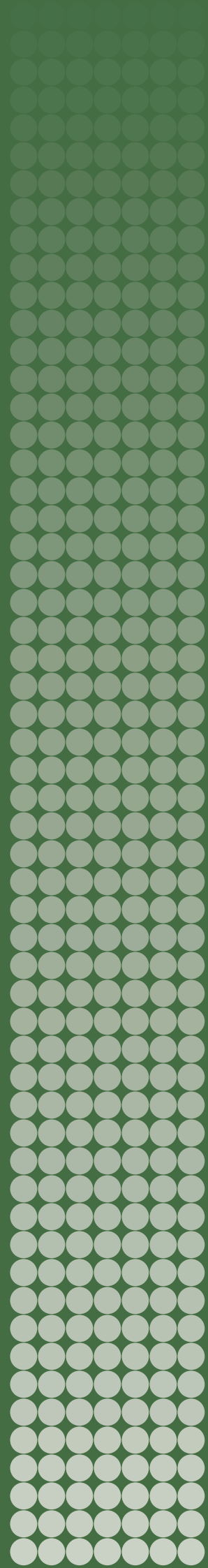# MAPPING THE DEVELOPMENT OF AUTONOMY IN WEAPON SYSTEMS

VINCENT BOULANIN AND MAAIKE VERBRUGGEN

MAPPING THE DEVELOPMENT OF AUTONOMY IN WEAPON SYSTEMS

VINCENT BOULANIN AND MAAIKE VERBRUGGEN

November 2017

sipri

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

Contents

Acknowledgements

About the authors

Dr Vincent Boulanin (France/Sweden) is a Researcher within Armament and Disarmament at SIPRI and the principal investigator of the SIPRI Mapping Study on the Development of Autonomy in Weapon Systems. He works on issues related to the production, use and control of emerging military and security technologies. Before joining SIPRI in 2014, he completed a PhD in Political Science at École des Hautes Études en Sciences Sociales (the School for Advanced Studies in the Social Sciences) in Paris.

Maaike Verbruggen (Netherlands) joined SIPRI in April 2016 to work as a research assistant with the SIPRI Mapping Study on the Development of Autonomy in Weapon Systems. She left SIPRI in November 2017 to work as a PhD researcher at the Vrije Universiteit in Brussels. Before joining SIPRI, she completed a Master of Philosophy degree in Peace and Conflict Studies at Oslo University.

Executive summary

This report presents the conclusions of a one-year mapping study on the development of autonomy in weapon systems. It is intended to provide diplomats and members of civil society interested in the issue of lethal autonomous weapon systems (LAWS) with a better understanding of (*a*) the technological foundations of autonomy; (*b*) the state of autonomy in existing weapon systems; (*c*) the drivers of, and obstacles to, further increasing autonomy in weapon systems; and (*d*) the innovation ecosystems behind the advance of autonomy in weapon systems.

I. Key findings

What are the technological foundations of autonomy?

Chapter 2 explores the technological foundations of autonomy. The main findings are as follows.

1. Autonomy has many definitions and interpretations, but is generally understood to be the ability of a machine to perform an intended task without human intervention using interaction of its sensors and computer programming with the environment.
2. Autonomy relies on a diverse range of technology but primarily software. The feasibility of autonomy depends on (*a*) the ability of software developers to formulate an intended task in terms of a mathematical problem and a solution; and (*b*) the possibility of mapping or modelling the operating environment in advance.
3. Autonomy can be created or improved by machine learning. The use of machine learning in weapon systems is still experimental, as it continues to pose fundamental problems regarding predictability.

What is the state of autonomy in weapon systems?

Chapter 3 explores the state of autonomy in deployed weapon systems and weapon systems under development. The main findings are as follows.

1. Autonomy is already used to support various capabilities in weapon systems, including mobility, targeting, intelligence, interoperability and health management.
2. Automated target recognition (ATR) systems, the technology that enables weapon systems to acquire targets autonomously, has existed since the 1970s. ATR systems still have limited perceptual and decision-making intelligence. Their performance rapidly deteriorates as operating environments become more cluttered and weather conditions deteriorate.
3. Existing weapon systems that can acquire and engage targets autonomously are mostly defensive systems. These are operated under human supervision and are intended to fire autonomously only in situations where the time of engagement is deemed too short for humans to be able to respond.
4. Loitering weapons are the only 'offensive' type of weapon system that is known to be capable of acquiring and engaging targets autonomously. The loitering time and geographical areas of deployment, as well as the category of targets they can attack, are determined in advance by humans.

What are the drivers of, and obstacles to, the development of autonomy in weapon systems?

Chapter 4 explores the key drivers and obstacles to an increase of autonomy in weapon systems. The main drivers identified by the report are as follows.

1. *Strategic*. The United States recently cited autonomy as a cornerstone of its strategic capability calculations and military modernization plans. This seems to have triggered reactions from other major military powers, notably Russia and China.

2. *Operational*. Military planners believe that autonomy enables weapon systems to achieve greater speed, accuracy, persistence, reach and coordination on the battlefield.

3. *Economic*. Autonomy is believed to provide opportunities for reducing the operating costs of weapon systems, specifically through a more efficient use of manpower.

The main obstacles identified by the report are as follows.

1. *Technological*. Autonomous systems need to be more adaptive to operate safely and reliably in complex, dynamic and adversarial environments; new validation and verification procedures must be developed for systems that are adaptive or capable of learning.

2. *Institutional resistance*. Military personnel often lack trust in the safety and reliability of autonomous systems; some military professionals see the development of certain autonomous capabilities as a direct threat to their professional ethos or incompatible with the operational paradigms they are used to.

3. *Legal*. International law includes a number of obligations that restrict the use of autonomous targeting capabilities. It also requires military command to maintain, in most circumstances, some form of human control or oversight over the weapon system's behaviour.

4. *Normative*. There are increasing normative pressures from civil society against the use of autonomy for targeting decisions, which makes the development of autonomous weapon systems a potentially politically sensitive issue for militaries and governments.

5. *Economic*. There are limits to what can be afforded by national armed forces, and the defence acquisition systems in most arms-producing countries remain ill-suited to the development of autonomy.

Where are the relevant innovations taking place?

Chapter 5 explores the innovation ecosystems that are driving the advance of autonomy. The main findings are as follows.

1. At the basic science and technology level, advances in machine autonomy derive primarily from research efforts in three disciplines: artificial intelligence (AI), robotics and control theory.

2. The USA is the country that has demonstrated the most visible, articulated and perhaps successful military research and development (R&D) efforts on autonomy. China and the majority of the nine other largest arms-producing countries have identified AI and robotics as important R&D areas. Several of these countries are tentatively following in the USA's footsteps and looking to conduct R&D projects focused on autonomy.

3. The civilian industry leads innovation in autonomous technologies. The most influential players are major information technology companies such as Alphabet

(Google), Amazon and Baidu, and large automotive manufacturers (e.g. Toyota) that have moved into the self-driving car business.

4. Traditional arms producers are certainly involved in the development of autonomous technologies but the amount of resources that these companies (can) allocate to R&D is far less than that mobilized by large commercial entities in the civilian sector. However, the role of defence companies remains crucial, because commercial autonomous technologies can rarely be adopted by the military without modifications and companies in the civilian sector often have little interest in pursuing military contracts.

II. Recommendations for future discussions on LAWS within the framework of the Convention on Certain Conventional Weapons (CCW)

The report concludes with eight recommendations that aim to help the newly formed Group of Governmental Experts on LAWS at the United Nations to find a constructive basis for discussions and potentially achieve tangible progress on some of the key aspects under debate.

1. Discuss the development of 'autonomy in weapon systems' rather than autonomous weapons or LAWS as a general category.

2. Shift the focus away from 'full' autonomy and explore instead how autonomy transforms human control.

3. Open the scope of investigation beyond the issue of targeting to take into consideration the use of autonomy for collaborative operations (e.g. swarming) and intelligence processing.

4. Demystify the current advances and possible implications of machine learning on the control of autonomy.

5. Use case studies to reconnect the discussion on legality, ethics and meaningful human control with the reality of weapon systems development and weapon use.

6. Facilitate an exchange of experience with the civilian sector, especially the aerospace, automotive and civilian robotics industries, on definitions of autonomy, human control, and validation and verification of autonomous systems.

7. Investigate options to ensure that future efforts to monitor and potentially control the development of lethal applications of autonomy will not inhibit civilian innovation.

8. Investigate the options for preventing the risk of weaponization of civilian technologies by non-state actors.

Key words: artificial intelligence, autonomy, Convention on Certain Conventional Weapons, existing capabilities, human control, innovation, lethal autonomous weapon systems, machine learning, mapping study, research and development, robotics, state of the art, weapon systems.

