictability.

The Gulf War of 1991 (Operation Desert Storm) is widely acknowledged as the first space war. Not because spacecraft engaged in battle, but because of the significant military advantage that harnessing satellite imaging power, signals intelligence, and GPS gave the coalition forces. Military space activities also played a large part in the success of the coalition war-fighting effort in the 2003 Iraq War (Operation Iraqi Freedom). For example, all secure communication between coalition partners went through space, space systems detected 26 rocket launches from Iraq, and all predator unmanned aerial vehicle (UAV) data transmissions went via satellite.

We know that the armed forces have an inherent interest in utilising space, either in support of conventional armed conflicts in already established war-fighting domains, or by defending space-based assets. What then is a military space operation? Unfortunately, there is no set definition of the term. It is clear that it must be wider than merely transposing the definition usually applied when discussing military operations under International Humanitarian Law into outer space.[2] Thus it has been suggested that adapting the broader NATO definition of operations to the space environment would be a good way forward: 'a sequence or coordinated actions with a defined purpose that is military in nature and which have a material nexus to outer space'.[3] This seems a sensible approach and one that I will adopt here. In addition to

[2] The ICRC commentary to Article 3 of Additional Protocol I defines 'military operations' as 'the movements, manoeuvres and actions of any sort, carried out by the military with a view to combat': Y Sandoz, et al, (ed.) *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (ICRC, 1987), para. 152.

[3] Kubo Mačák, 'Military Space Operations', ECIL Working Paper 2020/2, forthcoming, in: Sergey Sayapin (ed.), *An Introduction to International Conflict and Security Law* (T.M.C. Asser Press, 2020).

the use of space assets to support or enable conventional military activities (as previously discussed), that nexus might be established in one of four ways:

- military operations ***in*** space, e.g. on-orbit proximity operations;
- military operations ***from*** space, e.g. space-based ballistic missile interceptors:
- military operations ***to*** space, e.g. direct-ascent kinetic anti-satellite weapon (ASAT);
- military operations ***though*** space, e.g. ballistic missiles transiting space.

IN – On Orbit Proximity Operations. Rendezvous and proximity operations (RPO) are not new to space activities. They have been a part of human spaceflight since the very beginning of the Space Age. For example, they are what enabled the Apollo astronauts to land on the Moon; they were used to construct the International Space Station. Rendezvous and docking are necessary to transfer astronauts to and from space stations and space labs. Dozens of such RPO activities have been conducted by several spacefaring countries over the last sixty years. What is of concern however, is the increase in both Russian and Chinese satellite systems manoeuvring closely to, for example, classified US spy satellites and other assets in both low earth orbit (LEO) and geostationary (GEO) satellites– the lack of transparency has caused great concern among the international community and (obviously) the US military. While some see this as Russia using so-called 'inspector' satellites, trying to see what the capabilities of the US systems are, others are more concerned that this represents a continuance of the old Soviet Cold War co-orbital ASAT capability.[4]

FROM Space-Based Ballistic Missile Interceptors. Although not yet in development, they are actively being researched as ICBM missile interceptors. Imagine a satellite with a payload of approximately 10 rockets which would be launched to intercept an incoming missile, much like the land-based systems now (or Israel's Iron Dome) but from space.

TO – Kinetic ASAT. In 2019 India became the fourth State (following earlier tests by Russia, the US and China) to launch a direct-ascent kinetic anti-satellite missile. Such operations from Earth to space might be kinetic (in the case of these direct-ascent ASAT missiles) or virtual – hacking a satellite is easier and causes far less debris. The debris created from the destruction of FY-1C (by China in 2007) forced the Terra spacecraft to permanently change its orbit,

[4] In 2014, Russia launched the Olymp-K (or Luch) spacecraft into GEO orbit, though the international community has yet to accurately determine the name of the craft and verify its purported mission. Since its launch, Olymp-K has occupied 14 different positions in the highly congested GEO belt and, most notably, placed itself in a narrow window between two satellites belonging to Intelsat, a private satellite company. See Brian Weeden, 'Dancing in the Dark redux: Recent Russian Rendezvous and Proximity Operations in Space', in: *The Space Review*, 5 October 2015, 2, available at: <https://www.thespacereview.com/article/2839/2>.

and the destruction of Microsat-R (by India in 2019) placed the International Space Station in severe danger.

THROUGH – Ballistic Missiles Transiting Space. There is debate as to whether we would consider this a military space operation in *all* senses, as missiles traditionally do not complete a full orbit, but only a fraction of one. However, to the extent that the missile operates in space, they certainly have a nexus to space.

One must also remember that military space operations include the use of space assets (weather, reconnaissance, communications, etc.) to support or enable conventional military activities.

Having established what military space operations might entail, and the importance of maintaining space security for all of us, we now come to the laws governing these operations, and the processes that are ongoing in the international community – both multilaterally and bilaterally – to ensure that security. Whether that be in the Conference on Disarmament process for the 'prevention of an arms race in outer space', through the Committee on the Peaceful Uses of Outer Space (COPOUS), or the Artemis accords. And also those academic processes to clarify the applicable rules such as the Woomera Manual and McGill Manual. To do that, it is my great pleasure to introduce our panel of speakers. [*The chair then introduced the speakers and their topics*]

CONSTRAINTS RELATED TO THE USE OF WEAPONS IN OUTER SPACE UNDER IHL
LES CONTRAINTES POSÉES PAR LE DIH À L'UTILISATION D'ARMEMENTS DANS L'ESPACE EXTRA-ATMOSPHÉRIQUE

Wen Zhou

Legal Adviser, Arms and Conduct of Hostilities Unit, ICRC Geneva

Résumé

Wen Zhou est conseillère juridique au sein de l'unité « armes et conduite des hostilités » du CICR. Cette contribution examine les contraintes juridiques applicables à l'utilisation d'armements dans l'espace extra-atmosphérique.

La première partie de cette contribution montre comment le droit international et en particulier le DIH s'appliquent à l'utilisation des armes dans l'espace. L'utilisation d'armes dans l'espace est réglementée par le Traité de l'espace de 1967, la Charte des Nations unies et les règles de DIH régissant les moyens et méthodes de guerre. Le Traité de l'espace interdit d'y placer des armes de destruction massive. Son article III confirme l'applicabilité du droit international, et donc du DIH, à l'utilisation de l'espace en général. Toute utilisation de méthodes ou moyens de guerre dans, depuis, vers ou à travers l'espace doit donc respecter le DIH, notamment les principes de distinction, de proportionnalité et de précautions qu'une attaque doit respecter. Le DIH protège les biens indispensables à la survie de la population civile. Il interdit également les armes indiscriminées par nature ainsi que de nombreux types spécifiques d'armes. Ces interdictions ne sont en aucun cas limitées au domaine terrestre. La Cour internationale de justice a rappelé que les principes et règles établis du DIH s'appliquent «à toutes les formes de guerre et à tous les types d'armes», y compris «celles du passé, celles du présent et celles du futur»[1]. Enfin, l'article 49(3) du premier Protocole additionnel aux Conventions de Genève indique que ces règles sont conçues pour s'appliquer à tous les types de guerre susceptibles d'affecter les civils sur terre[2].

La deuxième partie aborde trois des nombreuses questions en débat concernant l'interprétation et l'application du DIH. Un risque majeur est de perturber, endommager, détruire ou mettre hors

[1] CIJ, 'Licéité de la menace ou de l'emploi d'armes nucléaires', Avis consultatif, 8 juillet 1996, para.86, p.37.

[2] Article 49(3) du Premier protocole additionnel : « Les dispositions de la présente Section s'appliquent à toute opération terrestre, aérienne ou navale pouvant affecter, sur terre, la population civile, les personnes civiles et les biens de caractère civil. Elles s'appliquent en outre à toutes les attaques navales ou aériennes dirigées contre des objectifs sur terre, mais n'affectent pas autrement les règles du droit international applicable dans les conflits armés sur mer ou dans les airs. »

service des objets spatiaux civils ou à double usage dont dépendent des services civils essentiels sur terre. Premièrement, s'il ne fait aucun doute que les opérations cinétiques visant un objet spatial constituent une attaque au sens du DIH, il n'existe pas de consensus concernant les opérations non cinétiques. Pour le CICR, le fait de rendre dysfonctionnel un objet spatial sans l'endommager physiquement constitue une attaque au sens du DIH. La position inverse se traduit par une incertitude juridique. Deuxièmement, les satellites peuvent avoir un double usage, civil et militaire. Les dommages directs et indirects causés aux civils et aux biens de caractère civils doivent être anticipés pour évaluer la licéité d'une attaque sur un tel satellite. Troisièmement les débris créés par une attaque restent présents dans l'espace pendant des décennies et risquent d'endommager d'autres satellites soutenant des activités et des services civils. Cela devrait être pris en considération.

Affirmer l'applicabilité du DIH à l'espace extra-atmosphérique ne revient pas à encourager ou à légitimer la militarisation de l'espace. Le CICR recommande que les futurs accords multilatéraux reconnaissent les conséquences humanitaires potentiellement graves que l'utilisation d'armes dans l'espace extra-atmosphérique pourrait entraîner pour les civils sur terre et la protection offerte par les règles du DIH, qui limitent le choix des moyens et méthodes de guerre des belligérants, y compris dans l'espace extra-atmosphérique.

Introduction

For several decades, military uses of space objects have been an integral part of warfare, for instance the use of satellite imagery to support the identification of enemy targets, the use of satellite communication systems for command and control, and more recently for remotely controlled means of warfare. The weaponisation/militarisation of outer space would increase the likelihood of hostilities in outer space, with potentially significant humanitarian consequences for civilians on Earth.

It is clear that the use of weapons in outer space – be it through kinetic or non-kinetic means, using space – and/or ground-based weapon systems – would directly or incidentally disrupt, damage, destroy or disable civilian or dual-use space objects. And safety-critical civilian activities and essential civilian services increasingly depend on these space objects.

Outline

The vulnerability of space-based systems that serve essential civilian activity on Earth presents significant challenges for complying with the IHL rules of distinction, proportionality and precautions in attack.

I will first discuss the applicability of International Humanitarian Law to space warfare, before turning to some selected challenges (by no means a comprehensive list), in particular the implications of non-kinetic operations under IHL, challenges surrounding satellites, and the issue of debris.

IHL Applicability

The use of weapons in outer space would not occur in a legal vacuum. It is constrained by existing law, notably the Outer Space Treaty, the UN Charter and IHL rules governing means and methods of warfare.

While the 1967 Outer Space Treaty clearly prohibits the placement of weapons of mass destruction in orbit, it does not expressly apply such prohibition to other weapons.

What is certain is that any hostile use of outer space in armed conflict – that is, any use of means and methods of warfare in, from, to or through outer space – must comply with IHL, in particular its rules of distinction, proportionality and precautions in attack. Furthermore, attacking, destroying, removing or rendering useless objects indispensable to the survival of the civilian population is prohibited. IHL notably prohibits weapons which are by nature indiscriminate, as well as a number of other specific types of weapons. These prohibitions are not limited to the terrestrial domains.

The applicability of IHL is confirmed by Article III of the Outer Space Treaty, which states that international law applies to the use of outer space, and IHL is obviously part of international law.

Furthermore, the International Court of Justice has recalled that the established principles and rules of humanitarian law applicable in armed conflict apply 'to all forms of warfare and to all kinds of weapons', including 'those of the past, those of the present and those of the future'.[3]

In terms of treaty law, the First Additional Protocol to the Geneva Conventions, which contains most rules on the conduct of hostilities applies 'to all cases of declared war or of any other armed conflict which may arise between two or more of the High Contracting Parties'.

[3] ICJ, 'Threat or Use of Nuclear Weapons', Advisory Opinion, 8 July 1996, para. 86, p. 37.

In establishing the scope of application of the Protocol's rules on the conduct of hostilities, Article 49(3) shows that they were meant to apply to all types of warfare which may affect civilians on land.[4]

Challenges that Outer Space Warfare Raises for the Interpretation and Application of IHL

Kinetic and Non-Kinetic Operations

There is no doubt that a kinetic operation against a space object constitutes an attack under IHL. However, a space object can also be disabled (rendered dysfunctional) without being physically damaged, for example by directed energy/laser weapons or a cyber-attack. In the ICRC's view, such non-kinetic operations constitute attacks under IHL, and therefore are governed by the above-mentioned rules.

Let me stress that the notion of attack under IHL is different from and should not be confused with the notion of 'armed attack' under Article 51 of the UN Charter.[5]

But the ongoing discussions about cyber warfare have demonstrated that there is no consensus on whether non-kinetic means can constitute an attack. In those discussions, some have expressed the view that an operation leading to a loss of functionality of an object or system without physically damaging it would not constitute an attack. Otherwise, non-kinetic attacks on space systems would not be governed by IHL, resulting in fewer and less precise rules governing such operations.

Satellites

Let me now turn to some challenges stemming from the particularities of satellites.

IHL forbids targeting civilian objects in outer space. However, civilian satellites or some of their hosted payloads may also be used by the armed forces and are hence of a 'dual-use' na-

[4] Article 49(3) of the First Additional Protocol reads: 'The provisions of this Section apply to any land, air or sea warfare which may affect the civilian population, individual civilians or civilian objects on land. They further apply to all attacks from the sea or from the air against objectives on land but do not otherwise affect the rules of international law applicable in armed conflict at sea or in the air". Given the increasing reliance on space assets by commercial and other civilian interests, there is little doubt that outer space warfare risks affecting civilians on land, through the disruption of services that space assets provide to civilians. Furthermore, outer space warfare might be waged at least partly from infrastructure located on land and/or against the ground component of the space system.

[5] Article 49 of the First Additional Protocol defines attacks under IHL as 'acts of violence against the adversary, whether in offence or in defence'.

ture. They may become military objectives, provided that their use for military purpose is such that they fulfil the definition under Article 52(2) of the 1977 First Additional Protocol. But, disabling the civilian functions of such satellites could disrupt large segments of modern-day societies, especially if they also support safety-critical civilian activities and essential civilian services on Earth. The foreseeable direct and knock-on (reverberating) incidental civilian harm and damage to civilian objects expected from an attack against a dual-use satellite must be considered when assessing the lawfulness of such attack under the IHL rules governing the conduct of hostilities.

Debris

Another issue of concern is the risk posed by space debris. Debris can be created by a host of space activities. A kinetic attack on a satellite, for example, risks causing far more debris than other space activities. Debris can travel in the orbits in which they are produced for decades or more. Given the speed at which they travel, debris risk damaging other satellites supporting civilian activities and services. This would have to be considered and may limit the choice of means and methods of warfare in outer space.

Conclusion

Let me again stress that asserting that IHL applies to outer space warfare is not an encouragement to weaponise outer space. It should not be understood either as legitimising outer space warfare in any way. Resort to force by States, in space as anywhere else, always remains governed by the UN Charter and the applicable international law on the use of force (*jus ad bellum*). Asserting that IHL applies only reaffirms that warfare in outer space would not occur in a legal vacuum.

The ICRC is concerned by the potentially high human cost of the use of weapons in outer space. It therefore submitted a working paper to the Group of Government Experts (GGE) on further practical measures for the prevention of an arms race in outer space in March last year.[6] The ICRC recommends that future multilateral processes acknowledge the potentially significant humanitarian consequences that the use of weapons in outer space could entail for civilians on Earth and the protection afforded by the IHL rules that restrict the belligerents' choice of means and methods of warfare, including in outer space.

[6] The GGE was established pursuant to Resolution A/RES/72/250 adopted by the United Nations General Assembly, 24 December 2017. The ICRC Working Paper can be found at: <https://undocs.org/GE-PAROS/2019/WP.1>.

BEYOND THE PEACEFUL USE OF OUTER SPACE: POTENTIAL CONFLICTS?
AU-DELÀ DE L'UTILISATION PACIFIQUE DE L'ESPACE EXTRA-ATMOSPHÉRIQUE : DES CONFLITS POTENTIELS ?

Mickael Dupenloup
French Ministry of Defence

Summary

Mickael Dupenloup is Legal Adviser with the French Ministry of the Armed Forces. The outer space environment became a strategic sector for States. It is increasingly the subject of economic, technological and military competition. The uncertainties and new ambitions of States are a factor of risk, tension and conflict escalation, outer space becoming a place of possible confrontation. In this context, this paper explores the limitations of international law's rules and instruments applicable to military space activities, as well as existing alternatives to regulate and limit the risks.

The first section analyses the international instruments and rules applicable in outer space. The main legal principles are broad enough to be consensual. As such, they are guiding principles more than constraints. The dedicated body of law is based on a 1963 United Nations General Assembly (UNGA) Resolution; the 1967 Outer Space Treaty is the core legal instrument, later complemented with other agreements, conventions and UNGA Resolutions. Nonetheless, outer space remains a poorly regulated environment. The main principle is that outer space 'shall be free for exploration and use by all States'. The absence of a strict legal definition generates a legal latitude for States. For instance, the absence of definition of the line between the Earth's atmosphere and outer space means that space objects can circulate freely close to the Earth. The 'free use' principle is constrained by the obligation to respect the freedom and 'interest of all States' as well as international law, but the divergent interpretations could translate into open conflicts between States. In addition, although international law provides that outer space is used for peaceful purposes, military space activities are not fully prohibited. The Outer Space Treaty allows the militarisation of outer space, provided that weapons of mass destruction are not deployed, that the Moon, celestial bodies and their orbits are demilitarised and that any use of force complies with the UN Charter. The available technology makes it possible for States to openly finance anti-satellite technologies under cover of civilian objectives. In comparison, being more regulated, the weapons of mass destruction appear as a lower threat in space.

