

LETHAL AUTONOMOUS WEAPONS SYSTEMS: ARTIFICIAL INTELLIGENCE AND AUTONOMY



Raine Sagrarsingh

IEEE // Washington Internships for Students in Engineering

About the Author

Raine Sagramsingh is a recent graduate of the FAMU-FSU College of Engineering. She graduated summa cum laude with a Bachelor of Science in mechanical engineering. She was an active member of the Florida State University Center for Leadership and Social Change's Service Scholar Program. Raine completed internships with the Air Force Research Lab and Northrop Grumman Aerospace Systems.

Raine will begin a full-time position as an electronics engineer with the 412th Test Wing at Edwards Air Force Base after completion of the WISE program.

About IEEE



IEEE and its members inspire a global community to innovate for a better tomorrow through its more than 423,000 members in over 160 countries, and its highly cited publications, conferences, technology standards, and professional and educational activities.

IEEE is the trusted “voice” for engineering, computing, and technology information around the globe.



IEEE-USA is an organizational unit of IEEE, created to support the career and public policy interests of IEEE's U.S. members.

Through its Government Relations programs, IEEE-USA works with all three branches of the federal government to help shape the workforce and technology policy to the benefit of members, the profession and the American public.

About the WISE Program



Founded in 1980 through collaborative efforts of several professional engineering societies, the Washington Internships for Students of Engineering (WISE) program has become one of the premier Washington internship programs. The WISE goal is to

prepare future leaders of the engineering profession in the United States who are aware of, and who can contribute to, the increasingly important issues at the intersection of science, technology, and public policy. The students spend 9 weeks living in Washington, D.C. during which they learn how government officials make decisions on complex technological issues, and how engineers can contribute to the legislative process and regulatory public policy decision-making.

Acknowledgments

The author would like to thank the WISE program and society sponsor, IEEE for the opportunity to participate in this research opportunity. A special thank you to Erica Wissolik, Diana Librizzi, Dr. Gil Brown, and Dr. Lydia Kostopoulos, for their continued guidance and support throughout this process. Finally, a special thank you to the WISE interns (and Jackie Johnson) who have been a sounding board for ideas and have made this a truly impactful 9-week program.

Contents

Acronyms	7
Tables & Figures	7
Executive Summary	8
Introduction	9
Background	12
Global Definitions of LAWS	12
US Definitions of Autonomy	13
Current U.S. Government Action	14
Current Global Discussions	16
UN GGE on LAWS	16
UNIDIR Project on The Weaponization of increasingly Autonomous Technologies.....	16
Campaign to Stop Killer Robots	18
Key Conflicts	19
Goals of AI.....	19
What is Intelligence?.....	21
Automation v. Autonomy & Autonomous	22
International Humanitarian Law	24
Policy Recommendations.....	26
I. Classify the Level of Autonomy and Intelligence of Current Weapons Systems	26
II. Limit the Use of the Term Artificial Intelligence	30
III. Differentiate Autonomous Systems from Autonomous Intelligent Agents	32
IV. Distinguish Between Semi- and Fully-Autonomous Systems.....	33
References	34

Acronyms

LAWS	Lethal Autonomous Weapons Systems
AI	Artificial Intelligence
OODA	Observe, Orient, Decide, Act
NGO	Non-Governmental Organization
IHL	International Humanitarian Law
UN	United Nations
GGE	Group of Governmental Experts
UK	United Kingdom
US	United States
DoD	Department of Defense
ROE	Rules of Engagement
USD(P)	Undersecretary of Defense for Policy
USD(AT&L)	Undersecretary of Defense for Acquisition, Technology, and Logistics
CJCS	Chairman of the Joint Chiefs of Staff
INP	Innovative Naval Prototypes
DE	Directed Energy
EW	Electronic Warfare
CCW	Convention on Certain Conventional Weapons
EU	European Union
UNIDIR	United Nations Institute for Disarmament Research
UNODA	United Nations Office for Disarmament Affairs
ICRC	International Committee of the Red Cross
IQ	Intelligence Quotient

Tables & Figures

Figure 1. OODA Loop	9
Figure 2. Google Term Trends for AI and Autonomy	10
Figure 3. Three Pillars of UNIDIR Phase 3	17
Figure 4. Campaign to Stop Killer Robots [6]	18
Figure 5. Howard Gardener's 9 Types of Intelligence [8]	21
Figure 6. Summary of Turing Test [9]	22
Figure 7. Support for Banning LAWS [6]	25
Figure 8. Diagram of Weapon Systems Currently in Use	28
Figure 9. Subfields of AI [16]	31
Table 1. Countries and Their Definition of LAWS	12
Table 2. U.S. Military Definitions of Autonomy	13
Table 3. Definitions from the 4 Categories of AI	20

Executive Summary

The terms in which the international community discusses lethal autonomous weapons systems (LAWS) can at times be ambiguous as there is an absence of a common vocabulary and common understanding of the terms autonomy and artificial intelligence (AI). LAWS are weapons systems that apply lethal force independently, with freedom from human control. AI has many subfields including machine learning, data mining, speech recognition, and image processing. Policymakers frequently refer to any of the subfields of AI as AI. By grouping these various terms together under the umbrella of AI, the term is frequently misused by technical experts, policymakers, and the general public.

As advances in technology lead to increased autonomy in systems, the terms AI and autonomy are frequently used together. Despite their relationship to each other, it is a mistake to use these terms interchangeably. Currently, the international discussion does not define the technical characteristics of LAWS, hindering the development of public policies, laws, and practices related to the use of autonomy and AI in weapons systems. To improve understanding of the technical characteristics of autonomy and AI in LAWS, this paper aims provides the following recommendations.

Recommendation 1: Classify the level of autonomy and intelligence in current weapon systems. Members of the international discussion on LAWS should evaluate where current weapon systems fall on the spectrum of intelligence and autonomy and should consider which of those should or should not be a part of the continuing discussion.

Recommendation 2: Limit the use of the term Artificial Intelligence. While potential exists for fully-autonomous intelligent weapon systems, they are not a current reality. Using the term AI to refer to machine learning, computer vision, natural language processing, and many other fields of study limits the development of public policy because there is no individual evaluation for their potential positive or negative impact on LAWS.

Recommendation 4: Distinguish autonomous systems from autonomous intelligent agents. An autonomous system does not imply the use of AI, whereas autonomous intelligent agent does. These terms represent significantly different technologies that warrant separate discussion.

Recommendation 5: Distinguish between semi- and fully-autonomous systems. To make progress developing public policies regarding LAWS, the discussion must distinguish semi-autonomous systems from fully-autonomous systems. Regulating the two technologies with the same set of policies ultimately will be unsuccessful.

Introduction

International discussions about lethal autonomous weapons systems (LAWS) have become more prevalent as nations attempt to come to terms with how these weapons work and how to regulate them. The terms in which the community discusses them can at times be ambiguous as there is an absence of a common vocabulary and common understanding of the terms autonomy and artificial intelligence (AI). A **weapon system** is comprised of four actions, searching for a target, detecting that target, making the decision to engage the target, and engaging the target. [1] Developed by Colonel John Boyd, the OODA loop is a model for competitive decision making, comprised of those four actions that define a weapon system: observe, orient, decide, and act. [1] In human operated weapons systems, a human operator completes all of these actions. In a weapon system with autonomy, the human operator relinquishes control over some part of the loop.

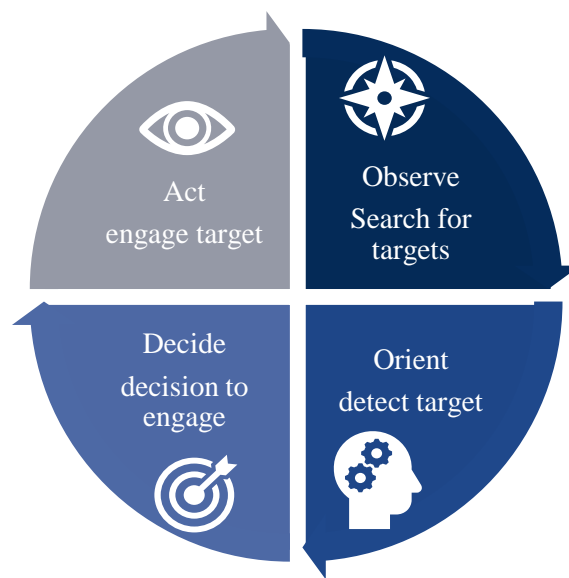


Figure 1. OODA Loop

The word autonomy comes from the early 17th century, Greek word, *autonomia*, stemming from *autonomos* which means, ‘having its own laws’. The Oxford English Dictionary defines **autonomous** as “freedom from external control or influence; independence.” Based on the above definition, **LAWS** are weapons systems that apply lethal force independently, with freedom from human control. This definition encompasses weapon systems from landmines¹ to unmanned aerial vehicles² to the Israeli Harpy³. While this range of weapons systems have autonomous aspects, they differ in the complexity of their algorithms and software, or their level of intelligence. Landmines merely sense pressure, movement, sound, magnetism or vibration while the Harpy, “detects, attacks and destroys enemy radar emitters, hitting them with high hit accuracy” [2]. The increasing complexity of weapons systems and their software led to the introduction of AI into the LAWS discussion.