Abbreviations

3-D	Three-dimensional
A2/AD	Anti-access/area-denial
AGI	Artificial general intelligence
AI	Artificial intelligence
APS	Active protection systems
ART2	Autonomous reliable teammate technology
ATGM	Anti-tank guided missile
ATR	Automatic or automated target recognition
BLADE	Battlefield Loitering Artillery Direct Effect
CARACaS	Control Architecture for Robotic Agent Command and Sensing
CCW	1980 United Nations Convention on Certain Conventional Weapons
CIMS	Counter IED and Mine Suite
CIWS	Close-in weapon system
CODE	Collaborative Operations in Denied Environment
CwC	Communicating with Computers
DARPA	Defense Advanced Research Projects Agency
DMZ	Demilitarized Zone
DOD	Department of Defense
DSTL	Defence Science and Technology Laboratory
EA Focus	Enhanced Awareness and Forward Operating Capability
EDRP	European Defence Research Programme
EU	European Union
FLIR	Forward-looking infrared
FP7	Seventh Framework Programme
FPI	Russian Foundation for Advanced Studies
GGE	Group of Governmental Experts
GPS	Global Positioning System
GSM	Global System for Mobile
H2020	Horizon 2020 Programme
HRI	Human–robot interaction
HRW	Human Rights Watch
ICRC	International Committee of the Red Cross
ICT	Information and communications technology
IDF	Israel Defense Forces
IEDs	Improvised explosive devices
IFF	Identification, friend or foe
IHL	International humanitarian law
IMU	Inertial measurement units
IR	Infrared
ISO	International Organization for Standardization
ISR	Intelligence, surveillance and reconnaissance
LADAR	Light detection and ranging
LAWS	Lethal autonomous weapon systems
LOCAAS	Low Cost Autonomous Attack System
LRASM	Long-Range Anti-Ship Missile
MASI	Microsoft Academic Search Index
MDARS	Mobile Detection Assessment and Response System
MIT	Massachusetts Institute of Technology
NATO	North Atlantic Treaty Association

NGO	Non-governmental organization
NLOS-LS	Non-Line-of-Sight Launch System
NSM/JSM	Naval Strike Missile/Joint Strike Missile
OODA	Observe, orient, decide, act
PLA	People's Liberation Army of China
PPP	Public–private partnership
R&D	Research and development
RFID	Radio-frequency identification
ROS	Robot Operative System
RPG	Rocket-propelled grenade
SEAD	Suppression of enemy air defences
SRC	Skolkovo Robotics Centre
TRACE	Target Recognition and Adaption in Contested Environments
UAE	United Arab Emirates
UAS	Unmanned aerial system
UAV	Unmanned aerial vehicle
UCAS	Unmanned combat aerial system
UCLASS	Unmanned carrier-launched strike and surveillance aircraft
UGS	Unmanned ground system
UGV	Unmanned ground vehicle
UN	United Nations
USAF	United States Air Force
UV	Ultraviolet
V&V	Validation and verification
ViDAR	Visual Detection and Ranging

1. Introduction

I. Background and objective

Since 2013 the governance of lethal autonomous weapon systems (LAWS) has been discussed internationally under the framework of the 1980 United Nations Convention on Certain Conventional Weapons (CCW), which regulates weapons that may be deemed to have an excessively injurious or indiscriminate effect.[1] After three years of informal expert discussions, states parties to the CCW agreed to formalize their discussion with the creation of a Group of Governmental Experts (GGE). The question of whether states parties to the CCW should take formal action on LAWS is not yet officially on the agenda, but it is bound to be a central point of discussion for the GGE. The Campaign to Stop Killer Robots—a coalition of non-governmental organizations (NGOs)—and 19 states are advocating the adoption of a pre-emptive ban on the development, production and use of LAWS.[2] However, at previous CCW meetings most other states have expressed that they are not yet ready to discuss this possibility as they are still in the process of understanding the full implications of increasing autonomy in weapon systems.

To support states in this process and also contribute to more concrete and structured discussions on LAWS at CCW meetings, the Stockholm International Peace Research Institute (SIPRI) conducted a one-year mapping study on the development of autonomy in military systems in general and weapon systems in particular. The rationale for conducting this study was that an assessment of the current state of development and use of autonomy in weapon systems could provide helpful insights for future CCW discussions on LAWS. Specifically, such an assessment could support delegates to (*a*) improve their understanding of the technological foundations of autonomy and obtain a sense of the speed and trajectory of progress of autonomy in weapon systems; (*b*) find concrete examples that could be used to start delineating the points at which the advance of autonomy in weapons may raise technical, legal, operational and ethical concerns; (*c*) investigate possible parameters for meaningful human control, using lessons learned from how existing weapons with autonomous capabilities are used or misused; and (*d*) identify realistic options for the monitoring and regulation of the development of emerging technologies in the area of LAWS.

II. Approach and methodology

This research report presents the key findings and recommendations of the SIPRI study. It maps the development of autonomy in weapon systems from four different perspectives: technical, operational, political and economic (see figure 1.1). The aim of this approach is to provide CCW delegates and interested members of civil society with a basic but comprehensive understanding of the development of autonomy in weapon systems. The report is structured around the following questions.

1. What are the technological foundations of autonomy?
2. What is the state of autonomy in weapon systems?
3. What are the drivers of, and obstacles to, the advance of autonomy in weapon systems?
4. Where are the relevant innovations taking place?

---

[1] 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 (CCW Convention, or 'Inhumane Weapons' Convention), with Protocols I, II and III, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983.

[2] Campaign to Stop Killer Robots, 'Country views on killer robots', 17 Oct. 2017.
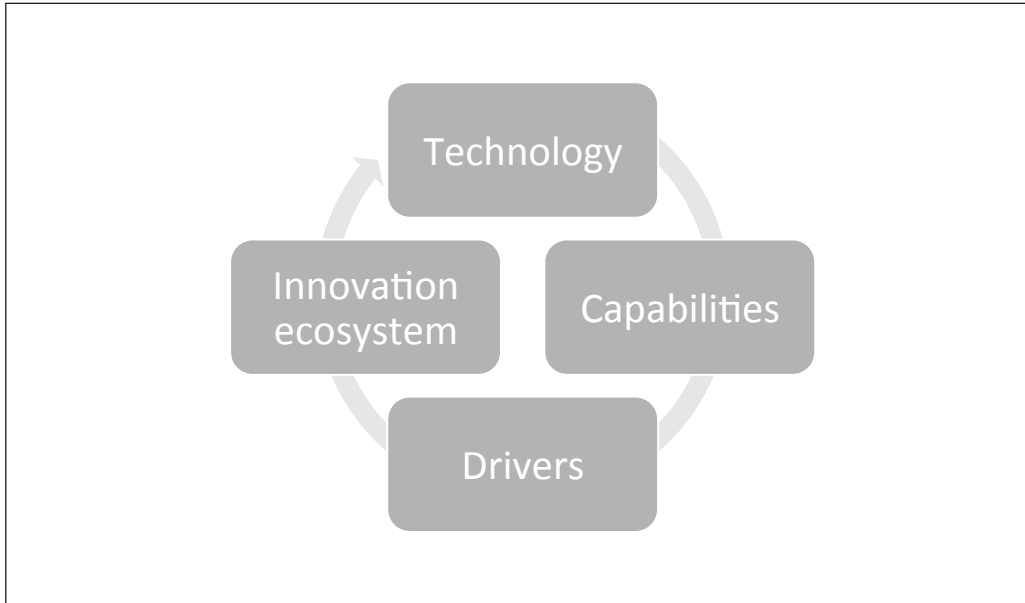
Figure 1.1. A comprehensive approach to mapping the development of autonomy in weapon systems

The analysis presented in this report is based on an extensive review of the literature on civilian and military development of autonomy, robotics, artificial intelligence (AI) and related topics, as well as on a series of in-depth background interviews with relevant experts. It also builds on two extensive and original data collection efforts: (*a*) a mapping of military research and development (R&D) projects that are active, or were recently completed, in the 10 largest arms-producing countries and China; (*b*) a (non-comprehensive) mapping of (unmanned) weapon systems and unarmed military robotic systems that feature autonomous functions that have been deployed or are under development in China, France, Germany, India, Israel, Italy, Japan, South Korea, Russia, Sweden, the United Kingdom and the United States. Information about the types, purposes, users, development status and autonomous capabilities of these systems was recorded and coded into a dataset, which, as of April 2017, consisted of 381 different systems.

III. Outline

Each of the four main chapters of the report (chapters 2 to 5) tackles one of the key questions mentioned above. Chapter 2 maps the conceptual and technical foundations of autonomy. It begins with a review of existing interpretations of the concept of autonomy. It then presents the underlying capabilities and technologies that enable autonomy, and concludes with a discussion of the difficulties involved in engineering autonomous capabilities.

Chapter 3 maps the current state of autonomy in existing weapon systems and military systems more generally. It presents the different functions and capabilities of autonomy in deployed systems and systems under development. It also reviews the characteristics and use of existing weapon systems that are known to have the capability to acquire, or possibly engage, targets autonomously.

Chapter 4 maps the factors driving the adoption of autonomy in weapon systems and examines some of the obstacles to this process. It discusses the extent to which major military powers have articulated a strategic reflection on the development of autonomy in weapon systems and maps out the spectrum of arguments that are commonly mobilized to justify the development of autonomy within weapon systems. It

sets out the variety of technical, political and economic hurdles to further increasing autonomy in weapon systems.

Chapter 5 explores the innovation ecosystem that is driving the advance of autonomy in weapon systems. It maps where relevant innovations are taking place from three different perspectives: a science and technology perspective; a geographical perspective; and an industry sector perspective.

The concluding chapter (chapter 6) summarizes the key findings of the report and returns to the CCW debate on LAWS with a series of practical recommendations that are intended to help the newly formed GGE to constructively advance debate on LAWS.

The report includes an appendix that contains original research material, as well as a glossary that provides working definitions of the key technical terms.

2. What are the technological foundations of autonomy?

I. Introduction

In order to understand the current state and future development of autonomy in weapon systems and military and civilian systems more generally, it is useful to clarify some basic facts about the conceptual and technological foundations of autonomy. This chapter aims to provide non-technical experts with answers to the following basic questions.

1. What is autonomy?
2. How does it work?
3. How is it created?

The chapter consists of four main sections. Section II maps existing interpretations of the concept of autonomy. Section III describes the underlying machinery of autonomy. Section IV discusses how autonomy is created and how difficult it is to engineer autonomous systems or systems with autonomous capabilities. The concluding section (section V) presents some takeaway points for future discussions on LAWS within the framework of the CCW.

II. Searching for a definition: what is autonomy?

Autonomy: a three-dimensional concept

In simple terms 'autonomy' can be defined as the ability of a machine to execute a task, or tasks, without human input, using interactions of computer programming with the environment.[1] An autonomous system is, by extension, usually understood as a system—whether hardware or software—that, once activated, can perform some tasks or functions on its own.