The second section sets out different initiatives and attempts to compensate for these limitations. Military space activities are discussed in two international fora: the UN Committee on the

Peaceful Uses of Outer Space (COPUOS) and the Conference on Disarmament (CD), with an ad hoc committee in charge of space arms control issues (the PAROS Committee). However, since the end of the Cold War, States have failed to agree on new binding legal instruments. In the absence of consensus between States, national or international agencies, experts and private companies have been attempting to clarify the applicable law. This endeavour supported by certain space powers, largely excludes the poorest States. Binding sub-State contractual relations, memorandums of understanding or agreements exist, but often remain flexible, referring to obligations of means rather than obligations of result. The Artemis Accords, recently signed by nine States, are an illustration. Other texts of a blurred nature have emerged: various codes of conduct since the early 2000s and more recently, Standard Procedures and Practices (SPP) developed by ad hoc groups, such as the Inter Agency Space Debris Coordination Committee, the Consortium for Execution of Rendezvous and Servicing Operations (CONFERS) and the Space Traffic Management Workshops. These initiatives are part of a bottom-up normative approach contributing to a flexible or soft law. Lastly, the space community has been developing doctrinal work. Two projects of manuals on international law applicable to military space operations were launched in 2015 and 2018: the Manual on International Law Applicable to Military Uses of Outer Space (MILAMOS) project and the Woomera Manual project. These handbooks are intended to identify and clarify the rules of international law applicable to the field, to inform the operational staff and contribute to the stability of international relations. Once published, these manuals, while not binding, are meant to become references for experts and practitioners. They could also lead some States to express their interpretation of the rules applicable to military space activities.

International legal framework governing military space operations: Need for more clarity, better compliance or more regulation? (Outer Space Treaty (WMD), UN Charter, IHL, the Woomera and Milamos projects)

Mesdames, Messieurs,

Je suis honoré de participer pour la première fois à cette manifestation prestigieuse qui me conduit au cœur de l'un des enjeux commerciaux, industriels et géostratégiques actuels : l'utilisation de l'espace extra-atmosphérique.

Comme beaucoup d'autres secteurs marqués par la technologie, l'espace connaît une révolution qui bouscule les équilibres en place : les ruptures technologiques et d'usage ainsi que les cycles accélérés des innovations font évoluer rapidement les critères qui fondent la puissance spatiale. Enjeux traditionnels de rivalité entre États, l'accès et l'utilisation des espaces stratégiques communs ou partagés, tels que l'espace extra-atmosphérique, font l'objet d'une compétition, d'abord économique et technologique puis militaire, dont l'intensité croît. Cette compétition est également caractérisée par l'augmentation rapide du nombre des acteurs, puissances établies ou émergentes, étatiques ou non étatiques, comme de leurs moyens d'action.

L'émergence de cette compétition nouvelle et de cette multipolarité se traduit par une remise en cause de certaines règles et par un contournement des institutions internationales qui encadrent juridiquement et régulent les activités spatiales depuis plus de cinquante ans.

L'incertitude, l'anxiété ou au contraire les ambitions nouvelles générées par cette situation mouvante sont en soi facteurs de risque, de tension et, le cas échéant, d'escalade. Longtemps perçu comme un sanctuaire, l'espace est désormais appréhendé par les principales puissances spatiales comme un espace de confrontation possible où des stratégies alternatives, en deçà ou au-delà du seuil des conflits armés, pourraient être développées.

Dans ce contexte, la question de la pertinence et de la suffisance du droit international des activités spatiales, en particulier des stipulations régissant les activités spatiales militaires, est régulièrement soulevée.

*

Enonçant des principes juridiques suffisamment généraux pour être consensuels, les instruments internationaux applicables aux activités spatiales ont été élaborés davantage pour guider que pour contraindre les activités spatiales des États. Ainsi, l'espace est un milieu peu régulé car régi, à titre principal, par un principe de liberté d'exploration et d'utilisation.

Historiquement, les activités spatiales ont connu un développement extrêmement rapide. Douze années séparent le lancement du premier satellite artificiel de la Terre, Spoutnik, de l'alunissage de MM. Armstrong et Aldrin. Entre temps furent également lancées des sondes spatiales et placés sur les orbites terrestres les premiers satellites d'observation de la Terre et de télécommunication. En parallèle de ces avancées technologiques, un corpus juridique spécifique a été élaboré pour encadrer et accompagner ces activités.

Ce corpus trouve son origine dans une résolution de l'Assemblée générale des Nations unies. Adoptée à l'unanimité des États en 1963, cette résolution énonce les principes juridiques régissant les activités des États en matière d'exploration et d'utilisation de l'espace extra-atmosphérique. Ces principes ont été réaffirmés dans un instrument juridiquement contraignant qui aujourd'hui encore constitue le fondement du droit international des activités spatiales : le Traité de l'espace[1].

Ce traité, « magna carta » du droit de l'espace, a, par la suite, été complété par d'autres instruments[2], en particulier l'accord sur le sauvetage des astronautes, le retour des astronautes et le retour des objets lancés dans l'espace extra-atmosphérique du 22 avril 1968, la Convention sur la responsabilité internationale pour les dommages causés par les objets spatiaux du 29 mars 1972, et la Convention sur l'immatriculation des objets lancés dans l'espace extra-atmosphérique du 14 janvier 1975. Il a également été complété par des Résolutions de l'Assemblée générale des Nations unies qui ont précisé les règles applicables à certaines activités spatiales telles que la télédétection[3].

Ce corpus juridique est organisé autour du principe selon lequel l'exploration et l'utilisation de l'espace extra-atmosphérique, déclarées « apanage de l'humanité toute entière[4] », s'effectuent librement. En application de ce principe, aucun État ne peut se voir imposer des restrictions ou des conditions par un autre État pour accéder à l'espace, l'explorer et l'utiliser conformément au droit international. L'espace extra-atmosphérique est libre d'accès et d'usage à la différence de l'espace aérien où s'exerce la souveraineté complète et exclusive de l'État sous-jacent.

La latitude dont disposent les États en matière spatiale résulte également de l'absence conjointe de délimitation de *l'espace* et de définition de ce que recouvre son *utilisation*.

L'espace, en particulier son seuil, n'est pas délimité. Faute de certitude scientifique et de consensus politique, la ligne de partage qui sépare l'atmosphère terrestre de l'espace n'a pas été définie. Dans ces conditions, l'accès et l'utilisation des orbites terrestres les plus basses sont demeurées libres, laissant aux États la possibilité d'y mener les activités spatiales de leur choix, notamment militaires, et d'y circuler librement.

[1] Traité du 27 janvier 1967 sur les principes régissant les activités des États en matière d'exploration et d'utilisation de l'espace extra-atmosphérique, y compris la Lune et les autres corps célestes de l'Espace.

[2] Voir <http://www.unoosa.org/pdf/publications/STSPACE11F.pdf> pour les traités et principes des Nations unies relatifs à l'espace extra-atmosphérique.

[3] Dont la résolution 41/65 du 3 décembre 1986 portant principes sur la télédétection ; la résolution 47/68 du 14 décembre 1992 portant principes relatifs à l'utilisation de sources d'énergie nucléaires dans l'espace ; la résolution 51/122 du 13 décembre 1996 portant déclaration sur la coopération internationale en matière d'exploration et d'utilisation de l'espace au profit et dans l'intérêt de tous les États, compte tenu en particulier des besoins des pays en développement ; la résolution 59/115 du 10 décembre 2004 relative à l'application de la notion d' « État de lancement » ; la résolution 62/101 du 17 décembre 2007 portant recommandations visant à renforcer la pratique des États et des organisations internationales inter-gouvernementales concernant l'immatriculation des objets spatiaux.

[4] Traité de l'espace, article I.

Tout en encadrant la pratique des États, le régime juridique de l'espace garantit de manière générique sa libre *utilisation*. Dans cette perspective, la recherche scientifique est libre et aucune application spatiale, civile comme militaire, n'est *a priori* interdite. Un tel régime favorise par essence l'initiative et la diversification des acteurs comme de l'offre de services spatiaux.

Cette libre utilisation n'est toutefois pas absolue. Elle est limitée par l'obligation de respecter, d'une part, la liberté et « l'intérêt de tous les États[5] » et, d'autre part, le droit international.

Ainsi, utiliser ou occuper l'espace ne saurait fonder un quelconque droit souverain au profit de l'État qui exerce cette liberté. Sur ce fondement, un égal accès et un partage équitable des positions orbitales et des fréquences radioélectriques sont assurés par l'Union internationale des télécommunications. Ce principe de non-appropriation est aujourd'hui contesté : des États considèrent qu'il ne s'applique pas aux minerais et autres ressources qui pourraient être extraits des corps célestes. Il est essentiel d'éviter que ces actuelles divergences d'interprétation ne deviennent à terme une source de conflits ouverts entre États.

Si l'espace doit être utilisé à des fins pacifiques, conformément au droit international, toute activité spatiale militaire n'est pas pour autant prohibée. Le Traité de l'espace permet la militarisation des orbites terrestres – c'est-à-dire le déploiement de satellites militaires –, voire leur arsenalisation – sous réserve que des armes de destruction massive n'y soient pas déployées – ainsi que le recours à la force, dans le strict cadre de la Charte des Nations unies. La Lune, les corps célestes et leurs orbites sont quant à eux totalement démilitarisés.

À l'heure actuelle, cette problématique de la militarisation et de l'arsenalisation de l'espace se pose en des termes renouvelés. Les progrès des techniques de rendez-vous dans l'espace, les capacités de robotique et de propulsion électrique permettent de réparer, de ravitailler en carburant voire de désorbiter des engins spatiaux. Sous couvert d'objectifs civils, des États peuvent donc financer ouvertement des technologies potentiellement antisatellites. Celles-ci permettraient la mise en service d'outils dont les actions sont plus difficiles à détecter, à suivre, à attribuer et, le cas échéant, à contrer que des actions antisatellites plus classiques telles que l'emploi de missiles, de lasers ou de brouilleurs.

Par contraste, la menace que représentent les armes de destruction massives dans l'espace paraît aujourd'hui davantage circonscrite puisque des instruments juridiques soit interdisent les essais et le placement d'armes nucléaires dans l'espace extra-atmosphérique soit limitent la possibilité pour les États de lancer de telles armes dans l'espace comme les vecteurs sus-

[5] *Ibid.*

ceptibles de les emporter : des zones exemptes d'armes nucléaires existent[6] ; des mesures de lutte contre la prolifération des armes de destruction massives [7] et des missiles balistiques[8] ont été mises en place.

*

De la même façon qu'à l'époque où le droit international des activités spatiales a été élaboré – celle de la guerre froide –, l'enjeu actuel du respect ou du développement des normes applicables à l'espace extra-atmosphérique vise à préserver la stabilité internationale. Ceci suppose l'obtention d'un consensus interétatique au sein des institutions internationales concernées. Or, à la lumière du contexte géostratégique actuel, force est de constater qu'un tel consensus, permettant soit la réouverture des traités existants soit la négociation de nouveaux instruments juridiques ou de nouveaux principes directeurs, fait défaut. Il en résulte un déplacement du lieu d'élaboration des règles applicables aux activités spatiales y compris militaires et l'émergence de nouveaux acteurs en la matière.

Les initiatives visant à encadrer les activités spatiales militaires sont débattues au sein de deux enceintes internationales : d'une part, le Comité onusien en charge de l'utilisation pacifique de l'espace extra-atmosphérique (CUPEAA) ; d'autre part, la Conférence du désarmement (CD) de Genève, dotée depuis 1985 d'un comité *ad hoc* en charge des questions de maîtrise des armements spatiaux (le comité PAROS). L'obtention d'un consensus sur les sujets spatiaux s'avérant particulièrement délicate au CUPEEA et à la CD, aucun instrument juridique multilatéral contraignant concernant les activités spatiales militaires n'a vu le jour depuis la fin de la guerre froide. À la CD, le projet sino-russe de traité visant à interdire le déploiement d'armes dans l'espace ainsi que l'utilisation de la force ou de la menace de la force dans l'espace achoppe depuis 2001, faute de définition agréée des armes spatiales et d'interdiction des armes antisatellites. Au CUPEEA, non seulement l'élaboration de résolutions ou de déclarations est privilégiée à la rédaction de projets de conventions, mais la politisation des débats retarde la finalisation des travaux : l'adoption d'un catalogue partiel de lignes directrices non

[6] Notamment en Amérique latine-Caraïbes, en Asie centrale et dans le Pacifique Sud. Voir : <http://www.un.org/press/fr/2001/CD236.doc.htm>.

[7] Créé en 1987, le Régime de contrôle de la technologie des missiles (MTCR) vise à freiner la prolifération des missiles, des véhicules aériens non pilotés et la technologie connexe pour les vecteurs d'une charge utile de 500 kilogrammes sur une distance d'au moins 300 kilomètres, ainsi que les vecteurs d'armes de destruction massive (ADM). Pour cela, il coordonne les efforts nationaux en matière de licences d'exportation. Il réunit 35 États membres dont la France. Voir : <http://mtcr.info/?lang=fr>.

[8] Créé en 2002, le Code de conduite de La Haye contre la prolifération des missiles balistiques (HCoC) établit des mesures de confiance et de transparence. Il regroupe 137 États signataires dont la France. Voir : <https://onu-vienne.delegfrance.org/Code-de-Conduite-de-La-Haye>.

contraignantes visant à soutenir la viabilité à long terme des activités spatiales s'est ainsi étalée de 2011 à 2019.

Dans ce contexte, à côté des États, de nouveaux acteurs émergent. Des agences nationales ou internationales, des experts et des entreprises privées interviennent de plus en plus pour clarifier le droit applicable et pallier le manque de règlementation internationale au moyen d'accords *ad hoc*, de règles de conduite, de procédures et de pratiques standardisées ou de manuels. Ce phénomène, parfois encouragé ou sponsorisé par certaines puissances spatiales, exclut cependant très largement les États les plus pauvres qui ne participent que très peu à ces activités.

A côté des sources classiques de droit international, se sont développées d'autres sources de droit. Des relations contractuelles infra-étatiques se sont établies de façon systématique au niveau des agences nationales. Des mémorandums d'accord ou d'entente sont notamment très utilisés. Les accords Artemis, signés cette semaine entre l'Australie, la Canada, les Émirats arabes unis, les États-Unis, l'Italie, le Japon, le Luxembourg, le Royaume-Uni, qui définissent les principes régissant la coopération dans le domaine de l'exploration et de l'utilisation de la Lune, de Mars, des comètes et des astéroïdes du système solaire en fournissent une illustration topique. Ils sont obligatoires en tant qu'accords mais l'esprit qui les conduit est le plus souvent assez souple, leurs stipulations faisant souvent référence à des obligations de moyens plutôt qu'à des obligations de résultat.

Outre ces mémorandums, le droit de l'espace connaît également le développement de textes à la nature peu précise. Ainsi, divers codes de bonne conduite ont été élaborés à compter du début des années 2000 pour lutter contre la prolifération des missiles balistiques, limiter les débris spatiaux ou favoriser la sécurité des activités menées dans l'espace extra-atmosphérique. Désormais, les acteurs, publics ou privés, semblent privilégier la promotion de procédures et pratiques standardisées (*Standard Procedures and Practices*) ou de normes de comportement. Dans cette perspective, divers groupes *ad hoc* ont vu le jour tels que le Comité inter-agence sur les débris spatiaux (*Inter Agency Space Debris Committee*), le Consortium établi pour définir des standards pour les activités de rendez-vous et de services en orbite (*Consortium For Execution of Rendezvous and Servicing Operations*) ou encore les ateliers de réflexion portant sur la gestion du trafic spatial (*Space Traffic Management Workshops*). Ces initiatives participent d'une approche normative ascendante, qui érige en normes internationales des lois et politiques nationales, ceci afin d'une part de mieux caractériser ce qui constituent des comportements irresponsables ou inamicaux dans l'espace et d'autre part de déterminer les réponses à y apporter conformément au droit international. Elles contribuent indéniablement à l'émergence d'un droit souple ou *soft law,* représentatif de la pratique des principaux acteurs, agences gouvernementales et industriels du secteur.

Enfin, les travaux doctrinaux se développent au sein de la communauté spatiale sur le modèle de ceux observés dans les autres domaines de conflictualité militaire. Sur le modèle des manuels de San Remo, d'Harvard et de Tallinn (1.0 et 2.0), respectivement consacrés au droit de la guerre maritime, au droit de la guerre aérienne et au droit des opérations cybernétiques, ont été lancés en 2015 et en 2018 deux projets de manuels consacrés au droit international applicable aux opérations spatiales militaires : respectivement le projet MILAMOS (*Manual on International Law Applicable to Military uses of Outer Space*) et le projet Woomera. Ces manuels sont censés recenser et clarifier les règles de droit international applicables au domaine, de façon à ce que cet état des lieux éclaire les opérationnels du secteur et contribue à la stabilité des relations internationales. Une fois publiés, ces manuels, sans pour autant être contraignants, auront vocation à devenir des ouvrages de référence, largement commentés et étudiés par la communauté des spécialistes et des praticiens (professeurs d'universités, étudiants, juristes des armées, décideurs politiques, etc.). Ils pourraient également amener certains États à exprimer leur interprétation des règles de droit applicables aux activités spatiales militaires. A la lumière de ce qui est observé pour le Manuel de Tallinn, une génération de praticiens et d'acteurs du secteur spatial pourrait à terme se former au droit international des activités spatiales à travers la lecture des règles de ces manuels et de leurs commentaires.

PROTECTING HUMANS ON EARTH FROM WAR IN SPACE[1]
PROTÉGER LES HUMAINS SUR TERRE CONTRE LA GUERRE DANS L'ESPACE

Jessica West
Canadian Peace Research Institute, Project Ploughshares

Résumé

Jessica West est chercheuse à Project Ploughshares, l'institut de recherche sur la paix du Conseil canadien des Églises, et responsable d'un projet d'index de sécurité spatiale. Sa contribution précise les risques liés aux développements des activités militaires et à l'arsenalisation dans l'espace, et encourage à travailler à des mesures de protection. Elle commence par rappeler l'importance des systèmes spatiaux, composés d'un ou plusieurs satellites en orbite, pour de nombreux services essentiels sur Terre : téléphones portables, trafic aérien, alertes aux catastrophes, production agricole, banque électronique, transport maritime, réseaux électriques, internet. La plupart des systèmes qui permettent le fonctionnement de services civils essentiels ont un usage multiple, en particulier un usage militaire, et certains satellites d'abord militaires sont également essentiels aux activités civiles. C'est le cas par exemple du système de localisation aux États-Unis (Global Positioning Sytem, GPS).

