

PEACE AND SECURITY

ADDRESSING THE THREAT OF AUTON- OMOUS WEAPONS

Maintaining Meaningful Human Control

**Matthew Breay Bolton, Matilda Byrne, Ryan Gariepy,
Emilia Javorsky, Volker Lehmann, and Laura Nolan**
January 2021



This paper argues for a legally-binding instrument on lethal autonomous weapons systems, such as an additional protocol to the Convention on Certain Conventional Weapons.



There is a need both for a legally-binding prohibition of certain autonomous weapons and for strong positive obligations to ensure meaningful human control of the use of force rooted in International Humanitarian Law and International Human Rights Law.



The Alliance for Multilateralism has endorsed voluntary Guiding Principles on LAWS. Its Member States should now lead the Principles' upgrade towards more robust international agreements.

PEACE AND SECURITY

ADDRESSING THE THREAT OF AUTON- OMOUS WEAPONS

Maintaining Meaningful Human Control

BACKGROUND AND INTRODUCTION

According to the International Committee of the Red Cross (ICRC), autonomous weapons systems are: »Any weapon system with autonomy in its critical functions – that is, a weapon system that can select (search for, detect, identify, track or select) and attack (use force against, neutralize, damage or destroy) targets without human intervention.«¹ Since 2015, efforts against lethal autonomous weapons systems (LAWS) have gained momentum among different stakeholders. Artificial intelligence experts, roboticists and scientists as well as high tech professionals and technology companies have expressed their reservations against their research, products and platforms being adapted for use in warfare.

The 2019 Meeting of High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) adopted 11 »Guiding Principles affirmed by the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.« (see Box 1). Some 40 UN Member States belonging to the »Alliance for Multilateralism« are now championing these 11 Guiding Principles and calling on States to contribute actively to the clarification and development of a normative and operational framework.²

The civil society-led Campaign to Stop Killer Robots is working for a preemptive ban on the development, production, and use of fully autonomous weapons. Similarly, UN Secretary-General António Guterres as well as his High Representative for Disarmament, Izumi Nakamitsu, have called for LAWS to be banned by international law.³ A growing num-

ber of Member States have also called for a prohibition of LAWS. At a minimum, there seems to be broad agreement that it is necessary that States have an obligation to maintain *meaningful human control* over the lethal use of force. Member States, with the support and active participation of the United Nations and other international organizations, civil society and the private sector, quickly need to reach common understandings on how to ensure human beings retain control over the use of force.

Improving communication between policymakers and scientific and technical experts is crucial given the dual-use nature of artificial intelligence. Indeed, a prohibition on weapons outside of human control would not be counterproductive to technological development. Rather, there is a need to fully harness technological progress while maintaining and advancing international law that safeguards humanitarian protections, human rights and international peace and security.

Therefore, in September and October 2020, the Friedrich-Ebert-Stiftung New York Office (FESNY) launched the »New Alliances for Meaningful Human Control« project aimed at building on the 11 Guiding Principles on LAWS and moving them to the next level, towards international legal instruments that regulate high-tech weapons and prohibit LAWS. Linked to the UN75 Global Governance Forum,⁴ FESNY convened three discussions with representatives of governments, the United Nations, academia, civil society and the private sector to discuss the moral, ethical, legal and humanitarian challenges posed by LAWS (see Annex I for list of participants). Participants, speaking under Chatham House Rule, offered many diverse viewpoints, representing different perspectives on LAWS and pathways to addressing the risk they pose.⁵

This paper is rooted in the discussions facilitated by FESNY. While it summarizes the agreements between the co-authors, who participated, it does not purport to represent

1 ICRC (2016), Views of the International Committee of the Red Cross on Autonomous Weapon Systems, <https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system> (May 06, 2020), p. 1.