¹ Landmines: explosive devices triggered by the presence of a person or vehicle.

² Unmanned Aerial Vehicle: an aircraft piloted by remote control or onboard computers. [3]

³ Harpy: a loitering munition that searches for and engages a target but requires no human approval. [2]

Computer scientist John McCarthy first used the term artificial intelligence in 1955. [3] The term's recent overwhelming use makes AI a point of discussion in many fields. Military and civilian applications of AI are broad, including commuting, banking, farming, and search and rescue efforts. As a buzzword for political, academic, and private spheres, AI is frequently used as an umbrella term for fields including machine learning, data mining, speech recognition, and image processing. By grouping such terms together, the term AI is frequently misused by technical experts, policymakers, and the general public. Figure 2 below shows Google's search interests over the last five years for the terms autonomy and artificial intelligence. An overall upward trend in the interest in AI is evident in the past three years.

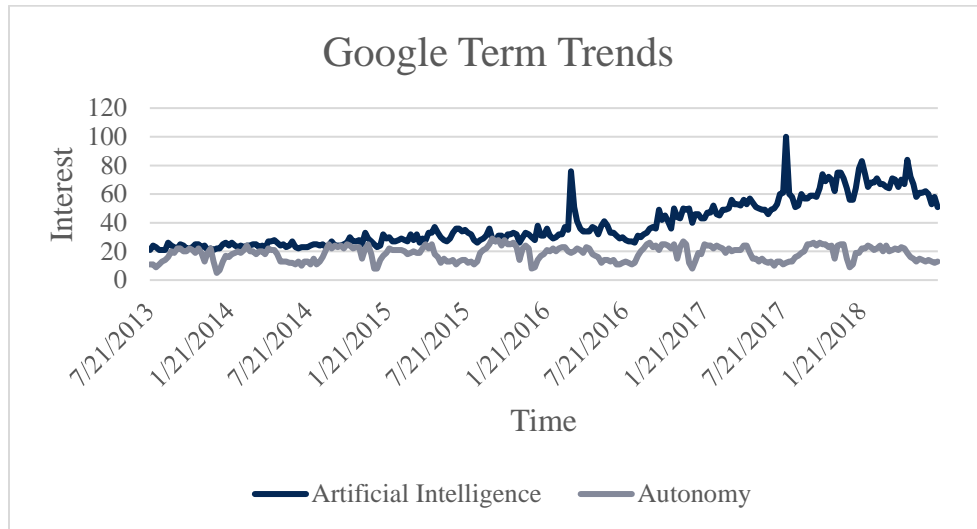


Figure 2. Google Term Trends for AI and Autonomy⁴

The international conversation on LAWS leaves several key issues unaddressed in the process of creating public policy regarding the use of autonomy and AI in weapons systems. As advances in technology lead to increased autonomy and AI, the scope of these technologies frequently overlaps, however it is a mistake to use the terms interchangeably because intelligence is not analogous with autonomy. A system does not need AI to be autonomous and the presence of AI does not imply the system is autonomous. To differentiate the terms, an understanding of the technical characteristics and bounds related to each is necessary. Another issue is a misunderstanding on how these technologies affect weapons systems. The lack of a universal definition of AI and autonomous weapon system leads to additional misunderstanding of the characteristics of AI and autonomy. Without comprehensive knowledge of the technologies, policymakers cannot understand what impacts these technologies will have on LAWS. The absence of a common consensus within the international community on LAWS hinders the development of public policies, laws, and practices related to the use of autonomy and AI in weapons systems.

⁴ Numbers represent search interest relative to the highest point on the chart for the U.S. over the last 5 years. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Experts and policymakers often discuss autonomy in weapons systems in terms of human-machine interaction. The level of autonomy being dependent on the amount of human input that the system requires to deploy lethal force. Intelligence, on the other hand, has no relation to human input. Instead, a machine's ability to decide based upon a certain input and preconfigured algorithm is the measure of its intelligence. Members of the AI community make a strong argument for keeping a human in the loop⁵ of decisions made by weapons systems, even intelligent ones. One example is the Aegis Combat System, used by the U.S. Navy⁶. This system has the ability to observe, track, target and engage without human input, demonstrating a high level of autonomy. However, human operators usually operate the system in a mode requiring human input at various points in the process. Even in a fully autonomous setting, the human operator can override the system's decision.

An international ban on LAWS currently has widespread support from members of academia, the private sector, and non-governmental organizations (NGOs). Several leaders in the fields of AI and autonomy argue against the use of these technologies in weapons systems. Not all parties in the international discussion on LAWS agree with a ban, especially a preemptive ban on technologies that do not exist yet. Although many arguments exist on whether to ban LAWS, one of the most significant is the discussion of International Humanitarian Law (IHL). States have the responsibility to uphold IHL, but the methods to do so differ between nations. The various definitions of LAWS only complicate the understanding of how the systems uphold IHL. For either side of the banning LAWS argument, a definition, or at least a common understanding of the terminology is paramount. Without a consensus on the technology there is no basis on which to develop a ban.

The United Nations (UN) Group of Governmental Experts (GGE) on LAWS will meet in August 2018 to continue the discussion on the future of LAWS. Participants in the GGE include over 40 states, and several NGOs. Documents from the last GGE meeting present varying definitions and system characterizations of LAWS from 10 different countries and numerous non-governmental sources. Each group, while providing their own analysis and proposed definition, agrees that a lack of a working definition impedes the progress of discussions on LAWS. These international discussions face obstacles because the parties involved lack an agreed common understanding of the technologies. Additionally, determining how to classify weapons systems as autonomous and how AI plays a role in that classification hampers discussions. Public policy decisions therefore face delay or total lack of action because the involved parties disagree on what the subject of such public policy action is.

⁵ In the OODA loop, when the system interacts with the human operator at any point within the loop, it is referred to as having a "human in the loop"

⁶ The Aegis System is "a centralized, automated, command-and-control (C2) and weapons control system that was designed as a total weapon system, from detection to kill." [1]

Background

Global Definitions of LAWS

The international discussion on LAWS faces added difficulty due to the various definitions of autonomous weapons systems adopted by the participating countries. Below is an outline of the adopted definitions by several countries, adapted from the 2017 GGE chair's document.⁷

Table 1. Countries and Their Definition of LAWS

Country	Definition of Lethal Autonomous Weapon System
Cuba	Fully autonomous weapons should be understood as those who act without human supervision once they are deployed. The semi-autonomous weapons must be understood as those with the supervision of a human operator, at least in its critical functions, that is to say for selection and attack of targets.
France & Germany	LAWS are defined as fully autonomous lethal weapon systems. Systems such as remotely piloted and automated systems (e.g. conventional charges exploding with a set timer), tele-operated (e.g. drones), automated missile defense systems, torpedoes, guidance and navigation systems, surveillance and detection systems are not considered as LAWS.
Netherlands	"A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention."
Russia (Ministry of Defense)	"Autonomous weapons system— an unmanned piece of technical equipment that is not a munition and is designed to perform military and support tasks under remote control by an operator, autonomously or using the combination of these methods"
Switzerland	Weapons systems that are capable of carrying out tasks governed by IHL in partial or full replacement of a human in the use of force, notably in the targeting cycle.
UK	An autonomous system is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.
US	A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation.

⁷ This document provides, "characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention".

While most countries have varying definitions of LAWS, one point they all agree on is the need for a unified outline or definition of LAWS. In the most recent GGE meeting on LAWS, several countries included in their working papers that a crucial step for the GGE’s conversation is to come to a consensus on the technologies that they plan to include as a part of the LAWS discussion.

US Definitions of Autonomy

The U.S. definition of LAWS, referred to by the U.S. as autonomous weapon system, presented above in table 1 is from the U.S. Department of Defense (DoD) Directive 3000.09. This directive also provides a distinguished definition for a semi-autonomous weapon system.

“semi-autonomous weapon system. A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. This includes:

1. Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement.
2. “Fire and forget” or lock-on-after-launch homing munitions that rely on TTPs to maximize the probability that the only targets within the seeker’s acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator.”

Despite this well-outlined definition of LAWS from the DoD, the U.S. Navy, air force, and army each provide varying definitions of autonomy and autonomous systems. Below is a list of the various definitions provided in official documents from each of the three branches.

