HEINRICH BÖLL STIFTUNG
DEMOCRACY

VOLUME 49

Autonomy in Weapon Systems

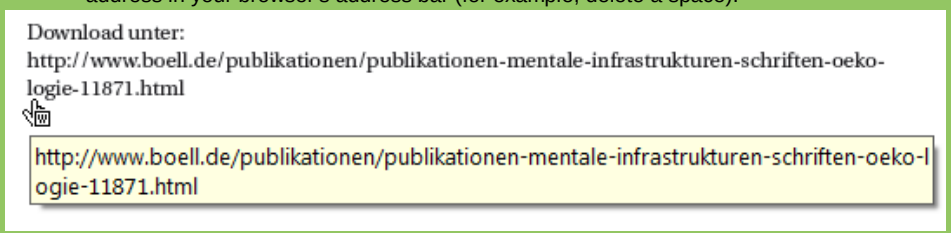
The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy

Edited by the Heinrich Böll Foundation

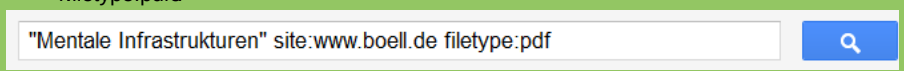
Note on web links

Based on technical changes caused by typesetting some links in this document may have gotten altered, for example through the addition of a separator, and may thus be faulty. Others may have become invalid due to changes on the respective websites.
To verify whether a link is complete, place the mouse pointer on the link and compare the displayed address with the one in the document.

In the document the link may appear to be complete, yet technically its validity may be compromised by a delimiter. You can correct such errors by simply editing the address in your browser's address bar (for example, delete a space).



In other cases it will be easier to locate the document through web search by entering the document's title and, if necessary, the name of the website and the file type. To do this, copy the document's title into your search window, for example the brochure "mental infrastructure", and, to narrow your search, add ñsite:nameofsite.orgò and ñfiletype:pdfò



We apologize that we are not always able to provide accurate links.
Your editorial team

AUTONOMY IN WEAPON SYSTEMS

HEINRICH BÖLL STIFTUNG
PUBLICATION SERIES ON DEMOCRACY
VOLUME 49

Autonomy in Weapon Systems

The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy

A Report by Daniele Amoroso, Frank Sauer, Noel Sharkey, Lucy Suchman and Guglielmo Tamburrini

Edited by the Heinrich Böll Foundation

Task Force on Disruptive Technologies and 21st Century Warfare
For inquiries, please contact Dr. Frank Sauer (Bundeswehr University Munich).
E-Mail: frank.sauer@unibw.de

Autonomy in Weapon Systems
The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy
A Report by Daniele Amoroso, Frank Sauer, Noel Sharkey, Lucy Suchman and Guglielmo Tamburrini
Volume 49 of the Publication Series on Democracy
Edited by the Heinrich Böll Foundation

Design: feinkost Designnetzwerk, Sebastian Langer (predesigned by blotto design)
Title photo: Civil Air Patrol (– flickr) (CC BY 2.0)
Printing: ARNOLD group, Großbeeren
ISBN 978-3-86928-173-5

This publication can be ordered from: Heinrich-Böll-Stiftung, Schumannstraße 8, 10117
T +49 30 28534-0 **F** +49 30 28534-109 **E** buchversand@boell.de **W** www.boell.de

CONTENTS

PREFACE

Two key areas of our recent foreign and security policy work have been the legal responsibility for war crimes and human rights violations, and the prospects of outlawing nuclear weapons under international law. As futile as both efforts may have seemed at the outset, we were pleasantly surprised when 122 UN member states adopted the Treaty on the Prohibition of Nuclear Weapons in June 2017, and the first ratifications followed shortly thereafter in September 2017. The *International Campaign for the Abolition of Nuclear Weapons* (ICAN) – our long-standing partner and recipient of the Nobel Peace Prize for its work in autumn 2017 – played a major role in bringing about this treaty.

With our *Task Force on Disruptive Technologies and 21st Century Warfare*, we have plunged ourselves into another seemingly hopeless battle: In response to the current rapid development of «artificial intelligence» (AI), we want to establish clear rules for the military use of AI and to advocate a global ban on autonomous weapon systems. In other words, a prohibition on all future weapons that, once activated, will automatically select their targets and complete their deadly mission without further human intervention.

Gregor Enste, who headed the Department of Foreign and Security Policy of the Heinrich Böll Foundation until October 2017, deserves the credit for setting up this innovative group of experts. We would also like to thank Dr. Frank Sauer with Bundeswehr University Munich for leading the task force as its scientific coordinator and for managing the contributions to this study.

In its report, our task force has laid out substantial objections to the further development of autonomous weapons. They are initially directed at German decision-makers, calling on them to take a clear stand against autonomous weapon systems, and culminate in the hope that these weapons will be banned within the framework of the United Nations.

Ten years ago, when ICAN first proposed banning nuclear weapons under international law, the idea was met with wry amusement by parts of the security establishment. Those voices became subdued, however, following the adoption of the Treaty on the Prohibition of Nuclear Weapons and the awarding of the Nobel Peace Prize. With its presence at this year's Munich Security Conference, ICAN's position has finally reached the security policy mainstream.

We hope that our task force and the activists of the global *Campaign to Stop Killer Robots* will be equally successful in the coming years in their mission to establish preventive controls for autonomous weapon systems. Important steps have been taken by stating the ethical, legal and security concerns in this paper, and by suggesting a framework to ensure meaningful human control over weapon systems. It is to be

hoped that these objections and suggestions will be heard by our national and international decision-makers and will focus the further discourse on autonomous weapon systems on what really matters: ensuring the inviolable nature of human dignity and the humanitarian principles of international law.

Berlin, May 2018

Giorgio Franceschini
Heinrich Böll Foundation, Foreign and Security Policy Department

VORWORT

Zwei zentrale Themen unserer außen- und sicherheitspolitischen Arbeit der letzten Zeit drehten sich um die rechtliche Verantwortlichkeit bei Kriegsverbrechen und Menschenrechtsverletzungen sowie die Perspektiven einer völkerrechtlichen Ächtung von Kernwaffen. So aussichtslos beide Unterfangen auch auf den ersten Blick erscheinen, so sehr wurden wir in einer Sache im letzten Jahr positiv überrascht: Im Juni 2017 verabschiedeten 122 UN-Mitgliedsländer einen Vertrag, der den Besitz und den Einsatz von Kernwaffen verbietet, und bereits im September 2017 erfolgten die ersten Ratifikationen. Maßgeblichen Anteil am Zustandekommen dieses Verbotsvertrags hatte unser langjähriger Kooperationspartner, die *International Campaign for the Abolition of Nuclear Weapons* (ICAN), die für ihr Engagement im Herbst 2017 sogar mit dem Friedensnobelpreis ausgezeichnet wurde.

Mit unserer *Task Force on Disruptive Technologies and 21st Century Warfare* haben wir uns in einen weiteren scheinbar aussichtslosen Kampf gestürzt: Wir wollen in einer Phase der rasanten Entwicklung «Künstlicher Intelligenz» (KI) klare Regeln für die militärische Nutzung der KI etablieren und uns für eine globale Ächtung autonomer Waffensysteme stark machen – also all jener Waffen der Zukunft, die – einmal aktiviert – vollkommen autonom ihre Ziele auswählen, ansteuern und ohne weiteres menschliches Zutun ihre tödliche Mission erfüllen.

Gregor Enste, der bis zum Oktober 2017 das Referat für Außen- und Sicherheitspolitik der Heinrich-Böll-Stiftung geleitet hat, gebührt das Verdienst, diese innovative Expert/innen-Gruppe ins Leben gerufen zu haben, und Dr. Frank Sauer von der Universität der Bundeswehr München unser Dank, die Task Force als wissenschaftlicher Koordinator geleitet und die Beiträge dieses Berichts koordiniert zu haben.

Unsere Task Force nennt in der vorliegenden Publikation gewichtige Gründe, warum der Entwicklung autonomer Waffen Einhalt geboten werden sollte. Ihr Appell richtet sich zunächst an die deutschen Entscheidungsträger/innen, sich bei dieser Frage klar gegen autonome Waffensysteme zu positionieren, und mündet in der Hoffnung auf eine Ächtung dieser Waffen im Rahmen der Vereinten Nationen.

Als ICAN vor zehn Jahren die völkerrechtliche Ächtung von Kernwaffen ins Gespräch brachte, wurden sie von Teilen des Sicherheitsestablishments noch milde belächelt. Nach der Verabschiedung des Vertrags für das Verbot von Atomwaffen und der Verleihung des Friedensnobelpreises wurden diese Stimmen jedoch leiser, und im Februar dieses Jahres war ICAN durch einen Redebeitrag bei der Münchner Sicherheitskonferenz schließlich im Herzen des sicherheitspolitischen Mainstreams angekommen.

Es ist unserer Task Force sowie den Aktivist/innen der globalen *Campaign to Stop Killer Robots* zu wünschen, dass ihrem Anliegen einer präventiven Kontrolle

autonomer Waffensysteme in den nächsten Jahren ein ähnlicher Erfolg beschieden sein wird. Wichtige Schritte dazu sind durch die Niederschrift der ethischen, völkerrechtlichen und sicherheitspolitischen Bedenken in diesem Papier und den darin ebenfalls enthaltenen Vorschlägen zur Wahrung menschlicher Verfügungsgewalt über Waffensysteme erfolgt. Es bleibt zu hoffen, dass diese Einwände und Empfehlungen bei unseren nationalen und internationalen Entscheidungsträger/innen Gehör finden und sie den weiteren Diskurs zu autonomen Waffensystemen auf das lenken, worauf es ankommt: die unantastbare Würde des Menschen und die Prinzipien des humanitären Völkerrechts.

Berlin, im Mai 2018

Giorgio Franceschini
Referat Außen- und Sicherheitspolitik der Heinrich-Böll-Stiftung

FOREWORD AND ACKNOWLEDGMENTS

Unmanned systems have been in use in militaries since at least World War I. However, recent advances in military robotics and especially in artificial intelligence (AI) represent a turning point in this process that gives rise to a number of grave legal, ethical and political concerns.

In 2017, the Heinrich Böll Foundation decided to establish a Task Force comprised of experts from various academic backgrounds, including anthropology, computer science, international law, philosophy, political science and robotics.

We, the members of this group, were tasked with analyzing the «disruptive» potential of the military application of AI and robotics in weapon systems. Our aim was to develop policy recommendations and suggest self-commitments for Germany's foreign and security policy. This report represents the final product of this joint endeavor led by Dr. Frank Sauer.

We would like to thank the representatives of the German Federal Foreign Office, the German Ministry of Defense and the defense industry who, over the course of 2017, took time out of their busy schedules to be our dialogue partners and engage in discussions with us.

For sharing their expertise with us at the two Task Force workshops in Berlin, we would like to thank Dr. Marcel Dickow (German Institute for International and Security Affairs SWP), Igor Fayler (Green Party Faction Berlin Neukölln), Prof. Dr. Robin Geiss (University of Glasgow), Lt. Col. André Haider (NATO Joint Air Power Competence Centre), Prof. Dr. Dominik Herrmann (University of Bamberg) and Christian Wussow (Green Party Faction to the German Bundestag). We would like to express special gratitude to PD Dr. Jürgen Altmann (Technical University Dortmund) for his feedback on an early draft version of the report. Jan Weiland deserves our warm thanks for his editorial assistance.

Lastly, we are of course deeply grateful to the Heinrich Böll Foundation for the generous financial and organizational support the Task Force has received. In particular, we would like to thank Gregor Enste and his successor Giorgio Franceschini, Stephanie Mendes-Candido, Doreen Beierlein and finally yet importantly the Böll Foundation's director Ellen Ueberschär. This research would not have been possible without their support. Any remaining mistakes in this document are ours alone.

Prof. Daniele Amoroso, *Università degli studi di Cagliari, Italy*
Dr. Frank Sauer, *Bundeswehr University Munich, Germany*
Prof. Dr. Noel Sharkey, *University of Sheffield, United Kingdom*
Prof. Dr. Lucy Suchman, *Lancaster University, United Kingdom*
Prof. Dr. Guglielmo Tamburrini, *Università di Napoli «Federico II», Italy*

ZUSAMMENFASSUNG

Früher, entschiedener und substantieller engagiert und Verantwortung übernehmend – so stellen sich seit einigen Jahren viele die zukünftige deutsche Außen- und Sicherheitspolitik vor. Es bleibt dabei jedoch eine offene und anhaltende Debatte, *wie* genau Deutschland seiner wachsenden Verantwortung gerecht werden soll, insbesondere in Bezug auf seine Streitkräfte – die Bundeswehr.

Der vorliegende Report geht davon aus, dass ein Nexus zwischen internationaler Sicherheit und neuen Technologien entsteht. Dieser bedeutet einen Lackmustest mit Blick auf die grundlegenden Normen und Werte, denen Deutschland und die Bundeswehr sich bei der Übernahme zusätzlicher Verantwortung verpflichten. Der Report schaut dabei im Speziellen auf die militärische Nutzung von Künstlicher Intelligenz (KI) und Robotik in Form sogenannter Autonomer Waffensysteme (AWS).

Als Arbeitsdefinition für AWS legt der Report, in Anlehnung an Definitionen des Internationalen Komitees vom Roten Kreuz (IKRK) und des US-Verteidigungsministeriums, den Schwerpunkt auf Autonomie in den kritischen Funktionen von Waffensystemen, also Zielauswahl und Zielbekämpfung. Der Sachverhalt wird jedoch durch die Einführung einer Skala mit 5 möglichen Stufen menschlicher Kontrolle und Aufsicht genauer aufgeschlüsselt. Dies ermöglicht eine präzisere Untersuchung menschlicher Verfügungsgewalt über AWS, samt der in der aktuellen Diskussion gängigen Vorstellungen von «appropriate levels of human judgement», «human oversight», «human in» oder «on the loop» sowie «meaningful human control».