However, autonomy is a relative notion: within and across relevant disciplines, be it engineering, robotics or computer science, experts have a different understanding of when a system or a system's function may or may not be deemed autonomous. According to Paul Scharre, these approaches can be divided into three categories: (*a*) the human–machine command-and-control relationship; (*b*) the sophistication of the machine's decision-making process; and (*c*) the types of decisions or functions being made autonomous.[2]

The human–machine command-and-control relationship

A very common approach for assessing autonomy relates to the extent to which humans are involved in the execution of the task carried out by the machine. With this approach, the systems can be classified into three categories. Systems that require human input at some stage of the task execution can be referred to as 'semi-autonomous' or 'human-in-the-loop'. Systems that can operate independently but are under the oversight of a human who can intervene if something goes wrong (e.g. a malfunction or systems failure) are called 'human-supervised autonomous' or 'human-on-the-loop'. Machines

---

[1] This definition is based on one previously proposed by Andrew Williams. Williams, A., 'Defining autonomy in systems: challenges and solutions', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO: Norfolk, VA, 2015).

[2] Scharre, P., 'The opportunity and challenge of autonomous systems', eds Williams and Scharre (note 1), p. 56.

that operate completely on their own and where humans are not in a position to intervene are usually referred to as 'fully autonomous' or 'human-out-of-the-loop'. The concept of 'sliding autonomy' is sometimes also employed to refer to systems that can go back and forth between semi-autonomy and full autonomy, depending on the complexity of the mission, external operating environments and, most importantly, legal and policy constraints.

The sophistication of the machine's decision-making process

A more technical approach to autonomy relates to the actual ability of a system to exercise control over its own behaviour (self-governance) and deal with uncertainties in its operating environment.[3] From this standpoint, systems are often sorted into three major categories: automatic, automated and autonomous systems. The label 'automatic' is usually reserved for systems that mechanically respond to sensory input and step through predefined procedures, and whose functioning cannot accommodate uncertainties in the operating environment (e.g. robotic arms used in the manufacturing industry). Machines that can cope with variations in their environment and exercise control over their actions can either be described as automated or autonomous. What distinguishes an automated system from an autonomous system is a contentious issue. Some experts see the difference in terms of degree of self-governance, and view autonomous systems merely as more complex and intelligent forms of automated systems.[4] Others see value in making a clear distinction between the two concepts. Andrew Williams, for instance, presents an 'automated system' as a system that 'is programmed to logically follow a predefined set of rules in order to provide an outcome; its output is predictable if the set of rules under which it operates is known'. On the other hand, an 'autonomous system':

is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system can 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.[5]

While the distinction between automatic, automated and autonomous can be conceptually useful, in practice it has proved difficult to measure and therefore determine whether a system falls within one of the three categories. Moreover, the definitions of, and boundaries between, these three categories are contested within and between the expert communities.

The types of decisions or functions being made autonomous

A third dimension to consider focuses on the types of decisions or functions that are made autonomous within a system. This 'functional' approach is not incompatible with the two other approaches; it acknowledges simply that referring to autonomy as a general attribute of systems is imprecise, if not meaningless, as it is the nature of the tasks that are completed autonomously by a machine that primarily matters, not the level of autonomy of the systems as a whole. Autonomy is best understood in relation to

---

[3] According to Thrun, 'Autonomy refers to a robot's ability to accommodate variations in its environment. Different robots exhibit different degrees of autonomy; the degree of autonomy is often measured by relating the degree at which the environment can be varied to the mean time between failures, and other factors indicative of robot performance'. Thrun, S., 'Toward a framework for human–robot interaction', *Human–Computer Interaction*, vol. 19, no. 1–2 (2004), pp. 9–24.

[4] Mindell, D., *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking: New York, 2015), p. 12.

[5] Williams (note 1).

the types of tasks that are executed at the subsystems/function level.[6] Some functions in weapon systems may be made autonomous without presenting significant ethical, legal or strategic risks (e.g. navigation), while others may be a source of greater concern (e.g. targeting).[7]

Autonomy in weapon systems: a situated approach

For its study, SIPRI favoured a 'functional approach' to autonomy. The notable merit of this approach is that it enables a flexible examination of the challenges posed by autonomy in weapon systems. It recognizes that the human–machine command-and-control relationship and the sophistication of a machine's decision-making capability may vary from one function to another. Some functions may require a greater level of self-governance than others, while human control may be exerted on some functions but not others depending on the mission complexity and the external operating environment, as well as regulatory constraints. Also, the extent of a human operator's control or cancel functions may change during the system's mission.

Thus, it could be said that the focus of the research presented in this report is on the development of autonomy *in* weapon systems rather than the development of autonomous systems per se.[8] The ambition is to discuss the development and application of autonomy in a large range of weapon systems in general, not just the few types of weapon systems that may be classified as autonomous according to some existing definitions (current definitions of autonomous weapon systems are presented in box 2.1; the types of weapon systems that are sometimes described as autonomous are presented in chapter 3).

III. Unravelling the machinery

How does autonomy work?

From a basic technical standpoint, 'autonomy is about transforming data from the environment into purposeful plans and actions'.[9] Regardless of the nature of the human–machine relationship, the degree of sophistication of the system or the type of task that is executed, autonomy (in a physical system) is always enabled by the integration of the same three fundamental capabilities: sense, decide and act.[10] These capabilities will be presented in turn.

---

[6] United Nations Institute for Disarmament Research (UNIDIR), *Framing Discussions on the Weaponization of Increasingly Autonomous Technologies*, UNIDIR Resources No. 1 (UNIDIR: Geneva, 2014).

[7] NATO, *Uninhabited Military Vehicles (UMVs): Human Factors Issues in Augmenting the Force*, RTO Technical Report TR-HFM-078 (NATO: 2007); Vignard, K., 'Statement of the UN Institute for Disarmament Research', CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems, Geneva, 12 Apr. 2016; and Gillespie, A., 'Humanity and lethal robots: an engineering perspective', eds G. Verdirame et al., *SNT Really Makes Reality, Technological Innovation, Non-obvious Warfare and the Challenges to International Law* (King's College London: London, forthcoming).

[8] For a number of experts, the term 'autonomous weapon systems' is actually a misnomer. Stensson and Jansson argue, for instance, that the concept of 'autonomy' is maladaptive as it implies, philosophically, qualities that technologies cannot have. For them, machines, by definition, cannot be autonomous. Stensson, P. and Jansson, A., 'Autonomous technology: source of confusion: a model for explanation and prediction of conceptual shifts', *Ergonomics*, vol. 57, no. 3 (2014), pp. 455–70. The concept of autonomous systems has also caused complex and contentious debate regarding the level at which a system may be deemed truly autonomous. In a report dated 2012, the US Department of Defense's Defense Science Board concluded that defining levels of autonomy was a waste of time and money, and tended to reinforce fears of unbounded autonomy. The report noted that discussion of levels of autonomy 'deflects focus from the fact that all autonomous systems are joint human-machine cognitive systems ... all systems are supervised by humans to some degree ... There are no fully autonomous weapons systems as there are no fully autonomous sailors, airmen, or marines'. US Department of Defense (DOD), Defense Science Board, *Task Force Report: Role of Autonomy in DOD Systems* (DOD: Washington, DC, 2012), pp. 23–24. See also Bradshaw, J. et al., 'The seven deadly myths of autonomous systems', *IEEE Intelligent Systems*, vol. 28, no. 3 (2013), pp. 54–61.

[9] Mindell (note 4), p. 12.

[10] US Department of Defense (DOD), Office of Technical Intelligence, Office of the Assistant Secretary of Defense for Research and Engineering, *Technical Assessment: Autonomy* (DOD: Washington, DC, Feb. 2015), p. 2.

Box 2.1. Existing definitions of autonomous weapon systems

Broadly speaking, the definitions of autonomous weapon systems can be classified into three groups.

1. The first category consists of definitions that are articulated around the nature of the human–machine command-and-control relationship. It includes the definition supported by the United States, which describes an 'autonomous weapon system' as 'a weapon that, once activated, can select and engage targets without further intervention by a human operator'.[a] It also encompasses the definition proposed by Human Rights Watch (HRW), the non-governmental organization that coordinates the International Campaign to Stop Killer Robots. HRW makes a distinction between human-in-the-loop weapons, human-on-the-loop weapons and human-out-of-the-loop weapons. Human-out-of-the-loop weapons are robots that are capable of selecting targets and delivering force without any human input or interventions.[b]

2. The second category includes definitions that are based on capability parameters. The United Kingdom's definition, for instance, defines an 'autonomous weapon system' as a system that is 'capable of understanding a higher-level of 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 be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be'.[c]

3. The definitions in the third category are structured along legal lines and lay emphasis on the nature of tasks that the systems perform autonomously. The definition favoured by the International Committee of the Red Cross presents 'autonomous weapons' as an umbrella term that would encompass any type of weapon with 'autonomy in its "critical functions", meaning a weapon that can select (i.e. search for or detect, identify, track) and attack (i.e. intercept, use force against, neutralize, damage or destroy) targets without human intervention'.[d] Switzerland's working definition describes 'autonomous weapon systems' as 'weapons systems that are capable of carrying out tasks governed by IHL [international humanitarian law] in partial or full replacement of a human in the use of force, notably in the targeting cycle', although it explicitly states that this should not necessarily be limited to the targeting cycle.[e]

This classification of definitions is, of course, hardly ideal and does not cover all definitions. The Holy See, for example, uses a mixture of definitions characterizing armed autonomous robots using '(1) the degree and duration of supervision, (2) the predictability of the behaviour of the robot, (3) and the characteristics of the environment in which it operates'.[f]

[a] US Department of Defense, Directive 3000.09 on Autonomy in Weapon Systems, 21 Nov. 2012.
[b] Docherty, B., *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch/International Human Rights Clinic: Washington, DC, 2012).
[c] British Ministry of Defence, Development, Concepts and Doctrine Centre (DCDC), *Joint Doctrine Publication 0.30.2: Unmanned Aircraft Systems* (DCDC: Shrivenham, Aug. 2017), p. 13.
[d] International Committee of the Red Cross (ICRC), 'Autonomous weapon systems: is it morally acceptable for a machine to make life and death decisions?', CCW Meeting of Experts on Lethal Autonomous Weapon Systems, Geneva, 13–17 Apr. 2015.
[e] Government of Switzerland, 'Towards a "compliance-based" approach to LAWS', Informal Working Paper, 30 Mar. 2016, CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems, Geneva, 11–15 Apr. 2016.
[f] Holy See, 'Element supporting the prohibition of LAWS', Working Paper, 7 Apr. 2016, CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems, Geneva, 11–15 Apr. 2016.