L'espace extra-atmosphérique se caractérise dès lors par l'absence de séparation entre zone militaire et zone civile. L'espace a également d'autres avantages et vulnérabilités spécifiques. Il est encombré de plus de 3.000 satellites en orbite. C'est un milieu fragile et peu protégé. En cas de collision, les débris risquent de créer des dommages en cascade. La destruction d'objets dans l'espace, y compris lors de tests antisatellites, est donc une source de risque supplémentaire à long terme. Les satellites permettent de surveiller les conditions météorologiques ou d'assurer la communication et la géolocalisation en cas d'urgence humanitaire. Ces besoins spécifiques sont reconnus notamment dans la Charte internationale « Espace et Catastrophes Majeures » et la Charte de connectivité de crise des Nations unies. Le commandement et le contrôle des armes nucléaires reposent également sur des équipements militaires dans l'espace. L'endommagement de ces équipements pourraient ainsi déclencher une frappe nucléaire accidentelle, si pas intentionnelle.

[1] This written contribution is based on Jessica West's presentation at the present Colloquium and was published in The Ploughshares Monitor, Volume 41 Issue 4 Winter 2020. Available at: <https://ploughshares.ca/pl_publications/protecting-humans-on-earth-from-war-in-space/?mc_phishing_protection_id=28047-bv76qo2du819b470da20>.

Malgré les risques, la réglementation des armements dans l'espace ne fait pas consensus. La Russie et la Chine sont en faveur d'une interdiction des armes spatiales, bien que la définition de ce que constitue une arme ne soit pas clairement établie. De nombreux États occidentaux craignent qu'une interdiction ne donne lieu à des abus. En parallèle, plusieurs États développent des armes antisatellites basées sur la Terre. Le Royaume-Uni a également proposé des normes de comportement responsable. En l'absence de consensus, les risques pour les civils demeurent. Dès lors, la communauté internationale travaille à développer des lois, normes et règles pratiques. Des manuels juridiques sont en cours d'élaboration, tels que les manuels Milamos et Woomera. Project Ploughshares vise à développer des normes et identifier des mesures de sécurité pérennes réduisant le risque d'escalade de conflits et permettant de mieux protéger les civils. Il est crucial que la communauté internationale restreigne les activités militaires risquant d'infliger des souffrances indiscriminées et développe des protections pour les infrastructures civiles essentielles. De plus, il faut également envisager comment utiliser l'espace de manière plus résiliente en cas de défaillance des mesures de protection. Plus d'informations sont disponibles sur le site de l'index de sécurité spatiale[2].

More States are preparing for war in outer space. The result could be accelerated, intensified conflict; environmental destruction; and nuclear winter. Even if we avoid the ultimate catastrophe, the consequences of war in space are serious. The destruction of space systems would harm every human on Earth. We must start working to protect civilians on Earth from such a fate.

Our Reliance on Space Systems

A space system is an assembly of one or more satellites and a ground station that uses communications links to collect and exchange data. There are now more than 3,000 satellites in orbit, with the number growing almost every week.

These systems form a meta-capability that enables almost all essential services on Earth. Consider cell phone connectivity, air traffic control, disaster warnings, agricultural production, electronic banking, shipping, power grids. The internet.

But most systems that support civilian functions are multi-use and also enable warfighting capabilities – on Earth and in space. Some States operate only a few satellites, which must meet military, government, and civilian needs. Commercial satellite operators often sell their

[2] Available at: < https://spacesecurityindex.org/ >

services to a variety of customers, including militaries. And some military satellites are essential to civilian life.

The United States Global Positioning System (GPS) – one of several global satellite navigation systems – is a case in point. GPS is the central nervous system of the US military. It provides precision timing, navigation, and targeting capabilities to military units and weapons systems. But GPS also communicates with individual wayfinding and fitness apps, and supports global travel, financial systems, civil communications, and power grids.

Civilian GPS signals are already a target of hostile forces, even during peacetime. For example, Russia has been accused of deliberately interfering with GPS signals in Finland and Norway, threatening the safety of passengers and crew on local airlines. Such interference, while targeted and temporary, is still dangerous. A greater use of force against critical military systems could be devastating.

A Crowded Battlefield

The outer space environment challenges any attempt to target only military objects. The portions nearest to Earth are crowded with military, civilian, and commercial satellites. There is no separate military zone.

Outer space is fragile and unprotected. Anything that is sent into space stays there. And when those objects break apart, the clouds of bits of debris that they create also stay there. These bits can then collide with other objects in space, creating a cascade of damage that not only harms other satellites, but makes surrounding orbits unusable.

While accidental collisions can and have occurred, the intentional destruction of objects is a key source of contamination. China's anti-satellite test in 2008 created the largest debris field to date. And all the pieces are still up there in space.

The Benefits of Space under Threat

A conflict in outer space would almost certainly disable essential civilian services that rely on satellites.

Earth observation satellites monitor and track weather patterns. Their ability to detect wildfires and monitor hurricanes and cyclones makes them indispensable for disaster early warning. They are also essential for disaster response. This need is recognised by the International Charter on Space and Major Disasters, which provides satellite data to help manage disasters.

Satellite-enabled communications meet the daily needs of billions of users on Earth and are even more critical during a disaster, when other ways of communicating are lost. Global Navigation Satellite System (GNSS) signals such as GPS are critical in establishing the precise location of those in need. The Crisis Connectivity Charter is designed to make satellite-based communications more readily available during disasters to those providing humanitarian aid and to affected communities.

Project Ploughshares is currently leading a project to advance the development of norms in space.

The command and control of nuclear weapons are also tied to military assets in space. Damaging those assets could trigger an accidental nuclear strike or provoke a deliberate one.

Pursuing Arms Control in Space

Arms control in outer space is a contentious international topic. Russia and China fear that the United States will develop space-based missile defences that might strike at Earth. They want a space weapons ban – although it is not clear what constitutes a weapon.

Many Western States fear that initiatives that either ban or pledge no-first-use of space weapons are open to abuse. And they do not trust Russia, which they believe already possesses a weapons capability in space that is directed at other satellites.

Meanwhile, a number of States are developing Earth-based anti-satellite weapons.

In response to all of this, the United Kingdom is championing a new initiative to reduce threats to space and the risk of armed conflict in space, by focusing on norms of responsible behaviour.

With no measures gaining consensus, civilians remain vulnerable.

A Protective Mesh

With no major agreement in sight, the international community, including civil society, must prepare to protect civilians through a combination of laws, norms, and practical measures.

Efforts are under way to develop appropriate legal manuals. Notable are the McGill Manual on International Law Applicable to Military Uses of Outer Space (MILAMOS) and the Woomera Manual on the international law of military space operations.

Project Ploughshares is currently leading a project to advance the development of norms in space, working with experts from around the world to identify existing safety and sustainability measures that can be used to inform security practices and reduce the risk of escalating conflict. Included for consideration are the civilian dimensions of conflict in space.

Going forward, the international community must begin work to restrict military activities that inflict indiscriminate harm both on Earth and in space, such as the intentional creation of space debris. We must develop protections for critical civilian infrastructure. And because we know that protection so often fails, we must also think about how to make our ability to use outer space more resilient.

To learn more, visit our updated Space Security Index website. Detailed accounts on outer space at this year's United Nations First Committee meeting by Jessica West are available in Reaching Critical Will's First Committee Monitor.

DISCUSSION

During this Q&A session, the panel discussed the following issues:

1. IHL and the Interaction of New Technologies in Outer Space

The moderator started the discussion with a question on the interaction between different technologies in outer space, as well as the interaction between different domains, e.g. outer space and cyberspace. She wondered what this might entail for the traditional IHL definitions, such as the definitions of an attack or the means and methods of warfare. In particular, it raises the question of whether different definitions are needed for different domains or whether existing definitions remain unchanged.

A panellist emphasised that the question of the impact on new technologies on IHL is not new. Throughout human history, new technologies have benefitted humankind enormously. At the same time, new technologies are often used in the military domain before being used for civilian purposes. This is still the case today, with for instance, cyber technologies being used in armed conflict. Some have argued that new technologies as such would help limit the effects of war on civilians if they are used properly or in a responsible manner. However, there are also challenges in terms of how these technologies can be used and whether human beings can make sure that the use of technologies do not pose additional risks and harm to civilians in armed conflict. In this regard, IHL provides for the legal review of new means and methods of warfare. This legal review applies regardless of the type of technology, i.e. including cyber technology or the weaponisation of outer space. The legal review aims at limiting or even predicting incidental effects, at understanding how to restrict human activities, for instance in space and armament, and how to repair damage or other harmful consequences after they have been caused. The legal review is therefore one of the effective tools that should be taken into consideration and which demonstrates the relevance of IHL in space warfare. However, the relevance and applicability of IHL should not set any limits if States were willing to develop additional limitations or restrictions in relation to the use of weapons in outer space. The moderator added that the IHL's requirements to take precautions in attack and to select means and methods of warfare may sometimes push towards one technology over and above another, for instance towards non-kinetic means rather than self-direct descent anti-satellite missiles.

2. Environmental Aspects

A participant raised a question on the possible increasing importance of the environmental aspect in the proportionality assessment when planning attacks in a space environment. A panellist confirmed that also in his view, the environmental aspect might gain importance in outer space. When considering a military attack or operation which may reach the level of attack as defined in Article 49 of the First Additional Protocol to the Geneva Conventions, the military consider the possible predictable effects, whether direct or indirect, of the attack they may carry out. In space, direct effects might be the debris, for instance. Some States called for the emergence of a standard practice aimed at prohibiting the destruction of satellites, apart from legitimate self-defence, in order to avoid the over-pollution of the Earth's orbits. Indirect effects should also be considered in this assessment. For instance, the assessment should include the effects that the loss of signal or the loss of functions by a satellite would have on populations on Earth, regardless of the nature of the loss, e.g. data supplies or transmission and see if the satellite is essential or contributes to life on Earth. This means that the military will have to refine and improve their systems for assessing potential collateral damage and effects, as well as determining military objectives. When contemplating a military action, the military commander will want to gather information on the object, whether the object has a military function only or, whether it is a dual-purpose object, i.e. which has both civilian and military functions. If so, the commander will certainly want to know as much as possible about the distribution of functions, i.e. the respective percentages of military and civilian uses, in order to take it into account in the proportionality assessment when the decision is made to target the object or not. Another panellist agreed with this answer and stressed that possible knock-on or irreversible effects must be included in the proportionality assessment. In addition, in practice, there are debates on determining what the natural environment is and how such rules would apply to States.

The moderator concluded that the Environmental Modification Convention (ENMOD) also specifically refers to outer space as part of its ambit. It is also worth noting that the environmental aspects refer to the environment on Earth but may also refer to the space environment itself, for instance when seeing the implications of the long-lasting debris circulating for years at massive rate of speed in orbit, including from World War II. Both the environment on Earth and in space are of interest and under at least some of the environment protections and the Geneva Conventions.

3. Ongoing Academic Processes

The moderator then moved on to a question on the ongoing academic processes, especially the McGill Manual (MILAMOS) and the Woomera Manual. Some have argued that one is a peace manual and the other is a war manual. However, this does not seem to be the case. For one of the panellists, the MILAMOS process showed that the rules of international law applicable to the military use of space are not limited to the law on the use of force by States (*jus ad bellum*). While the initial project was focused on this aspect, the MILAMOS team may now be considering a second volume, which would complement the document to be released in the next months. This second volume would also identify the international law that applies in the conduct of hostilities. The Woomera project had a different objective from the outset. The ambition was to embrace both the *jus ad bellum*, which relates more to peacetime considerations, and the *jus in bello*, in order to provide practitioners and operational staff with the most possible comprehensive view of the possible questions raised by the planning and conduct of military space operations. Indeed, the moderator confirmed that the Woomera Manual is a full spectre project going from peaceful uses of outer space, through the prohibition of the use of force and the *jus ad bellum* into the *jus in bello*. This is therefore certainly not just a manual on the use of military space operations.

4. Outer Space as a Strategic Priority

A participant raised a last question on States' military strategies: are States likely to adopt military strategies that prioritise having power in space, much like it was a priority to have the power of the skies in the early 20th century? A panellist said that indeed, this was beginning to unfold. The strategic importance of the space sector is nothing new and dates back to the beginning of the Space Age. However, the centrality of space power to overall military strategies is growing. The recently released United States Space Security Strategy is a case in point. Not only does it emphasise the importance of dominance in space and secured access to the use of space, but it situates this power and this ability as central to military power in all other domains. It is another shift that really speaks to the importance of assessing the implications of IHL and potential warfare and struggles on access and use of outer space. The moderator added that the capstone document that the European Union has put out also emphasises that drive. Nonetheless, the drive is more to retain dominance in other domains and to preserve assets within space. The US Chief of Space Operations Gen. John W. Raymond has gone on record as saying that although the US do not intend to fight a war in space, they must be ready should they have to. There is probably no State in the international system who would like a space war. Everybody recognises that armed conflict in space would be problematic for everyone: for civilians, for the military, for States who want to continue to operate in space. Nevertheless, the assets that are being developed in space enable dominances, including in other domains than space itself.

In conclusion, this session highlighted the remaining uncertainties and risks related to the use of outer space. A panellist concluded that the ICRC's concern is the potential very high human cost of the use of weapons in outer space. In that respect the ICRC encourages States and other stakeholders, including academic institutions and experts, to pay more attention to the subject. The ICRC hopes that this issue will generate some attention and will also help the humanitarian cause in the light of increasing competition among major powers, including beyond outer space domains.

Panel 5
New Technologies and Humanitarian Action
Nouvelles technologies et action humanitaire

INTRODUCTION

Massimo Marelli
ICRC Geneva

Résumé

Massimo Marelli est le chef du Bureau de la protection des données au Comité international de la Croix-Rouge (CICR). En tant que modérateur de ce panel, il en a présenté le thème, les nouvelles technologies et l'action humanitaire, ainsi que les panélistes.

S'écartant de la perspective traditionnellement juridique de ce Colloque, ce panel aborde le secteur humanitaire et l'utilisation de plus en plus importante qu'il fait des nouvelles technologies. Cette évolution est motivée par plusieurs facteurs, tels que la demande des donateurs de faire plus avec moins ; le devoir envers les populations affectées d'avoir le plus d'impact possible ; l'importance croissante des programmes de transfert de fonds, souvent associés à des données biométriques et une identité digitale, voire à la blockchain *; l'idée que la big data, l'intelligence artificielle (IA) et l'apprentissage automatique peuvent aider à identifier les acteurs pertinents pour négocier l'accès humanitaire et mener les dialogues bilatéraux confidentiels ; ou encore les difficultés croissantes à établir une proximité physique, la proximité digitale apparaissant comme une solution complémentaire.*

Cet accroissement de l'empreinte numérique dans le secteur humanitaire a de nombreuses implications. Certaines sont liées au DIH mais d'autres amènent également dans la discussion un nouveau domaine du droit : la protection des données personnelles et la notion de vie privée. Pour des organisations humanitaires telles que le CICR, protéger et respecter les données personnelles revient à protéger et respecter les personnes affectées. Le CICR assurait donc la protection des données depuis longtemps sans la nommer ainsi. Toutefois, la complexité du contexte technologique actuel implique que le secteur humanitaire s'appuie de plus en plus sur les lois et outils de la protection des données personnelles pour assurer la protection des droits et la dignité des individus lors du traitement de leurs données. En ce sens, la protection des données personnelles devient un outil permettant aux organisations humanitaires d'appliquer le principe de «ne pas

nuire » ('do no harm') *et de maintenir l'individu au centre de la réponse humanitaire, y compris dans un environnement numérique.*

Le CICR travaille depuis plusieurs années avec des experts en la matière, notamment à la rédaction d'un Manuel sur la protection des données dans l'action humanitaire (2ème édition)[1]*, auquel les membres de ce panel ont contribué. Le manuel contextualise les principes de protection des données dans un environnement humanitaire et en précise l'application lors de l'utilisation de technologies telles que l'IA, les données biométriques, les programmes de transferts de fonds, le* cloud *ou dans des domaines technologiques complexes, tels que l'identité numérique. Ce panel se penche plus particulièrement sur l'identité numérique.*

Massimo Marelli is the Head of the Data Protection Office at the International Committee of the Red Cross (ICRC). In his role as moderator of this panel, he introduced the theme of the session and the panellists.

The topic of today is quite new and unexpected for this Colloquium, which for 20 years has had a specific emphasis on International Humanitarian Law (IHL), focusing on the conduct of parties to conflict and actors in violence, with humanitarian organisations primarily observing, assessing, analysing whether a particular conduct, means or method adopted by a party to the conflict is compatible with the law. We have seen how technology advancements have been impacting this analysis very significantly in recent years, and this has indeed been the focus of the last few days.

Today's perspective is slightly different. Today we look at humanitarian action, and at how the humanitarian sector itself has increasingly been using new technologies with a view to making its work more effective and efficient.

This is driven by a number of key factors such as:

- the famous blanket problem: pressure from donors to do more with less;
- a duty towards affected populations to do whatever we can to make our work have more impact and reach further, and to leverage technology in so far as this can enable us to do so;
- conflicts last longer (the average conflict lasts 30 years) with people increasingly relying on assistance in the longer term. Cash transfer programmes become useful tools, and these

[1] ICRC, 'Handbook on data protection in humanitarian action' (2nd edition). Available at: <https://www.icrc.org/en/data-protection-humanitarian-action-handbook>.

are often linked with the use of biometrics and the creation of digital identities. The use of blockchain is often suggested as relevant to support such programmes;

- conflict theatres become more volatile and difficult to read as you no longer have two armies lining up, one facing the other in bright colours helping you to understand who is who and where. We nowadays deal mainly with non-international armed conflicts, with factions often involved in shifting alliances, splinter groups, radicalised groups, and it is difficult to know who to speak to if you are trying to negotiate access and carry out a bilateral confidential dialogue (big data, AI and machine learning can help);
- with increasing challenges relating to access, physical proximity becomes increasingly challenging, leading us to look for new ways of complementing it with digital proximity (drones, messaging apps).

There are many implications involved in this exponentially augmented digital footprint of the humanitarian sector. Some of them have a strong IHL connotation, but some others also bring into the heart of the discussion a new area of law for this community that is highly topical: personal data protection and privacy.

For humanitarian organisations, like the ICRC, that have at the heart of their mandate the protection of and assistance to affected populations, protecting and treating personal data with respect means protecting and respecting affected populations. This means that, without calling it so, we have been applying data protection for a long time.

The complexity of the new technological landscape, however, makes it such that the humanitarian sector must also rely more and more on personal data protection, and its body of laws and tools, to ensure the protection of the rights and dignity of individuals when processing their data.