Der Report argumentiert, dass die menschliche Verfügungsgewalt über kritische Waffensystem-Funktionen einer sorgfältigen Prüfung bedarf angesichts der rechtlichen, technischen, moralisch-ethischen und sicherheitspolitischen Implikationen von AWS.

Rechtlich gesehen deutet alles darauf hin, dass der Einsatz von AWS mit dem humanitären Völkerrecht zumindest in absehbarer Zukunft nicht in Einklang zu bringen ist und dass AWS bisher noch ungelöste Probleme mit Blick auf Rechenschaftspflichten und Verantwortungsübernahme bei der Anwendung militärischer Gewalt aufwerfen.

Technisch gesehen verfügen autonome Waffen nicht über die notwendigen Fähigkeiten, um Unterscheidungs- und Verhältnismäßigkeitsgeboten des humanitären Völkerrechts gerecht zu werden. Ihr Verhalten ist inhärent unvorhersehbar, insbesondere in Szenarien, in denen mehrere AWS interagieren.

Die im deutschen Grundgesetz in Artikel 1 verankerte Achtung der Menschenwürde schreibt, ebenso wie die internationalen Menschenrechte, vor, dass Maschinen die Entscheidungen über Leben und Tod prinzipiell nicht überlassen werden sollte.

Mit Blick auf globale Sicherheit birgt die Entwicklung von AWS schließlich ernsthafte Risiken für die regionale und globale Stabilität und setzt Proliferationsanreize, was ihre Nutzung durch solche Akteure befördert, die die völkerrechtlichen Rahmenbedingungen für den Einsatz militärischer Gewalt missachten.

Vor diesem Hintergrund entwickelt der Report folgende Empfehlungen an die deutsche Bundesregierung:

- Verfassen und Veröffentlichen eines nationalen Leitliniendokuments mit Blick auf die militärische Nutzung von autonomen Waffensystemen in der Bundeswehr.
- Übernehmen einer einfachen, am IKRK orientierten Definition von autonomen Waffensystemen mit dem Fokus auf Autonomie in den kritischen Waffensystem-Funktionen von Zielauswahl und Zielbekämpfung.
- Gesetzliche Verankerung einer Vorschrift zur Wahrung echter menschlicher Verfügungsgewalt («meaningful human control») über alle Waffensysteme der Bundeswehr, damit eine über Stufe 3 («Software selektiert Ziel vor, und Mensch muss Bekämpfung genehmigen») in Waffensystemen hinausgehende Autonomie vermieden und «vollautonome Waffensysteme» auf nationaler Ebene effektiv verboten werden.
- Kontinuierliches Erforschen der Interaktionen zwischen Menschen und autonomen Funktionen mit Blick auf zukünftige Waffensysteme der Bundeswehr, um die Wahrung menschlicher Verfügungsgewalt nachhaltig abzusichern; daneben adäquate Ausbildung der Soldatinnen und Soldaten der Bundeswehr im Hinblick auf die dazu erforderlichen Techniken, Verfahren und Taktiken.
- Fortgesetzte Nutzung und Regulierung defensiver SARMO-Systeme (Sense and React to Military Objects) durch die Bundeswehr, dabei die in diesem Report spezifizierten, strengen Rahmenbedingungen und Beschränkungen für Konstruktion und Betrieb anlegend.
- Fortführen und Intensivieren der deutschen Unterstützung für ein internationales, rechtsverbindliches und überprüfbares Verbot autonomer Waffensysteme, die der menschlichen Verfügungsgewalt entzogen sind.

Im Hinblick auf ein internationales AWS-Verbot könnte die Übernahme einer Führungsrolle und die noch aktivere und entschlossenere Arbeit auf Ebene der Vereinten Nationen ein beispielhafter Meilenstein in der neuen Außen- und Sicherheitspolitik Deutschlands sein. Deutschland würde damit ein deutliches Zeichen dafür setzen, dass es grundlegenden Normen und Werten und gewachsener internationaler Verantwortung gleichermaßen gerecht wird.

EXECUTIVE SUMMARY

Engaged and taking on responsibility earlier, more decisively, and more substantially – that is how German foreign and security policy has come to be envisaged over the last few years. However, it is an open and ongoing debate *how* Germany will meet its growing responsibilities, especially with regard to its armed forces – the Bundeswehr. This report argues that there is an emerging nexus of security and new technologies which provides a litmus test regarding the fundamental norms and values that Germany and the Bundeswehr heed whilst taking on additional responsibilities. It specifically focuses on the military use of artificial intelligence (AI) and robotics in so-called autonomous weapon systems (AWS).

The report adopts a working definition of AWS that, in accordance with definitions provided by the International Committee of the Red Cross (ICRC) and the US Department of Defense, puts emphasis on autonomy in a weapon system's critical functions, that is, target selection and engagement. However, the report reframes this issue in terms of 5 levels of human supervisory control. This allows for closely examining the governance of AWS and teasing out what is meant by notions such as «appropriate levels of human judgement», «human oversight», «human in» or «on the loop» and «meaningful human control».

The report argues that human supervisory control over critical functions must be subject to careful scrutiny considering the legal, technical, moral/ethical, and security implications of AWS.

Legally, all existing evidence indicates that the deployment of AWS could not comply with International Humanitarian Law (IHL) for at least the foreseeable future, and that they pose as yet unresolved problems regarding accountability and responsibility for the use of violent force.

Technically, autonomous weapons lack the necessary components to ensure compliance with the IHL requirements of distinction and proportionality. Their behavior is inherently unpredictable, particularly in scenarios where multiple AWS would interact.

Morally, the guiding principle of respect for human dignity, enshrined in German basic law («Grundgesetz») Article 1 (1) as well as in International Human Rights Law, dictates that machines should not be making life or death decisions regarding humans.

In terms of global security, the development of AWS poses serious dangers for regional and global stability and provides incentives for their proliferation, including their use by actors not accountable to legal frameworks governing the use of force. In light of this, the report develops the following recommendations to the German government:

- develop a national policy on the use of autonomy in weapon systems and make that document publicly available;
- adopt a simple, ICRC-like definition of autonomous weapon systems, based on the system's «task autonomy» in the performance of the critical functions of selecting and engaging targets;
- stipulate the legal requirement of «meaningful human control» over all Bundeswehr weapon systems, avoiding autonomy in target selection beyond level 3 («software selects the target and a human must approve it before the attack») thereby effectively prohibiting «fully autonomous weapon systems» at the national level;
- continuously investigate and map out levels of human control for autonomous functionality in future Bundeswehr weapon systems to allow the human reasoning and control needed (that is, levels 1-3 for critical functions), and train Bundeswehr personnel accordingly with regard to the required tactics, techniques and procedures;
- regulate the Bundeswehr's continued use of defensive SARMO (Sense and React to Military Objects) systems to satisfy the strict targeting limitations and constraints on design and operations specified in this report;
- continue and intensify support for an international, legally binding and verifiable ban on fully autonomous weapon systems, that is, weapon systems with autonomy in the performance of the critical functions of selecting and engaging targets.

Regarding the international prohibition on AWS, adopting a leadership role and working even more actively and decisively toward this goal at the United Nations would represent an exemplary milestone in Germany's new foreign and security policy. It would send a strong signal that Germany is heeding its fundamental norms and values whilst living up to its newly grown international responsibilities.

Introduction

It is by now commonplace to describe Germany as the key player in Europe, a politically stable economic powerhouse that has markedly gained in global status and influence since its reunification in 1990. Engaged earlier, more decisively, and more substantially – that is how German foreign and security policy has been envisaged in statements by prominent politicians and in key policy documents over the course of the last few years. Since the beginning of the US presidency of Donald J. Trump, Germany's increasing responsibility in matters of global affairs has been emphasized to an even greater degree both at home and abroad.

However, it is still very much an open and ongoing debate *how* Germany's foreign and security policy will go ahead in meeting this growing responsibility, especially with regard to the role Germany ascribes to its armed forces – the Bundeswehr.

In its 1994 landmark ruling, the German Federal Constitutional Court allowed for the German parliament to deploy armed forces within systems of collective security (United Nations, NATO) in «out of area» missions. Since then, taking on additional international responsibility increasingly included Bundeswehr missions abroad, culminating in the massive German engagement in Afghanistan. Germany still has multilateralism (and Europe) in its DNA, of course; in fact, many of the key foreign and security policy tenets that earned Germany the positive reputation of a «civilian power» (a term coined by German political scientist Hanns W. Maull) remain(ed) constant. At the same time, regarding the nature and scope of its military commitments, the last two decades represent nothing short of an epochal change in Germany's post-WWII history – and a drastic adjustment for the German armed forces themselves. In a still ongoing and challenging process, the Bundeswehr is rapidly transforming itself from a Cold War deterrent army into a modern combat-ready fighting force (Enskat/Masala 2015). Meanwhile the general population, it is crucial to point out, tends to remain deeply wary of additional military engagements.

With this in mind, we propose, first, that the self-commitments that Germany applies with regard to its military will be indicative for the trajectory of its new, more active and responsible-minded foreign and security policy at large. Second, the currently emerging nexus of security and new technologies provides a litmus test regarding the fundamental norms and values that Germany heeds while heading into this future.

This report analyzes the military use of artificial intelligence (AI) and robotics. It specifically focuses on «autonomous weapon systems» (AWS). We closely examine the «disruptive» potential of AWS from technical, legal, ethical, and political angles. In light of this analysis, the report develops a list of policy recommendations, suggesting

specific self-commitments for Germany to adopt with regard to the way its military engages with AI and robotics.

The commercial sector considers technologies disruptive when they yield new products, services and markets and disrupt previously existing structures by displacing established products and leading companies. The Internet is a prime example. From a military perspective, the very same technologies can spawn new weapon systems, practices and even entirely new operational domains. The commercial sector is also what currently drives progress in robotics and AI. They will yield disruptive effects on the battlefield. In addition, their dual-use character renders them extremely prone to proliferation. As a result, a new paradigm of warfare is currently emerging – a process that some observers have come to compare with the revolutions following the introduction of gunpowder, aircraft and the atomic bomb respectively (Future of Life Institute 2015; see also Allen/Chan 2017: 1-6, 10). The impact of technological progress on warfare is already generating sharp conflicts with existing international legal, ethical, and political frameworks. In the recent controversy surrounding the use of armed unmanned aerial vehicles («drones») these frictions have already become clearly visible (Sauer/Schörnig 2012) – even while, in this example, the leveraging of technologies and practices of automation and robotization as well as data science and machine learning for military purposes is only its infancy.

A specific concern motivates this report and unites the authors around a shared goal, which they have jointly pursued starting from their diverse backgrounds in various fields of research. While we hope for the peaceful uses of robotics and AI, and while we are aware that they can also benefit the military and law enforcement in many respects, we are deeply concerned about specifically those developments surrounding the growing autonomy in weapon systems. Our aim is thus to help safeguard against the negative effects of the military application of robotics and AI in autonomous weapons.

To connect our analysis to a German frame of reference, we looked for a basic principle to guide the overall thrust of our argument. We found this guiding principle in Germany's basic law (Grundgesetz) Article 1 (1) which states that «[h]uman dignity shall be inviolable. To respect and protect it shall be the duty of all state authority.»

We believe that human dignity is the Archimedean point of the AWS debate. For legal and political assessments of AWS may differ, but the German Grundgesetz provides the axiomatic ethical reminder that the dignity of all humans, including those that military violence is legitimately directed against, must be kept intact. Outsourcing the selection and engagement of targets to algorithms in military machines is out of the question for a society that accepts this imperative. Machines should not make life and death decisions. Instead, societies must safeguard meaningful human control over weapons and retain human governance within their military's decision-making processes on who is being targeted in war.

Since 2013, the parties forming Germany's government (Bundesregierung 2013: 124; 2018: 149) have stipulated in their coalition treaties that they will in fact not abdicate this human responsibility for targeting decisions. They also pledged to work for an international ban on AWS. German diplomats at the United Nations (UN) Convention on Certain Conventional Weapons (CCW) in Geneva have since echoed this at the diplomatic level, as have high-ranking military officials in Germany (such as Lieutenant General Ludwig R. Leinhos at the 2018 Munich Security Conference).

It is our hope that this report further informs and fosters the already ongoing public debate in Germany about the issue of autonomy in weapon systems. We especially hope that this report helps move German political and military decision-makers further toward developing, adopting and publishing an official Bundeswehr policy document on autonomy in weapon systems, one that makes the retention of meaningful human control over Bundeswehr weapon systems a legal requirement.[1] Lastly, we find that taking a leadership role by working even more actively and decisively toward a legally binding and verifiable prohibition on AWS at the UN level would be an exemplary milestone in Germany's new foreign and security policy. It would send a strong signal that Germany is heeding its fundamental norms and values whilst living up to its newly grown international responsibilities.

1 The recommendations contained in this report are forward-looking and not exhaustive in terms of the specific weapon systems discussed. We recommend a general prohibition on fully autonomous weapons, that is, of weapon systems that are not under meaningful human control, the requirements for which we examine in this report. In doing so, we note that one size of meaningful human control does not fit all weapon systems. In fact, we prominently discuss defensive SARMO (Sense and React to Military Objects) systems engaging unambiguous materiel targets (munitions) when there is no time for human intervention as requiring less strict human supervision conditions than other weapon systems in order for them to be under meaningful human control. Further discussions of specific weapon systems we consider to best be conducted at the implementation stage, given, of course, that the authors' general recommendation that meaningful human control must be retained over all weapon systems finds acceptance.