Sense

To complete a task autonomously a system needs to be able to perceive the environment in which it operates. For that, it requires sensors to collect data (the 'sense' part of perception) and a computer which uses a dedicated program—a sensing software—that can fuse and interpret the data (the 'think' part of perception).[11] The way sensing software works can vary significantly depending on the type of sensory data and the end use of the processed data. Many types of sensing software, notably computer vision software used for target detection, rely on pattern recognition: the software looks for predefined patterns in the raw data and compares them to example patterns stored in a computer memory, either on-board or off-board the system. It is worth emphasizing that computers identify patterns, such as for image or speech recognition, in a fundamentally different way from the way humans do. They use mathematical methods to

---

[11] Sensors may also be turned inwards to make the system capable of self-assessment, e.g. monitoring power resources or the state of physical components.

find relationships in the sensory data. This means that when computers make errors, they are very different from those that a human would make. Recent studies have shown that state of the art computer vision systems that display human-competitive results on many pattern recognition tasks can easily be fooled. One study illustrated that changing an image originally correctly classified (e.g. a lion) in a way that is imperceptible to the human eye can cause the computer vision software to label the image as something entirely different (e.g. mislabelling a lion as a library).[12] Another study demonstrated that it is easy to produce images that are completely unrecognizable to humans but that computer vision software believes to be a recognizable object with over 99 per cent confidence.[13]

Decide

The data that has been processed by the sensing software serves then as input for the decision-making process, which is assured by the control system. The way the control system determines the course of action towards the task-specific goal can differ greatly from one system to another. Drawing upon Stuart Russell's and Peter Norvig's classification of intelligent agents, two generic categories of control system (which themselves can be further divided into two types) can be identified: (*a*) reactive control systems (simple or model-based); and (*b*) deliberative control systems (goal-based or utility-based).[14] The decision-making processes presented by these categories differ radically from each other.

Reactive control systems can be divided into two subtypes: simple reflex-control systems and model-based reflex-control systems. Simple reflex systems follow a strict sense–act modality. They merely consist of a set of condition–action rules (also known as 'if–then rules') that explicitly prescribe how the system should react to a given sensory input. To take the example of a landmine, these rules would be: *if* the weight exerted on the mine is between X and Y kilogrammes, *then* detonate. These systems succeed only in environments that are fully observable through sensors.

Model-based reflex-control systems are slightly more complex in their design as they include a 'model of the world', meaning a knowledge base that represents, in mathematical terms, how the world works: how it evolves independently of the system and how the system's actions affect it (see figure 2.1). The additional information provided by the model helps to improve performance and reliability as it aids the control system to keep track of its percept history and parts of the environment it cannot observe though its sensors.[15] For instance, for an autonomous vacuum cleaner this information could simply be a map of the surface that has to be vacuumed. Like simple reflex-control systems, model-based reflex-control systems follow a fixed set of rules and their decision making is implemented in some form of direct mapping from situation to action.

---

[12] Szegedy, C. et al., 'Intriguing properties of neural networks', arXiv:1312.6199v4 [cs.CV], 19 Feb. 2014.

[13] Nguyen, A., Yosinski, J. and Clune J., 'Deep neural networks are easily fooled: high confidence predictions for unrecognizable images', Institute of Electrical and Electronics Engineers (IEEE), Computer Vision and Pattern Recognition, 2015.

[14] Russell and Norvig define 'agents' as 'anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators'; an agent can be a human, a robot or software. Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Pearson Education: Harlow, 2014), p. 35, p. 49. Note that other typologies could be used to categorize control systems. Using Albus's and Barbera's classification of planning algorithms, control systems could be sorted between those that use 'case-based planning' and 'search-based planning'. Albus, J. and Barbera, A., '4D/RCS reference model architecture for unmanned ground vehicles', eds R. Madhavan, E. Messina and J. Albus, *Intelligent Vehicles Systems* (Nova Science Publishers: New York, 2006), pp. 11–12.

[15] Russell and Norvig describe reflex agents that include a model of the world as model-based reflex agents. Those that do not have a model are referred to as a 'simple reflex agent'. Russell and Norvig (note 14).
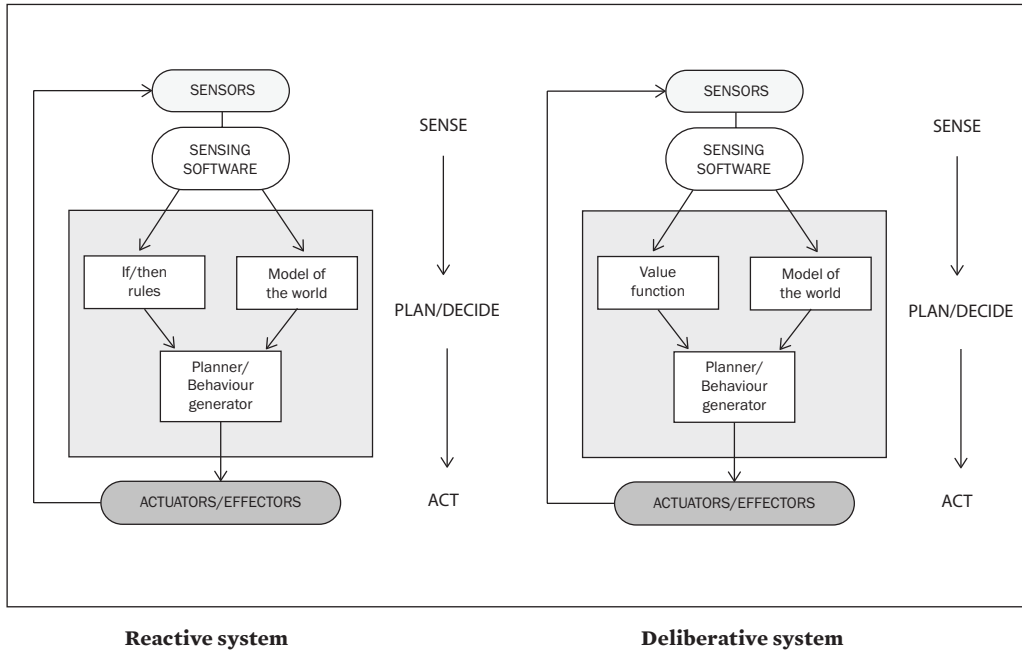
Figure 2.1. Anatomy of autonomy: reactive and deliberative systems

Deliberative control systems can govern their own actions by manipulating data structures, representing what Gerhard Weiss calls their 'beliefs', 'desires' and 'intentions'.[16] They combine a model of the world (belief about how the world works and the reactions to the system's actions), a value function that provides information about the desired goal (desire), and a set of potential rules that help the system to search and plan how to achieve the goal (intention) (see figure 2.1).[17] To make a decision, deliberative control systems weigh the consequences of possible actions and measure whether and to what extent they will serve the achievement of the goal. One concrete example would be the homing function in a beyond-visual-range air-to-air missile (e.g. the Meteor missile developed by the European producer MBDA). The desired goal of the missile is to attack a predetermined target. Combining input from sensors, information from the model of the world and the rules included in its utility function, the missile's control system can find the quickest and most energy-efficient route to approach the target. It can then track the target until it has an opportunity to attack it.

Deliberative control systems feature a level of deliberative intelligence or self-governance that reflex agents do not have. They do not simply go through a series of pre-scripted actions; they can reason about the possible consequences of actions and then act accordingly. Their main advantage is flexibility. They can handle scenarios that could not be foreseen in the design stage. This does not necessarily mean, however, that their behaviour is not predictable or that the systems are capable of free will. Control systems do only what they are programmed to do, regardless of the complexity of their programming.[18]

---

[16] Weiss, G., *Multiagent Systems*, 2nd edn (MIT Press: Cambridge, MA, 2013), pp. 54–55. Note that belief, desire and intention are expressed in numerical terms. The value function assigns numbers to an action to achieve a goal. E.g. a task to pick up an object would give 1 as a value for picking up the object and 0 otherwise, maybe 0.5 if the object is picked up but then falls.

[17] Control systems that only include goal information in their value function are counted as 'goal-based systems' under Russell's and Norvig's classification. Control systems that include information about utility of the action outcomes in their value function are called 'utility-based agents'. These agents can vector performance and efficiency factors to maximize their course of action. Utility-based agents are more intelligent and efficient than goal-based agents. They are preferable when meeting the goal cannot be achieved in a single action and the agent is required to plan a series of sequential actions. Russell and Norvig (note 14).

[18] Righetti, L., 'Emerging technology and future autonomous systems: speaker's summary', Autonomous Weapon Systems: Implication of Increasing Autonomy in the Critical Functions of Weapons, Expert Meeting, Versoix,

It should be mentioned that 'randomized' algorithms can be used in both reactive control systems and deliberative control systems. Randomized algorithms are 'non-deterministic' in that they allow systems to randomly pick a solution to solve a problem. In the context of a reactive agent, the use of randomized algorithms allows the agent to escape from an infinite loop (i.e. the situation when an agent endlessly repeats an action to meet a goal but the goal cannot be achieved by that action) by randomly picking between two predetermined alternatives. In the case of a vacuum cleaner, this could be randomly turning left or right when confronted by an obstacle. In deliberative control, the use of randomized algorithms is useful to prevent a system from having to search all possible combinations of actions. For some processes, the use of random algorithms provides the simplest or fastest way to achieve a result. The issue with the use of randomized algorithms is that it provides such systems with the potential to generate different behaviour under the same input condition. In other words, it introduces some unpredictability into the behaviour of the system. That is why the use of non-deterministic algorithms is rare in safety-critical systems (i.e. systems whose failure could result in loss of life, significant property damage, or damage to the environment), which include application areas such as medical devices, aircraft flight control, weapons and nuclear systems.[19]

Act

The decisions made by the control systems are then exerted in the real world through computational or physical means.[20] In the cyber-realm, for instance, this could be a software program that would implement a specific action such as blocking a malicious code. When discussing robotic platforms, the means through which the systems interact with the environment are commonly referred to as 'end-effectors' and 'actuators'. End-effectors are the physical devices that assert physical force on the environment: wheels, legs and wings for locomotion, as well as grippers and, of course, weapons. Actuators are the 'muscles' that enable the end-effectors to exert force, and include things such as electric motors and hydraulic or pneumatic cylinders. It should be noted that actuators and end-effectors might in some cases be coupled with sensors that will provide feedback information to the control systems concerning the task execution.