Table 2. U.S. Military Definitions of Autonomy

Military Branch	Definition of Autonomy
Air Force	“Systems which have a set of intelligence-based capabilities that allow it to respond within a bounded domain to situations that were not pre-programmed or anticipated in the design (i.e., decision-based responses) for operations in unstructured, dynamic, uncertain, and adversarial environments.” [4]
Army	“the level of independence that humans grant a system to execute a given task in a given environment. This independence is a point on a spectrum that can be tailored to the specific mission, level of acceptable risk, and degree of human-machine teaming.” [5]
Navy	“proposed by the U.S. Navy’s 28th Strategic Studies Group in its report <i>The Integration of Unmanned Systems into Navy Force Structure</i> : ‘a level of autonomy is more correctly addressed as a combination of a degree of human interaction with a degree of machine automation . . .’”

Current U.S. Government Action

The DoD Directive 3000.09 is the current instruction on autonomy in weapon systems. Its two main purposes are to assign responsibility for development and use of semi- and fully-autonomous functions in weapons systems, and to minimize probability and consequences of failures in these systems that could lead to unintended engagement. The directive has 5 main policies,

1. 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.
2. Persons who authorize the use of, direct the use of, or operate autonomous and semiautonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).
3. Autonomous and semi-autonomous weapon systems intended to be used in a manner that falls within the policies in subparagraphs 4.c.(1) through 4.c.(3) will be considered for approval in accordance with the approval procedures in DoD Directive 5000.01 (Reference (b)), DoD Instruction 5000.02 (Reference (c)), and other applicable policies and issuances.
4. Autonomous or semi-autonomous weapon systems intended to be used in a manner that falls outside the policies in subparagraphs 4.c.(1) through 4.c.(3) must be approved by the Under Secretary of Defense for Policy (USD(P)); the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)); and the Chairman of the Joint Chiefs of Staff (CJCS) before formal development and again before fielding in accordance with the guidelines in Enclosure 3, References (b) and (c), and other applicable policies and issuances.
5. International sales or transfers of autonomous and semi-autonomous weapon systems will be approved in accordance with existing technology security and foreign disclosure requirements and processes, in accordance with Directive-Type Memorandum 11-053 DoD Directive 5111.21 (Reference (e)).

In May 2018, the White House held a Summit on Artificial Intelligence for American Industry with senior government officials, technical experts, and business leaders. The purpose of the summit was to discuss AI and the public policies necessary to “maintain U.S. leadership in the age of artificial intelligence”. Congressman John K. Delaney (MD-6) launched the Artificial Intelligence Caucus, co-chaired by Congressman Pete Olson (TX-22) in 2017. The goal of the AI Caucus is to “inform policymakers of the technological, economic and social impacts of advances in AI and to ensure that rapid innovation in AI and related fields benefits Americans as fully as possible.” The caucus has 21 bipartisan members of Congress, and their discussions include experts from government, academia, and the private sector.

Representative Elise M. Stefanik introduced H.R. 5356, The National Security Commission Artificial Intelligence Act of 2018, in March 2018. The bill establishes the National Security Commission on Artificial Intelligence, an 11-member independent commission responsible for reviewing advances in AI and related technologies. The Commission must also assess the competitiveness of the U.S. among countries investing in AI, the means and methods to maintain a technological advantage, international trends, the means to foster research, workforce education, risks of military use of AI, ethical considerations, the means for data standards, and privacy- and security-protecting measures. The bill explicitly defines artificial intelligence as the following:

1. Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.
2. An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.
3. An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
4. A set of techniques, including machine learning, that is designed to approximate a cognitive task.
5. An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision-making, and acting.

Since autonomy is already an integral part of many military systems, the U.S. Government oversight of autonomous systems is located in appropriations to the DoD. From H.R. 5515, the National Defense Authorization Act for FY 2019, Congress charges the DoD with assessing the impact of emerging autonomous technology in the air and space, on and under sea, and on land. [6]

Current Global Discussions

UN GGE on LAWS

In 2014, the UN Convention on Certain Conventional Weapons (CCW) held their first meeting on LAWS to discuss emerging technologies in LAWS. Eighty-seven countries participated in the informal meeting. Additionally, representatives from the European Union (EU), United Nations Institute for Disarmament Research (UNIDIR), United Nations Office for Disarmament Affairs (UNODA), and International Committee of the Red Cross (ICRC), as well as a number of non-governmental organizations and research and academic entities participated. In 2016, the CCW established a Group of Governmental Experts on LAWS. The GGE's first meeting of 2018 took place in April 2018, and the second is in August 2018. The GGE meetings on LAWS addressed four main issues,

1. Characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the CCW;
2. Further consideration of the human element in the use of lethal force; aspects of human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems;
3. Review of potential military applications of related technologies in the context of the Group's work;
4. Possible options for addressing the humanitarian and international security challenges posed by emerging technologies in the area of LAWS in the context of the objectives and purposes of the Convention without prejudging policy outcomes and taking into account past, present and future proposals.

Characterization of LAWS has consistently been an issue for the GGE. Several individuals in attendance have echoed the idea that it took the first year of GGE meetings to establish that the conversation was not about drones. The slow progress of discussion within the GGE is partially due to the disagreement on the terminology and the technologies that fall within the purview of LAWS.

UNIDIR Project on The Weaponization of increasingly Autonomous Technologies

The UNIDIR Project on The Weaponizing of Increasingly Autonomous Technologies has three phases, two of which – the implications for security and arms control and addressing competing narratives – are complete. The third and final phase is promoting “a practical understanding among policymakers of the potential challenges raised by increasingly autonomous technological capabilities—and in particular learning systems—in the near to medium term”. This phase has three pillars, shown below in figure 3.

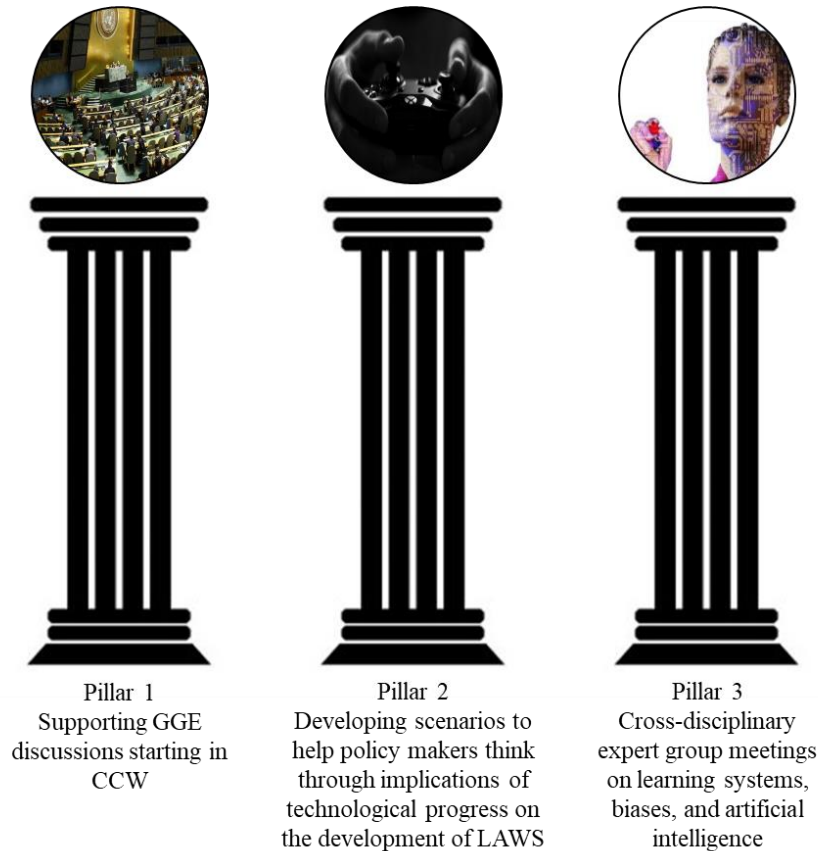


Figure 3. Three Pillars of UNIDIR Phase 3

In, *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches*, UNIDIR presents three commonly discussed definitional approaches to LAWS.

1. Technology-centric approach – a technical definition that physically describes an object.
2. Human-centered approach – a description that describes an autonomous weapons system in relation to a human user
3. Task/Functions approach – identifying tasks/functions delegated to a weapon which make it autonomous

While various nation states support each of these three approaches, UNIDIR argues that combining the three – starting with human-centric to reiterate human responsibility, then identifying tasks/functions of concern, and finally discussing the technology – is an acceptable and all-encompassing approach.