Personal data protection therefore becomes a tool for humanitarian organisations to apply the principle of 'do no harm' in a digital environment, and to keep the individual at the centre of a humanitarian response.

It all sounds good, but what does it mean in practice? For a few years now, we have been working with many experts in the field – some of them are with us today – to create guidance, for example with the Handbook on Data Protection in Humanitarian Action (Second Edition)[2]. The Handbook looks at the basic data protection principles contextualised in a humanitarian

[2] ICRC, 'Handbook on data protection in humanitarian action' (2nd edition). Available at: <https://www.icrc.org/en/data-protection-humanitarian-action-handbook>.

environment and at the declination of these principles in the use of certain specific technologies such as AI, biometric data, cash programmes, cloud, or other technology areas that are incredibly complex, such as digital identity.

It is indeed on the challenge of digital identity that we would like to focus today. Thankfully we are joined today by an incredible panel of experts in this area, to try to dissect from their perspectives the area of digital identity, to see how it is relevant in the humanitarian sector, what kind of issues it can raise and how we can navigate that complex environment.

So, without further ado, I will now turn to introducing our experts. The full biographies can be found in the Colloquium materials, so I will not go there, even just a quick introduction will show their great level of expertise.

Mr Edgardo Yu has been the Chief of Beneficiary Services within the Technology Division of the United Nations World Food Programme (WFP) since 2014. WFP recently won the Nobel Peace Prize. Congratulations Edgardo to you and all the colleagues at WFP for the great work you have been doing and the well-deserved recognition.

Prof. Carmela Troncoso is an assistant professor at EPFL (Ecole Polytechnique Fédérale de Lausanne, Switzerland) where she heads the SPRING Lab. Her research focuses on security and privacy. And speaking of awards, she was recently listed in the *Fortune 40 under 40* for her work, and particularly for the work she led recently on DP3T, the Privacy-Preserving Digital Contact Tracing protocol which has now been implemented by many countries.

Prof. Wojciech Wiewiórowski needs no introduction as he is the European Data Protection (EDP) Supervisor. He is the competent authority to supervise the application of data protection laws within the European Union institutions as well as one of the most knowledgeable experts in the field of personal data protection.

Ms Alexandrine Pirlot de Corbion is Director of Strategy at Privacy International (PI). Among all the great work that she has been performing in this role, she has been leading PI's work with the development and humanitarian sector. In 2018, she co-authored a report published by the ICRC and PI which explored the risks associated with the use of data and new technologies in the humanitarian sector, called 'The Humanitarian Metadata Problem: 'Doing No Harm' in the Digital Era'.

Ms Catherine Kayser works at the Humanitarian Action Desk at the Luxembourg Ministry of Foreign Affairs, having previously worked in the field as well as being a legal adviser at the

Ministry. In this role, she is a key player in Luxembourg's extraordinary support of the humanitarian partners' efforts to leverage the potential of new technologies while making sure that beneficiaries stay at the centre of each intervention.

What all the speakers have in common is that they and their organisations have all been actively contributing to the development of the Handbook on Data Protection in Humanitarian Action, which is now at its second edition, and which provides detailed and precise guidance in this area.

NEW TECHNOLOGIES. WHY IT IS RELEVANT AND USEFUL IN A HUMANITARIAN CONTEXT?

NOUVELLES TECHNOLOGIES : POURQUOI SONT-ELLES PERTINENTES ET UTILES DANS UN CONTEXTE HUMANITAIRE ?

Edgardo Yu

Technology Division, World Food Programme (WFP)

Résumé

Edgardo Yu travaille à la Division pour la technologie du Programme alimentaire mondial (PAM). Dans sa présentation, il montre en quoi les nouvelles technologies, et plus particulièrement les identités digitales, peuvent être pertinentes et utiles dans un contexte humanitaire.

Dans une première partie, Edgardo Yu explique l'intérêt du numérique dans le secteur humanitaire et plus spécifiquement pour le PAM. La portée des opérations du PAM est considérable. SCOPE, la plateforme d'information sur les bénéficiaires et de gestion des transferts du PAM, compte environ 60 millions d'identités, pouvant inclure des noms, des photographies, la composition des ménages, les niveaux de vulnérabilités et des données biométriques. Le PAM collecte deux types de données biométriques : les empreintes digitales et l'iris. Edgardo Yu liste cinq avantages de l'identité numérique dans un contexte humanitaire. Premièrement, les identités conservées sur la plateforme SCOPE permettent d'atteindre de manière plus efficace les zones difficiles d'accès. Deuxièmement, l'identité digitale diminue le risque d'erreurs et permet de mieux répondre aux besoins tout en réduisant les coûts de transaction. Troisièmement, une base unique de données d'identités permettrait aux partenaires humanitaires de fournir et de suivre l'assistance entre partenaires humanitaires de manière plus coordonnée et efficace, en minimisant la duplication et en permettant un usage optimal des ressources. Cela donnerait également davantage de garanties aux donateurs. Quatrièmement, l'identité fonctionnelle fournie par l'assistance humanitaire sert de véhicule d'inclusion, y compris financière, et facilite l'alphabétisation numérique. Cinquièmement, la numérisation via un système développé par le PAM permet une amélioration de la protection des données personnelles et de la vie privée des bénéficiaires en collectant uniquement l'information nécessaire. Au-delà de l'identité numérique, les nouvelles technologies présentent également plusieurs avantages pour le PAM : amélioration de l'analyse de la vulnérabilité ; efficacité programmatique ; responsabilité à l'égard des bénéficiaires, des partenaires et des donateurs (éviter la duplication et optimiser les ressources, suivre les transactions, mécanisme de responsabilité et de retour d'information plus efficace) ; meilleur accès humanitaire.

La deuxième partie détaille les éléments identifiés par le PAM comme essentiels à l'utilisation de l'identité digitale dans des contextes humanitaires :

- *adopter une approche éthique et normative solide, en insistant sur l'évaluation de l'impact sur la vie privée, les droits des personnes concernées et les mesures de sécurité des données ;*
- *veiller à ce que les identités numériques soient fiables, précises et neutres et qu'elles puissent être utilisées pour prouver l'identité de nombreux acteurs, y compris ceux qui ont peu de capacités numériques ;*
- *veiller à ce que les identités numériques permettent de mettre en place un canal de communication avec les bénéficiaires, notamment des mécanismes de plaintes ;*
- *recueillir uniquement les données essentielles et nécessaires pour fournir une assistance ;*
- *respecter les principes de protection des données, qui sont, selon le PAM : une collecte et un traitement de données légaux et équitables, une utilisation des données personnelles identifiables (DPI) dans un but légitime et spécifique, la garantie de la qualité des données, la participation de l'individu, la responsabilité de l'organisation concernant les informations sauvegardées, la garantie de la sécurité des données.*

Why Digital in the Humanitarian Sector? Why so Relevant and Important? To Do What?

In 2019, we reached 86.7 million beneficiaries in 83 countries. This number has further increased this year because of Covid-19. With greater demands and more people affected by crises, we have adapted the way we work – embracing digital tools as an enabler in improving the quality of our services to beneficiaries.

If I may, I would like to share with you the scale at which WFP operates.

Our beneficiary information and transfer management platform, called SCOPE, hosts about 60 million identities.

For many of the identities in SCOPE, we get to know their names, their pictures, their household compositions, their vulnerability levels and their biometrics (fingerprints and iris recognition are the only biometrics we collect).

Out of those 60 million, 14.73 million have received assistance this year alone. We have accomplished this across 42 countries in more than 1,400 interventions for a total value of 815 million United States Dollars (USD).

Hosting their identities in SCOPE, we are reaching the most remote areas in the world more efficiently.

Digital identities are effective in reducing errors in targeting, and enabling us to reach the right people with the right assistance while reducing the transactions costs of such assistance.

A unique identity base leads to more coordinated and efficient delivery and the tracking of assistance between humanitarian partners, minimising overlaps and enabling optimal use of resources across the humanitarian community, thus giving more reassurance to our donors in terms of the level of our duty of care with the resources that they provide.

In the increasingly digitised economy, the functional identity provided by humanitarian assistance has served as a vehicle for inclusion, including financial inclusion, and facilitating digital literacy amongst those left behind by development.

Digital technology has further facilitated an improvement in the protection and privacy of personal data of beneficiaries by capturing only the identifiable information required. WFP has been able to establish a system of capturing these at just the right time, giving us additional opportunities to protect such data.

In addition to contributing to setting identities, WFP's gains in the use of digital technology include:

- enhanced vulnerability analysis (e.g. mobile data analysis)
- programmatic efficiency (e.g. enhanced implementation and monitoring of assistance)
- accountability towards beneficiaries, partners and donors (e.g. deduplication and optimisation of resources, follow up of transaction throughout the course of the assistance; and more efficient accountability and feedback mechanism (e.g in Libya)
- humanitarian access (e.g. to areas inaccessible to m-VAM, WFP's mobile Vulnerability Analysis and Mapping)

What Are the Features of Digital Identity that Are Crucial in Humanitarian Settings? What Data? To Do What?

The essential features are the following:

- a solid ethical and normative approach, with particular emphasis on privacy impact assessment, data subjects' rights and data security measures. Policy comes first. To that end we strongly support the development of data protection frameworks and tools to build stronger responsibility and accountability

- we need to ensure that digital identities are robust, accurate and vendor or technology neutral, that it can be used for identity proofing by many actors including those that have basic or minimal digital capabilities
- we need to ensure that digital identities enable the provision of a communications channel with beneficiaries. 80% of WFP programmes are linked to beneficiary complaints and feedback mechanisms
- we must collect only the essential, minimal data needed to provide assistance
- we need to focus on humanitarian principles and humanitarian protection above all in adopting the right posture in the wide variety of contexts that we operate in
- we must uphold the principles of data protection, which to us are lawful and fair collection and processing, using personally identifiable information (PII) for a legitimate and specified purpose, ensuring data quality, ensuring the individual's participation and taking accountability for the information we safeguard, and ensuring that we secure the data

TECHNOLOGY FRAMEWORK: WHAT IS AUTHENTICATION AND IDENTIFICATION, WHAT IS FUNCTIONAL IDENTITY AND FOUNDATIONAL IDENTITY AND WHY DOES IT MATTER?

LE CADRE TECHNOLOGIQUE : QU'EST-CE QUE L'AUTHENTIFICATION ET L'IDENTIFICATION, L'IDENTITÉ FONCTIONNELLE ET L'IDENTITÉ FONDAMENTALE ET EN QUOI CES CONCEPTS SONT-ILS IMPORTANTS ?

Prof. Carmela Troncoso

EPFL

Résumé

Carmela Troncoso est professeure assistante à l'Ecole polytechnique fédérale de Lausanne où elle dirige notamment le SPRING Lab ('Security and Privacy Engineering Laboratory'). Dans sa présentation, elle questionne l'idée qu'utiliser l'identité numérique est nécessaire pour mener à bien les activités humanitaires.

Cette contribution commence par souligner deux idées reçues. D'abord, l'identité numérique n'est pas forcément nécessaire pour minimiser les erreurs ou pour éviter la duplication des services, et la numérisation n'est pas toujours la solution pour être plus efficace. De plus, Carmela Troncoso explique que lorsque la numérisation est utile – ce qui peut être le cas quand les besoins augmentent – une erreur commune est de reproduire le fonctionnement du monde hors ligne dans le monde en ligne. Or la technologie ne respecte pas les lois physiques. La technologie permet de faire des choses qu'il ne serait pas possible de faire dans le monde physique. Dès lors, même si cela semble plus simple et familier, numériser les processus tels qu'ils existent dans le monde réel n'est pas toujours pertinent et ce n'est certainement pas la meilleure solution en termes de protection des données.

Carmela Troncoso appelle à ce que le choix de numériser dépende de l'objectif visé. Par exemple, si le but est d'éviter la duplication, certaines technologies permettent de tester si une demande vient de la même personne sans avoir besoin de son identité. L'identité n'est pas nécessairement la solution. Elle reconnait qu'en terme d'efficacité, cette proposition peut sembler être un retour en arrière. Toutefois, elle rappelle que les bénéficiaires de l'action humanitaire sont plus vulnérables et que les risques potentiels sont bien plus élevés pour eux que pour les personnes utilisant des technologies similaires dans le monde occidental. Dès lors, les outils technologiques utilisés au quotidien ne sont pas nécessairement adaptés au secteur humanitaire.

Elle suggère ainsi qu'il pourrait être nécessaire de développer des technologies spécifiques au secteur humanitaire. Si l'objectif est de pouvoir intégrer la protection des données et d'éviter les dérives fonctionnelles, il existe des moyens de concevoir des systèmes dans lesquels la limitation de la finalité, et pas seulement la minimisation des données, est intégrée à la technologie elle-même. Le développement de technologies propres au secteur humanitaire risque certes de prendre plus de temps. Néanmoins cela permettrait d'avoir des solutions à long terme et d'éviter d'augmenter les risques pour les bénéficiaires. Le déploiement de technologies existantes peut en effet avoir des conséquences à long terme. Par exemple, plusieurs organisations se disent très attachées à la minimisation des données, mais en collectent elles-mêmes des millions sur des millions de personnes. Or, une fois que la base de données a été construite, qu'en fait l'organisation ? Ne pourrait-elle pas penser le problème autrement ? Carmela Troncoso conclut que si cette discussion sur l'objectif final et la nécessité de collecter ces données n'a pas eu lieu et qu'une organisation se contente de numériser ce qui se fait dans le monde hors ligne, il deviendra plus difficile par la suite de respecter les principes de protection des données.

My presentation will go to the controversial side. The first point is that we all start with the assumption that we need identity. But so far, none of the things I have heard today essentially requires identity. We want to produce an error targeting tool or to avoid duplication: none of these require identity *per se*. Identity is the mechanism we use in the offline world: an identity card for instance. We should not fall into the trap of thinking that we need to digitalise everything for effectiveness and efficiency. I would agree that it helps, especially as the needs increase. For instance, with the Covid-19 pandemic, we have started to use digital contact tracing and other technologies. That is very nice but digitalising and relying on technology do not mean that we need to mimic exactly in the online world what we would do in the offline world. Because technology has no physics. With technology we can do things that we could not do in the physical world. So sometimes the error is to digitalise processes as they are. This is the easiest and most familiar thing to do but may not be the most empowering or the best if we want to protect and embed data protection principles at the core of the system.

What we need to think about is: 'what is the goal'? If the goal is for instance non-duplication, we have technologies that allow us to test that a request does not come from the same person without learning the identity of the person. In the physical world, this has been done for years. When I was young and I would go to the cinema, they would give you a little paper, which did not have my identity and yet allowed the cinema to decide whether I am the same person or not or whether I am entering the cinema twice.

So, it is not necessarily the case that identity is the solution. And I think it is very important for the humanitarian sector to take a step back and think whether it is so.

I am very aware that if we look at effectiveness, what I suggest could be seen as taking a step back. However, we are dealing with people who are very vulnerable by nature and maybe are the most endangered of all the people we could have in these systems, as opposed to people who would use these systems in the Western world for shopping. Maybe the systems that are being developed for Western use, which are believed to be very safe and advanced technologies, are not the ones that we should use in the humanitarian sector. Maybe we need to develop new technologies. And that is going to be slower. But if the goal is to be able to embed data protection and avoid function creep, there are ways to design systems in which purpose limitation, and not only data minimisation, are given by the technology itself.

We need to open the discussion. Maybe we need to develop or deploy at a slower pace. Maybe indeed we need to take the hit at the beginning, so that in the long term we do not end up with a worse situation. Looking at the short term and just deploying technology may not be the solution. Some organisations say they care a lot about data minimisation and collection but at the same time they have millions of data of millions of people, with a lot of information about their lives. Once the database is there, what does the organisation do with that? The database can be used for many things. Now the question is: should the database ever be built? Could they have thought about the problem in another way? If the discussion is not happening and we just digitalise the work we do in the offline world, we are not going to end in a place in which fulfilling the principles of data protection is easy and many times may not even be possible.

DATA PROTECTION PROVISIONS FRAMING THE DIGITAL IDENTITY AND BIOMETRICS (AND THE MAIN PROBLEMATIC AREAS)
COMMENT LES DISPOSITIONS DE PROTECTION DES DONNÉES ENCADRENT L'IDENTITÉ NUMÉRIQUE ET LES DONNÉES BIOMÉTRIQUES. OÙ SE LOGENT LES PROBLÈMES ?

Wojciech Wiewiórowski
European Data Protection Supervisor

Résumé

Wojciech Wiewiórowski est le Contrôleur européen de la protection des données (CEPD). Dans le cadre de ses fonctions, il participe en tant que conseiller au processus législatif de tous les actes juridiques de l'Union européenne (UE) concernant tout type de données. Dans sa présentation, il souligne notamment la coopération entre le CEPD et les organisations humanitaires et l'organisation par le CEPD ces dernières années d'ateliers sur les organisations internationales, y compris humanitaires. Il rappelle que sur ces questions de protection des données dans le secteur humanitaire, il ne s'agit pas seulement de questions juridiques mais surtout de prévenir et de répondre aux dangers réels liés à l'utilisation de données sensibles par des organisations humanitaires.

Il explique que l'identité numérique a de profondes implications pour la protection des droits de l'homme et des personnes concernées. D'une part, l'identité numérique peut rendre les individus moins vulnérables, en leur permettant d'accéder à des services auxquels ils n'auraient pas accès autrement, d'autre part, elle représente un risque pour la protection des données personnelles individuelles. La coopération entre les autorités chargées de la protection des données et les organisations internationales vise à atténuer les risques. Cependant, ces risques ne vont pas disparaître. En particulier, il faut veiller à ce que les bénéficiaires ne deviennent pas des « cobayes » pour tester les nouvelles technologies numériques. De plus, les idées mises en place par des organisations fiables peuvent être réutilisées par des tiers, au bénéfice ou au préjudice de leurs propriétaires.

Dès lors, le principal défi est de trouver des solutions pratiques à ces risques, en gardant à l'esprit les principes de la législation sur la protection des données, tels que la limitation de la finalité et la minimisation des données. Il s'agit de stocker les données uniquement nécessaires pour fournir les services et de s'assurer de leur exactitude. Lors de l'introduction de l'identité numérique dans le cadre d'actions humanitaires, l'évaluation de l'impact sur la vie privée ou sur la protection des données est également une procédure très utile, devant intervenir avant même l'utilisation du système. L'identité numérique constitue un instrument très utile pour les actions

humanitaires, mais comme tout outil, elle peut être mal utilisée. Dans la plupart des endroits où les organisations humanitaires travaillent, il n'y a pas d'autorités locales qui s'occupent de la protection des données. Ce sont donc ces organisations humanitaires qui sont à la fois les responsables du traitement des données et les contrôleurs de la manière dont elles sont utilisées. Pour ces raisons, le CEPD se réjouit d'avoir pu contribuer au Manuel sur la protection des données dans l'action humanitaire, qui aborde ces sujets.