1. Concepts and definitions

Autonomous weapon systems (AWS)[2] have recently gathered widespread attention, particularly since more than 3,700 artificial intelligence (AI) and robotics researchers published an open letter in 2015 warning against an impending AI weapons arms race (Future of Life Institute 2015). This was followed in 2017 by an open letter to the United Nations (UN) Convention on Certain Conventional Weapons (CCW)[3] from 160 high profile CEOs of companies developing artificial intelligence technologies, calling for the UN «to work hard at finding means to prevent an arms race in these weapons, to protect civilians from their misuse, and to avoid the destabilizing effects of these technologies» (Future of Life Institute 2017).

Stigmatized as «killer robots» by opponents, AWS are widely regarded as harbingers of a paradigm shift in warfare (Geiss/Lahmann 2017). Prototypes of autonomous ground robots, fighter jets, submarines, ships and «swarms» are being developed and tested by technologically advanced nations. The US, Russia, China, and Israel are the frontrunners, with others, such as the UK and South Korea, following their lead.

There has been extensive discussion regarding how to define autonomous weapon systems at the UN's CCW in Geneva since the issue was first discussed in an Informal Meeting of Experts in 2014. Some confusion has arisen, for two main reasons: (1) a framing of the issue as a «levels of autonomy» problem, instead of looking at the human involvement with regard to critical weapon system functions; and (2) the concern of several nation states that a prohibitive treaty against autonomous weapons would result in the loss of important defensive weapons already in use. It is worth pointing out as well that excessive querying of definitions could be used by some states to deliberately slow the UN process from getting to the next stage.

Generally speaking, the AWS debate is organized around two basic understandings of autonomy.

On the one hand, we have definitions building on the assumption that, in order to qualify as «autonomous,» a weapon system should possess a «situational understanding» that is comparable to that of a competent human being (UK Ministry of Defense 2017). The focus here is on the machine's perceptual and evaluative capabilities or, in the words of a recent SIPRI Report, on the «sophistication of the machine's decision-making process» (Boulanin/Verbruggen 2017: 6). This requirement cannot be

2 AWS are currently also discussed under the acronym LAWS (Lethal Autonomous Weapon Systems). We are not adopting the acronym LAWS because we find that problems with autonomy in weapon systems are not necessarily dependent on their lethality.

3 The Convention's full title is «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 as amended on 21 December 2001».

satisfied by any existing weapon system, and it is difficult to make reasonable predictions about the prospects of constructing machines of this kind. Therefore, these notions of autonomy project AWS in some undetermined technological future and are not helpful to the debate and to growing concerns about autonomous decision-making and action by weapon systems, independently of whether these systems achieve situational understanding comparable to that of a competent human being.

On the other hand, autonomy has been defined on the basis of «the human-machine command-and-control relationship» (Boulanin/Verbruggen 2017: 5-6) in the fulfilment of a system's functions, especially the *critical functions* of selecting and engaging targets. According to the definition offered by the US Department of Defense (US DoD 2017: 13), an autonomous weapon system is thus a weapon that «once activated … can select and engage targets without further intervention by a human operator.» This aligns with the clear definition proposed by the International Committee of the Red Cross:

> **«Autonomous Weapon Systems are defined as any weapon system with autonomy in the critical functions of target selection and target engagement. That is, a weapon system that can select (i.e. detect and identify) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention»** (ICRC 2016: 8).

By shifting the focus from the endowment of human-like intelligence to mere «task autonomy» (Tamburrini 2016) and adopting definitions of this latter kind, some presently operating weapon systems have to be qualified as autonomous, their limited target baskets notwithstanding (for they perform critical functions of target selection and engagement without human intervention). This includes a number of already existing loitering munitions and fire-and-forget systems. For example, Harpy/Harop and other air-to-ground loitering munitions systems homing in on radar should accordingly be counted as AWS, as should the UK's Brimstone anti-tank missile when fired in lock-on mode.[4]

The adoption of the ICRC's definition – or one like it – is strongly advisable. Consistency requires forgoing the notion that AWS are mere «future weapons». Instead, we realize that we are already beginning to allow autonomy in critical weapon system functions. But more importantly, and independent of which precise terminology is

4 Brimstone is an air-to-ground anti-tank missile. In laser-guided mode, human operators pick out specific targets for the missile to destroy. In lock-on mode, Brimstone is loaded with targeting data, including data that serve to circumscribe the area within which it will then search for, select and attack – without any further intervention by human operators – armored enemy vehicles. Thus, Brimstone satisfies the DoD (and the ICRC) requirement for autonomy in lock-on mode, and fails to satisfy the same requirement in laser-guided mode. The operation mode is selected by human operators on the basis of available information about individual attack scenarios (e.g. by considering whether there are civilians or friendly forces in the vicinity of targets) (Boulanin/Verbruggen 2017: 48-50).

used to define AWS, the fact remains that the continued development and potential widespread use of weapon systems that independently identify, select and engage targets raise a number of unprecedented legal, ethical and political issues, which *urgently* require a concerted response by the international community.

One definitional question that raises concerns from some nation states, including Germany, is the status of current weapon systems that are already capable of operating without human intervention once activated. This especially concerns SARMO (Sense and React to Military Objects) weapon systems, which intercept high-speed inanimate objects such as incoming missiles, artillery shells and mortar grenades (Sharkey 2014). Examples include the «Nächstbereichschutzsystem» (NBS) MANTIS, which is deployed by the German Bundeswehr.

Systems like MANTIS have been deemed by Human Rights Watch (HRW 2012) as precursors to fully autonomous weapons. Others have tried to separate them from fully autonomous weapons by calling them automated or automatic systems (Sauer 2016). The US Department of Defense attempted to bound the scope of these weapons by suggesting that: «The automatic system is not able to initially define the path according to some given goal or to choose the goal that is dictating its path» (US DoD 2013: 66).

There are a number of common features for SARMO weapons that do in fact keep them distinct from those AWS that raise concerns. Ideally speaking, SARMO systems are:

- fully pre-programmed to automatically perform a small set of defined actions repeatedly and independently of external influence or control,
- used in highly structured and predictable environments that are relatively uncluttered with very low risk of civilian harm,
- operating from a fixed base – although some are used on naval vessels, they are «fixed» in the same sense as a robot arm mounted on a stationary platform on the same ship would be,
- unable to dynamically initiate a new targeting goal or change mode of operation once activated,
- constantly evaluated and monitored by human operators for rapid shutdown in cases of targeting errors, change of situation or change in status of targets,
- predictable in terms of system output and behavior,
- only used defensively against direct attacks by military objects uninhabited by humans when time is of the essence,
- constrained by design, that is, impossible to direct at human or human-inhabited targets after deployment.

Insofar as SARMO systems, such as the German MANTIS, operate with these features listed above, their use to protect soldier's lives is not problematic.

There is a risk, however, that these boundaries may be overstepped. A system such as MANTIS might be turned into – or embedded within – a system of systems that tracks the location of the attackers and fires back at them without human intervention. This would be problematic indeed. It would raise all the legal, ethical, and political issues commonly connected to AWS. We will discuss this again in more depth in section 6.

To sum up, concerns over the supposed extreme difficulty of reaching a clear definition of autonomy can be addressed by focusing the discussion on the functions of weapon systems, especially the critical functions of target identification and the initiation of violent force, and the associated requirement for human control.

With this in mind, we turn to discussing the legal, ethical, and political implications of AWS in more detail. We organize the discussion around four questions: Can AWS be guaranteed to comply with International Humanitarian Law (IHL)? Is it ethical to allocate to machines the determination of when to apply violent force? What issues of accountability and responsibility are raised by AWS? And what will AWS mean for global security?

2. Adherence to the principles of international humanitarian law (IHL)

Much of the discussion concerning the ethical and legal implications of AWS is about guaranteeing that any fielded system will comply with IHL. The guarantee of compliance is problematic for a number of reasons.

2.1 The principle of distinction between civilians and combatants

As clarified by the ICRC in its codification of customary IHL, the Principle of Distinction (Rule 1) establishes that «The parties to the conflict must at all times distinguish between civilians and combatants. Attacks may only be directed against combatants. Attacks must not be directed against civilians» (ICRC 2006: 3).

It is highly doubtful that AWS will in any foreseeable future be able to discern between civilians and combatants, as required by IHL. This is partly attributable to the types of sensing capabilities available for AWS and partly because it is very difficult to define the notion of civilian (Sharkey 2008).

There are systems that have a weak form of distinction. For example, the already mentioned Israeli Harpy/Harop is a loitering munition that detects radar signals. When it finds one, it looks at its database to find out if it is friendly – if not, it dive-bombs the signal's source. This type of discrimination does not meet the criteria required by the Principle of Distinction because the system cannot tell if the radar is, for example, positioned within a military anti-aircraft installation or on the roof of a school.

More generally speaking, AWS lack, for the foreseeable future, the main components required to ensure compliance with the Principle of Distinction.

AWS lack adequate sensory or vision processing systems for separating combatants from civilians, particularly in insurgent warfare, or for recognizing wounded or surrendering combatants. AWS may be equipped with various sensors such as cameras, infrared sensors, sonars, lasers, temperature sensors and lidars. But these sensors and the accompanying processing systems are unable to differentiate legitimate from non-legitimate (human) targets, particularly from a great distance or in the fog of war. While a computer can compute any given procedure that can be written down in a programming language, for discrimination we would need a clear specification of «civilian-ness». This simply does not exist. We also cannot derive one from IHL (1944 Geneva Convention and 1977 Protocol 1), at least not one that would be unambiguous

enough to enable a machine to make the distinction between civilians and combatants. After all, even for humans this distinction can be extremely difficult to make, especially if an adversary does not play by the rules.

Formidable challenges are posed especially by contemporary armed conflicts where the distinction between combatants and protected persons is no longer based on easily perceivable and distinctive signs such as military uniforms, but rather on people's «behaviour and actions on the battlefield» (NATO JAPCC 2016). In IHL, reference is made in particular to those cases of civilians losing protection from attacks as a consequence of their «direct participation in hostilities» or, in the context of non-international armed conflicts, their performance of «continuous combat functions». Distinction, in other words, is embedded in the broader requirement of situational awareness (Suchman 2016).[5] Situational awareness presupposes, in its turn, an open-ended repertoire of *in situ* human discriminatory capabilities, intuitions and experience, which is not attainable by AI and robotic systems for the foreseeable future. So while we may eventually move machines towards having some limited sensory and visual discrimination capability in certain narrowly constrained circumstances, AWS would still lack the required understanding of context, especially in the unstructured and unpredictable scenarios on the battlefield, and the necessary situational awareness, experience and common sense to allow for a discrimination decision.

2.2 Proportionality in attack

Under the principle of proportionality, it is forbidden to launch «an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated» (ICRC 2006). This principle, in other words, raises «the terrible and impossible problem» (Solis 2016: 293) of striking a balance between military gains expectedly deriving from some given course of action and harm to civilians ensuing from it (HRW 2016).

The prospect of developing AWS capable of assessing proportionality with sufficient competence prior to and during an attack appears at the present to be groundless, insofar as proportionality analysis, like distinction, relies heavily on qualitative elements and open-textured standards such as the judgment of a «reasonable military commander» (Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia 2000: para 50).

5 An Army Times article quotes US Army Chief of Staff Gen. Mark Milley speaking at the Future of War Conference in Washington, D.C. who comments that in future, increasingly urban, warfare «[s]oldiers will have to be highly trained in discriminating fire, able to quickly and effectively tell who is a combatant and who is a bystander … Our leaders at the pointy end of the spear are going to have very, very high degrees of ethical skill and resilience to be able to deal with incredibly intense issues in ground combat» (Myers 2017).

2.3 The prohibition of attacks against persons *hors de combat*

Structurally analogous to the rule of distinction is the prohibition, stemming from the principle of military necessity, against attacking persons recognized as *hors de combat*: namely anyone who is defenseless because of «unconsciousness, shipwreck, wounds or sickness» and «anyone who clearly expresses an intention to surrender» (ICRC 2006). The ability of AWS to ensure human-like compliance has been questioned in relation to this rule as well (Sparrow 2015). Indeed, the recognition of behaviors that convey unconventional surrender messages and fighting incapacitation poses formidable challenges for AWS programmers and developers.

Note that making reference to this prohibition in the discussion of AWS is of particular importance: unlike the rules of distinction and proportionality, this rule applies to every warfare scenario in which humans are involved. It therefore counters the argument that IHL would not pose any obstacle to the deployment of lethal AWS in a variety of scenarios where civilians or civilian objects are totally absent (e.g. a battleship in the high seas; Schmitt 2013).

2.4 The unpredictability of AWS

The nature of autonomous systems in general makes their behavior unpredictable, and nowhere more so than in open, unstructured environments replete with unanticipated circumstances (Sharkey 2012; see also, among others, Sparrow 2007; Grut 2013; Geiss 2015; HRW 2015; Egeland 2016). This is particularly problematic in a scenario where one autonomous system interacts with another. Any AWS will be equipped with algorithms relating its input/sensing to its output behavior (its movement, targeting and application of force). These algorithms will have to remain secret: otherwise opponents will easily be able to counter AWS. Consequently, if two AWS or swarms of AWS meet, the resulting behavior of both will be impossible to model adequately and predict with the accuracy required by IHL compliance, with escalation being one, worrisome and plausible outcome, as we will argue further below in section 5. The unpredictability of AWS makes *ex ante* guarantees of IHL compliance extremely challenging, if not impossible.