Campaign to Stop Killer Robots

A group of NGOs launched The Campaign to Stop Killer Robots, an international coalition to ban fully-autonomous weapons in 2013. Participating organizations include Human Rights Watch, Amnesty International, Harvard Law School International Human Rights Clinic, and the International Committee for Robot Arms Control. The coalition’s main concern is that weapons will operate on their own without “meaningful human control”. The Campaign aims to inform policymakers and the general public about the issues concerning LAWS, and makes statements and provides resources, such as the one shown below, to educate their audience.

STOP KILLER ROBOTS
WHY WE NEED TO PUT BRAKES ON THE DEVELOPMENT OF FULLY AUTONOMOUS WEAPONS BEFORE IT IS TOO LATE

WHAT ARE FULLY AUTONOMOUS WEAPONS?
Fully autonomous weapons - or killer robots - are weapons that can, without human control, detect, select and engage targets. They do not yet exist, but the rapid developments in robotics and autonomous technology, indicate that it is only a matter of time before fully autonomous weapons become an inhumane reality.

NOT JUST FICTION: SOME EXISTING (SEMI-)AUTONOMOUS SYSTEMS

<p>SAMSUNG SGR-A1 Type: Sentry gun Current use: uncertain, previously guarded Korean Demilitarized Zone Autonomous mode: Partial Feature: can track multiple moving targets</p>	<p>BAE SYSTEMS TARANIS Type: Unmanned Combat Air Vehicle Current use: demonstration only Autonomous mode: Yes Feature: capable of flying intercontinental missions</p>	<p>NORTHROP GRUMMAN X-47B Type: Unmanned Combat Air Vehicle Current use: demonstration only Autonomous mode: Yes Feature: autonomously land on aircraft carriers, one of the hardest manoeuvres</p>
--	---	--

£25 MILLION
EPSRC FUNDING TOWARDS RESEARCH INTO ROBOTICS & AUTONOMOUS SYSTEMS AT 20 UK UNIVERSITIES

THE PROBLEMS WITH KILLER ROBOTS

<p>The extent and nature of human control over fully autonomous weapons is unclear. States are unlikely to let killer robots roam freely but it is left to be determined what constitutes “meaningful” human control.</p>	<p>Concerns about about robots’ lack in human judgement and the ability to understand context necessary to make ethical choices on a dynamic battlefield and to distinguish between soldiers and civilians.</p>	<p>Killer robots would create an accountability gap. There is no clarity on who would be legally responsible for a robot’s actions: the commander, programmer, manufacturer, or the robot itself?</p>	<p>Replacing human troops with machines could make the decision to go to war easier, which could shift the burden of armed conflict even further onto civilians.</p>
---	---	---	--

THOUGHTS ON KILLER ROBOTS

<p>We need not wait for a weapon system to emerge fully before appropriate action can be taken to understand its implications and mitigate and eliminate unacceptable risks. Angela Kane, UN High Representative for Disarmament Affairs</p>	<p>It is possible to prevent the development and proliferation of autonomous robot weapons, but only if we act now before there is too much investment. Professor Noel Sharkey, Chair of the International Committee for Robot Arms Control</p>	<p>Machines lack morality and mortality, and as a result should not have life and death powers over humans. Christof Heyns, UN Special Rapporteur on extrajudicial, summary or arbitrary executions</p>
---	--	--

AOAV IS A MEMBER OF THE CAMPAIGN TO STOP KILLER ROBOTS. PLEASE VISIT STOPKILLERROBOTS.ORG.

Sources: AP / Article 36 / Campaign to Stop Killer Robots / Huffington Post / parliament.uk / QinetiQ / Reuters / UN Photo / Wired

Figure 4. Campaign to Stop Killer Robots [6]

Key Conflicts

Goals of AI

One of the impedances to developing public policy regarding AI is answering the question, *what is AI?* The Oxford English Dictionary defines **Artificial Intelligence** as, “They theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.” One of the difficulties with defining AI is that the term encompasses a variety of other fields such as machine learning, speech recognition, computer vision, robotics, language understanding and reasoning. As the technology and understanding of each sub-field developed, the term for the specific field became more widely used. Some define AI as the state of the art in developing systems that achieve the goal of thinking like humans. Therefore, AI is continually changing as the science and engineering behind the technologies develop.

Scholars have several methods of dividing or distinguishing AI such as “weak” and “strong AI”. “**Weak AI**” suggests that humans design machines to act as if they are intelligent. This means that the programmer hard-codes every decision into the system, and the machine has no capability to learn. “**Strong AI**” leads to machines that can be conscious. At which point, the system would be independently thinking and learning. Hard-coding of every decision would be unnecessary, and the system would be capable of evolving and improving its code based on its previous experiences. Distinguishing between narrow and general AI is also important. **Narrow AI** is considered to be a system that is able to complete one or a narrow set of tasks at the same level or better than humans. **General AI**⁸ is considered to be a system that is able to complete all tasks that a human can, while also being able to adapt to uncertain situations. This definition implies the system has machine learning and all of the sensory and decision-making abilities of a human.

Stuart Russell, a leader in the field of AI, presents four possible goals of AI. These goals represent the ultimate capabilities of AI and include systems that think like humans, systems that act like humans, systems that think rationally, and systems that act rationally. Russell pulls definitions from several AI textbooks that support each of the four goals in Table 3, below.

⁸ General AI is not a reality. This is the type of AI that is portrayed in science-fiction like *The Terminator's* Skynet.

Table 3. Definitions from the 4 Categories of AI

Think Like Humans	<ul style="list-style-type: none">• “The exciting new effort to make computers think...<i>machines with minds</i>, in the full and literal sense” (Haugeland, 1985)• “[Automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning” (Bellman, 1978)
Act Like Humans	<ul style="list-style-type: none">• “The art of creating machines that perform functions that require intelligence when performed by people” (Kurzweil, 1990)• “The study of how to make computers do things at which, at the moment, people are better” (Rick and Knight, 1991)
Think Rationally	<ul style="list-style-type: none">• “The study of mental faculties through the use of computational models” (Charniak and McDermott, 1985)• “The study of the computations that make it possible to perceive, reason, and act” (Winston, 1992)
Act Rationally	<ul style="list-style-type: none">• “A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes” (Schalkoff, 1990)• “The branch of computer science that is concerned with the automation of intelligent behavior” (Luger and Stubblefield, 1993)

The various classifications and methods of distinguishing AI make a general understanding and consensus on its definition problematic. Due to the constantly shifting nature of the technology, discussions on a definition of AI are seemingly always behind. To create public policy on AI, policymakers must have a better understanding of the bounds of the technology. A preliminary step may include determining the minimum capabilities of a system that parties involved in the LAWS discussion consider artificially intelligent. This may rely on the parties’ definition of intelligence, further complicating the conversation.

What is Intelligence?

No less disputed than the definition of AI is the definition of intelligence. While scholars debate what intelligence is, no standard definition exists. Like the definition of AI, the various definitions of intelligence have common themes. Legg and Hutter compiled a list of some of the most referenced definitions of intelligence, two of which are: “The ability to acquire and apply knowledge and skills” and “. . . the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system’s ultimate goal.” From their compilation of definitions, common themes of defining intelligence were:

1. intelligence is a property that an individual agent has as it interacts with its environment or environments
2. intelligence is related to the agent’s ability to succeed or profit with respect to some goal or objective
3. intelligence depends on how able the agent is to adapt to different objectives and environments.”

From these common themes they produced their own definition of **intelligence**: “Intelligence measures an agent’s ability to achieve goals in a wide range of environments.” [7]

In humans, IQ tests measure intelligence quotient, but some experts debate how meaningful these tests are in measuring one’s intelligence. Psychologist Howard Gardener developed a method of describing intelligence that included 9 different types of intelligence. The figure below outlines these 9 types of intelligence.