2 German Federal Foreign Office. (2019) »Six Initiatives for Multilateralism.« Available at <https://www.auswaertiges-amt.de/en/aussenpolitik/internationale-organisationen/vereintenationen/alliance-multilateralism/2250460>

3 UN. (2019) »Autonomous weapons that kill must be banned, insists UN chief.« Available at <https://news.un.org/en/story/2019/03/1035381>; UN. (2019) »Side event »The urgent need for a treaty to retain meaningful human control over the use of force.« Available at <https://www.un.org/disarmament/update/side-event-the-urgent-need-for-a-treaty-to-retain-meaningful-human-control-over-the-use-of-force/>

4 UN75 Global Governance Forum. (2020) »Designing a Roadmap to The Future We Want, The UN We Need.« Available at <https://www.stimson.org/2020/un-75-global-governance-forum/>

5 FESNY. (2020) »New Alliances for Meaningful Human Control.« Available at <https://www.fesny.org/article/new-alliances-for-meaningful-human-control/>

Box 1

Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System

It was affirmed that international law, in particular the United Nations Charter and International Humanitarian Law (IHL) as well as relevant ethical perspectives, should guide the continued work of the Group. Noting the potential challenges posed by emerging technologies in the area of lethal autonomous weapons systems to IHL, the following were affirmed, without prejudice to the result of future discussions:

- (a) International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems;
- (b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system;
- (c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole;
- (d) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control;

(e) In accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, a determination must be made whether its employment would, in some or all circumstances, be prohibited by international law;

(f) When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation should be considered;

(g) Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems;

(h) Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations;

(i) In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized;

(j) Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies;

(k) The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity and humanitarian considerations.

Source: CCW. (2019) *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*. Available from [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/4F3F92951E0022D9C12584F50034C2F4/\\$file/CCW+MSP+2019+9.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/4F3F92951E0022D9C12584F50034C2F4/$file/CCW+MSP+2019+9.pdf)

the views of all members of the discussions. For instance, participants disagreed on the precise meaning, scope and value of the 11 Guiding Principles. However, a significant majority agreed that good faith interpretation of the 11 Guiding Principles – particularly paragraphs b), c) and d) (see Box 1) – required progress toward negotiation of a legally-binding multilateral instrument mandating a positive obligation to maintain meaningful human control over the use of force, at the level of individual attacks.

Given the contested international political environment, progress toward negotiating a legally-binding instrument would also demonstrate the potential for effective global governance, as currently spearheaded by the Alliance for Multilateralism.

Towards this end, the subsequent paper argues that a potential venue for negotiating such a mandate could be within the CCW, in the form of a new protocol on meaningful human control. While positive obligations would be the most suitable starting point, a new protocol should also entail principles about technologies not to be developed and deployed. In what follows, two main issues will be addressed to help States as they move toward negotiations on a treaty

to ensure meaningful human control: First, meaningful positive obligations, and second, legally-binding control.

MEANINGFUL POSITIVE OBLIGATIONS

Many States have coalesced around the concept of meaningful human control as the basis for new international law on LAWS. However, meaningful human control is not explicitly referenced in the 11 Guiding Principles. States opted for less clear terminology like »human responsibility« and »human-machine interaction« in paragraphs b) and c). There is concern that such language, without elaboration and clarification, would add little meaningful constraint on State behavior.

To be *meaningful*, certain operational and technical requirements for human control need to be met. It must be active and involved (so-called »in-the-loop«) not passive (»on-the-loop« or »out-of-the-loop«). Human control needs to be maintained from the activation of a weapon until an attack is completed, aborted or terminated. This requires real-time human supervision at the level of the attack, including full knowledge of the weapon system's actions and a reliable

communications link between the weapon and its operator. It also requires allowing sufficient time in decision-making for a human operator to make meaningful decisions about targeting and a capacity to intervene and deactivate the system. A positive obligation to maintain a ratio of human operators/weapon systems greater than or equal to 1:1 may help in this regard.

Where automated systems suggest targets to human operators, those operators must be trained to understand the capabilities, limitations and known failure modes of the automated targeting system. Consideration must be given in the design of systems, processes, and operator education towards counteracting automation bias (the tendency for human operators to favor suggestions from automated decision-making systems and to ignore contradictory information from other sources).

Positive obligations for meaningful human control do not need to prohibit autonomous point-defense weapons, such as anti-missile defenses, as long as these a) include a routine non-autonomous mode where the weapon will not use force without a human confirmation; and b) when in autonomous mode, are continuously supervised by a responsible human who can interrupt its operation.

Thus, any legally-binding instrument on LAWS needs to include strong positive obligations on all people and institutions involved in the use of lethal force to maintain human control over weapons *capable* of operating autonomously. As an additional safeguard, a legally-binding instrument should also include an outright prohibition of weapons systems that are *incapable* of meaningful human control.