In summary, autonomy derives, from a technical standpoint, from the ability of a system to sense and act upon an environment and direct its activity towards achieving a given goal. Figure 2.1 represents in a simple fashion how these different capabilities interact with each other within a system that uses a (model-based) reactive control system or a deliberative control system.

What are the underlying technologies?

Anatomy of autonomy: underlying technology architecture

As implied by the previous description, autonomy is, at a fundamental level, always enabled by some type of underlying technology:

1. Sensors that allow the system to gather data about the world.
2. A suite of computer hardware and software that allows the system to interpret data from the sensor and transform it into plans and actions. The three most important

---

Switzerland, 15–16 Mar. 2016, p. 39.

[19] Knight, J., 'Safety critical systems: challenges and directions', Conference paper, 24th International Conference on Software Engineering, Orlando, Florida, 19–25 May 2002.

[20] Russell and Norvig (note 14), pp. 988–90.

technologies in this regard are computer chips, sensing software and control software that together form the 'brain' of the system.

3. Communication technology and human–machine interfaces that allow the system to interact with other agents, whether they be machines or humans.

4. Actuators and end-effectors that allow the system to execute the actions in its operating environment.

These different components form the underlying architecture of autonomy. The actual characteristics of these underlying technologies will be different depending on the nature of the task and the operating environment. It should also be noted that technologies may be integrated within a single machine (which could be described as 'self-contained autonomy') or distributed across a network of machines (which could be described as 'distributed autonomy').

Autonomy: a 'software endeavour'

Advances in autonomy in weapon systems are dependent upon technological progress in multiple areas. Advances in sensor technologies are certainly crucial as such technologies determine the accuracy of the data that systems can collect on their operating environments. Likewise, advances in computer processing technologies play an important role as they determine the speed at which the software part of a system can 'think' as well as the volume of data that it can efficiently handle. The design of the actuators and end-effectors will also affect the hardiness, endurance and cost of the systems.

The technologies that are deemed the most critical to autonomy, however, are the software elements. As a 2012 report by the Defense Science Board of the US Department of Defense (DOD) pointed out, autonomy is primarily a 'software endeavour'.[21] It is the complexity of sensing, modelling and decision-making software that actually determines the level of autonomy of a system. In other words, autonomy is a very 'diffuse' technology that does not easily lend itself to being tracked or measured because it fundamentally depends on the ingenuity of human programmers to find a way to break down a problem into mathematical rules and instructions that the computer will be able to handle. That being said, the state of the art is relatively well known. The following section describes what is currently feasible for humans to achieve in programming within the bounds of contemporary knowledge.

IV. Creating autonomy

This section takes stock of the extent to which autonomy remains an engineering challenge. It starts by discussing the variables that make autonomy difficult to engineer from a programming perspective. Next, it presents the state of enabling technology and what such technology allows through the development of machine perception, decision making and actuation. Finally, it discusses how autonomy is programmed and the extent to which the recent progress made in machine learning could fuel significant advances in autonomy in weapon systems.

How difficult is it to achieve autonomy?

Achieving autonomy is, by definition, not actually that difficult. According to Russell and Norvig, the extent to which it is feasible with today's technology depends on two

---

[21] US Department of Defense (DOD), Defense Science Board (note 8) p. 22.

interrelated variables: (*a*) the complexity of the task; and (*b*) the complexity of the environment (see figure 2.2).[22]

The complexity of the task

The complexity of a task primarily has to do with the extent to which it is possible to model the task mathematically and does not reflect how difficult its execution might be according to human standards. A famous paradox in the AI and robotics community—known as 'Moravec's paradox'—is that 'hard problems are easy and easy problems are hard'. According to Hans Moravec, 'it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility'.[23]

There are several variables that contribute to making a task complex from a programmer's point of view. The first variable is precision: how well defined is the task? Does the task follow programmable rules or a concrete logic? The more abstract or ill-defined the task specifications, the harder it is to formulate in terms of a mathematical problem and a solution.

The second factor is that of tangibility: can the expected outcome be quantified? Task executions that require qualitative judgement are often problematic because the outcome cannot be assessed in objective terms. It is debatable for instance whether the principles that govern the use of force in international humanitarian law (IHL)—notably proportionality and precaution in attacks—could, or should, ever be represented in terms that a computer could reason with. A third variable is dimensionality: can the task be executed in a single action or does it require sequential decisions and actions? How many possibilities are the systems facing to execute each action? The combined answers to these two questions determine the number of possibilities that the systems might have to process to take a decision. The more possibilities that exist, the more advanced the programming needs to be and the more computing power is necessary to engineer optimal solutions to a problem. A fourth variable is interaction: does execution of the task require interaction with other autonomous agents (e.g. humans)? What is the nature of the interaction: are agents competing, collaborating or simply communicating? Modelling interaction with other agents, particularly humans, in either a competitive or collaborative context is fundamentally difficult as human behaviour is often unpredictable.

The complexity of the environment

The complexity of the environment derives from several elements. Is the environment fully observable or partially observable through sensors? Is it a known or well-understood environment? Is it structured or unstructured? Is it cluttered or uncluttered? Is it static or dynamic? Is it a deterministic or stochastic environment (i.e. does the system's action always produce the same effects on it?) Is it an adversarial environment where actors may actively seek to defeat the system? All these variables affect the extent to which the environment is predictable and can be modelled in advance either explicitly (e.g. a map showing what the environment looks like precisely) or implicitly (rules about how it works, e.g. rules of the road). The less predictable the environment, the harder it is to model and therefore the harder it is to create autonomous capabilities within systems, at least those that are effective and reliable.

---

[22] Russell and Norvig (note 14).

[23] Pinker, S., *The Language Instinct* (Harper Perennial: New York, 2007), pp. 190–91; and Moravec, H., *Mind Children* (Harvard University Press: Cambridge, MA, 1988).
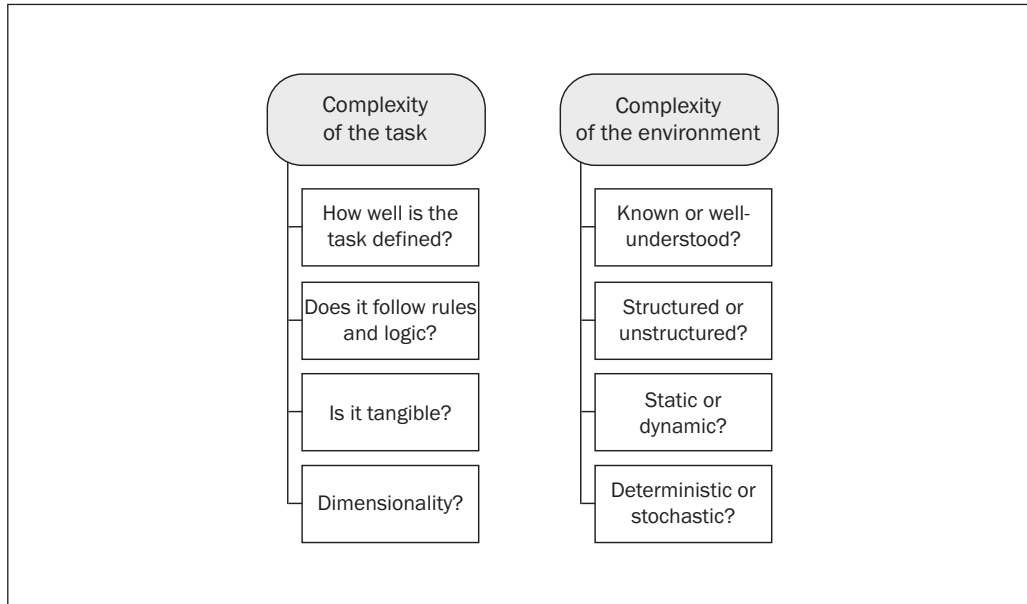
Figure 2.2. Complexity factors in creating autonomy

Source: Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Pearson Education: Harlow, 2014).

The case of navigational autonomy in robotic platforms provides a good illustration of the challenges posed by varying levels of complexity in different environments. Navigational autonomy is comparatively easy to create for systems operating in the air or underwater for the simple reason that generally these two domains are uncluttered: they feature a limited number of possible obstacles. In addition, the laws of physics in these two domains are well understood. Hence, they can be easily represented in mathematical terms. The land domain, on the other hand, offers greater complexity in many regards: the structure of the terrain may vary greatly, the systems may face many different types of obstacles and have to interact with other autonomous agents—either other machines or humans—whose behaviour might be unpredictable. Engineers know very well how to make self-driving vehicles that can operate within constrained and structured environments (within a factory or on the tarmac of an airport) or unpopulated or sparsely populated semi-structured environments (such as a motorway) because these can easily be explicitly mapped or implicitly modelled in advance. Making self-driving vehicles capable of operating in highly diverse human environments, such as a city centre, and various weather conditions is much more challenging because it is difficult—if not impossible—for a programmer to develop a model that will capture all possible combinations of events. By definition, a model is a simplified version of the world; there is always a risk of a corner case (i.e. a problem or situation that has not been represented and planned for in the model of the world).