Thank you for the possibility to be with you here. I am Wojciech Wiewiórowski, the European Data Protection Supervisor, which means that I am the supervisor of the European Union (EU) institutions, bodies and agencies, and – what is more important for today's discussion – I am taking part as an advisor in the legislative process of all the legal acts of EU law which pertain to any kind of data.

The reason I am here and taking part in this meeting is the already fruitful cooperation that we have with humanitarian organisations in this field. Let me thank you, Massimo, for all your efforts to get the people involved in the activities of these organisations and public authorities at the same table to discuss the subjects of privacy and humanitarian activities. Discuss them in a useful and very practical way. In recent years, the European Data Protection Supervisor has been organising workshops for international organisations, including humanitarian organisations. I have to say that those who are dealing with humanitarian aid are usually the best prepared to deal with personal information in their everyday work. This of course requires not only a knowledge of international law which drives us, but, first of all, of the practical dangers that are connected with the fact that lots of sensitive data are used by humanitarian organisations.

The digital identity has profound implications for the protection of human rights and individuals concerned. On the one hand, digital identity can make individuals less vulnerable, by enabling them to access services that they might not access otherwise, on the other hand, they represent a risk for the protection of the individual personal data. Of course, the cooperation between the data protection authorities, also the European Data Protection Supervisor, and the international organisations aims at mitigating the risks that may exist. But the risks are here to stay. We need to ensure that the recipients do not become guinea pigs for new technologies in a digital world, which are deployed during humanitarian crises in 'training mode'. There might be many things which can go wrong. The temporal and functional ideas which are issued by the trusted organisations may be re-purposed by third parties to the advantage or the disadvantage of the owners. Those functional ideas unintentionally become the foundational ideas in many places. The re-use and dual-use are at the same time challenges and risks.

No matter if we think about sophisticated hacking attacks or ‘only’ about the leaking of data, we know data is not secured in the right way.

The most important challenge is to find practical solutions for these dangers, bearing in mind the principles of data protection law such as purpose limitations or data minimisation: how to store data only until ‘it is necessary to provide the services’, and how to keep the accuracy of the data.

We should always stress that when introducing any kind of digital identity for the purpose of humanitarian actions, the privacy impact assessment or data protection impact assessment becomes a very useful tool to be applied before the systems are used. That will enable us to assess the impact of the different technologies that are going to be used. Digital identity is a very important asset for humanitarian actions, but like all tools it can be misused and in most of the places in which humanitarian organisations are working we do not have local authorities dealing with data protection. So actually, these are the organisations who at the same time are the controllers of the data but also the supervisors of the way the data is used. For these reasons we were very happy to take part in the preparation of the Handbook for Data Protection in Humanitarian Action which is touching on these subjects.

A CIVIL SOCIETY PERSPECTIVE: DIGITAL IDENTITY IN HUMANITARIAN ACTION: WHAT ARE THE IMPLICATIONS FOR PEOPLE AND THEIR RIGHTS?
LA PERSPECTIVE DE LA SOCIÉTÉ CIVILE : L'IDENTITÉ NUMÉRIQUE DANS L'ACTION HUMANITAIRE : QUELLES SONT LES IMPLICATIONS POUR LES PERSONNES ET LEURS DROITS ?

Alexandrine Pirlot de Corbion
Privacy International

Résumé

Alexandrine Pirlot de Corbion est la Directrice de la stratégie de Privacy International. Dans cette contribution, elle recommande de s'interroger sur la nécessité d'utiliser l'identité numérique et présente plusieurs risques que cela peut entraîner pour les personnes et leurs droits.

Tout d'abord, elle note que l'identité numérique fait déjà partie des technologies utilisées par les organisations humanitaires mais recommande de continuer à s'interroger sur la nécessité d'un système d'identité. Une question qui devrait être au centre de toutes les décisions est de savoir si les identifiants fournis par les organisations humanitaires sont nécessaires à un octroi efficace de l'aide. Elle recommande aux organisations humanitaires de se poser les questions suivantes : ont-elles besoin de savoir qui est un bénéficiaire en particulier ? Le déploiement d'un système est-il proportionné et nécessaire à l'objectif poursuivi, à savoir fournir une aide d'urgence à ceux qui en ont besoin ? Ce test de nécessité et de proportionnalité, censé vérifier l'objectif et la légitimité du système d'identification, est souvent négligé au profit de la question de savoir comment le mettre en oeuvre et comment atténuer les risques. Or, certains risques ne peuvent être atténués. Dès lors, une organisation humanitaire doit-elle poursuivre cette démarche tout en ayant conscience des conséquences négatives possibles ou doit-elle arrêter avant que le mal ne soit fait ? Comment les risques inhérents à de tels systèmes interagissent-ils avec le principe de « ne pas nuire » qui sous-tend l'action humanitaire ?

La contribution précise ensuite les limites et les risques de l'utilisation de l'identité numérique. La rigidité des systèmes d'identification rend difficile de garantir l'universalité, l'inclusion, la sécurité et l'exactitude. De plus, les conséquences d'une erreur peuvent être très importantes. Privacy International a identifié plusieurs risques : exclusion et discrimination ; utilisation de données à des fins autres que celles pour lesquelles elles ont été traitées (risque qui est également lié aux demandes d'interopérabilité de la part des donateurs) ; vue à 360 degrés permettant de surveiller un individu tout au long de sa vie ; risque de prendre des décisions sur la base de données inexactes ou obsolètes ; difficulté pour un individu de sortir d'une catégorie qui lui est assignée ; risque de

priver un individu ayant des identifications différentes de la possibilité de relier celles-ci en fonction des contextes. L'auteure donne l'exemple d'une situation au Kenya : plus de 40 000 personnes ne peuvent pas s'inscrire pour obtenir un Huduman Numba, la carte d'identité nationale en cours de déploiement, car aux yeux du gouvernement kenyan, elles sont reconnues comme des réfugiés inscrits dans le système proGres du HCR. Bien qu'un processus de déduplication soit en cours, cela laisse des milliers de personnes dans l'incertitude et l'impossibilité de jouir de droits fondamentaux tels que l'éducation, la santé, le vote et les opportunités économiques.

Alexandrine Pirlot de Corbion conclut que les organisations souhaitant déployer un système d'identité numérique doivent donc, après en avoir justifié la nécessité, intégrer la protection au cœur du système. Ces risques ne peuvent être ignorés et doivent éclairer les décisions prises par les organisations humanitaires. De plus, ces risques s'inscrivent dans des contextes particuliers, où la surveillance et l'exploitation des données ne sont pas nécessairement réglementées. Les décisions relatives à la création d'une identité numérique doivent être prises à la lumière de ce contexte particulier. Cela exige de connaître les besoins mais aussi les éléments de vulnérabilité de leurs propres organisations et des personnes qu'elles aident et de se préparer à répondre à un problème, par exemple une violation ou une utilisation abusive des données.

Digital Identity in Humanitarian Action: What Are the Implications for People and their Rights?

Humanitarian organisations have been deploying new technologies and making use of data-intensive systems in their programmes for many years now, and one of the many new technologies being seen deployed in the humanitarian sector is digital identity.

Many humanitarian and development organisations have taken it upon themselves to provide identity credentials to the affected individuals they assist as a pre-requisite to providing them with assistance. Identification systems have begun to be seen as an end in themselves, rather than as a tool for achieving a certain goal. This has emerged with the push from donors for the sector to be more accountable about how money is spent and that those who most deserve it receive assistance, and also supposedly as part of anti-fraud measures.

When it comes to the application of a digital identity in the humanitarian sector, one question which is key and should be at the centre of all decision making is whether providing that 'legal identity' or those 'identity credentials' is necessary for the effective provision of humanitarian assistance. The following are questions humanitarian organisations need to ask themselves. Do we need to know who a particular affected person is? Is it not enough to know that they have been identified as needing humanitarian assistance? Is the deployment of a digital identity scheme in a humanitarian context proportionate and necessary to the aim pursued, i.e. to deliver emergency assistance to those who need it?

We must continue to question the need for an identity system.[1] Often the 'necessity and proportionality test', which asks what the purpose of the identity document (ID) system is and whether the aim is legitimate, is missing, and often there is a leap to the how can we make it happen and how do we mitigate the risks.[2] Whilst some features of an identity system can integrate privacy-friendly solutions, it must be recognised that some risks cannot be mitigated, and so what does that mean? Does a humanitarian organisation proceed knowing some will be negatively impacted and accept that trade off, or do they need to stop the idea at the concept level before harm is done? And importantly, how do the risks that emerge from such systems interplay with the principle of 'do no harm' that underpins humanitarian action.

So, what harms and risks are we talking about?[3]

Human identities are complex, multifaceted and fluid, and yet from a technological perspective the rigidity of the designs of identification systems make it difficult to ensure universality, inclusion, security and accuracy. The consequences of getting these things wrong are huge, which is why those wanting to deploy a digital identity system, once they have justified the need for it, must design to protect, and protect by design.

Here are some of the risks we have observed with regards to digital identity systems, and which have implications for their deployment in humanitarian action.

Exclusion and Discrimination

By their very nature and concept, identity systems create risks for those who have access to identity credentials, in the form of an identity card or part of a system such as a cash transfer programme or an app, as well as those who do not. This creation of categories establishes the foundation for exclusion and discrimination.[4] Despite all the claims for universality, there

[1] Privacy International, 'Understanding Identity Systems Part 1: Why ID?'. Available at: <https://privacyinternational.org/explainer/2669/understanding-identity-systems-part-1-why-id>.

[2] Centre for Humanitarian Data, 'Data Impact Assessments', in: *Guidance Notes Series: Data Responsibility in Humanitarian Action*, published in collaboration with the International Committee of the Red Cross, UN Global Pulse and Privacy International, July 2020. Available at: <https://privacyinternational.org/sites/default/files/2020-07/guidance_note_data_impact_assessments.pdf >.

[3] Privacy International, 'Understanding Identity Systems Part 3: The Risks of ID'. Available at: <https://privacyinternational.org/explainer/2672/understanding-identity-systems-part-3-risks-id>.

[4] Privacy International, 'Understanding Identity Systems Part 1: Discrimination and Identity'. Available at: <https://privacyinternational.org/explainer/2670/understanding-identity-systems-part-2-discrimination-and-identity>.

will be some people who do not have access to identity credentials, or those who cannot use their identity credentials, and as a result are denied access to goods and services.[5] And this is heightened when identity becomes mandatory. This risk is particularly concerning given that those seeking humanitarian assistance and related support like other marginalised people in vulnerable populations are often the least likely to have proof of their identity, while also the most in need of the protection and services linked to their identification. In a humanitarian setting, does this mean that someone in need of assistance, based on those credentials, will be given assistance or not?

Mission or Function Creep

One of the hardest risks to regulate is mission or function creep, i.e. data being used for another purpose than that for which it was processed in the first place, and how the principle of purpose limitation is enforced in practice[6]. The mere availability of data is tempting to many and it can be seen as an opportunity. When it comes to digital identity systems, this results in the spread of an identity system to more and more aspects of people's lives. Once the system exists, it is very easy for new purposes to be added to it and with both policy and tech the options are endless as to what a system could be used for. In the humanitarian sector we have seen how donors play a role in maximising the existence of a system for others to also use and ensure interoperability. The sequence of the identity then being used for multiple purposes is fast. What happens when an identity system provided by humanitarians as a functional identity, i.e. which is used for a particular purpose or function, becomes a foundations identity, i.e. used by an authority as identity credentials.

A 360-Degree View

In the way identity systems are currently being designed and implemented, they give a 360-degree view of the person, which raises concerns in particular for the long term. The very creation of this supposedly essential 'unique identifier' sets the foundation for and is the facilitator of surveillance by giving the ability to link and monitor information about an individual throughout their lives and at key stages, depending on how well or not data protection and other safeguards are respected. There is a risk that life-changing decisions are made on inaccurate or outdated data as our identities evolve from one moment in our lives to another. For example, what happens when a person who was previously a refugee is denied a bank loan following resettlement in a new home country or once they are back in their home countries

[5] Privacy International, 'Exclusion and identity: life without ID', 14 December 2018. Available at: <https://privacyinternational.org/long-read/2544/exclusion-and-identity-life-without-id>.

[6] See: 'Chapter 2.5.2 The purpose limitation principle', in: Kuner, C., and Marelli, M (eds.), *Handbook on Data Protection in Humanitarian Action*, Second Edition, 2020. Available at: <https://www.icrc.org/en/data-protection-humanitarian-action-handbook>.

because of their all well too documented past which limits their opportunities for the future on the basis of arbitrary criteria which categorise them as undesirable customers?

Getting Out of the Box

Identity systems are often built as rigid systems where individuals are categorised according to a set of criteria ranging from gender or age, to location, etc. and they are provided with a unique identifier. And once you are in a box it is very hard to get out of it, and it is almost impossible for you to claim that you should not be in it, and also it denies a person the possibility for different identities to relate to one another in different situations.

When it comes to concrete examples of when a digital identity system can go wrong in the humanitarian action, the current fiasco which has been unfolding in Kenya for many years now is very illustrative.[7] An estimated (although the number could be much higher) 40,000 people are unable to register for a Huduman Numba, the national ID being rolled out in Kenya[8], because in the eyes of the Kenyan Government they are recognised as refugees as they are enrolled in United Nations High Commissioner for Refugees (UNHCR)'s proGres system. A process of deduplication is underway but for now thousands of people are left in limbo, unable to access and enjoy fundamental rights such as education, health care, the right to vote, as well as economic opportunities.

The reason why all of these risks cannot be ignored and must inform the decisions made by humanitarian organisations is that their activities are playing out in a broader ecosystem where surveillance and the exploitation of data is rampant, unregulated and builds up a reliance on technology and data which is fragile and uncertain.[9] Humanitarian organisations must understand this broader ecosystem to identify what internal and external measures they must put in place to protect their own organisations, the people they assist, and the people they work with.[10]

This means that decisions about creating a digital identity must be embedded and informed by this particular context, not only to question the need and proportionality for it as previously

[7] Weitzberg, K., 'In Kenya, thousands left in limbo without ID cards', 13 April 2020. Available at: <https://www.codastory.com/authoritarian-tech/kenya-biometrics-double-registration/>.

[8] Privacy International, 'Why the Huduma Namba ruling matters for the future of digital ID, and not just in Kenya', 6 February 2020. Available at: <https://privacyinternational.org/news-analysis/3350/why-huduma-namba-ruling-matters-future-digital-id-and-not-just-kenya>.

[9] See: <www.privacyinternational.org>.

[10] International Committee of the Red Cross (ICRC) and Privacy International, 'The humanitarian metadata problem: "Doing no harm" in the digital era', October 2018. Available at: <https://privacyinternational.org/report/2509/humanitarian-metadata-problem-doing-no-harm-digital-era>.

noted, but also to see how the benefits and, importantly, the risks could play out – and in particular this requires knowing, understanding and responding to the needs but also possible issues of vulnerability of their own organisations and the very people whom they assist, and being prepared for when things go wrong such as a breach or a misuse of data.

If not, there is a risk for humanitarian organisations not only doing more harm than good, but also not providing the assistance urgently needed by millions.

A DONOR'S PERSPECTIVE
LA PERSPECTIVE D'UN DONATEUR

Catherine Kayser
Luxembourg Ministry of Foreign Affairs

Résumé

Catherine Kayser est attachée aux affaires humanitaires au Directorat pour la coopération au développement et les affaires humanitaires du ministère des Affaires étrangères et européennes du Luxembourg. Lors de son intervention, elle a présenté la perspective d'un donateur sur la question de l'identité numérique dans l'action humanitaire en situations de conflit et souligné qu'il est important pour la communauté humanitaire, y compris les donateurs, de ne pas se tromper dans ce domaine.

Dans une première partie, cette contribution met en évidence l'importance que le Luxembourg accorde à l'innovation numérique mais aussi à la protection des données dans l'action humanitaire. Il est favorable à ce que la communauté humanitaire tire parti du potentiel des nouvelles technologies et des données numériques pour améliorer l'action humanitaire tout en respectant les principes humanitaires fondamentaux, en protégeant la dignité et en préservant la confiance des personnes affectées. En effet, l'objectif commun de l'action humanitaire est d'assister les populations les plus vulnérables de manière digne et éthique sans causer de dommages ou de risques supplémentaires.

Dans une deuxième partie, Catherine Kayser souligne les risques associés à l'utilisation de solutions numériques telles que les outils d'identité numérique. Le fait d'opérer dans des zones touchées par des conflits ajoute un niveau de complexité à l'utilisation des solutions numériques. Il est essentiel d'être conscient des implications pratiques dans de tels contextes et de savoir si ces outils sont adaptés pour servir la population touchée sans causer de dommages supplémentaires. Les risques doivent donc être évalués de manière adéquate, afin d'identifier et d'appliquer des mesures de protection appropriées pour atténuer ou éliminer ces risques. De plus, Catherine Kayser note qu'il n'existe pas d'approche unique, du fait de la diversité des contextes. Un risque important est le possible détournement ou la mauvaise utilisation des données personnelles collectées. Ainsi, la protection des données concernant l'identité numérique dans les situations d'urgence humanitaire nécessite une approche particulièrement prudente et flexible lors de l'application des principes de protection des données.

Dans une troisième partie, elle précise le rôle et la responsabilité que peuvent avoir les donateurs pour sensibiliser aux risques potentiels et aider à les atténuer. Les conflits prolongés incitent la communauté humanitaire, y compris les donateurs, à réfléchir à la manière d'améliorer l'efficacité des interventions humanitaires. Toutefois, cela ne doit pas compromettre la protection des bénéficiaires. En ce qui concerne le lien entre protection et identité numérique, les donateurs doivent donc écouter leurs partenaires et les experts. Les partenariats sont un outil essentiel dans ce domaine. Ainsi, le Luxembourg soutient les initiatives visant à atténuer les risques auxquels est exposée la protection des données, et encourage et promeut l'utilisation responsable et éthique des technologies et des données dans l'action humanitaire. La communauté des donateurs peut influencer de manière significative les pratiques humanitaires. Elle doit aider à définir et à atteindre les normes éthiques les plus élevées possibles afin de protéger les populations concernées. Catherine Kayser conclut qu'il relève d'une responsabilité collective de mettre la protection des personnes au premier plan et d'adhérer au principe 'ne pas nuire (numériquement)'.