Building on Common Article 1 of the Geneva Conventions – which obligates states »to respect and ensure respect« for international humanitarian law (IHL) »in all circumstances« – might serve as a potential model. This duty requires states both to abide by the rules established by IHL (»respect«) and to encourage others – whether their own troops, allied states or non-state actors – to do so too (»ensure respect«). According to legal commentary by the International Committee of the Red Cross (ICRC), Common Article 1 thus »comprises both a negative and a positive obligation.« States must not »encourage, nor aid or assist in violations« of IHL. But they must also »do everything reasonably in their power to prevent and bring such violations to an end.«⁶ This extends to preventing »violations when there is a foreseeable risk that they will be committed.«⁷ Common Article 1 is »not a loose pledge but a commitment vested with legal force,«⁸ con-

firmed by the International Court of Justice.⁹ The phrase »in all circumstances« extends the duty to ensure respect into both wartime and peacetime.¹⁰ Building on the precedent of Common Article 1, a legally-binding instrument on LAWS could specify a duty to maintain and ensure meaningful human control over individual attacks in all circumstances.

It is hoped that the CCW process will, through establishing legally-binding prohibitions, prevent LAWS from ever being developed or used. However, States should also consider establishing positive obligations to address harm resulting from potential violations of a norm against LAWS, rooted in International Humanitarian Law (IHL) and International Human Rights Law (IHRL). A legally-binding instrument might establish a positive obligation on States to pursue effective remedy, including assistance to victims. Precedent exists in several humanitarian disarmament treaties, including CCW Protocol V on Explosive Remnants of War, which includes provisions on victim assistance, demining and risk reduction education.

LEGALLY-BINDING CONTROL

Further progress toward stigmatization of LAWS will require engaging with multiple stakeholders, including industry, academia and civil society. However, inclusion of a broad range of stakeholders should not distract from the onus of responsibility for action remaining on states. Expecting the private sector to establish and maintain voluntary guidelines or codes of conduct on meaningful human control is unrealistic, given that states are the customers of weapons contracts and stipulate their expectations to the private sector. In fact, technology companies themselves have stressed the need for clear guidelines from states to help engineers, designers and technology workers make moral, ethical, and legal judgements about the systems they build.¹¹

Effective control of the humanitarian, human rights and security risks posed by LAWS will require legally-binding obligations on states negotiated in a multilateral forum. To date, the most likely venue for negotiating such a mandate would be within the CCW, in the form of negotiating a new protocol. However, the CCW's consensus rules of procedure have been interpreted as requiring agreement of all states – effectively granting a veto to the most intransigent and often resulting in lowest common denominator decisions. Given this context, more ambitious states and other actors may in time consider other potential avenues, including a UN General Assembly-mandated process or one analogous to the Ottawa and Oslo processes on landmine and cluster munitions.

6 ICRC. (2016) »Commentary of 2016: Article 1: Respect for the Convention.« Para. 154. Available at <https://ihl-databases.icrc.org/ihl/full/GCI-commentaryArt1>; Knut Dormann & Jose Serralvo. (2014) »Common Article 1 to the Geneva Conventions and the obligation to prevent international humanitarian law violations.« *International Review of the Red Cross*. 96(895/896). p. 707.

7 ICRC. (2016) »Commentary.« Para. 164.

8 ICRC. (2016) »Commentary.« Para. 170.

9 ICJ. (1986) *Military and Paramilitary Activities in and against Nicaragua case*. Merits, Judgment. Para. 220. Available at <http://www.icj-cij.org/docket/?sum=367&p1=3&p2=3&case=70&p3=5>

10 ICRC. (2016) »Commentary.« Paras. 184–191.

11 Future of Life Institute. (2018) *Lethal Autonomous Weapons Pledge*. Available at <https://futureoflife.org/lethal-autonomous-weapons-pledge/?cn-reloaded=1>

THE WAY FORWARD

For the regulation of LAWS, the 11 Guiding Principles have become the most agreed-upon framework, albeit one that requires more work. For instance, the Guiding Principles lack reference to human rights. As LAWS potentially pose a threat to the human rights to life and dignity, more discussion is needed to clarify the distinction between people and weapons, humanity and technology. The European Union's General Data Protection Regulation (GDPR) could be a potential source of better rights-based language. Article 22 affirms that people »shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.«¹² The use of violent force certainly »significantly affects« the targeted person.