What are the bottlenecks?

Presenting the current state of autonomy in a single description is difficult because the description depends upon the types of tasks, systems and environments that are of interest. Advances in autonomy in the context of weapon systems will be discussed in the next chapter. Several general observations can be made, however, with regard to the advances and limitations of underlying capabilities: perception, decision making and actuation.

Perception

Advances in machine perception are key to the progress of machine autonomy. In many respects, it is the limitation of perceptual intelligence that is today the most important obstacle to the development and use of robotic technologies outside simple, predictable or well-controlled environments.

Computers are increasingly efficient at sensing and making sense of the world. The ability of computer vision systems to recognize objects and people, and scenes and events, continues to improve. Speech recognition technologies are also increasingly efficient at recognizing spoken words and sentences. Computers still struggle, however, with interpreting the wider context. State of the art computer vision software may identify that a person is walking, but it is unable to determine why the person is walking. Likewise, state of the art speech interfaces can recognize a complex spoken sentence (what is said) but are unable to determine or recognize the topic of the conversation (what is being discussed). A computer's lack of contextual understanding derives from the fact that it remains very complex for engineers to represent in a model the abstract relationship between objects and people in the real world.[24]

From the perspective of autonomy, a fundamentally problematic consequence of a computer's perceptual intelligence limitations is that the systems or system functions can easily be tricked and defeated by a malevolent actor or unforeseen situations in the system's operating environment.

Decision making

Part of the limitations of machine perception derives from the limitation of synthetic reasoning. Advances in computer processing technology enable computers to perform calculations that are far beyond human capabilities. They are powerful, fast and precise. However, computers only excel at deductive reasoning, whereas humans are also able to conduct inductive and adductive reasoning. Computers still have major difficulties inferring general rules from single real-life cases (they need evidence of a large number of similar situations in order to learn). This is one reason why fielding autonomous robots in unknown and uncertain environments is currently so problematic. Because they cannot as yet generalize from previous experiences and adapt to novel situations, they can only function reliably in situations that the programmers have prior knowledge of.[25]

Designing autonomy for general tasks that demand a complicated combination of subtasks, planning and motion planning—for example, making a humanoid robot cook a meal—continues to be a fundamentally complex endeavour as it is difficult to model all the decision-making parameters, and it requires a significant volume of calculation for the systems to find the optimal solutions. Despite many significant technological advances, the current state of computer processing remains an obstacle to the execution of such tasks: it might take a long time or a lot of computer processing resources to solve every facet of the mathematical problem that these tasks involve.

Hardware problems

Advances in autonomy are hindered not only by the limitations of computer processing technology and software engineering but also by hardware weaknesses, with power sources posing a particular challenge. For many robotic systems (e.g. humanoid

---

[24] Karpathy, A., 'The state of computer vision and AI: we are really, really far away', Andrej Karpathy Blog, 22 Oct. 2012.

[25] Endsley, M. R., *Autonomous Horizons: System Autonomy in the Air Force: A Path to the Future, Volume 1: Human-Autonomy Teaming* (United States Air Force, Office of the Chief Scientist: Washington, DC, 2015), p. 5; and Cummings, M., *Artificial Intelligence and the Future of Warfare*, Research Report (Chatham House: London, 2017).

Box 2.2. Machine-learning methods

According to Nilsson, 'a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves. Some of these changes, such as the addition of a record to a database, fall comfortably within the province of other disciplines and are not necessarily better understood for being called learning. But, for example, when the performance of a speech recognition machine improves after hearing several samples of a person's speech, we feel quite justified in that case to say that the machine has learned'.

A machine can learn on the job (online learning) or during a training phase (offline) with a wide spectrum of methods that can be sorted into four generic categories: reinforcement learning, supervised learning, unsupervised learning and semi-supervised learning.

1. *Reinforcement learning*. The machine receives some reward for its action. It obtains more rewards when the outcome is closer to the desired outcome. This motivates it to find the most suitable solution. The desired outcome is never presented to the machine.

2. *Supervised learning*. The machine learns by comparing example inputs with desired outputs. The data is labelled with the correct answer. Examples include systems that learn image recognition by scanning databases with tagged images.

3. *Unsupervised learning*. The machine is only presented with raw data and it must find patterns in the data itself. It is the most difficult method of learning and the one that currently shows the least mature results.

4. *Semi-supervised learning*. The machine is presented with both labelled and unlabelled examples of data.

In practice, the distinctions between the categories are not always clear-cut and different methods may be used to train a system.

Source: Nilsson, N. J., *Introduction to Machine Learning: An Early Draft of a Proposed Textbook* (Stanford University: Stanford, CA, 1998), p. 1.

---

robots), the heavy weight and limited durability of batteries are fundamental obstacles to their viable use in outdoor and unstructured environments.

Handcraft programming versus machine learning

Currently, most software is handcrafted, meaning that human programmers are entirely responsible for defining the problems to be solved by the software and the way in which it solves those problems. This requires a great deal of research on how the world works. Engineers developing autonomous systems often cooperate with scientists from other scientific fields, notably the natural sciences (e.g. neurosciences and physics) and the social sciences (e.g. psychology, linguistics and sociology), in order to develop the model and rules that will govern the behaviour of the systems, whether for perception or decision making.

Handcraft programming has limitations, particularly when tasks and operating environments are too complex for a human to model them completely.[26] This is one of the reasons why in many areas of AI and robotics research—two disciplines that are directly involved in the development of autonomy—programmers now rely extensively on machine learning to develop their systems.[27]

Machine learning is an approach to software development that consists of building a system that can learn and then teaching it what to do using a variety of methods (see box 2.2). This is a complex and data-heavy undertaking. Machines learn by abstracting statistical relationships in data. To be taught, they need to be provided with large amounts of training data (real-world examples) and rules about the data relationship. The main advantage of machine learning compared with traditional programming is that humans do not have to explicitly define the problem or the solution; instead, the machine is designed to improve its knowledge through experience.

[26] Kester, L., 'Mapping autonomy', Presentation at the CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems, Geneva, 11–15 Apr. 2016.
[27] Russell and Norvig (note 14), p. 56.

Box 2.3. Deep learning

Deep learning is a type of representation learning, which in turn is a type of machine learning. Machine learning is used for many but not all approaches to artificial intelligence.

Representation learning is an approach to machine learning whereby the system 'learns' how to learn: the system transforms raw data input to representations (features) that can be effectively exploited in machine-learning tasks. This obviates manual feature engineering (whereby features are hard-coded into the system by humans), which would otherwise be necessary.

Deep learning solves a fundamental problem in representation learning by introducing representations that are expressed in terms of other, simpler representations. Deep learning allows the computer to build complex concepts from simpler concepts. A deep-learning system can, for instance, represent the concept of an image of a person by combining simple concepts, such as corners and contours.

Deep learning was invented decades ago but has made important progress in recent years, thanks to improvements in computing power and increased data availability and techniques to train neural networks.

Source: Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning* (MIT Press: Cambridge, MA, 2016), p. 8.

Machine learning: opportunities and challenges

Machine learning has been around for decades but has made great strides in recent years, notably due to improvements in computer power and developments in deep learning—a specific technique based on neural networks, which draws on knowledge of the human brain, statistics and applied maths (see box 2.3).[28] These recent advances have created both important opportunities and challenges for the development of autonomy in weapon systems.

Recent advances in machine learning have proved to be very useful for machine perception. They allow the programmer to design sensing software that features remarkable capabilities in terms of pattern recognition (whether objects, faces or radio signals).[29] They create improvement opportunities in all application areas of autonomy in weapon systems, from target recognition to navigation.

Machine learning also poses a number of practical challenges. First, machine learning is data intensive: in order to learn, the systems must be supplied with large volumes of training data. For many tasks, including targeting, the lack of high-quality training datasets remains a fundamental problem. This has led some experts to speculate that 'datasets—not algorithms—might be the key limiting factor to development of human-level artificial intelligence'.[30]

A second fundamental challenge concerns the predictability of systems.[31] Machine-learning systems, particularly those that run on deep neural networks, could be said to operate like 'black box' systems: the input and output of the system are observable but the process leading from input to output is unknown or difficult to understand. It is particularly difficult for humans to understand what such systems have learned and hence how they might react to input data that is very different from that used during the training phase.[32] Likewise, unless the system's learning algorithm is frozen at the end of the training phase, once deployed, it might learn something it was not intended to learn or do something that humans do not want it to do.[33] These are some of the reasons why the use of machine learning in the context of weapon systems has been limited to experimental research. The introduction of machine-learning capabilities in deployed systems is unlikely in the near future unless the engineer community manages to solve some of the methodological problems that

---

[28] Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning* (MIT Press: Cambridge, MA, 2016); and Murnane, K., 'What is deep learning and how is it useful?', *Forbes*, 1 Apr. 2016.

[29] Gershgorn, D., 'See the difference one year makes in artificial intelligence research', *Popular Science*, 31 May 2016.

[30] Wissner-Gross, A., 'Datasets over algorithms', *Edge*, 13 June 2017.

[31] Righetti (note 18).

[32] Postma, E., 'Deep learning: the third neural network wave', Data Science Center Tilburg Blog, Feb. 2016.

[33] Roff, H. and Singer, P. W., 'The next president will decide the fate of killer robots—and the future of war', *Wired*, 6 Sep. 2016.

learning systems, particularly those that can learn online, pose to existing methods of verification (i.e. methods that are used to ensure that a system conforms with a regulation, requirement, specification or imposed condition; the issue of verification is further discussed in chapter 4).

V. Conclusions

The key conclusions from this brief introduction to the technological foundation of autonomy can be summarized in two points.