2.5 The inadequacy of Article 36 reviews of AWS

A review of new weapons, means or methods of warfare to ensure that they can comply with IHL is required under Article 36 of Additional Protocol 1 to the Geneva Convention. In reality only around 12 states actually have the necessary facility to conduct such reviews (Boulanin 2016). When reviews are conducted, their results are never made public. An additional problem for a possible review of a fully computer-controlled system such as an AWS is that it is extremely difficult to formally verify its behavior in all circumstances (see subsection 2.4 above). There are currently no known methods to formally verify an autonomous system.

What this means for AWS reviews is that the only method available to measure IHL compliance is empirical testing. This can only be definitive, however, if the test results in a negative outcome, determining that the system does not comply. This is the case because it is not possible to test the indeterminate and unbounded number of unanticipated circumstances that can occur in the use of force in conflict, so that a system that complies under certain circumstances may not under others.

Some authors argue that, at this stage, the possibility cannot be excluded that someday in the future AWS will be able to comply with IHL principles (e.g. Schmitt 2013: 16-21; Anderson/Waxman 2017: 1109-1110). While this is correct in a very weak sense (how can one exclude the possibility that such goal can *ever* be achieved on the basis of an educated guess), three important clarifications are in order.

First, the «translation» of IHL principles and rules into machine algorithms (if something like that should ever be possible) must never be made at the cost of their oversimplification. This requirement is particularly significant with regard to the rule of proportionality. Contrary to what has been suggested (Schmitt 2013), AWS cannot simply be «pre-programmed» to carry out proportionality analysis on the basis of the assignment by human operators of a given value to each military objective (e.g. a tank or a military base) in terms of admissible collateral damage. After all, the proportionality principle must be respected not only in the planning phases of an attack, but also throughout its execution. Proportionality must be continually reassessed in the light of the circumstances on the ground, which may well change since the attack was launched. Determinations of military advantage and its proportionality, in short, are made on a case-by-case basis in the context of a dynamically changing environment. Therefore, the value, in terms of acceptable collateral damages, assigned to military objectives cannot be determined once and for all by the commander (and even less by the weapon's designers and producers). Rather, it has to be constantly adjusted to match the dynamic features of the environment in which the AWS has been deployed. This requirement is in line with the principle of precaution, whereby «[i]n the conduct of military operations, *constant care* shall be taken to spare the civilian population, civilians and civilian objects» (Additional Protocol I to the Geneva Conventions, Art. 57(1), our emphasis). Should those adjustments be carried out *in real time* by human operators, weapon systems would hardly qualify as autonomous: after all, target selection and engagement would then be subject to high levels of human control. If, on the other hand, AWS were to be endowed with the capability to make proportionality assessments throughout the execution phase, their algorithms would need to be far more sophisticated than the pre-assigned collateral damage value envisaged above which, in turn, would take us back to the problems already discussed in that regard above (in subsections 2.1-2.3).

Second, it should have been made sufficiently clear by this point that the development of AWS able to comply with IHL in cluttered scenarios poses formidable technological challenges, the solution to which is not the expected outcome of any existing research program (in robotics and AI). Accordingly, those who wish to claim that IHL-compliant AWS can be manufactured in a not-so-distant future must bear the formidable burden of proving the scientific basis of their contention.

Third and finally, even if we were to assume, for the sake of the argument, that AWS might someday become capable of human-like performances with respect to adherence to IHL, AWS breaches of IHL are still possible. These persisting problems of accountability and responsibility are what we turn to now.

3. Accountability and responsibility

An argument often raised against autonomy in weapon systems is that it is bound to create accountability gaps (Sparrow 2007; HRW 2015). Indeed, even some proponents of AWS are compelled to admit that, no matter how accurate, these systems will never operate completely flawlessly. Consequently, it is quite possible for an AWS to act in breach of IHL and even to commit acts amounting – at least materially – to war crimes. But then who will be personally accountable for these acts? Since AWS obviously cannot be held responsible as direct perpetrators, responsibility for their actions should be traced back to some persons in the decision-making chain. But this is where the problems begin.

3.1 The «many hands» scenario

At the outset, one should note that the list of potentially responsible individuals is quite long, as it includes «the software programmer, the military commander in charge of the operation, the military personnel that sent the AWS into action or those overseeing its operation, the individual(s) who conducted the weapons review, or political leaders» (Wagner 2016: mn. 22; see also Heyns 2013: para. 77), as well as «the manufacturer of the AWS» (Jain 2016: 321-324) and «the procurement official» (Corn 2016: 230-238). Far from facilitating the task of identifying the responsible individuals, this list raises the familiar «many hands» problem. This problem commonly occurs in software-related accidents, where while a group of people can be blamed collectively for a determined outcome, it is often the case that none of them can individually be held responsible. To the extent that no one actually pushes the «fire» button, and hence assumes at least *prima facie* responsibility in case of wrongdoing, AWS technology will put those involved in their use in the position to «pass the buck» to others (for the discussion of such a scenario see Amoroso/Tamburrini 2017: 7).

3.2 Implications of the unpredictability of AWS for accountability

An additional source of accountability problems lies in the fact that, as noted above, the behavior of an AWS cannot be predicted by its users. How this impacts on individual accountability is easy to grasp. There may be uncontroversial cases, such as that of a machine deliberately pre-programmed to carry out war crimes, or that of a commander who deploys an AWS in a context different from the one that it was designed for, and where it subsequently «commits» war crimes. In many conceivable

circumstances, however, the complexities of weapon autonomy and the resulting behavioral unpredictability in partially structured or unstructured warfare scenarios are likely to afford a powerful defense against criminal prosecution. Indeed, in most cases it would be impossible to ascertain the existence of intent, knowledge or recklessness, which is required under international criminal law (ICL) to ascribe criminal responsibility. As a consequence, it would be highly probable that no one person would be held criminally liable, even if the result of the military operation were to undeniably amount to a war crime.

3.3 Inadequacy of proposed solutions to problems of accountability/ responsibility

None of the proposals put forth to avoid an accountability gap proves able to address this problem adequately. Let us analyze each of them in turn.

Command responsibility. Some have argued that no accountability gap would arise in relation to the use of AWS, by relying on the doctrine of «command responsibility» (Schmitt 2013; NATO JACPP 2016). Under this doctrine, «[c]ommanders and other superiors are criminally responsible for war crimes committed by their subordinates if they knew, or had reason to know, that the subordinates were about to commit or were committing such crimes and did not take all necessary and reasonable measures in their power to prevent their commission, or if such crimes had been committed, to punish the persons responsible» (ICRC 2006: Rule 153). On this basis, it is submitted, the officer who decides to deploy an AWS will be held criminally responsible for war crimes perpetrated by it, if she or he has failed in her or his duty as commander.

However, the doctrine of «command responsibility,» as it currently stands, is of little help in relation to weapons that select and engage targets without human intervention. Let us start with observing that this doctrine is built upon the commander's knowledge of the subordinate's behavior, or at least upon its predictability. Yet, as we have already pointed out in subsection 2.4, there are good reasons to maintain that AWS may well take unforeseeable courses of action, which would make it particularly challenging to establish criminal responsibility.

One could object, in this respect, that human soldiers are «autonomous» as well (Corn 2016: 221), and can act in no less unpredictable ways than AWS, e.g. by disobeying orders. This would overlook, however, a number of important aspects which differentiate the superior-subordinates relationship from the one existing between an AWS and the employing commander.

First, human soldiers are subject to a continuous training process, aimed at instilling awareness of their obligations under IHL (Art. 87(2) I AP), which continues even when they are fielded into a combat scenario (Corn 2016: 222-223). An autonomous weapon, in contrast, could, in a legal sense, only be fielded when first deemed IHL-compliant. Consequently, it would have to be «trained» to respect IHL beforehand, that is, before being used in combat scenarios. Thus, the military commander

would have little or no chance to influence its behavior once it is deployed on the battlefield (Corn 2016: 214).

Second, in case of misconduct by human soldiers, the commander may (and indeed must) exercise her or his punitive power over them – an option that is clearly precluded when the «wrongdoer» is an AWS to which «punishment» is a meaningless concept.

Third, as the International Criminal Court made clear, in order to establish criminal responsibility under the doctrine of «command responsibility», the «superior must have had effective control over the perpetrator at the time at which the superior is said to have failed to exercise his powers to prevent or to punish» (ICC 2009: para 418). This requirement is very unlikely to be satisfied in the case of AWS: after all, no longer having to constantly exert human «effective control» is exactly what AWS are all about, and in the great majority of cases their faster-than-human reaction times would make a commander's intervention impossible (see section 6).

Responsibility of developers or procurement officials. Others have suggested shifting the accountability focus for AWS from the deployment to the development/procurement phase because, at that stage, it would still be possible to ensure that AWS are effectively equipped with all of the cognitive and evaluative capabilities that are needed to faithfully respect IHL principles. Accordingly, responsibility for AWS' war crimes should primarily lie with «military procurement managers, weapons developers and legal advisors» (Corn 2016: 224). However, shifting accountability to the development/procurement phase does not address the issues of unpredictability set out above (in subsection 2.4). If deployed in a dynamic environment, an AWS is capable of taking courses of action whose reason may be unfathomable «even to the system's designers» (Crootof 2016: 1373). Under these circumstances, it seems highly unlikely that those involved in the procurement/development phase could effectively be held responsible.

Opaque recklessness. Another proposal to avoid AWS-related accountability gaps is to lower the *mens rea* threshold, by introducing opaque recklessness as a culpable state of mind in relation to AWS-related war crimes. Under this provision, the defendant would act with recklessness only where he or she «knows his or her conduct is risky but either fails to realize or consciously disregards the specific reasons for the riskiness» (Jain 2016: 317). This would in fact allow holding the commander/field officer/deploying soldier criminally accountable for AWS war crimes, even if she or he was «unaware of the exact risk of harm posed by the AWS's conduct» and even if the latter's actions were «uncertain and unpredictable,» provided that she or he was aware that there was «a substantial and unjustified risk» of some unspecified «dangerous occurrence» (Jain 2016: 318).

The problem with this proposal and others like it is that in dealing with accountability issues, one should always take care not to confuse the fight against impunity with «scapegoat[ing] proximate human beings» (Liu 2016: 341). To the extent that the notion of opaque recklessness stretches culpability to the outer limits of strict liability, it risks the substitution of scapegoating for accountability.

Collective responsibility (State responsibility and corporate product liability). A last attempt to fill AWS-related accountability gaps relies on existing forms of collective responsibility for wrongful acts, namely State responsibility (e.g. Hammond 2015) and corporate product liability (NATO JAPCC 2016: 29-30). This proposal is largely unsatisfactory in that it fails to provide an adequate legal response to serious violations of international law such as war crimes (Chengeta 2016: 49-50). It should be recalled, in this respect, that in any legal system *individual criminal* responsibility performs a crucial, two-fold function. On the one hand, the threat of a punishment *deters* individuals from committing crimes (deterrent function); on the other hand, the actual imposition of a criminal sanction provides an adequate *retribution* to the offender for the harm done (retributive function). Both functions cannot be performed in the same way by collective responsibility, for the well-known reason that international crimes «are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced» (International Military Tribunal Nuremberg 1946: 447).

4. Human dignity, humanity, and public conscience

At a more general level, the process of «dehumanization» of the use of force represented by the development of lethal AWS has been variously stigmatized. This point has been made on the basis of the principle of human dignity and of the Martens Clause.

4.1 Human dignity

Some have claimed that autonomy in lethal weapon systems would run contrary to the principle of human dignity (ICRC 2018: 2). This claim is more far-reaching than the previous ones, as it is built upon a principle that is both foundational and open-textured in ethical and legal contexts alike. Indeed, the former Special Rapporteur on Extrajudicial, Summary or Arbitrary executions, Christof Heyns, characterized the appeal to human dignity as «an overriding consideration,» which would justify a ban on «the deployment of [fully autonomous weapons], no matter the level of technical competence at which they operate» (Heyns 2013: 17). In an oft-quoted passage of his Report on Lethal Autonomous Robotics, Heyns condensed the reasons for this conclusion as follows: «[m]achines lack morality and mortality, and should as a result not have life and death powers over humans» (Heyns 2013: 17).

More analytically, it is possible to distinguish two arguments supporting this view, one of which is centered on agent-relative duties, and the other one on patient-relative rights. The first variant moves from the assumption that the action of suppressing a human life is legally justifiable only if it is non-arbitrary, namely it is based on a considered and informed decision. In order to be non-arbitrary (and here is where the principles of humanity come in), the act of killing must be grounded on human judgement, for only human decision-making guarantees the full appreciation of «the value of individual life [and] the significance of its loss» (HRW 2014: 3). According to the second variant, human dignity would be blatantly denied if people were subject to robotic lethal decision-making, because this would place them in a position where they «have no avenue, futile or not, of appealing to the humanity of the enemy» (Heyns 2017: 156). Indeed, the decision to kill or not would be taken on the basis of hypotheticals set in advance in the AWS programming phase, or developed by the machine itself as rules of behavior extrapolated from its past experience. The ensuing life-or-death decision could, by definition, not be overridden when the AWS is actually releasing force, with the consequence that the human target would essentially be written off without (even the slightest) hope of changing her or his fate.

The upshot of both variants of this argument is that respect for human dignity affords a distinctive moral reason to forbid the use of AWS, which cannot be overridden by any envisaged technological developments that may occur in the future, even by technological developments that might lead to improved performances in AWS's critical targeting and engagement functions. More generally speaking, violations of human dignity cannot be justified by appeal to any other allegedly good consequence deriving from their perpetrations.