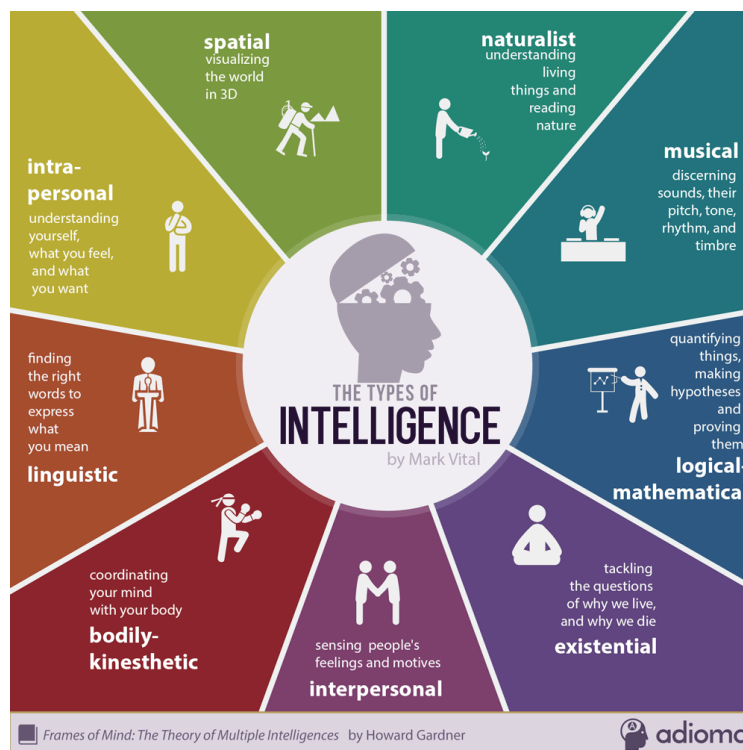


Figure 5. Howard Gardener's 9 Types of Intelligence [8]

Regarding AI, discussions of intelligence raise two additional questions, what is intelligence in machines...for example, is a machine intelligent if the system lacks linguistic intelligence and therefore cannot verbally express what it means? How do we measure a machine's intelligence? Alan Turing, in 1950, tackled this question and developed the Turing Test, one of the most well-known measures for machine intelligence. Turing designed the test to determine if a computer is capable of thinking like a human being. Below is an overview of the Turing Test.

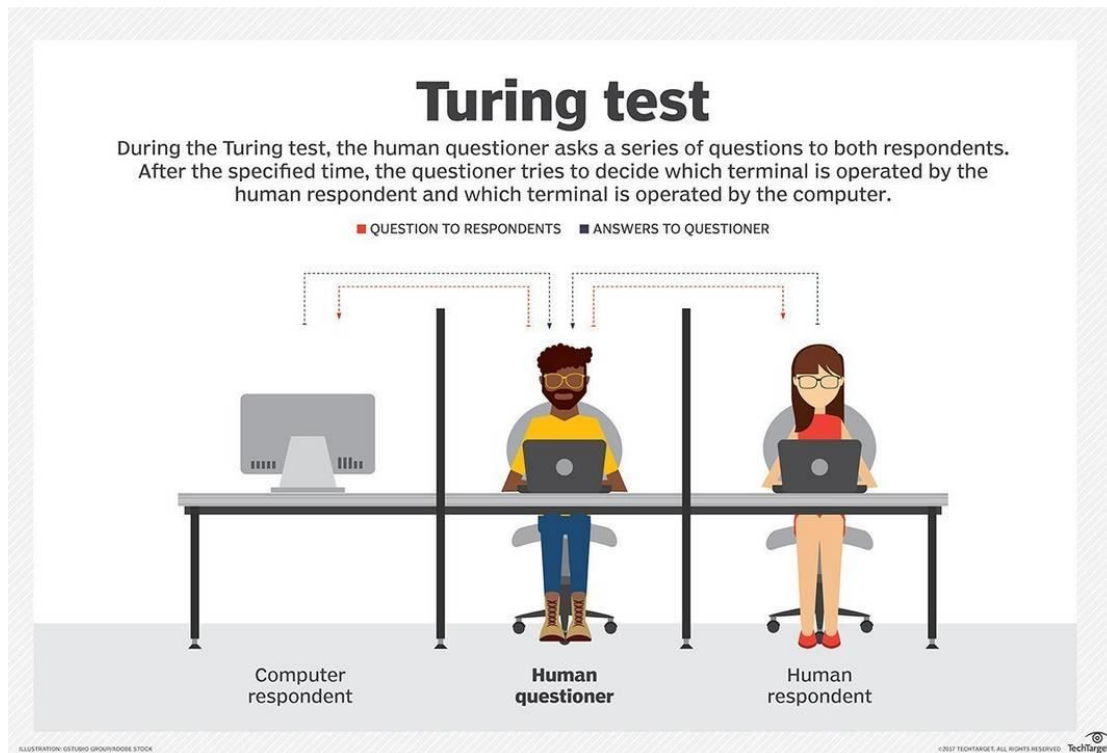


Figure 6. Summary of Turing Test [9]

Despite the existence of such “computer intelligence tests”, experts still dispute how to define intelligence in both humans and machines, and whether a machine is only artificially intelligent if the system possesses all of the components of intelligence that humans do.

Automation v. Autonomy & Autonomous

The Oxford English Dictionary provides the following definitions for automation, autonomy, and autonomous.

Automation: “The use or introduction of automatic equipment in a manufacturing or other process or facility.”

Autonomy: “Freedom from external control or influence; independence.”

Autonomous: “Denoting or performed by a device capable of operating without direct human control.”

Each of the three previous terms has a unique definition and meaning, yet frequently policymakers use them interchangeably. Currently, automated weapons are a staple of almost every military entity, and many militaries use systems that have some autonomy. One example is the Aegis Combat System, used by the U.S. Navy⁹. This system has four modes of operation,

1. Manual – the system’s radar tracks objects and once detected, a human must directly engage
2. Semi-Auto – the system generates a firing solution after tracking an object and a human operator makes the final decision on engaging
3. Auto SM – the system fully develops the engagement process however engagement requires positive human action
4. Auto-Special – when the system detects a threat that meets given parameters, it will automatically fire

Current use of autonomous weapons systems depends on one’s definition of autonomous. Stuart Russell uses a “stronger” view of an autonomous system. Instead of autonomous being the ability for a system to operate without the input of a human, Russell defines autonomous by the system’s programming. He states the following,

“If the agent’s actions are based completely on built-in knowledge, such that it need pay no attention to its percepts, then we say that the agent lacks autonomy.” [10]

If the programmer hard-codes every decision, or “move”, into the system, then that system has no autonomy – only the reflection of the original coder or operator’s intentions. For the system to be a truly autonomous intelligent agent, the system should be able to learn and adapt, ultimately making decisions and learning without written code for each and every step. Based upon his interpretation, a system cannot be autonomous without at least a basic form of machine learning. The system’s ability to adapt to unpredictable and unknown situations without the input of a human operator leads to an autonomous intelligent system.

Using Russell’s ideas, a LAWS would be more like a “killer robot”. A system that, given any task, would determine the best course of action to complete the task, learning and evolving over time. His definition of a LAWS is what many people, in the public and political world, think of when they hear lethal autonomous robot. Influences of popular culture include films such as *The Terminator* and *I, Robot* and books including *When HARLIE Was One* and even *Frankenstein*. The fear of uncontrollable “killer robots” that would make decisions about when and whom to kill without any input from a human and without a countermeasure for humans to intervene has garnered support from groups advocating for a ban on all LAWS.

⁹ The Aegis System is “a centralized, automated, command-and-control (C2) and weapons control system that was designed as a total weapon system, from detection to kill.”

International Humanitarian Law

The misunderstanding and mischaracterization of AI in LAWS creates a situation where governments and their militaries have the potential to violate IHL. While these laws outline the intent and impact of actions in wartime, they do not necessarily outline who or what must take the responsibilities for such actions. Therefore, while using LAWS can be within IHL, steps and precautions are necessary to make sure the system is operating within IHL. The Geneva Convention, Protocol I, Article 36 which basically says that lawful weapons are weapons that can be “targetable”¹⁰ and can “discriminate”¹¹. It concludes that,

- The High Contracting Parties are obliged to determine the legality or illegality of the use of any new weapon introduced into their armed forces.
- This obligation only concerns the normal use of the weapon as seen at the time of the evaluation, whether it were to be used in some or all circumstances (and not possibly misused).
- If a weapon is found to be illegal by a State, this does not by itself create a mandatory rule of international law vis-à-vis third parties, even for the State first mentioned, nor is there an obligation for this State to make its findings public. Consequently the High Contracting Parties are not bound to reveal anything regarding new weapons which are being developed or manufactured.
- However, Article 36 does imply the obligation to establish internal procedures with a view to elucidating the problem of illegality and therefore the other Contracting Parties can ask for information on this point.

For example, with regards to the principle of proportionality, if a LAWS is able to follow the guideline – “Launching 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, is prohibited” [11] – then using LAWS is legal.

Determining where the ultimate responsibility of an attack lies is dependent upon the understanding and trust in a system. In a system with “weak AI” the responsibility might fall on the computer programmers and the operators. Since every decision made is hard-coded, the programmers are responsible for understanding how much trust can be placed on the sub-systems of the weapon. They should then program decisions that will uphold IHL with an added margin of error. On the other hand, a “strong AI” system would make the decision whether an attack meets the requirements of IHL. The system would learn what attack options met all strategic goals while upholding legal and ethical boundaries. Semi-autonomous systems that require humans to analyze information and make decisions places the responsibility to uphold IHL back on the operator, while fully-autonomous systems potentially have no human oversight – no human in the loop – that could even take responsibility if there were to be a mistake.