First, the study of autonomy as a general attribute of a weapon system is imprecise and potentially misleading. Autonomy may serve very different capabilities in different weapon systems. For each of these capabilities the parameters of autonomy, whether in terms of the human–machine command-and-control relationship or the sophistication of the decision-making process, might vary greatly, including over the duration of a mission. In this regard, the continued reference to the concept of LAWS in the framework of the CCW may be deemed problematic. It has trapped states and experts into a complex and contentious discussion about the level at which a system might be deemed autonomous, while in reality the concerns—be they from a legal, ethical or operational standpoint—need to be articulated on the use of autonomy for specific functions or tasks. Future CCW discussions could, therefore, usefully benefit from a conceptual reframing and a shift from a platform- or system-centric approach to a functional approach to autonomy. Focusing on function and capabilities of 'autonomy in weapon systems' rather than the development of LAWS as a category of weapon could foster a much more consensual and constructive basis for discussion.[34]

If there is one technological development that future GGE discussion should focus on it is machine learning. Learning is often described as an increasingly important, if not the defining, feature of the future of autonomy in weapon systems. There seems to remain a lack of understanding, and occasionally some confusion, among CCW delegates about what machine learning actually is, how it works and to what extent it could unlock significant advances in autonomy in weapon systems. It would, therefore, be useful if the GGE could focus some of its work on machine learning's potential and the limitations of its algorithms with regard to further advancing autonomy in weapon systems. Clarifications about the difference between 'offline' and 'online' learning—whether in terms of potential, limitations or risks—would be particularly welcome. There is also one near-term development that deserves particular scrutiny: the use of deep-learning algorithms for the training of automatic or automated target recognition (ATR) systems (discussed in more detail in chapter 3). These are likely to be used to make ATR systems learn to differentiate between military and civilian objects. It would be useful to know what the implications of such a development would be, as they could be a key factor in assessing the legality of a system under IHL when conducting weapon reviews pursuant to Article 36 of Additional Protocol I of the 1949 Geneva Conventions (see chapter 4).[35]

---

[34] This view is also shared by a number of experts that have studied the development of autonomy in weapon systems, including Kerstin Vignard, Chief of Operations at the UN Institute for Disarmament Research (UNIDIR). Vignard stressed this point in her statement at the 2016 CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems in Geneva in Apr. 2016. Vignard (note 7).

[35] On autonomy and Article 36 see Boulanin, V., 'Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems', SIPRI Insight on Peace and Security, no. 2015/1, Nov. 2015.

3. What is the state of autonomy in weapon systems?

I. Introduction

This chapter provides a factual overview of the current state of autonomy in weapon systems. It aims to aid policy makers and the interested public to gain a more concrete sense of (*a*) the actual functions and capabilities of autonomy in weapon systems; and (*b*) how autonomy is currently used. The chapter proceeds as follows. The remainder of the introduction presents the dataset developed by SIPRI for the purpose of its mapping exercise. Section II maps the existing application area of autonomy in current weapon systems. Section III presents the major types of weapon systems that may be deemed autonomous according to some definitions.

Introducing the SIPRI dataset on autonomy in weapon systems

In order to obtain an overview of the state of autonomy in existing weapon systems, SIPRI designed and populated a dataset, which will be made publicly available on the SIPRI website in November 2017.

The dataset provides general information on the types, purposes, origins (companies/countries), users and development status of a sample of military systems that include autonomous functions in at least one of the following capability areas: mobility, targeting, intelligence, interoperability and health management. The dataset is not intended to be comprehensive. SIPRI has focused its data collection efforts on weapon systems and unarmed military robotic systems that have been deployed or are under development in the countries identified by SIPRI as among the largest producers of arms in the world—namely, the USA, the UK, Russia, France, Italy, Japan, Israel, South Korea, Germany, India, Sweden and China.[1]

The data was collected from a variety of sources, including industry guides, newspaper articles, company websites, press releases, defence publications, reports from NGOs, interviews, scientific articles and YouTube videos. Attempts were made to collect a minimum of three independent sources on each system to verify information. However, it has proved difficult to find and verify data for many of the systems due to the lack of details available and uncertainty as to the reliability of certain information (either because the source could be biased or because of translation issues). For these reasons, the dataset features a colour code that grades the reliability of the information collected.

As of April 2017, the dataset consisted of 381 different systems, including the following.

1. Unmanned weapon systems that feature some autonomy in their critical functions—that is, they can autonomously search for, detect, identify, select, track or attack targets.[2]

---

[1] Countries listed by size of share of arms sales of companies listed in the SIPRI Top 100 arms-producing companies and military service companies for 2014. The SIPRI Top 100 lists the world's 100 largest arms-producing companies and military services companies (excluding China). These are ranked by volume of arms sales. While China is not covered by the SIPRI Top 100 due to the lack of data on arms sales, it is believed to be one of the largest arms-producing countries. SIPRI considers that at least 9 of the 10 major state-owned conglomerates under which the Chinese industry is organized would be listed in the Top 100 if official data was available. Fleurant, A. et al., 'The SIPRI Top 100 Arms-Producing Companies and Military Services Companies, 2014', SIPRI Fact Sheet, Dec. 2015.

[2] A 'weapon system' is understood to be a system that may consist of multiple physical platforms, including carrier and launch platforms, sensors, fire control systems and communication links needed for a weapon to engage a target.
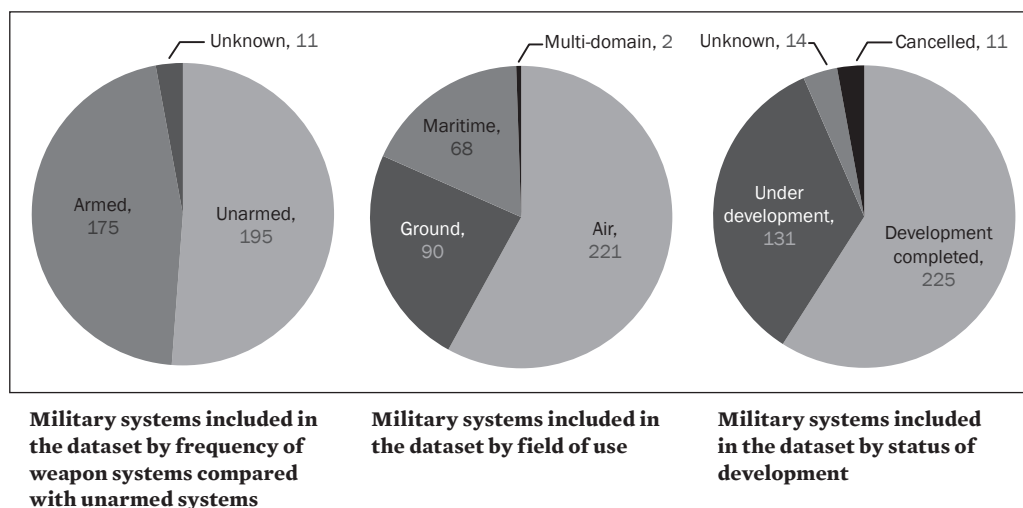
Figure 3.1. Military systems included in the SIPRI dataset by (*a*) frequency of weapon systems compared with unarmed systems; (*b*) field of use; and (*c*) status of development

Source: SIPRI dataset on autonomy in weapon systems.

2. Unmanned weapon systems that do not have autonomy in their critical functions but feature autonomous functions in any of the other capability areas covered by the study—namely mobility, intelligence, interoperability and health management.

3. Unmanned and unarmed military systems—uses of which include (but are not limited to) intelligence, surveillance and reconnaissance (ISR) missions or logistics (supply) missions—that feature any of the capability areas covered by the study.

Systems within the last two categories were included to provide a broader picture of the state of autonomy in unmanned military systems, and because these systems could eventually be weaponized (in the case of unarmed systems) or fitted with autonomous targeting capabilities in the future.[3]

It should be noted that SIPRI focused its mapping exercise on weapon systems rather than individual munitions. Guided munitions such as sensor-fused munitions, cruise missiles and torpedoes were excluded, primarily for reasons of data collection feasibility. Providing a detailed mapping of existing guided munitions would have been a study in itself.

Overall, the dataset contains 195 unarmed systems, 175 weapon systems and 11 systems whose armed status is unclear (see figure 3.1). Aerial systems make up the largest proportion of the systems included in the dataset, and development has been completed for the majority of systems covered.

II. Existing functions and capabilities

What is the state of autonomy in military systems today? Extensive research shows that existing military systems already include multiple autonomous functions. These functions can be divided into five capability areas, which are here presented in order of recurrence: (*a*) mobility; (*b*) targeting; (*c*) intelligence; (*d*) interoperability; and (*e*) health management (see figure 3.2). This section examines each capability area based on the following two questions.