The Guiding Principles therefore are not a ceiling, but a steppingstone towards a more legally binding agreement that ensures binding human control, has meaningful positive obligations, and respects IHL and IHRL. In this regard, it is good news that the Principles' political visibility increased after the endorsement of the Alliance for Multilateralism. Member States of the Alliance should therefore use this political momentum and lead the effort to negotiate an additional protocol to the CCW. After the November 2020 meetings of the GGE had to be cancelled due to Covid-19, attention is now on the next meeting of the High Contracting Parties to the CCW and the CCW Review Conference scheduled for 2021. Tangible progress by then is sorely needed to achieve some form of arms regulation in an area of rapid technological development.

Moreover, progress toward new international law would demonstrate the potential for effective global governance, called for by the Alliance for Multilateralism. The Alliance itself is still in a phase of self-definition and is a rather loose collection of states. In fact, among the countries who supported the Alliance's Guiding Principles, there are also 20 countries that have explicitly endorsed the call for a ban on LAWS. Progress on an additional protocol with meaningful obligations for human control of LAWS would demonstrate the bridge-building capacity of the Alliance and would increase its standing as an actor for effective global governance.

¹² European Union. (2016) General Data Protection Regulation (GDPR), Article 22. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

ANNEX I: FURTHER RESOURCES

Article 36. (2019) *Critical Commentary on the »Guiding Principles.«* Available at: <http://www.article36.org/wp-content/uploads/2019/11/Commentary-on-the-guiding-principles.pdf>

Article 36. (2019) *Targeting People: Key issues in the regulation of autonomous weapons systems.* Available at: <http://www.article36.org/wp-content/uploads/2019/11/targeting-people.pdf>

Campaign to Stop Killer Robots. (2019) *Key Elements of a Treaty on Fully Autonomous Weapons.* Available at: <https://www.stopkillerrobots.org/wp-content/uploads/2020/03/Key-Elements-of-a-Treaty-on-Fully-Autonomous-Weapons.pdf>

CCW. (2020) *National commentaries on the 11 guiding principles of the GGE on LAWS.* Available at: <https://meetings.unoda.org/section/group-of-governmental-experts-gge-on-emerging-technologies-in-the-area-of-lethal-autonomous-weapons-systems-laws-documents-4929-documents-4947/>

Human Rights Watch and Harvard Law School's International Human Rights Clinic. (2020) *The Need for and Elements of a New Treaty on Fully Autonomous Weapons.* Available at: https://www.hrw.org/sites/default/files/media_2020/06/202006arms_rio_autonomous_weapons_systems_2.pdf

IEEE. (2017) »Reframing Autonomous Weapons Systems.« *Ethically-Aligned Design.* Version 2. Pp. 113–130. Available at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

ICRC & SIPRI. (2020) *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control.* Available at: <https://www.icrc.org/en/document/limits-autonomous-weapons>

PAX. (2017) *Where to Draw the Line: Increasing Autonomy in Weapon Systems – Technology and Trends.* Available at: <https://www.paxforpeace.nl/publications/all-publications/where-to-draw-the-line>

ANNEX II: PARTICIPANTS IN WORKSHOPS OF THE »NEW ALLIANCES FOR MEANINGFUL HUMAN CONTROL« PROJECT

Name	Title, Organization
Niels Annen	Minister of State, Federal Foreign Office, Germany
Matthew Breay Bolton (Notetaker)	Director of the International Disarmament Institute, Pace University, New York City
Michael Bröning	Executive Director, FES NY
Matilda Byrne	Coordinator, Australia Stop Killer Robots, SafeGround Inc.
Amy Dowler	Political Affairs Officer, UN Office for Disarmament Affairs (UN ODA)
Ryan Gariepy	Co-Founder and CTO, Clearpath Robotics
Emilia Javorsky	Director, Scientists Against Inhumane Weapons (SAIWE) & Policy Team, Future of Life Institute
Robert H. Latiff	Maj. Gen. (Ret.), US Air Force, Research Professor at George Mason University
Volker Lehmann (Facilitator)	Senior Policy Analyst, FES NY
Pamela Moraga	First Secretary, Disarmament, Global Security and Humanitarian Affairs Unit, Permanent Mission of Chile to the International Organizations in Geneva
Laura Nolan	Computer Engineer, International Committee for Robot Arms Control (ICRAC)
John Reyels	Deputy Head of Division Conventional Arms Control and CSBM, Federal Foreign Office, Germany
Stuart Russell	Professor of Computer Science, Smith-Zadeh Professor in Engineering, University of California, Berkeley

ABOUT THE AUTHORS

Matthew Breay Bolton is associate professor of political science at Pace University in New York City. He is an expert on global peace and security policy, focusing on multilateral disarmament and arms control policymaking processes. He has been an active participant in global campaigns on landmines, cluster munitions, nuclear weapons, killer robots and the arms trade.