This line of ethical and legal reasoning strongly resonates with the basic principle of the inviolability of human dignity as enshrined in Germany's basic law («Grundgesetz») Article 1 (1). It is notably exemplified in a 2006 decision of the German Constitutional Court. The Court ruled that the Defense Minister cannot order the German Air Force («Luftwaffe») to shoot down a hijacked passenger airplane even when there is evidence that the airplane will be used by hijackers as a weapon to kill people on the ground. The basis for this decision was in fact «Grundgesetz» Article 1 (1). According to the Court, by shooting down the hijacked airplane, the State would treat its passengers as mere objects and not as persons, as instruments to achieve some admittedly worthy goal (i.e., saving the life of other people on the ground). By doing so, however, the State would not recognize the special worth as human beings of the airplane passengers, thereby violating their dignity (Bundesverfassungsgericht 2006).

4.2 The Martens Clause

The Martens Clause made its first appearance in international legal parlance in 1899, when it was inserted – on the proposal of the Russian publicist Fyodor Fyodorovich Martens (from whom it takes its name) – in the Preamble of the Second Hague Convention containing the Regulations on the Laws and Customs of War on Land. Subsequently incorporated into a number of IHL treaties (including in the Preamble to the Convention on Certain Conventional Weapons), in its modern formulation the Martens Clause states that, absent any specific regulation, «the civilian population and the combatants shall at all times remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience» (CCW 2001: Fifth Preambular paragraph).

In relation to AWS, the Martens Clause has been invoked to contend that the deployment of weapon systems enabled to take life-or-death decisions without human supervision would run contrary to both ‹the principles of humanity› and ‹the dictates of public conscience› (most recently: Sparrow 2017). As the HRW report pointed out: «The Blinding Lasers Protocol set an international precedent for preemptively banning weapons based, at least in part, on the Martens Clause. Invoking the clause in the context of fully autonomous weapons would be equally appropriate» (HRW 2016: 17).

Reliance on the «principles of humanity» aligns with the view that AWS would be prohibited under the principle of human dignity (see subsection 4.1 above). The «dictates of public conscience» prong of the Martens Clause, however, adds a valuable new element, in that it grounds the abstract concepts of «humanity» and «human

dignity» in the reactions of the international community to certain means and methods of warfare.

Admittedly, this prong of the Clause is the more controversial one, as it is far from certain whose conscience should be taken into consideration and how it would be queried. Nevertheless, the idea that machines should not take life-or-death decisions has been gaining consensus within the international community at large. Evidence of this may be found, in particular, in the declarations rendered by States at the Human Rights Council in reaction to the presentation of the Heyns' Report on LAWS (Campaign to Stop Killer Robots 2013), at the UN General Assembly First Committee on Disarmament and International Security (Campaign to Stop Killer Robots 2018), and during the CCW Informal Meetings of Experts (Lewis/Blum/Modirzadeh 2016: Appendix); in parliamentary initiatives specifically addressing this matter;[6] in reports issued by international human rights supervisory bodies (Heyns 2013; African Commission on Human and Peoples' Rights 2015, para. 35; Kiai/Heyns 2016); in the (qualified) criticism voiced in the aforementioned Open Letters signed in 2015 and 2017, respectively, by renowned experts in the fields of robotics and Artificial Intelligence (AI) and by founders and CEOs of AI and robotics companies; as well as in opinion surveys showing a spreading hostility to non-human lethal decision-making (Ipsos 2017) (especially – and this seems worthy of note – among members of the armed forces).[7] While it would be highly speculative to draw any final conclusions, there is clear evidence for an emerging global norm against weapon autonomy.

6 These include a resolution of the European Parliament (2014) calling on «the High Representative for Foreign Affairs and Security Policy, the Member States and the Council to: […] (d) ban the development, production and use of fully autonomous weapons which enable strikes to be carried out without human intervention». For other parliamentary initiatives, see Campaign to Stop Killer Robots (2017).

7 This is evidenced, in particular, in a survey conducted by YouGov America on the US public (Carpenter 2013a; 2013b; see also Openroboethics 2015 for a globally conducted online survey which underlines these results; cf. Horowitz 2016 for a critical perspective).

5. Global security and stability

A criticism commonly levelled against advocates of a ban on AWS is that they ignore the positive impact that autonomy in weapon systems could have on the protection of innocent civilians, and the respect for IHL in general. Would it not be beneficial, it is asked, if AWS were to become more accurate than human soldiers in targeting military objectives? After all, unlike human soldiers, the argument continues, AWS are utterly unconstrained by the need for self-preservation and immune from human passions (such as anger, fear and vengefulness) (Arkin 2013: 2; Schmitt 2013: 23; Sassòli 2014: 310; Anderson/Waxman 2017: 1108).

We judge the argument about the immunity to human passions a misleading anthropomorphic view of weapon systems. But if we were for a moment, and for the sake of the argument, to say that AWS would in fact be immune to anger, fear and vengefulness, then they would be equally immune to comradery, empathy and compassion. Consequently, such notions are best left aside. The remaining part of the overall argument is the possibly improved protection of innocent civilians. Our in-depth discussion (in section 2) of the very low likelihood that AWS can be compliant with IHL weakens this line of argument from the get go. Nevertheless, let us assume – again, for the sake of argument – that a future deployment of AWS might reduce casualties among belligerents and non-belligerents in a restricted battlefield scenario.

This line of argument is a narrow appraisal that only captures a fraction of the overall picture. It screens off more pervasive effects that are likely to flow from an increased use of AWS. A balanced assessment of expected costs and benefits, in contrast, requires one to take into account the wider landscape of geopolitical implications. Such a more realistic, broader approach would take into account, in addition to local battlefield implications, expected geopolitical consequences of AWS deployment, which range from proliferation to regional and global instabilities and unintended escalation of conflicts. In this wider context, as we shall see below, the expected costs of AWS deployment outweigh by far their alleged benefits in restricted battlefield scenarios.[8]

8 The following paragraphs draw heavily on Altmann/Sauer 2017. See also Tamburrini 2016 for the related distinction between narrow and wide appraisals of the consequences of AWS production, deployment and use.

5.1 Proliferation and arms races

AWS need not take the shape of one specific weapon system, for example a drone or a missile. AWS also do not require a specific military technology development path, in the way that, for example, nuclear weapons do. As the underlying technologies mature and begin to pervade the civilian sphere, militaries will be able to increasingly make use of them for their own purposes, as the development of information and communication technology as a whole suggests. Clearly, as with other military adaptations of dual-use technologies, there are many specific military requirements that do not exist in a civilian environment, or are less relevant for mass markets. Nevertheless, AWS development will profit from the implementation or mirroring of a variety of civilian technologies (or derivatives thereof) and their adoption for military purposes – technologies which are currently either already available or on the cusp of becoming ready for series production in the commercial sector – in this way continuing a trend that is already observable in armed drones. It is fair to say that the autonomy arms race is already underway. In contrast to arms races of the past, however, research and development for AWS-relevant technology is distributed over countless defense contractors, university laboratories and commercial enterprises, making use of economies of scale and the forces of the market to spur competition, lower prices and shorten innovation cycles. This renders the military research and development effort in the case of AWS quite different from those of past hi-tech conventional weapon systems, let alone nuclear weapons. So while the impact of AWS might be revolutionary in terms of the implications for warfare, their development within the military is rather accelerating an already existing trend to replace labor with capital, automate «dull, dirty and dangerous» tasks, and leverage AI for military purposes more generally.[9] Compared to nuclear weapons and the past efforts to curb nuclear arms races and proliferation, moreover, AWS are easy to obtain and harder to regulate, their proliferation more difficult to control.[10] They don't require ores, centrifuges, high-speed fuses or other comparably «exotic» components, put together and tested in a clandestine manner. Consequently, there are comparatively few chokepoints for non-proliferation policies to set their sights on, which renders AWS potentially available to a wide range of state and non-state actors, not just nation states able and determined to muster up the considerable resources required to conduct the robotic equivalent of a Manhattan program. This is why arms control with regard to AWS cannot rely on traditional, quantitative measures but rather requires establishing a norm for meaningful human control. We return to discuss this further below. The less scrupulous an actor is about IHL compliance, the easier AWS development gets. Comparably crude AWS, those not living up to the standards of a professional

9 See, in connection to this, a recently compiled dataset on the developing role of artificial intelligence in weapon systems (Roff/Moyes 2016; Roff 2016).

10 This was also pinpointed by the drafters of the Future of Life Institute (2015) Open Letter, who underscored that «[u]nlike nuclear weapons, [AWS] require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce», rendering them the «the Kalashnikovs of tomorrow».

military, could be put together with technology available today by less technologically advanced state actors or even non-state actors. After all, converting a remotely controlled combat drone to autonomously fire a weapon in response to a simple pattern-recognizing algorithm's command is already doable. It is, at the same time, unlikely that the technological edge regarding sophisticated AWS, desired especially by the US, can be kept over a longer term. The US officially declared AI and robotics the cornerstones of its new «third offset» strategy to counter rising powers, with former Secretary of Defense Ashton Carter seeking closer ties with Silicon Valley to hasten the incorporation of technological innovations into the US military (Hagel 2014; Lamothe 2016). While sensor and weapon packages to a large degree determine the overall capabilities of a system, implementing autonomy comes down to software alone. Software can easily be replicated; at the same time, it is also at a unique risk of being stolen via computer network operations. So while the development of AWS clearly presents a larger technical challenge to less technologically advanced actors, obtaining AWS with at least a noteworthy military capability is certainly not a distant goal for any country already developing remotely controlled UAVs. Proliferation of AWS would of course also occur via exports, including gray and black ones. AWS could in this way fall not only into the hands of less technologically advanced state actors, but also into those of non-state actors, including extremist groups. Hamas, Hezbollah and the Islamic State are examples of three non-state actors that already have deployed and used armed drones. With the increasing trend to miniaturize sensors, electronics and autonomy coming down to software, small and easily transportable systems can be made autonomous with respect to navigation, target recognition, precision and unusual modes of attack. Terrorist groups could also get access to comparably sophisticated systems that they could never develop on their own. Again, autonomy here does not necessarily have to be military-grade but can follow a «quick and dirty» approach. In fact, it stands to reason that terrorist groups would use autonomous killing capabilities indiscriminately as well as, if available to them, in a precise fashion for targeted assassinations.[11] As of yet, it is still unclear how the development of unmanned systems on the one hand, and specific countermeasures on the other will play out. Traditional-aircraft-sized drones such as the US's X-47B testbed or the British prototype Taranis are obviously susceptible to already existing anti-aircraft systems. Regarding smaller-sized systems, various avenues, from microwaves to lasers to rifle-sized radio jammers for disrupting the control link, are currently being developed as countermeasures. Simpler, less exotic methods such as nets, fences or even trained hunting birds might also have an effect – for remotely controlled and autonomous systems alike.

What is quite clear, however, is that saturation attacks have been identified as a key future capability for defeating a wide range of existing and upcoming defensive systems – human operated and also automatic ones (Scharre 2016). And it is especially with regard to the latter that swarming is currently being intensively researched

11 A scenario of rampant AWS proliferation is depicted in the video «Slaughterbots», see YouTube or autonomousweapons.org.

as the key solution. But military systems operating at very high speeds and in great numbers or swarms are bound to generate specific new instabilities, to which we turn now.

5.2 Instability and (unintended) escalation

The case of SARMO systems demonstrates that as a result of increasing operational speeds, the human involvement in AWS would eventually be limited to, at best, general oversight. Hence the actions and reactions of individual AWS as well as AWS swarms[12] would have to be controlled autonomously by software. After all, «winning in swarm combat may depend upon having the best algorithms to enable better coordination and faster reaction times, rather than simply the best platforms» (Scharre 2015). With swarms deployed in close proximity to each other, control software would have to react to signs of an attack within a very short, split-second timeframe by evading or, possibly counter-attacking in a «use-them-or-lose-them» situation. Indications of an attack – a sun glint interpreted as a rocket flame, sudden and unexpected moves of the adversary, or a simple malfunction – could trigger escalation. As already discussed in subsection 2.4, interactions among autonomous systems only increase AWS-inherent unpredictability and are associated with a higher probability of an escalation from crisis to war or, within armed conflict, to higher levels of violence. Comparable runaway interactions between algorithms are empirically observable already. Most famously, considerable havoc was caused in the New York Stock-Exchange «flash crash» of 6 May 2010 in which computerized high-frequency trade played an essential role. Within minutes the prices of many equity products fell by nearly 6%, stock indices and important industry stocks collapsed. In this case «circuit breakers» of the monitoring authorities set in, suspending high-speed trading and preventing further avalanche effects, with fast recovery. These oversight and intervention mechanisms have been improved since then, but debate continues as to whether they are sufficient to prevent another big flash crash, and mini-crashes and interventions occur daily (Shorter/Miller 2014). Humans – or fast reaction mechanisms put in place by humans – can act as a fail-safe in the stock-exchange case because, unlike in international politics, there is human oversight as well as an overarching authority enforcing a shared set of rules. During the Cold War's superpower standoff and afterwards, several well documented instances of erroneous indications of nuclear attack occurred in the US as well as the USSR (Sagan 1993; Blair 1993: Rosenbaum 2011; Schlosser 2013). They varied from sunlight reflected off clouds to magnetic training tapes fed into the early-warning system by accident. In all of these cases human reasoning led to restraint instead of escalation; double checks ended up revealing that the alarm had been false. At the time,

12 In AI and robotics, a swarm intelligence system consists of a population of agents which interact locally with each other on the basis of relatively simple rules of behavior. Swarms do not rely on any centralized control structure. Rather, the global behavior of the swarm results from the composition of the simple local interactions between its members. The principles of swarm AI and swarm robotics were originally inspired by the organization and behavior of biological systems such as ant colonies, bird flocks, and fish schools.

double checking and reconsidering was possible due to missile flight times between several hours (in the case of bombers and cruise missiles) and 30 minutes down to 10 minutes (for ballistic missiles covering intercontinental ranges, or launched from submarines), as well as due to systems put in place for preventing unwanted crisis escalation, with the «hot line» for communication between Moscow and Washington established after the Cuban Missile Crisis being the most prominent example. But with full weapon system autonomy, tried and tested mechanisms for double checking and reconsidering, with humans functioning as fail-safes or circuit-breakers, would be discontinued. This, in combination with unforeseeable algorithm interactions and thus unforeseeable outcomes of military confrontations, increases crisis instability and is unpleasantly reminiscent of Cold War scenarios of «accidental war». Along with the increasing risk of unintended escalation, AWS are bound to introduce stronger incentives for premeditated (including surprise) attacks. This is due to a combination of three factors: casualty avoidance, cost reduction, and the implications of weapon swarming.