¹⁰ The ability of the system to be selected as the aim of an attack. [12]

¹¹ Recognize a distinction; differentiate [12]

Additionally, many experts and scholars support banning LAWS completely. However, the lack of a concrete definition on what constitutes a LAWS in the eyes of the international community causes a problem for creating a ban. Primarily, what technologies should the international community ban?



Figure 7. Support for Banning LAWS [6]

Fully-autonomous, intelligent weapons systems do not exist yet. The “killer robot” or “Terminator” is not yet a reality. Predicting exactly what the effects of these technologies will be on weapons systems is difficult. Some experts argue that the potential for LAWS to improve the accuracy and reduce the civilian casualties in war means the international community should not ban them. They might have some positive functions in the future, with proper regulation and international discussion. If that is the case, some scholars argue that banning the technology and its development would be irresponsible. Implementing the technology that will cause the least amount of damage in international conflict would be our responsibility. Does restricting them now possibly limit to what extent we can use AI and autonomy to improve conflicts for both warfighter and civilian? Supporters of a ban want to ban fully-autonomous systems, but what about current semi-autonomous systems such as missile defense systems, many of which engage a target without the direct input of the human operator. Banning a technology so widely used by militaries across the world may prove difficult, but not an impossible task, as exemplified by the ban on anti-personnel landmines as well as other weapons.

Policy Recommendations

I. Classify the Level of Autonomy and Intelligence of Current Weapons Systems

Defining the technical characteristics of autonomy and AI challenges the conversation, and many states agree that a universal understanding will help progress public policy. The first recommendation is to use existing weapons systems to build understanding of how autonomy and AI impact weapons systems and their use. While autonomy and intelligence impact LAWS in separate and distinct ways, and are not analogous, they are both significant factors in the regulation and understanding of LAWS. Parties can consider both AI and autonomy on a scale that represents the level of each that a system has. These scales are not the same scale, signifying that the two technologies are separate and provide unique functions to the system. The scales also cannot be parallel, because this implies that as the level of autonomy increases, so does the level of intelligence. This is sometimes true but overall is a viewpoint that is detrimental to the advancement of the discussion. Instead, this paper recommends that policymakers consider the two technologies on a perpendicular scale.

Figure 8 shows several weapons systems, listed and briefly described below, on a scale of autonomy and intelligence. The weapons systems presented have a diverse set of technical characteristics. Adopted from various sources, the definitions used to place the weapons systems on the scale are below.



(1) Hand Gun – A short-barreled firearm designed to be fired with only one hand.



(2) Anti-personnel landmines – An explosive mine laid on or just under the surface of the ground, triggered by weight or pressure.



(3) Machine Gun – An automatic gun that fires bullets in rapid succession for as long as the trigger is pressed.



(4) Sentry Gun (ex. Phalanx CIWS) – a weapon that is automatically aimed and fired at targets that are detected by sensors.



(5) Armed UAVs (ex. MQ-1 Predator, MQ-9 Reaper) – Remotely piloted aircraft that serve in mostly ISR missions, armed with Hellfire missiles.



(6) Anti-Missile Defense System (ex. C-RAM) – A missile with a high, arching trajectory, which is initially powered and guided but falls under gravity on to its target.



(7) Loitering Munitions (ex. Israeli Harpy) – A weapon system category in which the munition loiters around the target area for some time, searches for targets, and attacks once a target is located.



(8) Terminal High Altitude Area Defense (THAAD) – Defends against short and medium range ballistic missiles through interception of missiles inside and outside of the atmosphere.

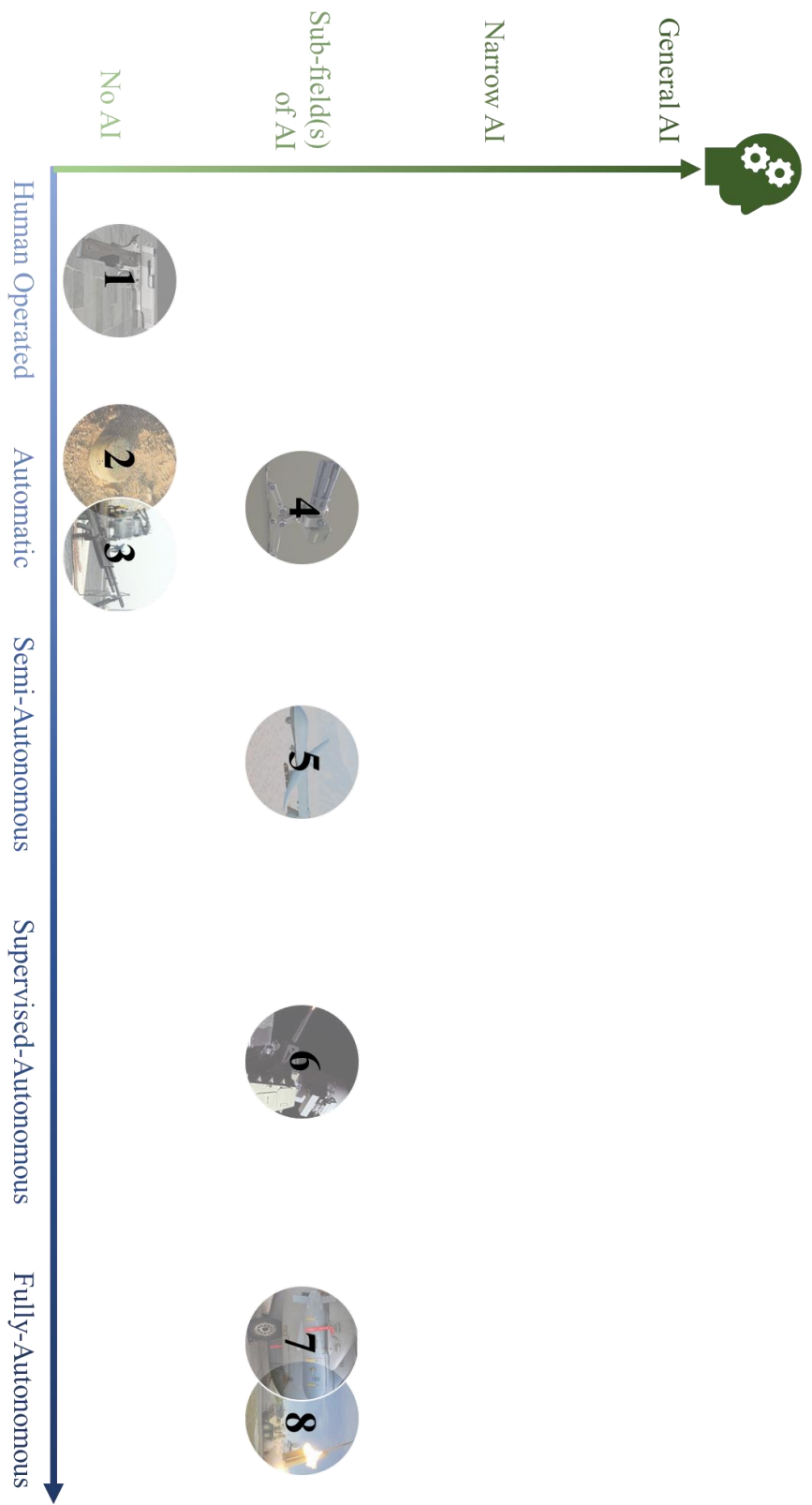


Figure 8. Diagram of Weapon Systems Currently in Use

Term Definitions:

- Human-operated – operated under direct human control [1]
- Automatic – a mechanical process that occurs by itself with little or no direct human control [2]
- Semi-Autonomous – a weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. [3]
- Supervised-Autonomous – An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur. [3]
- Fully-Autonomous – a weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation. [3]
- Narrow AI –an artificial intelligence system that addresses specific application areas such as playing strategic games, language translation, self-driving vehicles, and image recognition. [4]
- General AI –a notional future artificial intelligence system that exhibits apparently intelligent behavior at least as advanced as a person across the range of cognitive, emotional, and social behaviors. [4]
- Machine Learning – the capacity of a computer to learn from experience, i.e. to modify its processing on the basis of newly acquired information. [2]

While assessing the autonomy and intelligence of currently deployed weapons systems may assist policymakers in understanding the real-life implications of adopting systems with these technologies, categorizing weapons poses a risk. Defining LAWS based on level of autonomy does not address meaningful human control¹² of weapons. Many states participating in the GGE agree, the definition should be based upon the human interaction with the weapon and the ultimate implications of its use.