Matilda Byrne is National Coordinator of Australia Campaign to Stop Killer Robots at SafeGround Inc. She is also SafeGround's Disarmament Programme Lead, which focuses on legacy and indiscriminate weapons.

Ryan Gariepy is the CTO and co-founder of Clearpath Robotics and OTTO Motors. He serves on the board of directors for the NSERC Canadian Robotics Network, and is a co-founder of ROSCon, the first and largest conference for robotics developers. Ryan also led the effort resulting in Clearpath being the first for-profit company to publicly speak up about the dangers of lethal autonomous weapons systems.

Emilia Javorsky is a physician-scientist focused on the invention, translation, and development of medical technologies. Her work focuses on the topic of lethal autonomous weapons, leading advocacy efforts at the Future of Life Institute and directing Scientists Against Inhumane Weapons.

Volker Lehmann is a Senior Policy Analyst with the FES New York office. His activities and publications focus on the role of the UN in international peace and security, on the nexus of peace and development, and on the various aspects of reforming the UN.

Laura Nolan is a senior software engineer in industry with a focus on reliability in automation and complex software systems. She is a member of the International Committee for Robot Arms Control and a member of the board of the USENIX Association.

NEW YORK OFFICE

The office of Friedrich-Ebert-Stiftung (FES) in New York serves as the liaison for FES offices worldwide with the United Nations (UN) in New York and the international financial institutions (International Monetary Fund and World Bank) in Washington, D.C. The office addresses peace, justice and economic issues, working closely with academia, civil society, multilateral institutions and their Member State governments to convene multi-stakeholder debates. The formats

IMPRINT

Friedrich-Ebert-Stiftung | Global Policy and Development
Hiroshimastr. 28 | 10785 Berlin | Germany

Friedrich-Ebert-Stiftung | New York Office
747 Third Avenue, Suite 34D | New York, NY 10017 | USA

Responsible:
Michael Bröning | Executive Director | FES New York
Phone +1-212-687-0208

www.fesny.org

To order publications:
linklaar@fesny.org

Commercial use of all media published by the Friedrich-Ebert-Stiftung (FES) is not permitted without the written consent of the FES.

ADDRESSING THE THREAT OF AUTONOMOUS WEAPONS

Maintaining Meaningful Human Control



Since 2015, efforts against lethal autonomous weapons systems (LAWS) have gained momentum among different stakeholders. Artificial intelligence experts, roboticists and scientists, as well as high tech professionals and technology companies have expressed their reservations against their research, products and platforms being adapted for use in warfare.

Moreover, civil society organizations, UN Secretary-General António Guterres, his High Representative for Disarmament Izumi Nakamitsu, and a growing number of Member States have called for a prohibition of LAWS.



However, to date the only international agreement reached were the voluntary Guiding Principles on LAWS that the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) adopted in 2019.

In Fall 2020, to explore possibilities for a more robust international legal framework, the Friedrich-Ebert-Stiftung-New York Office (FESNY) launched the »New Alliances for Meaningful Human Control.« This paper is a first result of its work. The authors advocate for moving the voluntary 11 Guiding Principles on LAWS to the next level, potentially by negotiating a new, additional protocol



to the CCW. Towards this end, the authors discuss how to ensure binding human control, meaningful positive obligations, and respect for humanitarian law and human rights law.

Going forward, the Alliance for Multilateralism, which endorsed the Guiding Principles in September 2019, should also lead the effort to negotiate an additional protocol to the CCW. Progress on an additional protocol with meaningful obligations for human control of LAWS would demonstrate the bridge-building capacity of the Alliance and would increase its standing as an actor for effective global governance.

For further information on this topic:
<https://www.fesny.org/topics/sustaining-peace/>