First, while unmanned systems, generally speaking, keep soldiers out of harm's way, they also reduce the political risk of military endeavours (especially in democracies where public opinion is a concern). The current generation of remotely controlled combat drones are a case in point. They already make it easier and less costly for states to infringe on the territories of other states. This trend will only intensify with faster, smaller, stealthy and, eventually, autonomous unmanned systems.

Second, AWS need not necessarily be big, costly high-tech weapon systems. Instead, they can be cheap and disposable, 3D-printed units gaining strength from numbers, their «intelligence» residing in a distributed fashion in the swarm or, if outside communication is an option, at some higher level within the military «system of systems» at large.

Third, and in close connection to that, swarms would make mounting a successful defense especially difficult due to their resilience and their capability of attacking from many directions simultaneously in an overwhelming fashion.[13] Small and very small AWS (those sized in the range of centimeters to tens of centimeters) would suffer from limited power supply on board, but could be brought closer to the target by riding along on «motherships». While with payloads of grams to hundreds of grams the amount of destructive power would be limited for small and very small drones, if directed at political or military leadership or sensitive military infrastructure they could produce targeted damage and provide entirely new means for assassinations.

The combination of these three factors – with autonomous swarms as the new, hard-to-defend-against capability – presents a strong incentive to seize the advantage of being the first on the offensive. Swarms of AWS could also be used to attack nuclear weapon delivery systems, command and control systems and sensitive infrastructure components such as antennas, sensors or air intakes. Even though an attacker

13 It is worth mentioning in that regard that swarms inherently represent a challenge to human supervision and control because a human would never be able to control every single attack but can at best supervise «on the loop» what the swarm in its entirety is doing.

could actually have little interest and confidence in the success of such a disarming first strike, the novel possibility alone would increase nervousness between nuclear armed adversaries.

Moreover, AWS will likely continue and intensify the trend towards overlap between conventional and nuclear arsenals, not least by opening up new possibilities for tracking and holding nuclear submarines carrying sea launched ballistic missiles at risk (Brixey-Williams 2016).[14] When nuclear weapons or strategic command and control systems are, or are perceived to be, put at greater risk, such new conventional capabilities end up increasing instability at the global, strategic level.

In sum, today's unmanned systems already increase the risk that military force is used in scenarios where manned systems would previously have presented decision-makers with bigger, caution-inducing hurdles – a connection recently also confirmed in war gaming exercises (CNAS 2016). Swarming AWS need not lead to escalation necessarily and under all conditions, of course. In asymmetric settings, against adversaries lacking AWS capabilities, the escalatory mechanisms developed above would not take effect. But in symmetric settings, they would certainly exacerbate the overall development toward an increased risk of crisis instability and escalation.

After weighing the overall costs and benefits of a possible deployment of AWS, opting for a ban on AWS appears tantamount to choosing the collective rule of behavior that is most consistent with the maintenance of security and stability in a global geopolitical context (Tamburrini 2016: 137-141).

14 For the destabilizing effects on deterrence it is not relevant if the autonomous underwater vehicles are armed themselves or whether they just serve to localize and track enemy submarines for them to be attacked by other weapons.

6. Safeguarding human control over and responsibility for targeting decisions

The considerations in this report so far provide a wide variety of substantive reasons supporting public statements made by nation states such as Germany, but also the United Kingdom and the United States, that there should and will always be a «human in the loop» for all lethality decisions.

In the United Kingdom, the Parliamentary Under Secretary of State, Lord Astor of Hever (as cited in Sharkey 2016: 25), has stated that: «[T]he MoD [Ministry of Defense] currently has no intention of developing systems that operate without human intervention […] let us be absolutely clear that the operation of weapon systems will always – always – be under human control».[15] In the US's (US DoD 2017: 2) policy on autonomous weapons, the authors state: «Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force».

Germany has stated at the CCW in Geneva that «the red line leading to weapon systems taking autonomous decisions over life and death without any possibility for a human intervention in the selection and engagement of targets should not be crossed,» adding that Germany «will not accept that the decision to use force, in particular the decision over life and death, is taken solely by an autonomous system» (Federal Republic of Germany 2015).

A slightly different approach has been taken by the Dutch parliament in discussing a «wider loop» for AWS where humans will be involved in planning attacks before the AWS is launched (AIV/CAVV 2015: chapter 4).

What has not been made absolutely clear in the United Kingdom, however, is exactly what type of human control will be employed. Nor has the US DoD made any attempt to define what is meant by «appropriate levels of human judgment.» Without

15 See also, more recently, the UK's Joint Doctrine Publication JDP 0-30.2 (UK Ministry of Defense 2017: 43, para. 4.18): «The UK Government's policy is clear that the operation of UK weapons will always be under human control as an absolute guarantee of human oversight, authority and accountability. Whilst weapon systems may operate in automatic modes there is always a person involved in setting appropriate parameters».

addressing these points there is no transparency in the operation of computerized weapons.[16] To say that there is a human in the control loop does not clarify the degree of human involvement.[17] It could simply mean that a human has programmed the weapon system for a mission or that a button has been pressed to activate it. It does not necessarily mean that human judgment will be exercised in determining the legitimacy of targets before initiating any individual attack.

The UK-based NGO Article 36 (2014: 2) observes that «[t]he exercise of control over the use of weapons, and concomitant responsibility and accountability for consequences are fundamental to the governance of the use of force and to the protection of the human person.» In the context of the deliberations of the CCW, they have called for an analysis of how *meaningful human control* is exercised over existing weapon systems, as a basis for assessing the feasibility of extending those protocols to AWS.

We can elaborate the requirements for human control of computerized weapon systems through an examination of existing research on human supervisory control. This allows for the development of a classification system such as the five levels of control shown below. These levels should not be taken as absolutes, but rather as a basis for discussion and an initial effort towards the development of a common understanding for all stakeholders (Sharkey 2014; 2016).

Table 1: A classification for levels of human supervisory control of weapons[18]

1. human deliberates about a target before initiating any and every attack
2. software provides a list of targets and human chooses which to attack
3. software selects target and human must approve before attack
4. software selects target and human has restricted time to veto
5. software selects target and initiates attack without human involvement

The importance of the precautionary principle cannot be overstressed in the case of what is referred to as «supervised autonomy». It is thus essential that we avoid the erosion of human involvement and insist upon the legal principle of human control.

An example of the danger of such an erosion can be discussed by using the extension of a SARMO system (such as the German NBS MANTIS or its future derivatives). SARMO systems typically use the speed and trajectory of an incoming munition to calculate where to fire intercepting projectiles (or directed energy) into its path.

16 See Knuckey (2016) for a detailed discussion about transparency.

17 See Saxon (2016) for an excellent analysis of the problems and the vagueness of the DoD's Directive 3000.09.

18 This simple table is adapted from early work on general (non-military) supervised control with ten levels of human supervisory control by Sheridan/Verplank (1978). We discuss these levels and their importance for an understanding of what meaningful human control of weapon systems comprises in detail below.

However, it is a reasonable assumption that such (future) systems would also be capable of locating the source of the incoming fire. Clearly, all that can be detected from the speed and trajectory of the incoming munitions is from where they were fired. It should not be assumed that the assailants are present at that exact location. It is possible, for example, that the incoming munition – such as a mortar – was set in a civilian urban area and fired by remote control. Yet, the information generated by the SARMO system could be used to automatically and almost instantaneously counter-attack the area from where the munitions were launched (with another military asset). This is an obvious and technically feasible extension – but still it needs to be ruled out. After all, it would mean stepping outside of SARMO functionality, as detailed in section 1 above. In other words, the generated information can certainly be useful – but it would have to first be made available to a human commander to assess whether or not there are legitimate targets at the designated spot, whether or not an attack on them would be required by military necessity and what would be the proportionate response.

SARMO weapon systems may be operated with human control and supervision most of the time (in fact, the German MANTIS is said to routinely operate on level 3, with a human approving every attack), but they (or at least many of them) are reasonably placed at level 5 in the hierarchy developed above due to their capability to operate autonomously. Indeed, if push comes to shove humans hardly ever have a sufficient time frame available to veto SARMO actions, as required at level 4. In this respect, the possibility of shutting down SARMO-type systems when human operators realize that something is going, or might soon go, wrong falls short of exercising a veto power on each individual attack. We will return to SARMO systems as a «level 5 exception» to the rule below.

Given the technical, legal, moral and security challenges for the fielding of autonomous weapon systems, it should be clear that meaningful human control of weapons should be the goal of all nation states. In section 1 we proposed that talking about levels of autonomy of weapons both complicated and confused the way forward. A useful reframing is thus to discuss the issues in terms of the types of human control required to conform to international law.

Having fleshed out the importance of human control, we now turn to a fuller discussion of the levels of human control outlined in Table 1 above, as a basis for assessing the compliance of existing and proposed AWS with both IHL and IHRL.

Level 1: A human deliberates about the target before initiating any and every attack

While this level of human control is difficult in many circumstances, it is critically important to aim for the ideal of adhering to the strict requirements specified in Level 1 whenever possible. A human commander (or operator) must have contextual and situational awareness of the target area at the time of a specific attack and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack. There must be active deliberative participation in the attack and sufficient time for judgement on the nature of the target, its significance in terms of the necessity and proportionality of attack, and the likely effects of the attack

beyond engagement with its target. There must also be a means for the rapid suspension or abortion of the attack.

Level 2: Software provides a list of targets and a human chooses which to attack
This type of control could be acceptable if a human in control of an attack was in a position to assess whether an attack is necessary and proportional, whether all (or indeed any) of the suggested alternatives are permissible objects of attack, and which target may be expected to cause the least civilian harm. Without sufficient time or in a distracting environment, the illegitimacy of a target could be overlooked. A rank ordered list of targets is particularly problematic as there would be a tendency to accept the top ranked target unless sufficient time and attentional space is given for deliberative reasoning.

Level 3: Software selects the target and a human must approve it before the attack
This type of control has been experimentally shown to create what is known as automation bias or complacency, in which human operators come to accept computer-generated solutions as correct and disregard, or do not search for, contradictory information. For example, in an experiment on the control of Tomahawk missiles, Cummings (2004) found that operators working under Level 3 had a significantly decreased accuracy when computer recommendations were wrong compared to operators working under Level 2. Decades of experience with the PATRIOT missile defense system confirm this finding (Hawley 2017).

Level 4: Software selects the target and the human has restricted time to veto
This option is unacceptable because it does not require human target identification. Providing only a short time to veto reinforces Level 3 automation bias and leaves no room for doubt or deliberation. As the attack will take place unless a human intervenes, this undermines well-established presumptions under international humanitarian law that promote civilian protection.[19]

Level 5: Software selects the targets and initiates attacks without human involvement
From our earlier arguments, it should be clear that weapon systems controlled at Level 5 are unacceptable, except in the very narrowly bounded circumstance of SARMO. The acceptance of SARMO systems is contingent on the limitations of the system's targeting, that is, it only targeting unambiguously materiel (munitions) if there is no time for human intervention. In addition, the SARMO system has to be constrained by design, that is, it must be near to impossible to tinker with it in the field to make it fire on other targets. It is also essential that utmost vigilance is maintained during the

19 A cautionary tale regarding the errors connected to level 4 control is the 2003 Iraq war in which a US Army PATRIOT missile defense system engaged in fratricide, shooting down a British Tornado and two US F/A-18, killing the pilots (Cummings 2006).

system's operation and that there is a means for rapidly deactivating it if required (see the list of SARMO features in section 1).

This classification of levels of human control is just a beginning. We need to map out exactly the role that the human commander/supervisor plays for each supervised weapon system. Research is urgently needed to ensure that human supervisory interfaces allow the level of human reasoning needed to comply with the laws of war in all circumstances (similarly: Santoni de Sio/van den Hoven 2018).

It is important to underline that, with regard to a possible legally binding instrument to ban AWS – that is, weapons that are not under effective human control – SARMO systems specifically could be allowed to preserve their form of reactive autonomy (and to admit a refined level 5 in the human control hierarchy set out above) even in light of the best ethical and legal arguments against AWS generally.

In turn, however, other AWS should be prohibited (and submitted to human control at a minimum of level 3).

7. Summary

The world's governments are at a crucial juncture with respect to the automation of weapon systems. Further automation is put forward by proponents of AWS as the technological solution to problems created by the increasing speed of war fighting, itself a function of previous waves of automation. A majority of governments and numerous civil society actors are calling for an interruption of this trajectory, based on the proposition that these emerging weapon systems pose fundamental threats to the international norms and to both moral/ethical and legal frameworks governing the conduct of war and the use of violent force.