My intervention on 'New Technologies and Humanitarian Action: Digital Identity' at this Colloquium will focus on Luxembourg's perspective as a donor on the issue of digital identity in conflict situations in humanitarian action and why it is important for the humanitarian community, including donors, to 'get it right' in this area.

I would like first to briefly underline the importance that Luxembourg has devoted to digital innovation in humanitarian action before focusing on the opportunities and risks associated with the use of digital solutions such as digital identity tools. Lastly, I will provide Luxembourg's perspective on the role and responsibility of donors to raise awareness and help mitigate potential risks.

Importance of Digital Innovation in Humanitarian Action

In its humanitarian policy, Luxembourg has been a strong advocate for and supporter of digital innovation as well as protection in the digital era and has recognised the need to harness the potential of technological solutions as critical enablers of the humanitarian response. But it is critical that people in need of humanitarian assistance and affected communities remain at the centre of all interventions, be they of digital or human nature.

Luxembourg is therefore thankful for the tireless efforts of the humanitarian community to leverage the potential of new technologies and data to improve humanitarian action while respecting fundamental humanitarian principles, protecting the dignity of affected populations and preserving their trust. Indeed, the common and ultimate objective of humanitarian action is and must be to assist the most vulnerable populations in a dignified and ethical manner without causing additional harm or protection risks.

Implications of the Use of Digital Solutions and Digital Identity in Conflict Situations

The issue of digital identity in conflict situations is a critical and complex one that does not only offer widely acknowledged opportunities but also presents potential protection risks to affected persons. In addition, it is clear that operating in conflict-affected areas adds a layer of complexity to the use of digital solutions, and the data collected and generated through them, and raises many concerns that require our attention.

It is therefore crucial to be aware of the practical implications of digital identity tools in volatile contexts, and to know whether they are 'fit for purpose' to serve the affected population without causing additional harm. This means that risks must be adequately assessed in order to identify and apply appropriate safeguards to mitigate or eliminate these risks. In this regard, it is important to note that there is no 'one size fits all' approach as every context and conflict is different and consequently requires a different approach.

Luxembourg is mindful of the existence of potential constraints and risks in the (mis-)use of digital identity to leverage humanitarian effectiveness, especially in terms of protection of the most vulnerable populations and in volatile contexts. The possible re-identification of vulnerable persons, including refugees and other people fearing persecution, due to the potential misuse of their personal data collected to provide them with humanitarian assistance is of particular concern.

This is also where the issue of personal and sensitive data comes into play. Indeed, special data protection challenges regarding digital identity in (humanitarian) emergencies require a particularly careful and flexible approach when applying data protection principles.

Protection Risks and the Role and Responsibility of Donors to Raise Awareness and Help Mitigate these

As a donor increasingly supporting innovative initiatives with and through its trusted partners, Luxembourg has recognised the need to develop and implement data-driven solutions in a way so as not to undermine the dignity, safety and rights of affected and vulnerable populations. On data protection and responsibility more generally, this means that when engaging with partners, it is very important for Luxembourg that the challenges and risks associated with using humanitarian data, including personal data of crises-affected individuals, are part of the discussion.

The issue of protracted conflicts and consequently protracted humanitarian contexts must lead the humanitarian community, including donors, to reflect further on how to advance the effectiveness of our humanitarian interventions without compromising the protection of those we are seeking to assist. In the face of protection crises around the world, contributing to exacerbating the vulnerabilities of affected populations by implementing digital tools without properly assessing the risks is not an option.

In terms of the important linkage between protection and digital identity, and understanding its serious implications for people affected by conflict, donors must therefore listen to their partners and experts, and keep asking the right and often difficult questions, many of which remain unanswered.

As such, Luxembourg remains firmly committed to supporting and exploring initiatives that aim to mitigate protection risks associated with the use of digital solutions, and will continue to encourage and promote the proactive advancement of the responsible and ethical use of data and technology in humanitarian action, including in the critical area of digital identity. To this end, valuable and trusted partnerships in the field of data protection in humanitarian action are key.

The donor community being a key stakeholder in humanitarian action, which can significantly influence humanitarian practices, not only needs to help set high technical standards but also strive for the highest ethical standards regarding digital identity tools in order to protect affected populations. It is our collective responsibility to put the protection of people first and to adhere to the 'do no [digital] harm' principle, even when new technologies and digital solutions are developed with the 'sole' purpose of improving humanitarian assistance to people in need.

DISCUSSION

The discussion allowed the following topics to be raised and discussed in depth

1. Some Concrete Examples of the Risks Related to Digital Identity in Humanitarian Settings

The moderator opened the discussion by asking for concrete examples of the risks related to the use of digital identity in humanitarian settings. Before sharing examples, a first panellist pointed out that in many cases, follow-up or monitoring mechanisms to see when things go wrong are not established in the humanitarian sector or they are not systematic, or they are not connected, for example, between decision making at headquarters and field work. However, for an organisation which does not have indicators, triggers or factors to know what looks right or to indicate that something has gone wrong, the red flags on the risks will not come up. The limited internal culture around reporting, documenting and learning of incidents explains the difficulty to identify examples, not because there are not things that go wrong but because nobody is monitoring these types of incidents.

The panellist then gave two examples. A first example relates to the design of an identity system. Identity systems are quite rigid by design and they categorise people based on names, vulnerability assessment and other criteria. The panellist explained that 'people are put in these different boxes', on which decisions are based. However, once someone is in a specific category, it becomes very hard to get out of it. It is then up to the individual to provide information and make a claim that they no longer belong in that category. Such a case has been unfolding in Kenya for many years now and very little has been done so far to improve the situation: an estimated 40,000 people had been unable to register for an identity card because in the eyes of the Kenyan Government they are registered as refugees and enrolled in the UNHCR proGres system. These 40,000 individuals are not refugees even though they have been registered as such. However, they are not being recognised by the Kenyan Government and cannot access the identity system in Kenya which is mandatory to access a variety of services. In practice, this means that these individuals are not able to actively participate in public life in Kenya, to go to university or to access healthcare, for example. The panellist added that there is an ongoing identification process but there are still thousands of people left in limbo and the identification process is impacting different minority groups.

The panellist gave a second example. There is a platform called RedRose which is widely used in the humanitarian sector for beneficiary tracking. This platform was actually exposed because of insecurity and vulnerability, i.e. a third party was able to access the cloud server of an organisation which contained names, photographs, family details and phone numbers. The panellist specified that thankfully there was not a militia actor among these third parties and that steps were taken to address the breach. However, what would have happened if the third party was a militia, if these data were falling into the wrong hands? What would be the implications for some of the registered individuals?

Lastly, the panellist mentioned a recent guide[1] which explores some of those risks around the right to privacy, data protection, the use of biometrics and the impact on other fundamental rights such as human dignity, the right to life, equality and non-discrimination. It also goes through the legal arguments that were presented in legal challenges against national ID systems being deployed in a non-humanitarian setting. The panellist suggested that some of these discussions would be useful for a humanitarian organisation as well.

2. Preventing and Mitigating the Risks

Building on the risks that were previously highlighted, the moderator asked what was done and could be done to avoid and to mitigate those risks, including with regards to transparency or accountability, and to ensure continuous improvement and learning of humanitarian organisations' best practices.

The panellist first recognised that humanitarian organisations need identifiers, rather than an identity system *per se*. The identifier is supposed to be a neutral construct. In terms of specific measures or practices, a humanitarian organisation may have preventive processes and inclusive approaches within its infrastructure. It can aim to avoid or minimise both inclusion and exclusion errors. The panellist acknowledged that it may be challenging for a humanitarian organisation that operates on a very broad scale to make progress quickly on these topics precisely because it operates in so many different contexts. In fragile or crisis situations, there is a significant risk of data exposure. This means that in some countries, an organisation may have to restrict acquiring eID in order not to expose the personnel to the pressure of parties who may want to obtain this information. On the technical front, the panellist suggested that an organisation may implement the so-called Hippocratic database, which supports anonymised operations, and which may incorporate data privacy by design. However, the panellist also recognised that those are emerging new approaches. In some contexts, this may entail significant risks, for instance a risk that local authorities try to link registries to election registries. In such cases, an international organisation may assert the extraterritorial nature of the data. However, the panellist suggested that this remains challenging in practice. The panellist concluded that overall, the best tool is to have a robust, effective and updated

[1] Privacy International and the International Human Rights Clinic at Harvard Law School, *A Guide to Litigating Identity Systems*, September 2020

database and an impact assessment process, not at a single point in time but continuously. According to the panellist, this is the best tool humanitarian organisations have to inform the operations of the opportunities and risks as well as to be able to inform the plans of action in case of incidents.

The moderator emphasised the benefits of occasionally relying on privileges and immunities and extraterritoriality in dealing with situations where it is necessary to protect individuals. It is one of the lines of work that the ICRC has been focusing on, as the privileges and immunities could very well be the first line of protection of individual data. This brings a series of interesting reflections on how privileges and immunities enable an organisation to perform its mandate in an independent manner and *vis-à-vis* whom and for whose benefit in relation to the handling of personal data of individuals in humanitarian action. This is one of the key elements of safeguarding individuals and ensuring the delivery of humanitarian aid in a neutral, independent and impartial manner.

3. Challenging the Necessity and Relevance of Digital Identity

A panellist commented that some organisations talk about identifiers. However, it is unclear how different an identifier is from an identity, especially from the point of view of the risk it creates. It has been reported that local authorities sometimes ask for the digital identities collected for humanitarian purposes to be linked with their own database. However, local authorities do not need identity to link a database, they need identifiers. Furthermore, real identities are fluid and changing constantly, and digital identities fix these fluid real identities. An identifier attaches a feature to an individual. The panellist concluded that a digital identity may not be a proper way of identifying people, in addition to bringing with it significant risks. While many in the discussion emphasised that such risks exist, the panellist suggested that more concrete examples are needed to fully understand these risks in practice.

The moderator complemented this remark with a question on what would be technologically feasible in practice in terms of identifiers. In the offline non-humanitarian world, an individual who would need to demonstrate their identity would show their passport or other identity documents. However, if an individual in the digitalised world is asked about the possibility to authenticate or demonstrate that they have certain features that entitle them to an item or interaction – e.g. that they are below a certain age or are coming from a region with a particular aid programme – how, with technology, can they overcome the need to have to give an organisation a pack of information and instead just assert towards the organisation the features that are going to enable an interaction? The panellist answered that these technologies exist or are being developed. They allow an individual to prove attributes about themselves, e.g. to prove that they belong to a specific group or they are entitled to have help. It would even be possible to encode that someone is entitled to get two items on a specific day or that someone is entitled to only one item, and then check if someone has already collected the items they are entitled to without ever knowing who the individual was.

A participant also asked whether there would be an advantage in dealing with group identity rather than individuals. The participant wondered whether for instance there would be specific groups or regions that need help, especially as one could consider that International Humanitarian Law is based on groups and not individual identity, e.g. the wounded or sick. The panellist noticed that what is called group identity in this question is in fact not an identity. When someone wants to prove that they belong to a group, it means being part of a group of people that have particular needs, not identity. On the idea of group identity, another panellist later commented that theoretically, the problem is that no one wants to discriminate by saying that people are only the members of a group. Ideally, an organisation would rather address the individual needs of the persons, not the fact that they have this ethnic origin or that they live in this city or this province.

A panellist concluded that there are many applications for technology that the humanitarian community may not be aware of, and conversely, there are also many humanitarian issues that technologists are not aware of. The panellist suggested that more interactions may be needed. In particular, the technological sphere, which is more used to talking with big private companies, could talk more with humanitarian organisations to help them solve their problems. The moderator agreed with the panellist on the importance of the tech sector and of discussing further how to approach partnerships that can be shaping identity programmes.

4. The Role of Donors

The moderator then raised two questions on the role of donors. A first question related to due diligence processes and to how donors could help humanitarian actors to ensure accountability and transparency. Another question from a participant related to possible recommendations to convey to younger generations of diplomats who would like to leverage technology for humanitarian action.

Building on previous comments and questions, a panellist explained that donors themselves should take part and should help decrease vulnerabilities and not increase them. In that sense donors also play an important role in ensuring that the humanitarian sector is accountable and transparent, especially to the beneficiaries. One of the difficulties is to keep up with adequate protection mechanisms and safeguards. There is the common idea that the humanitarian sector needs to keep the same pace or keep up with the pace of digital transformation. However, the panellist agreed with previous comments that it may be time to take a step back and see what could be done to help beneficiaries without causing more harm.

The panellist also noted that this question forces donors to reflect on their own due diligence and on how to help mitigate potential protection risks. A donor who supports innovative solutions and initiatives with and through its partners needs to push for them to be implemented in a way as not to undermine the rights, dignity and safety of affected populations. The panellist suggested that as the ICRC has emphasised, there is a digital dilemma[2]: many times, beneficiaries are not given the choice about their data being collected, even if they know the potential risks and if they understand how these data could do more harm than good. Therefore, on data protection and data responsibility, and when donors engage with their trusted partners, it is very important that the challenges and risks associated with humanitarian data are part of the discussion so that they really try to ensure the highest ethical standards. Luxemburg for instance remains firmly committed to supporting and investing in and exploring initiatives that aim to mitigate these risks associated with the use of digital solutions. Accountability is a very important question for donors, but some trade-off needs to be made in order to avoid oversharing vulnerable data. It is important to contextualise due diligence requirements to the field and to the given environment.

The panellist concluded that this might be a message that could be conveyed to the younger generation of diplomats. Donors should be encouraged to really think about what kind of data they need as donors, what they really need to know. More specifically, do they need to have all the specifics? Would it not be enough that donors know the person needs humanitarian assistance and to focus on having more flexible requirements and on adapting these to a given context?

Another panellist stressed that investments are needed to implement solutions. If donors ask for humanitarian organisations to be transparent and accountable, effective and efficient and at the same time to be respectful of the beneficiaries and to avoid doing harm, there needs to be investments in the tools. The humanitarian community, including donors, may need to accept the fact that new, different and specific technologies are needed because the tools that exist do not fit the needs and requirements. This entails investment of time and delaying the deployment of technologies in the field. However, if the humanitarian community cannot afford all the requirements because humanitarian organisations deal with emergency situations, or if these investments in specific tools are not put in place, then the humanitarian

[2] ICRC website, Digital Dilemmas. Available at: <https://digital-dilemmas.com/>. This ICRC initiative aims to give a graphic representation of the risks and issues that are escalated by the digital dimensions in conflict.

community must also accept and recognise that its action is not fully respectful and harmless and that it is resorting to tools and processes that bear risks. This should be a conscious choice, which entails a privacy impact assessment and the knowledge that the risk exists. The impact assessment does not mean that the risk is avoided: some risks are present, so that the humanitarian community has to accept that it is doing harm and that it must weigh the benefits against the harm. However, this is another discussion.

5. Data Protection and Children

The moderator then invited the panellists to answer several questions that related to the protection of children’s data and how child biometrics and data protection are viewed in international law.

A panellist said that on the legal side, the Convention on the Rights of the Child includes specifically a right to privacy and the protection of the personal data of children. That could be one avenue to challenge the use of specific data technologies with children. The panellist noted that new technologies are now being deployed, particularly in school systems. For instance, biometrics can be used for paying for meals as part of the broader digitalisation of access to social protection that might be managed through schools. Biometric fingerprinting of students and teachers is also used for school attendance in some countries. The panellist said that this relates back to the question of the purpose behind such a program, and especially of whether it is necessary and proportionate to what an organisation is trying to do. For instance, in a situation in which children who had not signed to the plans try to access the meals: should it be considered a question of fraud at that level or a question of misuse of the resources that were being given? In addition, these systems might not even be helping to address these problems, while at the same time creating more risks. The answer may also differ depending on whether it is a one-to-one biometric system or one-to-many, whether it is checking if that individual child is in the system or whether it is cross-checking that individual child to a database where all the children who are registered are allowed to benefit from the system. From the perspective of the panellist, this is just one of unfortunately many examples where the techno-solutions are aimed at a problem that is not even that big or that relevant and are creating more risks.

The panellist emphasised that normalising these practices may not even be desirable. Children who start to use biometrics from a very early age might think that this is normal. Fingerprinting biometrics may not be as invasive as having a biometric scan of their iris, nonetheless, this contributes to a normalisation in seeing this kind of intrusive technology in many areas of people’s daily lives.

Another panellist added the example of a situation which happened in one of the schools linked to the European Union institutions – about which there is a decision of the data protection authority. Theoretically the system which was introduced in the school was not a mandatory system, but a voluntary one. Children had the option to opt out of the biometric data system. The result was that all the children who opted out were able to get their meals, however, they had to wait at the end of the queue, they were the last to be served and were also stigmatised as 'those who do not want to use these new technologies'. The panellist noted that this example refers to the previous comment on getting used to new technologies from early childhood. This also relates to proportionality. Proportionality means that involvement in sensitive data technology can be done only when there is no possibility to achieve the purpose without it. The mere fact that an organisation is giving the choice is not necessarily the response to all the questions.

6. Lessons Learned from Non-Humanitarian Settings

The moderator noted that the European Union (EU) has been at the forefront for quite some time in the area of data protection, including in the development of detailed analysis and guidance on the application of the principles of data protection in relation to specific technologies. Many developments are happening at the EU level in the area of digital identity. The moderator therefore asked whether there would be lessons that the humanitarian sector could learn from the current developments and proposals at the EU level.

A panellist confirmed that the humanitarian sector could learn from other policy experiences. However, the panellist suggested that these lessons would be better coming from a national level. This relates to the fact that there are different systems and different sets of data across EU Member States. For instance, some of the data which are collected in the Belgium or Polish systems cannot be collected in Germany. As a result, there are significant differences between the countries. The panellist mentioned as an example the German experience in the use of a semi-identifying ID which allows the humanitarian organisation to read only those features and data that are necessary for an operation without revealing the identity of the individual. Other examples can be drawn from challenges by the data protection authorities and by the data protection activists about any way enabling the State to keep track of where the person is located or where the person is using the services.