¹² meaningful human control has three essential components: Human operators are making informed, conscious decisions about the use of weapons; human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action; and the weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.” [54]

II. Limit the Use of the Term Artificial Intelligence

“AI currently encompasses a huge variety of subfields, from general-purpose areas such as perception and logical reasoning, to specific tasks such as playing chess, proving mathematical theorems, writing poetry, and diagnosing diseases.” [5]

This recommendation is to separate the AI subfields from the term AI. The general public, experts, and policymakers widely use the term AI to describe each subfield, leading to an international discussion on LAWS filled with confusion and misunderstanding about the technical scope of AI in LAWS. As subfields continue to progress and new subfields develop, the conversation should distinguish each technology. Particularly in the political and regulatory world, referencing technology in a more direct and narrow manner will help to promote concise public policy. Some subfields to reference separate of AI and their general definition are:

- Machine Learning – The capacity of a computer to learn from experience, i.e. to modify its processing on the basis of newly acquired information. [2]
- Computer Vision – concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding. [5]
- Perception – The neurophysiological processes, including memory, by which an organism becomes aware of and interprets external stimuli. [2]
- Natural Language Processing/Speech Processing – The application of computational techniques to the analysis and synthesis of natural language and speech. [2]
- Data Mining – The practice of examining large pre-existing databases in order to generate new information. [2]
- Information retrieval – the recovery of information stored in a computer system. [2]
- Knowledge Representation – the representation of a particular area of knowledge, esp. in a computer system; (more commonly) the branch of artificial intelligence concerned with this. [2]
- Expert Systems – a program or group of programs designed to store and apply the knowledge of experts in a given field, so that others can use it for deciding, evaluating, or inferring in that field. [2]

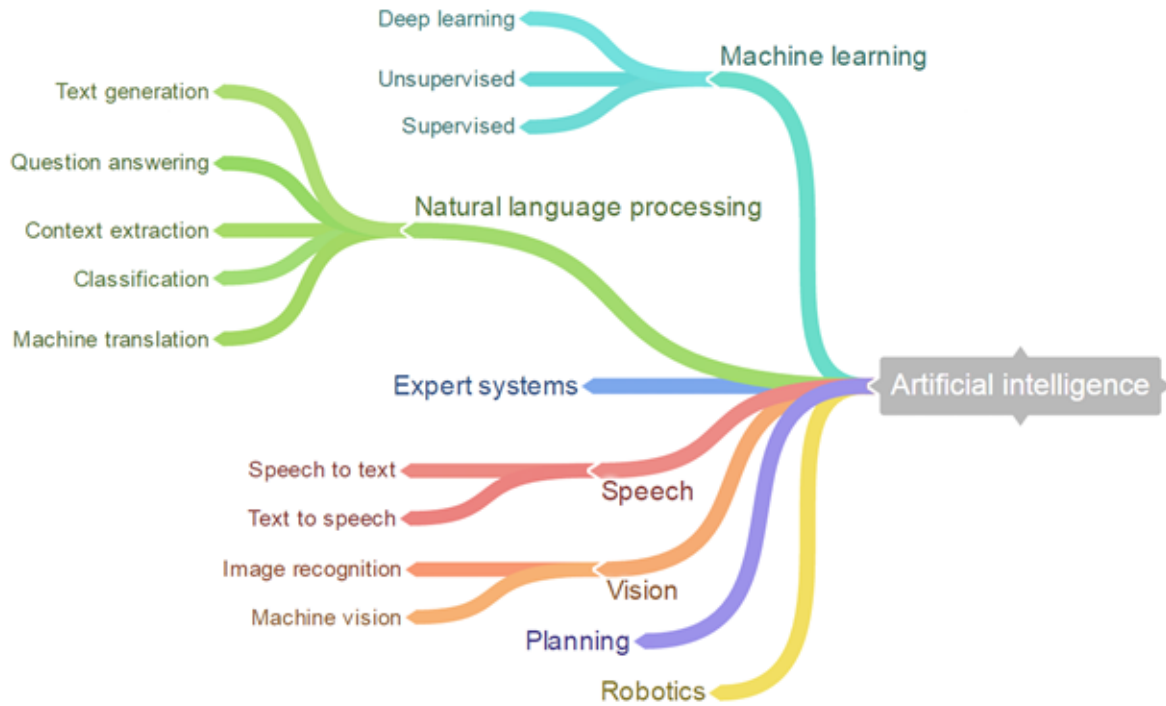


Figure 9. Subfields of AI [16]

The use of the term AI therefore would represent systems made up of all of the above sub-systems. The GGE should consider artificially intelligent systems to be systems that take in percepts, make decisions based on those percepts and previous knowledge, and act accordingly to accomplish a certain goal. The discussion should still differentiate between narrow AI and general AI, distinguishing the ability of a system to perform a single or narrow set of tasks, as opposed to fully thinking and behaving like a human, the premise of general AI.

The potential difficulty of separating out the subsystems of AI is further complication and confusion of the technology and terminology. As more subsystems develop, the continued disagreement on the definitions of the terms as well as boundaries of the technology will remain. The abundance of subsystems in AI means that referring to each individually may be overwhelming for policymakers, creating more difficulty with enacting public policy that promotes the advancement of technologies that impact AI.

III. Differentiate Autonomous Systems from Autonomous Intelligent Agents

A disagreement exists on the definition of autonomy and autonomous systems amongst the contributors to the international discussion on LAWS. One of the most significant separations is what an autonomous system is. Two thought processes are present in the discussion:

1. Autonomous Systems: some members of the discussion do not define an autonomous system based on its intelligence. Instead, these participants characterize an autonomous system based on the human-machine interaction and the amount of human input that the system allows or requires. An autonomous system has no explicit implication of machine learning or percept analyses.
2. Autonomous Intelligent Agents¹³: this definition of an autonomous system implies that the system is intelligent, meaning that it has general AI, and can operate without human input on a broad range of activities, making its own decisions and learning as time progresses. Members of the AI community, including Russell, accept the idea that autonomous systems are intelligent. Russell states, “If the agent’s actions are based completely on built-in knowledge, such that it need pay no attention to its percepts, then we say that the agent lacks autonomy.” Therefore, the ability to make decisions based on knowledge the system acquires defines a LAWS.

These two thought processes conflict particularly in the eye of the general public. A LAWS that is an autonomous intelligent agent might look more like a “killer robot” that has the potential to make unpredictable and incomprehensible decisions in lethal situations. On the other hand, an autonomous system that is not necessarily intelligent could look more like a loitering munition. The system may have percepts and use information from those percepts to search, identify, and engage a target. Additionally, without distinguishing between fully-autonomous, semi-autonomous, and supervised-autonomous, this idea of an autonomous system might even include systems that search for and identify a target, then require human input before engaging with the target.

Even though many discontinuities exist in the understanding and defining of terms related to LAWS, this opposing approach to the capabilities of autonomous systems creates a big hurdle for parties on either side to overcome. Those involved in the discussion on LAWS must acknowledge the existence of the two separate and valid thoughts on the fundamental definition of an autonomous system. One idea is not better than the other, or more correct, but they are unambiguously different. The ideas about public policy regarding these two thought processes would likely be completely different.

¹³ “An agent is something that perceives and acts in an environment. We split an agent into an architecture and an agent program.” [10]

IV. Distinguish Between Semi- and Fully-Autonomous Systems

Several states involved in the international discussion on LAWS provide a distinction between semi-autonomous and fully-autonomous weapons systems. Cuba, France, Germany and the U.S. each distinguish these two systems in their statements to the GGE. Other states have not distinguished between the two. Instead, their definitions group semi- and fully-autonomous systems together, or only cover fully-autonomous systems. The distinction in the technologies is most commonly based on the level of human input in the system, especially with critical decision-making on lethal actions.

Many parties agree that defining what a fully-autonomous weapon system is, might be counterproductive to the mission of the international discussion¹⁴. While this is a valid concern with defining an autonomous weapon, a distinct set of legal and ethical concerns exist with semi-autonomous and fully-autonomous weapons systems. The GGE outlined four approaches to characterization, each of which would differ when discussing semi-autonomous and fully-autonomous weapons.

1. Separative approach: An approach whereby characteristics and concepts not relevant to the objectives and purposes of the CCW are set aside (“via negative”), while gathering the characteristics and concepts that are definitely relevant to the objectives and purposes of the CCW (“via positive”).
2. Cumulative approach: An approach whereby categories of characteristics are added to a master list and then concepts and characteristics therein are evaluated against certain technical, legal, humanitarian or political-security criteria to assess their relevance to the objectives and purposes of the CCW. Such categories could include physical performance, targeting performance and other technical characteristics. They could also include characteristics that are related to the human-machine interface or secondary characteristics such as reliability, and predictability.
3. Accountability approach: considers a set of characteristics related to the functions and type of decisions handed over to machines, and which avoids using levels of autonomy and other technical characteristics or categories related to the loss of human control. This approach would depend rather on the context and scenario in which the systems under consideration would be used and would involve a combination of technical and human-interface evaluations centered on accountability of States and humans.
4. Purpose oriented and effect-based approach: This approach focuses on desirable and undesirable consequences of possible lethal weapons systems based on emerging autonomous intelligent systems and technologies

Acknowledging the differences in the technologies implies the need for different public policies and regulations for each. Therefore, this paper recommends that the international discussion on LAWS create separate opportunities to address the emerging technologies contributing to LAWS as well as the humanitarian and security challenges posed by these distinct technologies.