Our working definition of autonomous weapon systems, in accordance with definitions provided by ICRC and the US Department of Defense, calls out the *critical functions of target selection and the application of force*. At the same time, we found it useful to reframe the issue in terms of human supervisory control. This enables us to more closely examine their governance and tease out what exactly is meant by notions such as «appropriate levels of human judgement», «human oversight», or even «human in» or «on the loop». It is thus the (meaningful) human supervisory control over critical functions that must be subject to careful scrutiny with respect to the legal, technical, moral/ethical, and security implications of AWS development.

Legally, all existing evidence indicates that the deployment of AWS could not comply with International Humanitarian Law for at least the foreseeable future, and that they pose as yet unresolved problems regarding accountability and responsibility for the use of violent force.

Technically, autonomous weapons lack the necessary components to ensure compliance with the IHL requirements of distinction and proportionality. Their behavior is inherently unpredictable, particularly in scenarios where multiple AWS would interact.

Morally, the guiding principle of respect for human dignity, enshrined in German basic law («Grundgesetz») Article 1 (1) as well as in International Human Rights Law, dictates that machines should not be making life or death decisions regarding humans.

IHL, as well as wider public sentiment, underwrites the moral argument that the authority to use lethal force cannot be legitimately delegated to a machine, but rather must remain the responsibility of an accountable human with the duty to make a considered decision regarding necessity and proportionality in the use of force.

In terms of *global security*, the development of AWS poses serious dangers for regional and global stability and provides incentives for their proliferation, including their use by actors not accountable to legal frameworks governing the use of force, as well as their prospective use in civilian settings for policing and the suppression of

dissent. Interactions among AWS, moreover, increase the likelihood of precipitous or unintended escalation of violent conflict.

Recommendations

The last two coalition treaties negotiated by Christian Democrats (CDU/CSU) and Social Democrats (SPD) both pledged that the German government will work toward a ban on AWS at the international level (Bundesregierung 2013: 124; 2018: 149). In accordance with that, Germany has been playing a positive and productive role in the diplomatic deliberations at the CCW in Geneva, inter alia by chairing the last informal meeting of experts in April 2016, by convening a panel of international experts in Berlin to advise the deliberation process (iPRAW 2017), and by presenting a joint paper with France presented at the CCW's first Group of Governmental Experts Meeting in November of 2017 (France and Germany 2017).

The overall slow diplomatic progress at the CCW in Geneva, however, means that the onus rests on national policies more than ever. But while the US, the UK and the Netherlands, to name just three examples, have introduced national policies on the military use of autonomy in weapon systems, Germany has yet to formulate its own doctrine on AWS. Against this background, and recognizing the fact that diplomacy and policy-making risk being outpaced by the speed of military technology development, the time to move is now.

In addition, the AWS debate is comparably young and characterized by the fact that the international debate and national policy-making processes are still in their infancy and closely intertwined; in other words, by establishing norms, national policies can inform, and to a considerable degree steer, the course of the ongoing international arms control debate in Geneva. There is ample room for courageous policy-making aimed at taking a leading role and staking out the territory of future debates. Germany should view this – and seize this – as an opportunity to live up to its grown international responsibilities.

Our summary recommendation, in accordance with the ICRC and other international bodies, is that for legal, ethical and strategic reasons with regard to global security and stability, meaningful human control over weapon systems and the use of force must be retained. For weapons used by the German Bundeswehr this means a level of supervisory control in which a human deliberates about a target before initiating any individual attack. Weapon systems that do not enable this level of human control should be either closely regulated (in the case of SARMO systems) or (in the case of weapons directed at human or human-inhabited targets) pre-emptively banned. In short, we recommend that the German government:

- develop a national policy on the use of autonomy in weapon systems and make that document publicly available;

- adopt a simple, ICRC-like definition of autonomous weapon systems, based on the system's «task autonomy» in the performance of the critical functions of selecting and engaging targets;
- stipulate the legal requirement of «meaningful human control» over all Bundeswehr weapon systems, avoiding autonomy in target selection beyond level 3 (software selects the target and a human must approve it before the attack) thereby effectively prohibiting «*fully* autonomous weapon systems» at the national level;
- continuously investigate and map out levels of human control for autonomous functionality in future Bundeswehr weapon systems to allow the human reasoning and control needed (that is, levels 1-3 for critical functions), and train Bundeswehr personnel accordingly with regard to the required tactics, techniques and procedures;
- regulate the Bundeswehr's continued use of SARMO systems so as to satisfy the strict targeting limitations and constraints on design and operations as these are specified in sections 1 and 6 above;
- continue and intensify support for an international, legally binding and verifiable ban on fully autonomous weapon systems, that is, weapon systems with autonomy in the performance of the critical functions of selecting and engaging targets.

REFERENCES

AFRICAN COMMISSION ON HUMAN AND PEOPLES' RIGHTS 2015: General Comment No. 3 on the African Charter on Human and Peoples' Rights. The Right to Life (Article 4), 57th Ordinary Session, 4-18 November, Pretoria.

AIV/CAVV 2015: Autonomous Weapon Systems. The Need for Meaningful Human Control (No. 97 AIV / No. 26 CAVV), Den Haag.

ALLEN, GREG/CHAN, TANIEL 2017: Artificial Intelligence and National Security. Harvard Kennedy School. Belfer Center for Science and International Affairs, in: http://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf; 14.7.2017.

ALTMANN, JÜRGEN/SAUER, FRANK 2017: Autonomous Weapon Systems and Strategic Stability, in: Survival 59: 5, 117-142.

AMOROSO, DANIELE/TAMBURRINI, GUGLIELMO 2017: The Ethical and Legal Case Against Autonomy in Weapons Systems, in: Global Jurist online first. DOI: https://doi.org/10.1515/gj-2017-0012.

ANDERSON, KENNETH/WAXMAN, MATTHEW C. 2017: Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation Under International Law, in: Brownsword, Roger/Scotford, Eloise/Yeung, Karen (Hrsg.): The Oxford Handbook of Law, Regulation, and Technology (OUP), 1097-1117.

ARKIN, RONALD 2013: Lethal Autonomous Systems and the Plight of the Non-Nombatant, in: AISB Quarterly 137, 4-12.

ARTICLE 36 2014: Key Areas for Debate on Autonomous Weapons Systems. Memorandum for Delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).

BLAIR, BRUCE G. 1993: The Logic of Accidental Nuclear War, Washington, D.C.

BOULANIN, VINCENT 2016: Mapping the Development of Autonomy in Weapon Systems. A Primer on Autonomy. SIPRI Report, in: https://www.sipri.org/sites/default/files/Mapping-development-autonomy-in-weapon-systems.pdf; 17.01.2017.

BOULANIN, VINCENT/VERBRUGGEN, MAAIKE 2017: Mapping the Development of Autonomy in Weapon Systems. SIPRI Report, in: https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_0.pdf; 22.03.2018

BRIXEY-WILLIAMS, SEBASTIAN 2016: Will the Atlantic Become Transparent? (Second Edition), British Pugwash, in: http://www.basicint.org/sites/default/files/Pugwash_TransparentOceans_update_nov2016_v1%281%29.pdf; 10.04.2017.

BUNDESREGIERUNG 2013: Deutschlands Zukunft gestalten. Koalitionsvertrag zwischen CDU, CSU und SPD. 18. Legislaturperiode, in: https://www.cdu.de/sites/default/files/media/dokumente/koalitionsvertrag.pdf; 21.02.2014.

BUNDESREGIERUNG 2018: Ein neuer Aufbruch für Europa. Eine neue Dynamik für Deutschland. Ein neuer Zusammenhalt für unser Land. Koalitionsvertrag zwischen CDU, CSU und SPD. 19. Legislaturperiode, in: https://www.cdu.de/system/tdf/media/dokumente/koalitionsvertrag_2018.pdf?file=1; 22.03.2018.

BUNDESVERFASSUNGSGERICHT 2006: Urteil des Ersten Senats vom 15. Februar 2006, in: BVerfGE 115 115, 118-166, in: http://www.bverfg.de/entscheidungen/rs20060215_1bvr035705.html; 28.1.2018.

CAMPAIGN TO STOP KILLER ROBOTS 2013: Consensus: Killer Robots Must Be Addressed, in: https://www.stopkillerrobots.org/2013/05/nations-to-debate-killer-robots-at-un/; 17.09.2017.

CAMPAIGN TO STOP KILLER ROBOTS 2017: Parliamentary Actions, in: https://www.stopkillerrobots.org/2017/04/parliaments/; 17.09.2017.

CAMPAIGN TO STOP KILLER ROBOTS 2018: Chronology, in: https://www.stopkillerrobots.org/chronology/; 19.03.2018.

CARPENTER, R. CHARLI 2013a: Beware the Killer Robots. Inside the Debate over Autonomous Weapons. Foreign Affairs, in: http://www.foreignaffairs.com/articles/139554/charli-carpenter/beware-the-killer-robots#cid=soc-twitter-at-snapshot-beware_the_killer_robots-000000, 24.07.2013.

CARPENTER, R. CHARLI 2013b: US Public Opinion on Autonomous Weapons. Duck of Minerva, in: http://duckofminerva.dreamhosters.com/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf; 17.09.2017.

CCW 2001: 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 as amended on 21 December 2001, in: https://www.unog.ch/80256EDD006B8954/(httpAssets)/40BDE99D98467348C12571DE0060141E/$file/CCW+text.pdf; 13.01.2018.

CHENGETA, THOMPSON 2016: Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law, in: Denver Journal of International Law and Policy 45: 1, 1-50.

CNAS 2016: Game of Drones: Wargame Report. Center for a New American Security, Washington, D.C., in: http://drones.cnas.org/reports/game-of-drones/; 23.06.2016.

COMMITTEE ESTABLISHED TO REVIEW THE NATO BOMBING CAMPAIGN AGAINST THE FEDERAL REPUBLIC OF YUGOSLAVIA 2000: Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, in: http://www.icty.org/x/file/Press/nato061300.pdf; 12.03.2018.

CORN, GEOFFREY S. 2016: Autonomous Weapons Systems. Managing the Inevitability of ‹Taking the Man Out of the Loop› in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems: Law, Ethics, Policy (CUP), 209-242.

CROOTOF, REBECCA 2016: War Torts. Accountability for Autonomous Weapons 164 University of Pennsylvania Law Review 164: 6, 1347-1402.

CUMMINGS, M.L. 2004: The Need for Command and Control Instant Message Adaptive Interfaces. Lessons Learned from Tactical Tomahawk human-in-the-loop Simulations, in: CyberPsychology & Behavior. The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 7: 6, 653-661.

CUMMINGS, M.L. 2006: Automation and Accountability in Decision Support System Interface Design, in: Journal of Technology Studies 32: 1, 23-31.

EGELAND, KJØLV 2016: Lethal Autonomous Weapon Systems under International Humanitarian Law, in: Nordic Journal of International Law 85: 2, 89-118.

EUROPEAN PARLIAMENT 2014: Resolution on the Use of Armed Drones (2014/2567(RSP)), in: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+MOTION+P7-RC-2014-0201+0+DOC+XML+V0//EN; 20.03.2018.

FEDERAL REPUBLIC OF GERMANY 2015: General Statement at the 2015 CCW Informal Meeting of Experts, in: https://www.unog.ch/80256EDD006B8954/(httpAssets)/97636DEC6F1CBF56C1257E26005FE337/$file/2015_LAWS_MX_Germany.pdf; 22.03.2018.

FRANCE AND GERMANY 2017: Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, in the Context of the Objectives and Purposes of the Convention. For Consideration by the Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS) (CCW/GGE.1/2017/WP.4), in: http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2017/gge/documents/WP4.pdf; 22.03.2018

FUTURE OF LIFE INSTITUTE 2015: Autonomous Weapons. An Open Letter from AI & Robotics Researchers, in: http://futureoflife.org/AI/open_letter_autonomous_weapons#signatories; 31.08.2015.

FUTURE OF LIFE INSTITUTE 2017: An Open Letter to the United Nations Convention on Certain Conventional Weapons, in: https://futureoflife.org/autonomous-weapons-open-letter-2017/; 08.01.2018.

GEISS, ROBIN 2015: The International-Law Dimension of Autonomous Weapons Systems, Friedrich Ebert Stiftung Study, in: http://library.fes.de/pdf-files/id/ipa/11673.pdf; 22.03.2018.

GEISS, ROBIN/LAHMANN, HENNING 2017: Autonomous Weapons Systems: A Paradigm Shift for the Law of Armed Conflict?, in: Ohlin, Jens David (Ed.): Research Handbook on Remote Warfare. Edward Elgar Publishing, 371-404.

GRUT, CHANTAL 2013: The Challenge of Autonomous Lethal Robotics to International Humanitarian Law, in: Journal of Conflict & Security Law 18: 1, 5-23.

HAGEL, CHUCK 2014: Reagan National Defense Forum Keynote, http://www.defense.gov/News/Speeches/Speech-View/Article/606635; 29.07.2016.

HAMMOND, DANIEL N. 2015: Autonomous Weapons and the Problem of State Accountability, in: Chicago Journal of International Law 15: 2, 652-687.

HAWLEY, JOHN K. 2017: Patriot Wars: Automation and the Patriot Air and Missile Defense System, in: https://www.cnas.org/publications/reports/patriot-wars; 22.03.2018.

HEYNS, CHRISTOF 2013: Report by the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, UN Doc. A/HRC/23/47, in: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf; 22.03.2018.