However, the panellist warned against transferring examples from one continent to another continent, as this may lead to other mistakes. For instance, when comparing the solutions used for contact tracing in the context of the Covid-19 pandemic, similar applications which had been developed in the United States or in Europe were used in Africa several years ago. They did not work with African societies because there are different ways of using mobile phones and different ways of identifying the person through the use of the device. The devices are very personal in Europe or in America, while it is not necessarily the case in Africa. Therefore, the panellist stressed the importance of the differences between countries and cultures.

7. Data in Humanitarian Action and Sanctions

A participant raised a last question focusing on the challenges arising from restrictive measures, counter-terrorism and sanctions. The moderator more specifically asked to what extent existing policies, techniques and approaches take into account and address concerns that data related to humanitarian action may be sought by parties or donors to implement restrictive measures such as counter-terrorism. The moderator recalled that a panellist had mentioned the occasional reliance on privileges and immunities, which is also a very important factor for the ICRC. This was indeed an important issue that was discussed at the last International Conference of the Red Cross and the Red Crescent, which encompasses all the State Parties to the Geneva Conventions as well as all the components of the Movement. The Conference led to the adoption of a resolution on restoring family links, and data protection and privacy[3]. The moderator explained that there was a strong emphasis on ensuring that the data that has been collected exclusively for humanitarian purposes is respected and that this exclusivity is adhered to. This is important because it is a fundamental data protection principle, but it is also a question of trust and access to humanitarian services, particularly in conflict areas. The moderator then asked to what extent these concerns are taken into account in other international organisations' working terms.

The panellist specified that in terms of process on the measures and checks to be made. WFP, for instance, checks those against the United Nations' lists. The department of the UN maintains those lists and there is a process for checking vendors. For beneficiaries, the moderator noted that it is not done. It is not asserted as such because they do not gather those data for this purpose. This has come up in the past in WFP's operations, when there was an inquiry, as sometimes there is an inquiry from a donor. However typically it is not part of what the organisation does.

As a Concluding Remark, a panellist pointed out that there is a lot to be done in the context of a widely changing dynamic environment. Laws are changing, technology is changing, and donors' expectations are also changing. Ensuring the humanitarian abilities to address all of these and the continuous focus on beneficiaries is a tremendous challenge. A panellist em-

[3] Resolution 4 of the 33rd International Conference of the Red Cross and Red Crescent, 'Restoring family links while respecting privacy, including as its relates to personal data protection', 2020. Available at: <https://international-review.icrc.org/sites/default/files/pdf/1590391258/irc101_2/S1816383120000090a.pdf>.

phasised that technologies are not toys, they are tools, and they should be used as tools, in a careful manner. This is especially true for the biometric ones, which can stay longer than the problem that humanitarians are trying to solve. Another panellist suggested thinking again, as identity might not be the automatic solution to all the problems the humanitarian community is facing. Another panellist added that it is not because organisations can do something that they should, while yet another panellist concluded on the importance of not losing sight of realities on the ground.

CLOSING REMARKS

Knut Dörmann

Ladies and Gentlemen, dear Colleagues,

It is now time to wrap up this Colloquium. I hope that you agree with me that we have had a Colloquium which was really rich in terms of an exchange of views, expertise, and ideas. It has certainly been a new experience because we have had to run the Colloquium in a virtual way, but that provided the opportunity to reach a broader audience than we did in a physical setting.

There were two overarching topics that permeated the discussions this week. First the interaction between new technologies and humans, as well as the way this affects International Humanitarian Law (IHL). Second, the clearly expressed need for an increased educational effort. Indeed, the technologies we discussed are already in use or about to be deployed while it seems that the users and the decision makers themselves do not sufficiently know or understand the characteristics and potentialities of these technologies.

It is, of course, not possible to give an extensive overview of the discussions of this week in a few minutes, but the full proceedings will be published in the College of Europe's journal, 'Collegium'. I would however like to highlight a few points from the different sessions of this week.

The first session, which I was pleased to moderate, underlined the need to have in-depth discussions on cyber warfare, which has become a pressing issue with the digitalisation of our societies. This is even more crucial as States barely acknowledge using such cyber operations, whilst at the same time developing their cyber capabilities that could have harmful effects on civilian infrastructure and services.

Our experts provided insights into how different States position themselves on questions regarding the applicability of IHL to cyber operations during armed conflict and also a number of reasons for their positioning. What I have found especially valuable was the open exchange among participants about how existing rules or principles of IHL apply to cyber operations during armed conflict and how we may address issues on which different interpretations of IHL exist.

On Tuesday, when discussing autonomous weapons, the panellists focused on legal and ethical concerns around the need to ensure human control and to put strict limits on the autonomy of such weapons. Although concerns are widely shared, States often have differing views on how to address them. It was interesting to note, for instance, that for France, lethal autonomous systems do not exist and that, for now, they do not seem desirable for the military. In France's view, without human supervision these weapons would not pass a legal review. The discussions around ethics were also equally thought provoking. not least because of different ethical perspectives. Beyond the legal, ethical, and operational aspects, the speakers and the audience acknowledged that the discussion on AWS encompasses broader geopolitical and security considerations. Overall, the discussion demonstrated divergent viewpoints on the relationship between AWS and IHL, as well as several concerns regarding the potential impact on civilians, calling for further legal and political discussions and clarifications on AWS.

Moving to the third session. on Artificial Intelligence (AI), a core characteristic of AI is that it develops its own analysis based on the data it receives. AI represents both a potential for better decision-making and potential risks, as it might come with a loss of human judgement and an increasing speed of decision-making in warfare that goes beyond human capability.

A key take-away from the discussion is that the important challenge is to translate very complex IHL requirements in technical terms. This challenge is not only a legal and operational one, but also a political one. One speaker stressed that AI emerged in a new setting of world power competition and in a competitive strategic environment, bearing new risks and new threats. The core challenge is therefore no longer to develop technologies but rather to conduct a 'technological adaptation race' to the existing legal frameworks and to adopt reforms where needed. It was also emphasised that the military has a shared interest with civilians when it comes to the predictability, the reliability, and the traceability of their equipment. It is clear that, from an operational as well as from a governance perspective, trustworthiness and reliability of AI are core characteristics that need to be further explored, among others for the purpose of defining accountability. Other important points of discussion related to the potential biases in algorithms and their consequences on some groups of population, as well as the potential opportunities that AI could offer to reduce unintended consequences on civilians.

With yesterday's panel, we travelled to space. The reminder of how much space and space activities impact our everyday life, including military activities, was impressive. The comparison of the ways the US used space in the 1991 and then in the 2003 Gulf Wars, and took advantage of it, illustrated how much the use of space for military purposes has increased in just a few years.

The technological, economic, military, and strategic competition among States related to the use of space was highlighted, and with it the increasing potential risks of confrontation, as well as emerging alternative strategies. Respecting and developing norms applicable to space activities is a huge challenge, mainly because this requires consensus among States and in the current state of affairs, this consensus is quite often non-existent, as illustrated by the recent Artemis Agreement signed last week with only a handful of States, and some major players being absent. Noteworthy is that in such an environment a growing body of soft law and guidelines is developing, as for example in the two dedicated manuals: Woomera and MILAMOS. They may clarify some rules and offer a valuable resource for practitioners, but they obviously do not constitute law or address diverging views.

As was also mentioned yesterday, the more than 3,000 satellites currently in orbit are very vulnerable and there is a risk that they become targets. As space systems are connected to billions of people around the world, supporting a wide range of critical services, from airline traffic to banking, internet or disaster warning, the consequences in terms of human lives if they were targeted would be enormous. Their destruction or disruption could also have long-lasting effects due to the debris that would remain in outer space with no geographical boundaries. In this context, the normative framework on the protection of the natural environment also has a role to play.

Yesterday’s panel ended with an interesting reflection that preserving dominance in space increasingly becomes a strategic objective, not because States are willing to fight a war in outer space, but because they are willing to protect their assets in space, especially the ones which support and enable dominance in other domains.

During today’s session on the use of new technologies and humanitarian action, we heard how useful these technologies could be for the humanitarian sector, not least to combine digital proximity with physical proximity. New technologies have entered the toolbox of humanitarian organisations with all the new questions that it raises and the complexity it brings. The discussions that ensued were a good testimony of this reality.

In conclusion, we have covered a wide range of issues during this ‘Bruges week’, from legal to technical, ethical, and political point of views. I hope sincerely that you enjoyed this exercise as much as I did, really a new experience after many Colloquiums that I attended in my previous function.

Let me now conclude by thanking the many people who have worked extremely hard to make this series of events possible. I will start, first and foremost, with our moderators and speakers who provided a wealth of information and reflection in a very dynamic way, then our friends from the College of Europe, and in particular our two colleagues Maureen Welsh and Jonas Corneille who gave their constant and efficient support for this event to happen, our dear interpreters Nanaz Shahidi-Chubin, Olga Zalogina, Nataliya Kataeva and François Butticker without whom such an inclusive event across regions would never be possible. Then of course, the ICRC team from the Arms and Conduct of Hostilities unit at the Legal Division in Geneva, who helped designing the programme, and last but not least the Bruges Colloquium team from the ICRC Brussels delegation who all worked with great dedication, also trying to adapt to the ever-changing situation as the consequence of the Covid-19 pandemic, so thank you to Olga Peykrishvili, Eva Houtave, Charlotte Giauffret and Stéphane Kolanowski.

With this, I close the 21st Bruges Colloquium and I hope to see you all next year, for the 22nd edition of our annual Bruges rendezvous, that will be held, hopefully in Bruges, on the 21st and 22nd of October 2021.

Many thanks to all of you, stay safe in the extremely demanding time we are going through.

PARTICIPANTS LIST
LISTE DES PARTICIPANTS

430 participants from following countries joined this virtual Bruges Colloquium:

1. Albania
2. Algeria
3. Argentina
4. Australia
5. Austria
6. Azerbaijan
7. Belarus
8. Belgium
9. Canada
10. Chili
11. China
12. Denmark
13. France
14. Georgia
15. Germany
16. Greece
17. Ireland
18. Italia
19. Kazakhstan
20. Kyrgyzstan
21. Latvia
22. Luxembourg
23. Mali
24. Moldova
25. Netherlands
26. Norway
27. Pakistan
28. Poland
29. Philippines
30. Romania
31. Russia
32. Slovakia
33. Spain

34. Sweden
35. Switzerland
36. Turkey
37. UK
38. Ukraine
39. USA
40. Uzbekistan

College of Europe
Collège d'Europe
Brugge
Natolin

CICR

New Technologies on the Battlefield: Friend or Foe

21st Bruges Colloquium – Virtual Edition, 12-16 October 2020

Simultaneous translation into French / English / Russian
Traduction simultanée en anglais/français/russe

Monday, 12th October 14:00 – 15:45 (CET)

Opening Statements
Federica Mogherini, Rector of the College of Europe
Gilles Carbonnier, Vice-President of the ICRC

Panel 1: **Cyber Operations during Armed Conflict**
Les cyber-opérations en temps de conflit armé

14:15 – 14:25 **Introduction to the topic by the Chair: Knut Dörmann**, Head of the ICRC Delegation in Brussels

14:25 – 14:35 **Thresholds and Applicability of IHL Rules**
Speaker: **Prof. Vera Rusinova**, High School of Economics University, Moscow

14:35 – 14:45 **Challenges and Responses: Cyber Operations and the Notions of 'Attacks' and 'Objects' under International Humanitarian Law**
Speaker: **Prof. Hongsheng Sheng**, Shanghai University of Political Science and Law, China

14:45 – 14:55 **Improving Transparency: International Law and State Cyber Operations**
Speaker: **Prof. Duncan Hollis**, Temple Law School

14:55 – 15:45 **Discussion**

Tuesday, 13th October 14:00 – 15:30 (CET)

Panel 2: **The Use of Autonomous Weapon Systems: A Challenge to the International Rule of Law?**
L'utilisation des systèmes d'armes autonomes: un défi pour l'état de droit international?

14:00 – 14:10 **Introduction to the topic by the Chair: Maya Brehm**, ICRC Geneva

14:10 – 14:20 **Military Use of AWS Must Comply with International Law and, in Particular, with IHL**
Speaker: **Colonel Rudolph Stamminger**, French Ministry of Defense

14:20 – 14:30 **Ethics and Autonomous Weapon Systems**
Speaker: **Thompson Chengeta**, University of Southampton

14:30 – 14:40 **What are the Elements of Human Control and how can they Inform the Setting of Limits on Autonomous Weapon Systems?**
Speaker: **Netta Goussac**, SIPRI

14:40 – 15:30 **Discussion**

Wednesday, 14th October 14:00 – 15:30 (CET)

Panel 3: **Artificial Intelligence (AI) and Machine Learning**
Intelligence artificielle (IA) et apprentissage automatique

14:00 – 14:10 **Introduction to the topic by the Chair: Neil Davison**, ICRC Geneva

14:10 – 14:20 **AI-Supported Decision Making in Warfare: Far-reaching Implications**
Speaker: **Edward Hunter Christie**, NATO

14:20 – 14:30 **A Key Set of IHL Questions Concerning AI-supported Decision-making**
Speaker: **Dustin Lewis**, Research Director, Harvard Law Schoon Program and International Law and Armed Conflict

14:30 – 14:40 **Efforts to Govern (Military Applications of) AI**
Speaker: **Pauline Warnotte**, UNIDIR and University of Namur

14:40 – 15:30 **Discussion**

Thursday, 15th October 14:00 – 15:30 (CET)

Panel 4: **Military Space Operations: Constraints under International Law and Potential Humanitarian Consequences**
Opérations militaires spatiales : contraintes imposées par le droit international et conséquences humanitaires potentielles

14:00 – 14:10 **Introduction to the topic by the Chair: Heather Harrison Dinniss**, Swedish Defence University

14:10 – 14:20 **Constraints Related to the Use of Weapons in Outer Space under IHL**
Speaker: **Wen Zhou**, ICRC Geneva

14:20 – 14:30 **Beyond the Peaceful Use of Outer Space: Potential Conflicts?**
Speaker: **Mickael Dupenloup**, French Ministry of Defence

14:30 – 14:40 **Protecting Humans on Earth from War in Space**
Speaker: **Jessica West**, Canadian Peace Research Institute, Project Ploughshares

14:40 – 15:30 **Discussion**

Friday, 16th October 14:00 – 15:45 (CET)

Panel 5: **New Technologies and Humanitarian Action**
Nouvelles technologies et action humanitaire

14:00 – 14:10 **Introduction to the topic by the Chair: Massimo Marelli**, ICRC Geneva

14:10 – 14:15 **New Technologies: Why it is Relevant and Useful in a Humanitarian Context?**
Speaker: **Edgardo Yu**, Technology Division, Worl Food Programme (WFP)

14:15 – 14:20 **Technology Framework: What is Authentication and Identification, What is Functional Identity and Foundational Identity and Why does is Matter?**
Speaker: **Prof. Carmela Troncoso**, EPFL

14:20 – 14:25 **Data Protection Provisions Framing the Digital Identity and Biometrics (and the Main Problematic Areas)**
Speaker: **Wojciech Wiewiórowski**, European Data Protection Supervisor

14:25 – 14:30 **A Civil Society Perspective: Digital Identity in Humanitarian Action: What are the Implications for People and their Rights?**
Speaker: **Alexandrine Pirlot de Corbion**, Privacy International

14:30 – 14:35 **A Donor's Perspective**

Speaker: **Catherine Kayser**, Luxembourg Ministry of Foreign Affairs

14:40 – 15:30 **Discussion**

15:30 – 15:45 **CLOSING REMARKS**
Knut Dörmann, ICRC Brussels

SPEAKERS' BIOGRAPHIES
CURRICULUM VITAE DES ORATEURS

Day 1: Monday, 12 October

Welcome Addresses
Allocutions de bienvenue

Mrs Federica Mogherini has been the Rector of the College of Europe since September 2020. She has co-chaired the United Nations High Level Panel on Internal Displacement since January 2020. From 2014 to 2019, she served as the High Representative of the European Union for Foreign Affairs and Security Policy and Vice-President of the European Commission. Prior to joining the EU, she was the Italian Minister for Foreign Affairs and International Cooperation (2014), and a Member of the Italian Chamber of Deputies (2008-14). In her parliamentary capacity, she was head of the Italian delegation to the NATO Parliamentary Assembly and Vice-President of its Political Committee (2013-14); a member of the Italian delegation to the Parliamentary Assembly of the Council of Europe (2008-13); Secretary of the Defence Committee (2008-13); and a member of the Foreign Affairs Committee. She also coordinated the Inter-Parliamentary Group for Development Cooperation. Federica Mogherini is a fellow of the Harvard Kennedy School. She is also a member of the Board of Trustees of the International Crisis Group, a fellow of the German Marshall Fund, a member of the Group of Eminent Persons of the Preparatory Commission for the Comprehensive Nuclear Test Ban Treaty Organization, a member of the European Leadership Network for Multilateral Nuclear Disarmament and on Proliferation, and a member of the Board of Directors of the Italian Institute for Foreign Affairs (IAI). Federica Mogherini has a degree in Political Science from the University of Rome 'La Sapienza'. She was born in Rome in 1973, she lives in Belgium and has two daughters.

Dr Gilles Carbonnier is the Vice-President of the International Committee of the Red Cross (ICRC). (Appointed in 2018). Since 2007, Dr Carbonnier has been a professor of development economics at the Graduate Institute of International and Development Studies (Geneva), where he also served as Director of Studies and President of the Centre for Education and Research in Humanitarian Action. His expertise is in international cooperation, the economic dynamics of armed conflict, and the nexus between natural resources and development. His latest book, published by Hurst and Oxford University Press in 2016, is entitled 'Humanitarian Economics: War, Disaster and the Global Aid Market'. Prior to joining the Graduate Institute, Dr Carbonnier worked with the ICRC in Iraq, Ethiopia, El Salvador and Sri Lanka (1989–1991), and served as an economic adviser at the ICRC's headquarters (1999–2006). Between 1992 and 1996, he was in charge of international trade negotiations (GATT/WTO) and development cooperation programmes for the Swiss State Secretariat for Economic Affairs.