¹⁴ The purpose of the UN’s GGE on LAWS is to discuss the ethical and moral implications of increasingly autonomous weapon systems and their use in international conflict.

References

- [1] International Committee of the Red Cross, "Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects," International Committee of the Red Cross, Geneva, 2014.
- [2] OED Online, Oxford University Press, 2018.
- [3] *DoD Directive 3000.09 Autonomy in Weapon Systems*, U.S. Department of Defense, 2012.
- [4] U. Senate, *S.2217 - FUTURE of Artificial Intelligence Act of 2017*, Congress.gov, 2017.
- [5] The British Machine Vision Association and Society for Pattern Recognition, "What is computer vision?," [Online]. Available: <http://www.bmva.org/visionoverview>. [Accessed July 2018].
- [6] V. Boulanin and M. Verbruggen, "Mapping the Development of Autonomy in Weapon Systems," Stockholm International Peace Research Institute, Solna, 2017.
- [7] "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," in *International Joint Conference on Artificial Intelligence*, Stockholm, 2015.
- [8] M. A. Lorell, J. F. Lowell, M. Kennedy and H. P. Levaux, "Commercial Insertion and the Question of Weapon System Performance," in *Cheaper, Faster, Better? Commercial Approaches to Weapons Acquisition*, Santa Monica, RAND Corporation, 2000, pp. 57-69.
- [9] A. Ilachinski, "Artificial Intelligence & Autonomy Opportunities and Challenges," Center for Naval Analyses, Arlington, 2017.
- [10] The United Nations Institute for Disarmament Research, "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches," UNIDIR, Geneva, 2017.
- [11] R. Parasuraman, T. B. Sheridan and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286-297, 2000.
- [12] A. Singh Gill, "Chair's summary of the discussion on Agenda item 6 (a) 9 and 10 April 2018, Agenda item 6 (b) 11 April 2018 and 12 April 2018, Agenda item 6 (c) 12 April 2018, Agenda item 6 (d) 13 April 2018," in *Group of Governmental Experts on Lethal Autonomous Weapons Systems*, Geneva, 2018.
- [13] United States of America, "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate

Effects," in *Group of Governmental Experts on Lethal Autonomous Weapons Systems*, Geneva, 2018.

- [14] "Characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention," in *Group of Governmental Experts on Lethal Autonomous Weapons Systems*, Geneva, 2018.
- [15] United Nations Institute for Disarmament Research, "The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence," UNIDIR, Geneva, 2018.
- [16] Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), "Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)," Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 2017.
- [17] M. Biontino, "Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)," Meeting of Experts on LAWS, Geneva, 2016.
- [18] J.-H. Simon-Michel, "Report of the 2014 informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)," Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects , Geneva, 2014.
- [19] United Nations Institute for Disarmament Research, "Framing Discussions on the Weaponization of Increasingly Autonomous Technologies," UNIDIR, Geneva, 2014.
- [20] International Panel on the Regulation of Autonomous Weapons, "Focus on the Human-Machine Relation in LAWS," International Panel on the Regulation of Autonomous Weapons, Berlin, 2018.
- [21] S. Legg and M. Hutter, "A Collection of Definitions of Intelligence," IDSIA, Manno-Lugano, 2007.
- [22] S. Legg and M. Hutter, "A Formal Measure of Machine Intelligence," IDSIA, Manno-Lugano, 2006.
- [23] I. R. McKay, *U.S. Statement on Possible Options for Addressing the Humanitarian and International Security Challenges Posed by Emerging Technologies in the Area of LAWS*, Geneva: Convention on Certain Conventional Weapons (CCW), Group of Government Experts on Lethal Autonomous Weapons (LAWS), 2018.
- [24] U.S. Delegation to the GGE on LAWS, "CCW GGE: U.S. Slide Presentation – Counter Rocket, Artillery and Mortar System (C-RAM)," 11 April 2018. [Online]. Available:

<https://geneva.usmission.gov/2018/04/13/u-s-slide-presentation-at-ccw-gge-counter-rocket-artillery-and-mortar-system-c-ram/>. [Accessed 13 July 2018].

- [25] "Treaties, States Parties and Commentaries," International Committee of the Red Cross, [Online]. Available: <https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/vwTreatiesByTopics.xsp>. [Accessed 10 July 2018].
- [26] Center for Strategic and International Studies, "Tomahawk," Center for Strategic and International Studies, [Online]. Available: <https://missilethreat.csis.org/missile/tomahawk/>. [Accessed July 2018].
- [27] B. W. Bennett, "Why THAAD Is Needed in Korea," The RAND Blog, 7 August 2017. [Online]. Available: <https://www.rand.org/blog/2017/08/why-thaad-is-needed-in-korea.html>. [Accessed July 2018].
- [28] Campaign to Stop Killer Robots, Campaign to Stop Killer Robots, [Online]. Available: <https://www.stopkillerrobots.org/>. [Accessed July 2018].
- [29] Faculty Development and Instructional Design Center, "Howard Gardner's Theory of Multiple Intelligences," Northern Illinois University.
- [30] M.-H. Wang, *Artificial Intelligence and Subfields*, Santa Clara University, 2017.
- [31] *Hearing Before the Subcommittee on Space, Science, and Competitiveness of the Committee on Commerce, Science, and Transportation*, United States Senate, 2016.
- [32] J. P. Gunderson and L. F. Gunderson, *Intelligence - Autonomy - Capability*, Gamma Two, Inc..
- [33] K. Fox, *Autonomy vs. Automation*, Raytheon.
- [34] T. McFarland, "Factors shaping the legal implications of increasingly autonomous military systems," *International Review of the Red Cross*, vol. 97, no. 900, pp. 1313-1339, 2015.
- [35] "H.R.2810 - National Defense Authorization Act for Fiscal Year 2018," 12 December 2017. [Online]. Available: <https://www.congress.gov/bill/115th-congress/house-bill/2810>.
- [36] 115th Congress (2017-2018), *H.R.5356 - National Security Commission Artificial Intelligence Act of 2018*, Congress.gov, 2018.
- [37] United Nations Institute for Disarmament Research, *The Weaponization of Increasingly Autonomous Technologies (Phase III)*, Geneva: United Nations Institute for Disarmament Research, 2018.
- [38] United Nations Institute for Disarmament Research, "The Weaponization of Increasingly Autonomous Technologies (Phase III)," 2018. [Online]. Available:

<http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii>. [Accessed July 2018].

- [39] 115th , *H.R. 5515: John S. McCain National Defense Authorization Act for Fiscal Year 2019*, govtrack.us, 2018.
- [40] The White House Office of Science and Technology Policy, "Summary of the 2018 White House Summit on Artificial Intelligence for American Industry," Executive Office of the President of the United States, 2018.
- [41] G. A. Dorais, R. P. Bonasso, D. Kortenkamp, B. Pell and D. Schreckenghost, *Adjustable Autonomy for Human-Centered Autonomous Systems*, NASA.
- [42] G. Galdorisi, R. Laird and R. Volner, *Taking the Next Step: From "Unmanned" to True Autonomy*, San Diego: Space and Naval Warfare Systems Center Pacific, 2012.
- [43] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, New Jersey: Prentice-Hall, Inc., 1995.
- [44] A. M. Turing, *Computing Machinery and Intelligence*, *Mind* 49, 1950.
- [45] L. Kostopoulos, "Drivers for the Deployment of Lethal Autonomous Weapons Systems," *Medium*, 22 December 2017. [Online]. Available: <https://medium.com/@lkcyber/drivers-for-the-deployment-of-lethal-autonomous-weapons-systems-ae1dd6278a35>. [Accessed June 2018].
- [46] The United Nations Office at Geneva, "2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)," [Online]. Available: [https://www.unog.ch/__80256ee600585943.nsf/\(httpPages\)/7c335e71dfcb29d1c1258243003e8724?OpenDocument&ExpandSection=6%2C7%2C1%2C8%2C3#_Section6](https://www.unog.ch/__80256ee600585943.nsf/(httpPages)/7c335e71dfcb29d1c1258243003e8724?OpenDocument&ExpandSection=6%2C7%2C1%2C8%2C3#_Section6). [Accessed June 2018].
- [47] International Committee of the Red Cross, "Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects," ICRC, Geneva, 2014.