HEYNS, CHRISTOF 2017: A Human Rights Perspective on Autonomous Weapons in Armed Conflict. The Rights to Life and Dignity, in: Auswärtiges Amt (Ed.): Lethal Autonomous Weapons Systems. Technology, Definition, Ethics, Law & Security. German Federal Foreign Office, in: https://www.auswaertiges-amt.de/blob/204830/5f26c2e0826db0d000072441fdeaa8ba/abruestung-laws-data.pdf; 01.03.2018, 148-159.

HOROWITZ, MICHAEL C. 2016: Public opinion and the Politics of the Killer Robots Debate, in: Research & Politics 3: 1, in: http://rap.sagepub.com/content/3/1/2053168015627183; 22.03.2018.

HRW – HUMAN RIGHTS WATCH 2012: Losing our Humanity: The Case against Killer Robots, in: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots; 22.03.2018.

HRW – HUMAN RIGHTS WATCH 2014: Shaking the Foundations: The Human Rights Implications of Killer Robots, in: https://www.hrw.org/sites/default/files/reports/arms0514_ForUpload_0.pdf; 01.03.2018.

HRW – HUMAN RIGHTS WATCH 2015: Mind the Gap: The Lack of Accountability for Killer Robots, in: https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_0.pdf; 31.08.2015.

HRW – HUMAN RIGHTS WATCH 2016: Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban, in: https://www.hrw.org/sites/default/files/report_pdf/arms1216_web.pdf; 22.03.2018.

ICC – INTERNATIONAL CRIMINAL COURT 2009: Decision Pursuant to Article 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor Against Jean-Pierre Bemba Gombo, in: https://www.icc-cpi.int/CourtRecords/CR2009_04528.PDF; 22.03.2018.

ICRC – International Committee of the Red Cross 2006: Customary International Humanitarian Law (edited by Jean-Marie Henckaerts and Louise Doswald-Beck), Vol. I: Rules, Cambridge University Press.

ICRC – International Committee of the Red Cross 2016: Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons, https://shop.icrc.org/autonomous-weapon-systems.html?___store=default; 11.04.2016.

ICRC – International Committee of the Red Cross 2018: Ethics and Autonomous Weapon Aystems: An Ethical Basis for Human Control?, in: https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf; 04.11.2018.

INTERNATIONAL MILITARY TRIBUNAL NUREMBERG 1946: The Trial of German Major War Criminals, Judgment of October 1, in: https://crimeofaggression.info/documents/6/1946_Nuremberg_Judgement.pdf; 22.03.2018.

IPRAW 2017: International Panel on the Regulation of Autonomous Weapons (iPRAW), in: https://www.ipraw.org/; 21.03.2018.

IPSOS 2017: Three in Ten Americans Support Using Autonomous Weapons, in: https://www.ipsos.com/en-us/news-polls/three-ten-americans-support-using-autonomous-weapons; 17.09.2017.

JAIN, NEHA 2016: Autonomous Weapons Systems: New Frameworks for Individual Responsibility, in: Nehal Bhuta et al. (Eds.): Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge University Press, 303-324.

KIAI, MAINA/HEYNS, CHRISTOF 2016: Joint Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association and the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions on the Proper Management of Assemblies, UN Doc. A/HRC/31/66, in: https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwi8zY_60YDaAhXCMewKHcDiDcwQFggzMAA&url=http%3A%2F%2Fwww.ohchr.org%2FEN%2FHRBodies%2FHRC%2FRegularSessions%2FSession31%2FDocuments%2FA.HRC.31.66_E.docx&usg=AOvVaw3IaVel6cHCV7Np0-YE76oJ; 22.03.2018.

KNUCKEY, SARAH 2016: Autonomous Weapons Systems and Transparency: Towards an International Dialogue, in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge University Press, 164-184.

LAMOTHE, DAN 2016: Pentagon Chief Overhauls Silicon Valley Office, Will Open Similar Unit in Boston, Washington Post, in: https://www.washingtonpost.com/news/checkpoint/wp/2016/05/11/pentagon-chief-overhauls-silicon-valley-office-will-open-similar-unit-in-boston/; 25.07.2016.

LEWIS, DUSTIN A./BLUM, GABRIELLA/MODIRZADEH, NAZ K. 2016: War-Algorithm Accountability. Harvard Law School Program on International Law and Armed Conflict, Research Briefing + Appendices, in: https://goo.gl/7oHhYU; 22.03.2018.

LIU, HIN-YAN 2016: Refining Responsibility. Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems, in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge University Press, 325-344.

MAYER, MEGHANN 2017: Milley: Future Conflicts Will Require Smaller Army Units, More Mature Soldiers. ArmyTimes, in: https://www.armytimes.com/news/your-army/2017/03/21/milley-future-conflicts-will-require-smaller-army-units-more-mature-soldiers/; 22.03.2018.

NATO JOINT AIR POWER COMPETENCE CENTRE (JAPCC) 2016: Future Unmanned System Technologies. Legal and Ethical Implications of Increasing Automation, in: https://www.japcc.org/wp-content/uploads/Future_Unmanned_System_Technologies_Web.pdf; 22.03.2018.

OPENROBOETHICS 2015: The Ethics and Governance of Lethal Autonomous Weapons Systems. An International Public Opinion Poll, in: http://www.openroboethics.org/wp-content/uploads/2015/11/ORi_LAWS2015.pdf; 12.11.2015.

ROFF, HEATHER M. 2016: Weapons Autonomy is Rocketing. Foreign Policy, in: http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/; 05.10.2016.

ROFF, HEATHER M./MOYES, RICHARD 2016: Project: Artificial Intelligence, Autonomous Weapons, and Meaningful Human Control. Arizona State University - Global Security Initiative - Autonomy, Robotics & Collective Systems, in: https://globalsecurity.asu.edu/robotics-autonomy; 05.10.2016

ROSENBAUM, RON 2011: How the End Begins: The Road to Nuclear World War III, London.

SAGAN, SCOTT D. 1993: The Limits of Safety: Organizations, Accidents and Nuclear Weapons, Princeton, NJ.

SANTONI DE SIO, FILIPPO/VAN DEN HOVEN, JEROEN 2018: Meaningful Human Control over Autonomous Systems: A Philosophical Account, in: Frontiers in Robotics and AI 5: 15, in: https://www.frontiersin.org/articles/10.3389/frobt.2018.00015/full; 25.03.2018.

SASSÒLI, MARCO 2014: Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified, in: International Law Studies 90: 308-340.

SAUER, FRANK/SCHÖRNIG, NIKLAS 2012: Killer Drones - The Silver Bullet of Democratic Warfare?, in: Security Dialogue 43 (4), 363–380.

SAUER, FRANK 2016: Stopping ‹Killer Robots›: Why Now Is the Time to Ban Autonomous Weapons Systems, in: Arms Control Today 46 (8): 8-13.

SAXON, DAN 2016: A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of Appropriate Levels of Human Judgment over the Use of Force, in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge University Press, 185-208.

SCHARRE, PAUL D. 2015: Counter-Swarm: A Guide to Defeating Robotic Swarms. War on the Rocks in: http://warontherocks.com/2015/03/counter-swarm-a-guide-to-defeating-robotic-swarms/; 23.06.2016.

SCHARRE, PAUL D. 2016: Autonomous Weapons and Operational Risk. CNAS Working Papers, Center for New American Security, Washington, D.C., in: http://www.cnas.org/sites/default/files/publications-pdf/CNAS_Autonomous-weapons-operational-risk.pdf, 01.03.2016.

SCHLOSSER, ERIC 2013: Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety, London.

SCHMITT, MICHAEL N. 2013: Autonomous Weapon Systems and International Humanitarian Law. A Reply to the Critics, in: Harvard National Security Journal Features: 1-37.

SHARKEY, NOEL 2008: Grounds for Discrimination: Autonomous Robot Weapons, in: RUSI Defense Systems 11: 2, 86-89.

SHARKEY, NOEL 2012: The Evitability of Autonomous Robot Warfare, in: International Review of the Red Cross 94: 886, 787-799.

SHARKEY, NOEL 2014:. Towards a Principle for the Human Supervisory Control of Robot Weapons, in: Politica & Società 2: 305-324

SHARKEY, NOEL 2016: Staying in the Loop. Human Supervisory Control of Weapons, in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge University Press, 23-38.

SHERIDAN, T.B./VERPLANK, W. 1978: Human and Computer Control of Undersea Teleoperators, Man-Machine Systems Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

SHORTER, GARY/MILLER, RENA S. 2014: High-Frequency Trading. Background, Concerns, and Regulatory Developments, Congressional Research Service, Washington D.C., in: http://fas.org/sgp/crs/misc/R43608.pdf; 22.07.2016.

SOLIS, GARY D. 2016: The Law of Armed Conflict: International Humanitarian Law in War, Cambridge University Press.

SPARROW, ROBERT 2007: Killer Robots, in: Journal of Applied Philosophy 24: 1, 62–77.

SPARROW, ROBERT 2015: Twenty Seconds to Comply. Autonomous Weapons Systems and the Recognition of Surrender, in: International Law Studies 91, 699-728.

SPARROW, ROBERT 2017: Ethics as a Source of Law: The Martens Clause and Autonomous Weapons, in: Humanitarian Law & Policy. International Committee of the Red Cross Blog, in: http://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/; 14.11.2017.

SUCHMAN, LUCY 2016: Situational Awareness and Adherence to the Principle of Distinction as a Necessary Condition for Lawful Autonomy, in: Geiss, R./ Lahmann, H. (Eds.), Lethal Autonomous Weapon Systems: Technology, Definition, Ethics, Law & Security. German Federal Foreign Office, in: https://www.auswaertiges-amt.de/blob/204830/5f26c2e0826db0d000072441fdeaa8ba/abruestung-laws-data.pdf; 01.03.2018, 273-283.

TAMBURRINI, GUGLIELMO 2016: On Banning Autonomous Weapon Systems. From Deontological to Wide Consequentialist Reasons, in: Bhuta, Nehal et al. (Eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge University Press, 122-141.

UK MINISTRY OF DEFENCE 2017: Joint Doctrine Publication 0-30.2. Unmanned Aircraft Systems, London, in: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf; 22.03.2018.

US DEPARTMENT OF DEFENSE 2013: Unmanned Systems Integrated Roadmap. FY 2013-2038, in: http://www.dtic.mil/dtic/tr/fulltext/u2/a592015.pdf; 01.12.2018.

US DEPARTMENT OF DEFENSE 2017 (2012): Directive 3000.09: Autonomy in Weapon Systems, Washington.

WAGNER, MARKUS 2016: Autonomous Weapon Systems, in: Wolfrum, Rüdiger (Ed.): Max Planck Encyclopedia of Public International Law, Oxford University Press, in: https://goo.gl/CFh3sH; 22.03.2018.

ABOUT THE AUTHORS

Prof. Dr. Daniele Amoroso (Università degli studi di Cagliari, Italy) has published on a variety of issues of international law, including human rights and state responsibility. Amoroso's current research focuses on the legal implications of autonomy in weapon systems.

Dr. Frank Sauer (Bundeswehr University Munich, Germany) headed the Böll Foundation's ‹Task Force on Disruptive Technologies and 21st Century Warfare› that produced this report. Sauer has worked on issues of international security, particularly regarding the role of technology in the modern military, for over a decade. He is currently also a member of the International Panel on the Regulation of Autonomous Weapons (iPRAW).

Prof. Dr. Noel Sharkey (University of Sheffield, United Kingdom) has published extensively on autonomous weapon systems and has a long-standing track record of assessing the ethical, legal and technical issues regarding robots on the battlefield. Sharkey is an Emeritus Professor of AI and Robotics at the University of Sheffield. He is the co-founder of the International Committee for Robot Arms Control (ICRAC) and the Foundation for Responsible Robotics (FRR) and was among the first scholars to publicly raise awareness about the issue of AWS.

Prof. Dr. Lucy Suchman (Lancaster University, United Kingdom) has published widely in both computing and social sciences, from human-computer interaction to contemporary warfighting. Formerly for more than 20 years at Xerox's Palo Alto Research Center, Suchman is a leading scholar in the critical study of artificial intelligence and robotics, with an emphasis on the problem of automating situational awareness.

Prof. Dr. Guglielmo Tamburrini (Università di Napoli Federico II, Italy) has published and held academic seminars on ethical, legal, and social aspects of robotic and AI technologies, including the ethical implications of autonomous robotic weapons. Tamburrini was the coordinator of the first European project on the ethics of robotic systems (Ethicbots project: 2005-08) which addressed such ethical issues as human autonomy and responsibility in human-robot interaction, dual use and social desirability of robotic systems.

The authors are members of the International Committee for Robot Arms Control (ICRAC). Learn more about ICRAC on the web at www.icrac.net and on Twitter @icracnet

DEMOCRACY
VOLUME 49

Autonomy in Weapon Systems

The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy

The future international security landscape will be critically impacted by the military use of artificial intelligence (AI) and robotics. With the advent of autonomous weapon systems (AWS) and a currently unfolding transformation of warfare, we have reached a turning point and are facing a number of grave new legal, ethical and political concerns.

In light of this, the *Task Force on Disruptive Technologies and 21st Century Warfare*, deployed by the Heinrich Böll Foundation, argues that meaningful human control over weapon systems and the use of force must be retained. In their report, the task force authors offer recommendations to the German government and the German armed forces to that effect.

The report argues that in following these recommendations the German government would send a strong signal that Germany is heeding its fundamental norms and values whilst living up to its newly grown international responsibilities.

ISBN 978-3-86928-173-5

Heinrich-Böll-Stiftung Schumannstraße 8, 10117 Berlin
The Green Political Foundation 030 285340 info@boell.de www.boell.de