Session One: Cyber Operations during Armed Conflict
Première table ronde : les cyber-opérations en temps de conflit armé

Dr Knut Dörmann has been Head of Delegation of the ICRC Brussels delegation to the EU, NATO and the Kingdom of Belgium since June 2020. Previously he was ICRC's Head of the Legal Division and Chief Legal Officer (December 2007 - May 2020), Deputy Head of the Legal Division (June 2004 – November 2007) and Legal Adviser at the Legal Division (December 1998 - May 2004) (in charge of, among others, the law applicable to the conduct of hostility, cyber warfare, the protection of the environment, international criminal law). He holds a Doctor of Laws (Dr. jur.) from the University of Bochum in Germany (2001). Prior to joining the ICRC, he was Managing Editor of Humanitäres Völkerrecht - Informationsschriften (1991-1997), Research Assistant (1988-1993) and Research Associate (1993-1997) at the Institute for International Law of Peace and Armed Conflict, University of Bochum. Dr Dörmann has been a member of several groups of experts working on the current challenges of international humanitarian law. He has extensively presented and published on international humanitarian law, international law of peace and international criminal law. He received the 2005 Certificate of Merit of the American Society of International Law for his book 'Elements of War Crimes under the Rome Statute of the International Criminal Court', published by Cambridge University Press.

Prof. Vera Rusinova is a Professor of International Law at the Faculty of Law of the National Research University Higher School of Economics in Moscow (Russia), where she heads the School of International Law. The main fields of her research activities comprise international human rights law, international humanitarian law, use of force, the theory of international law, and application of international law to cyber operations. Vera Rusinova has published more than 75 scientific articles, book chapters, and papers on international law. Her last monograph is titled 'Human Rights in Armed Conflicts: Problems of Relationship between Norms of International Humanitarian Law and International Human Rights Law' (Moscow, 2017). She is a co-chair of the International Law Association's Committee on Use of Force: Military Assistance on Request. She is also a member of the Editorial Groups of 'International Justice' and the 'Journal of International Humanitarian Legal Studies', and a member of the Editorial Board of the 'International Cybersecurity Law Review'. In 2020 Vera Rusinova was leading the Research and Study Group working on the project 'Reshaping Public International Law in the Age of Cyber: Values, Norms, and Actors'.

Prof. Hongsheng Sheng is professor of Public International Law at Shanghai University of Political Sciences and Law, China, Ph.D supervisor at Guanghua Law School of Zhejiang University, China. He was British Chevening Scholar in 2000 and awarded LL.M (International Criminal Justice and Armed Conflict) at the University of Nottingham, UK in 2001, and Ph.D (Public International Law) at Wuhan University, China in 1996. From April 2004 to April 2005, he was the United Nations Expert on Mission for the MONUC in the Democratic Republic of the Congo, serving as a team leader of Military Observers, and as a senior liaison officer. He was also appointed by the Chief of the Mission Chair of the Independent Board of Enquiry to review international criminal cases. In April 2005, he was granted a United Nations Medal (In the Service of Peace). In June 2011, he was granted the title 'Qianjiang Professorship' by the People's Government of Zhejiang Province, China. He is Chief Expert for the joint international programme 'UK-China Collaboration for Conflict Prevention'. Up till now, he has published six titles including 'Challenges and Responses in International Criminal Law' (2017), 'Legal Aspects of Armed Conflict in Early 21st Century' (co-author, 2014), 'State Responsibility under International Law in Anti-Terrorism Campaign' (2008), 'United Nations Peacekeeping Operations: Legal Aspects' (2006), and over eighty articles in leading academic journals at home and abroad. His academic interests focus on international law, international relations, international organisation, international humanitarian law and international criminal justice.

Prof. Duncan Hollis is Laura H. Carnell Professor of Law at Temple University Law School in Philadelphia. He is editor of the award-winning Oxford Guide to Treaties (2012, 2nd ed., 2020) & (with Allen Weiner) International Law (2018). Professor Hollis has authored a series of articles on securing cyberspace, including (with Martha Finnemore) 'Constructing Norms for Global Cybersecurity', in the American Journal of International Law, and the forthcoming European Journal of International Law article, 'Beyond Naming & Shaming: Accusations and International Law in Cyberspace'. Professor Hollis is a non-resident scholar at the Carnegie Endowment for International Peace and an elected member of the American Law Institute. In 2016, Professor Hollis was elected by the General Assembly of the Organization of the American States to a four-year term on the OAS's Inter-American Juridical Committee. There, he serves as a rapporteur for a project on improving the transparency of States' understanding of how international law applies in cyberspace.

Day 2: Tuesday, 13 October

Session Two: The Use of Autonomous Weapon Systems: A Challenge to the International Rule of Law?

Deuxième table ronde : l'utilisation des systèmes d'armes autonomes : un défi à l'ordre juridique international ?

Ms Maya Brehm is a legal advisor in the Arms and Conduct of Hostilities Unit (Legal Division) of the ICRC. Her present work focuses on IHL questions raised by developments in the technologies of warfare, notably increasing autonomy in weapons systems, and on the promotion of responsible arms transfers. Before joining the ICRC, Maya Brehm worked as a researcher, lecturer and policy adviser for academic institutions, civil society organisations and UN bodies in the fields of humanitarian disarmament and human rights. She holds a MA in international relations and an LLM in International Humanitarian Law.

Colonel Rudolf Stamminger joined the French Air Force in 1996. As a legal advisor and specialist in the law of armed conflict, he has participated in numerous external operations (mainly in former Yugoslavia and Africa). He was assigned as Operation Legal Advisor to the NATO Joint Force Command Staff in Naples and as Legal Advisor for the Law of Operations to NATO's Allied Command Transformation (ACT) in the US. He was also the legal advisor to the Air Defence and Air Operations Command (CDAOA) in Paris where he dealt with the legal aspects of air operations conducted worldwide by the Air Force. Colonel Stamminger is also the director of the Advanced Course on the Law of Armed Conflict at the International Institute of Humanitarian Law in San Remo and teaches this subject at the University of Paris-Nanterre. Since the summer of 2019, he has been the head of the Office of the Law of Armed Conflict within the Directorate of Legal Affairs of the French Ministry of the Armed Forces, where he is in charge of the EWIPA (explosive weapons in populated areas) and SALA (systems d'armes létaux autonomes – autonomous lethal arms systems) files. He is also a Knight of the National Order of Merit.

Dr Thompson Chengeta is a European Research Council Fellow on Drone Violence and AI Ethics at the University of Southampton where he undertakes research and project-related leadership on autonomous weapon systems (AWS). His PhD thesis (University of Pretoria) was on international law and ethics relevant to the governance of AWS while his LLM thesis (Harvard Law School) was on elements that define human control over AWS. He has widely published on international law and AWS and participated in the research and writing of the 2013 Report on Lethal Autonomous Weapon Systems that was submitted to the UN Human Rights Council by the then UN Special Rapporteur on extrajudicial executions, Professor Christof Heyns. Thompson is an executive board member of the Foundation for Responsible Robotics and serves as a legal expert member of the International Panel on the Regulation of AWS and the International Committee for Robot Arms Control. He is also the African Region Lead for the Campaign to Stop Killer Robots. Dr Chengeta is the Director and founder of the Chengeta Diversity and Inclusion Consultancy (CDIC) and he is also a founding member of the Campaign to Stop Killer Robots' Working Group on Intersectionality (WGI).

Ms Netta Goussac is an Associate Senior Researcher at the Stockholm International Peace Research Institute (SIPRI)'s Armament and Disarmament areas, and a Special Counsel with Lexbridge. Ms Goussac has worked as an international lawyer for over a decade, including for the ICRC and the Australian Government's Office of International Law, and as a lecturer at the Australian National University. She has expertise in legal frameworks related to the development, acquisition and transfer of weapons. Ms Goussac has provided legal and policy advice related to new technologies of warfare, including autonomous weapons, military applications of artificial intelligence and cyber and space security. Since 2017, she has participated in the UN's Group of Governmental Experts on Lethal Autonomous Weapon Systems. She is one of the co-authors of the 2020 report 'Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control'.

Day 3: Wednesday, 14 October

Session Three: Artificial Intelligence and Machine Learning

Troisième table ronde : l'intelligence artificielle et l'apprentissage automatique

Dr Neil Davison is a senior adviser in the Arms and Conduct of Hostilities Unit within the Department of International Law and Policy at the ICRC's headquarters in Geneva. He represents the organisation on a range of weapons and disarmament issues, with a current focus on new technologies of warfare including autonomous weapons and AI. Prior to joining the ICRC in 2011, Dr Davison led initiatives on international security and diplomacy at the Royal Society (UK's national academy of science) and carried out research on arms control at the University of Bradford. His initial training was as a biologist. He holds a PhD in Peace Studies.

Mr Edward Hunter Christie is Deputy Head of the Innovation Unit at NATO Headquarters. He is an experienced public policy professional, with 20 years of work experience. He has held a succession of roles in research, industry, EU affairs, and NATO. His current work focuses on the development of NATO policies regarding emerging and disruptive technologies. He is the author of NATO's White Paper on Artificial Intelligence.

Dr Dustin Lewis is the Research Director of the Harvard Law School Program on International Law and Armed Conflict (HLS PILAC). With a focus on public international law sources and methodologies, Mr Lewis leads research into several wide-ranging contemporary challenges concerning armed conflict. Among his current areas of focus, Mr Lewis heads the research for the project on 'International Legal and Policy Dimensions of War Algorithms: Enduring and Emerging Concerns'.

Ms Pauline Warnotte is a researcher at the Security and Technology Programme of the United Nations Institute for Disarmament Research (UNIDIR) in Geneva (Switzerland) and a teaching assistant at the Law Faculty of the University of Namur (Belgium). Her areas of research and expertise include the law of armed conflict, the various aspects of weapons law, the legal review of new weapons, as well as the use of artificial intelligence and new technologies in the military domain. She is regularly invited as a speaker by academic institutions, civil society organisations and UN bodies in the fields of IHL and weapons law, and she recently co-authored a book on 'Robotisation des armées : enjeux militaires, éthiques et légaux' (ISBN: 9782717870954). Ms Warnotte studied law at the University of Liège, Belgium (2008) and holds a LL.M. in Human Rights from Saint-Louis University, Brussels, Belgium (2011). Her previous experience includes the IHL unit of the Belgian Ministry of Justice (2011-2013), the International and Humanitarian Law unit of the Belgian Ministry of Defence (2014-2019) and the chair of law of the Royal Military Academy (2019-2020). As part of her military duties, Ms Warnotte has been deployed in Qatar (2017) and Mali (EUTM Mali, 2017 and MINUSMA, 2018).

Day 4: Thursday, 15 October

Session Four: Military Space Operations: Constraints under International Law and Potential Humanitarian Consequences

Quatrième table ronde : les opérations militaires spatiales : les contraintes imposées par le droit international et les conséquences humanitaires

Dr Heather A. Harrison Dinniss is a Senior Lecturer at the Centre for International and Operational Law at the Swedish Defence University. Dr Harrison Dinniss' research focuses on the impact of modern warfare on international humanitarian law, on emerging military technologies such as cyber warfare, advanced and autonomous weapons systems and the legal aspects of human enhancement techniques on members of the armed forces. She is the author of 'Cyber War and Laws of War' (Cambridge University Press, 2012) which analyses the status and use of cyber operations in international law and the law of armed conflict. Dr Harrison Dinniss has served as a member of advisory groups to the Swedish Government on autonomous weapons systems and cyber operations, a member of the International Law Association's Study Group on Cyber Terrorism and International Law (2014-2016) and as a core expert for two projects to produce manuals on International Law Applicable to Military Uses of Outer Space (MILAMOS, 2016-18, & Woomera, 2018).

Dr Wen Zhou is Legal Adviser of the Arms and Conduct of Hostilities Unit of the Legal Division at the ICRC Headquarters in Geneva, Switzerland. Wen Zhou functions include, among others, coordinating, developing and representing the ICRC's legal and policy positions on a range of weapons issues, notably legal reviews of new weapons, weaponisation in outer space, and conventional weapons (e.g. landmines and cluster munitions). Prior to working at the ICRC headquarters in 2018, Dr Zhou was Head of Legal Department at the ICRC Regional Delegation for East Asia based in Beijing, where she worked in particular on the promotion and implementation of international humanitarian law in China, the Republic of Korea, the DPRK and Mongolia. Prior to joining the ICRC in 2013, she worked as Associate Legal Counsel in the World Bank Group (Washington D.C.) and Assistant Professor of Law at the Chinese Academy of Social Sciences (Beijing). Dr Zhou holds a Ph.D. in International Law and a Bachelor of Law from Peking University (China), and a Master of European Law from University of Paris I Panthéon-Sorbonne (France). She also studied at The Hague Academy of International Law and the Raoul Wallenberg Institute of Human Rights and Humanitarian Law of Lund University (Sweden).

Major Mickaël Dupenloup currently holds the position of Deputy Head of the Office of the Law of Armed Conflict within the Legal Affairs Directorate of the French Ministry of the Armed Forces. He has previously served as a legal adviser in the Air Force and the joint environment. In this role, he specialised in the fields of targeting, air operations, ballistic missile defence and space capabilities. He was successively deployed in Afghanistan, Germany, Italy, Chad, Qatar and Kuwait where he served as an operational legal advisor in support of air operations conducted by the French armies. In parallel to his professional activity, Major Dupenloup carries out teaching and research work within the academic world. He holds a Master I in Public Law, a Master II in Space and Telecommunications Law, as well as a Master I and II in Legal Advice to Commanders.

Dr Jessica West is a senior researcher at the Canadian peace research institute 'Project Ploughshares' and Managing Editor of the international Space Security Index project. Her research and policy work are focused on technology, security, and governance. She is currently developing a map of the existing normative landscape that shapes outer space activities. Dr West interacts regularly with key United Nations bodies tasked with space security and space safety issues. She holds a PhD in Global Governance and International Security from the Balsillie School of International Affairs (Ontario, Canada).

Day 5: Friday, 16 October

Session Five: New Technologies and Humanitarian Action

Cinquième table ronde : nouvelles technologies et action humanitaire

Mr Massimo Marelli is Head of Data Protection Office at the ICRC. He is a member of the Brussels Privacy Hub Advisory Board and is the co-editor of the Brussels Privacy Hub/ICRC Handbook on Data Protection in Humanitarian Action. He is also a member of the Advisory Board and a fellow of the European Centre on Privacy and Cybersecurity at the University of Maastricht. Before taking on this role at the ICRC headquarters in Geneva, Mr Marelli worked as an ICRC delegate in the field and as a legal adviser at the ICRC headquarters. Prior to joining the ICRC, he worked as a lawyer at the UK Office of Fair Trading, as a registrar at the EU General Court and as a lawyer in private practice.

Mr Edgardo Yu has served as the Chief of Beneficiary Services within the Technology Division of the United Nations World Food Programme since 2014. Responsible for transforming WFP's digital approach to beneficiary management and providing food assistance, his team leverages responsible data, digital and financial inclusion tools to help reduce hunger, while enhancing food security, nutrition and sustainable agriculture. Edgardo Yu's 30 years of experience includes serving in WFP as Chief of IT Policy, Architecture and Strategy, Manager for Infrastructure Development and Head of Information Security, and he has held various roles working towards systems coherence in United Nations' technology fora. Prior to joining the UN, Mr Yu pioneered several Internet service start-ups in the Philippines. He holds degrees in Physics and Computer Engineering from Ateneo de Manila University and has been a fellow for Digital Vision at Stanford University.

Prof. Carmela Troncoso is an assistant professor at EPFL, Switzerland, where she heads the SPRING Lab. Her research focuses on security and privacy. She holds a Master in Telecommunication Engineering from the University of Vigo (2006) and a PhD in Engineering from the KU Leuven in 2011. Before arriving at EPFL, she was a faculty member at the IMDEA Software Institute in Spain for two years; the Security and Privacy Technical Lead at Gradiant, working closely with industry to deliver secure and privacy-friendly solutions to the market for four years, and a postdoctoral researcher at the COSIC Group. Her work on Privacy Engineering received the CNILINRIA Privacy Protection Award in 2017, and recently she has led the DP3T effort towards privacy preserving digital contact tracing which is implemented by many countries.

Prof. Wojciech Wiewiórowski has been the European Data Protection Assistant Supervisor (EDPS) since December 2019. In addition to this, he is an adjunct-professor at the Faculty of

Law and Administration of the University of Gdańsk. Before his appointment, he served as the Assistant European Data Protection Supervisor (2014 to 2019) and as Inspector General for the Protection of Personal Data at the Polish Data Protection Authority (2010-2014). He was also Vice-Chair of the Working Party Article 29 Group in 2014. His areas of scientific activity include Polish and European IT law, processing and security of information, legal information retrieval systems, informatisation of public administration, and application of new IT tools (semantic web, legal ontologies, cloud, blockchain) in legal information processing.

Ms Alexandrine Pirlot de Corbion is Director of Strategy at Privacy International (PI). She manages and oversees the delivery of PI's strategic portfolio aimed at ensuring that innovative solutions serve both individuals and civil society while protecting their dignity rather than protecting State power and corporate interest. She has also been leading PI's work within the development and humanitarian sector. In 2018, she co-authored a report published by the ICRC and PI which explored the risks associated with the use of data and new technologies in the humanitarian sector, called 'The Humanitarian Metadata Problem: "Doing No Harm" in the Digital Era'. She also served as a member of the Advisory Group for the second edition of the 'Handbook on Data Protection in Humanitarian Action'. Previously, she was engaged in research and advocacy on issues relating to human rights, irregular migration, security sector reform, gender, conflict management, and human security. Ms Pirlot de Corbion graduated from the University of Birmingham with an MSc in Conflict, Security and Development following an LLM in International Law at the University of Westminster following a BA in Law with International Relations from Oxford Brookes University.

Ms Catherine Kayser obtained an LLM in Public International Law from Leiden University in 2015.

After spending a year with the United Nations' World Health Organisation (UNWHO) in Ethiopia to work on gender equality in relation to health, she decided to pursue her interest in international humanitarian law and protection issues by entering the Luxembourg Ministry of Foreign and European Affairs as a legal advisor before joining the humanitarian action team, where she helps support and promote Luxembourg's and its humanitarian partners' efforts to leverage the potential of new technologies while making sure that beneficiaries stay at the centre of each intervention.