---

[3] Cockburn, A., *Kill Chain: Rise of the High-Tech Assassin* (Picador: London, 2015).
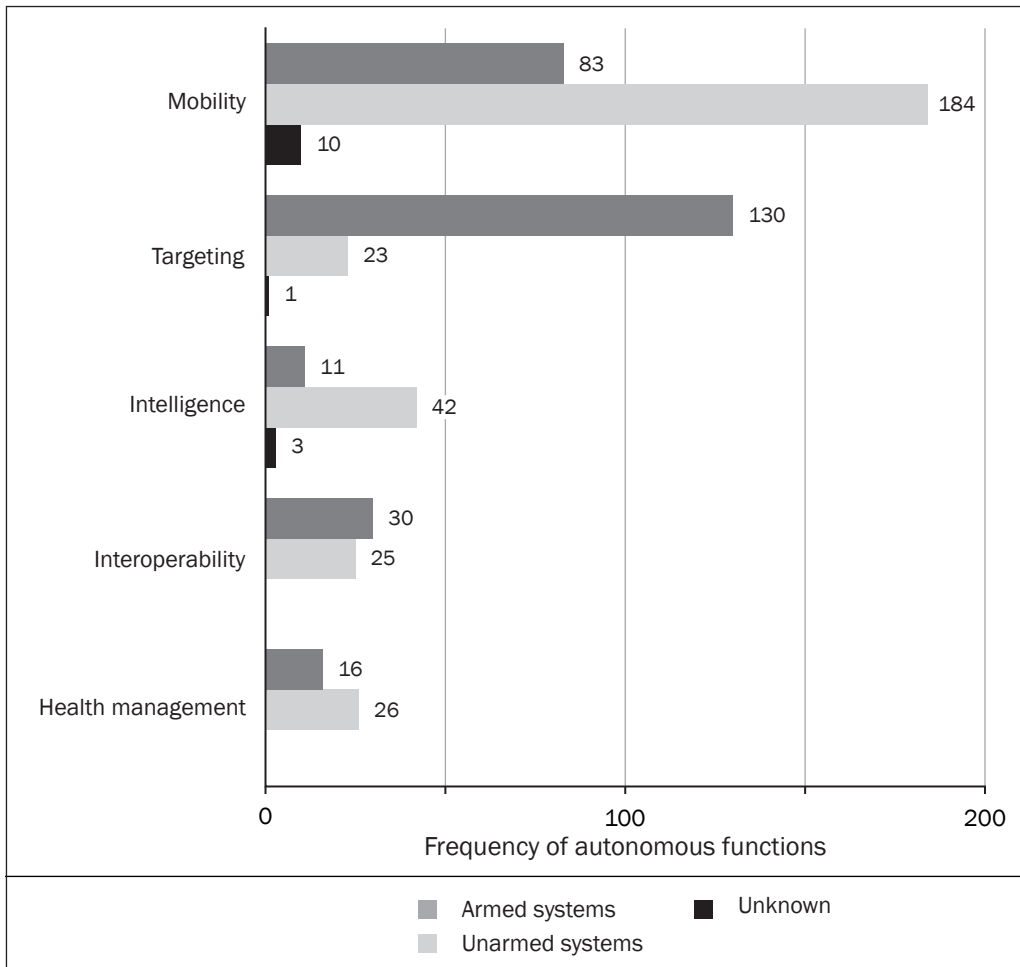
Figure 3.2. Autonomous functions in existing military systems, by capability area

Source: SIPRI dataset on autonomy in weapon systems.

    1. What can military systems do and not do autonomously?
    2. What is the nature of the human–machine command-and-control relationship when the systems execute the relevant capability autonomously?

III. Autonomy for mobility

The predominant application area for autonomy in military systems is mobility. SIPRI has identified 277 military systems (out of the 336 in the dataset that can be deemed mobile) that include functions which allow the system to govern and direct its own motion within its operating environment without direct involvement of a human operator.[4]

Functions and capabilities

Mobility-related autonomous functions that can be found in existing systems vary greatly in terms of capability and technological sophistication. The most noteworthy functions include (*a*) homing/follow-me; (*b*) autonomous navigation; and (*c*) take-off and landing.

---

[4] Note that the autonomous capabilities of munitions as launched by air defence systems were not included in this study, as the dataset focuses only on complete military systems.

Homing and follow-me

Homing and follow-me are, from a technical standpoint, simplistic forms of self-direction, which work on the same principle but have different purposes. Homing is a capability that is usually associated with missile technology; it requires that the system can find and track its targets, while follow-me refers to the ability of an unmanned system to follow another system or a soldier. In both cases, the system directs its motion towards a specific object or person that it detects and tracks through a radar, acoustic or electromagnetic signal, or an electro-optical (visual) or infrared (IR) (heat) signature. The signal or signature that the system follows is pre-programmed in advance and stored in the system memory. Existing systems have no ability to pick up new signals once activated and deployed. When operated in a cluttered environment a system might include an automatic sense-and-avoid capability to prevent collisions with possible obstacles.

Autonomous navigation

Autonomous navigation is the most crucial capability when it comes to system self-direction. It ensures that the system can accurately ascertain its position, and plan and follow a route on its own.

Most military systems that reportedly feature an autonomous navigation capability are arguably not truly autonomous in the sense that they rely on 'waypoint navigation': the system merely follows a series of geodetic coordinates that are entered by a human operator. Some systems, notably newer systems such as the MQ-4C Triton, an unmanned aerial system (UAS) developed by Northrop Grumman for long-term ISR missions, can autonomously plan a route, but the general navigation parameters (e.g. speed, altitude and mission objective) are still set by a human operator.[5]

The actual navigational autonomy of existing systems is also relative to the complexity of their operating domain (i.e. whether the system is operating on land, in the air or at sea, and whether or not the operating domain is adversarial).

The technical requirements are generally lower for aerial systems and maritime systems than ground systems, for the simple reason that the air and sea domains are typically far less complex than the land domain. The air and sea domains feature few, if any, obstacles and fewer unforeseeable environmental variations. In theory, waypoint navigation and a simple sense-and-avoid capability may be sufficient to ensure that an aerial or maritime system can navigate in complete autonomy for extended periods.

The land domain, especially in a military context, displays far greater complexity: (*a*) the structure of the terrain can vary markedly; (*b*) the domain may include many different types of obstacles; and (*c*) the system may need to interact with other autonomous agents—either other machines or humans—whose behaviour might be unpredictable. To navigate autonomously and to identify paths and obstacles, ground systems need to include advanced vision-based guidance systems or inbuilt pre-mapping of the environment or both. Existing ground systems that have an autonomous navigation capability tend to rely heavily on pre-mapping, partly because the state of the art vision-based guidance technology is not sophisticated enough. This means that most current ground systems are only capable of navigating autonomously if an area is known in advance and not subject to major changes, which drastically restricts the type of mission that they can perform autonomously. Such missions could include perimeter surveillance (around borders, military bases or critical infrastructure) and logistics.

---

[5] Rogoway, T., 'The Navy has the ultimate MH370 search tool, it's just not operational', Foxtrot Alpha, 18 Mar. 2014.

With the notable exception of missile systems and guided munitions—which are generally non-recoverable systems—military systems that feature an autonomous navigation capability are intended to operate in such a mode only in non-adversarial conditions. They do not have sufficient perception or decision-making capabilities to cope with adversaries that might actively seek to defeat their guidance system. One of the key vulnerabilities of these systems is that they typically rely on Global Positioning System (GPS) guidance, which makes them vulnerable to GPS jamming technologies. However, interest in systems capable of operating in GPS-denied environments is high, and GPS anti-jamming protection and non-GPS-based guidance systems seem to be important features in the latest generation of unmanned systems. In addition to jamming, enemies can also use strategies such as spoofing and cyber-attacks.[6]

Take-off and landing

An increasingly common feature among aerial systems is autonomous take-off and landing. From a technical standpoint, it is perhaps more appropriate to describe this capability in military systems as 'automatic take-off and landing' since these systems follow a very strict set of predefined rules, with the entire procedure operated by an algorithm. Reportedly, the technology has reached the point where machines outperform humans in terms of precision and reliability. One study notably found that the accident rate is lower when these phases of the flight are automated rather than being remotely operated by a human.[7]

Human–machine command-and-control relationship

Autonomy as a complement of remote control

The nature of the human–machine command-and-control relationship varies from one system to another. It is important to note that the aforementioned autonomous functions are most often used to complement remote control. Autonomous navigation, homing and follow-me are usually used to discharge humans from operating the system during phases of the mission where human cognitive capabilities are not essential or not the most appropriate. Autonomous take-off and landing capability is aimed at reducing the risk of accident when a system is supposed to take off or land in conditions that require high precision (e.g. take-off from or landing on an aircraft carrier). These features are also used to improve recoverability of systems in case of loss of communication, as they may be used to make the system 'return to base' or proceed to an emergency landing.

Mission autonomy

Existing systems that, once launched, navigate in complete autonomy, with little or no direct human supervision, can be divided into the following three categories.

1. Aerial, land and maritime systems that are deployed to conduct pre-programmed manoeuvres in known and semi-structured environments. Examples include the Amstaff, a tactical unmanned ground system (UGS) developed by Automotive Robotic Industries (Israel), which is capable of conducting perimeter protection operations autonomously.

---

[6] A 'spoofing' attack involves tricking a system's sensors using false information in order to alter the system's behaviour to the attacker's advantage. See Samuelson-Glushko Technology Law and Policy Clinic, 'Jamming and spoofing attacks: physical layer cybersecurity threats to autonomous vehicle systems', Submission to National Highway Traffic Safety Regulations, Washington, DC, 21 Nov. 2016, p. 5.

[7] Williams, K. W., *A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications* (Office of Aerospace Medicine: Washington, DC, 2004).

2. Unmanned systems that are intended to conduct long-term ISR missions in an environment where communications are difficult (e.g. underwater).

3. Missile systems and unmanned combat systems that are intended to strike targets in communication-denied environments.

IV. Autonomy for targeting

The second most notable application area of autonomy in weapon systems is targeting. SIPRI found that autonomy is used in at least 154 systems to support some, if not all, of the steps of the targeting process (at the tactical level), from identification, tracking, prioritization and selection of targets to, in some cases, target engagement.[8] Rather than discussing the systems themselves (they will be presented in more detail in sections VIII to XIII), this section focuses on the technology that supports the advance of autonomy for targeting.

Function and capabilities

'Autonomous' or 'automated' target recognition?

There is an open debate over whether it is appropriate to discuss autonomy in the area of targeting because the software technology that existing weapon systems use to find and attack targets is, from a technical standpoint, closer to basic automation than autonomy.

Target recognition software, often labelled as 'automatic or automated target recognition software' (ATR software), was invented in the 1970s and has relied on the same principle ever since: pattern recognition. Such software is programmed to recognize target types based on predefined target signatures. The decision-making process is simple: the target signature either matches or does not match a template that is stored in the target identification library.[9] When multiple targets can be identified, the systems also prioritize between them based on strict predefined criteria, which are likely to vary depending on the operational situation.[10]

It is important to note that ATR software has no deliberative autonomy. It can only identify and fire upon target types that have been predetermined by the human operator, and has no capability to learn new target signatures once deployed.

Automated target recognition

The target identification capabilities that can be found in existing weapon systems are, all in all, rather rudimentary.

In the majority of cases, ATR software can only recognize large and well-defined military objects: tanks, aircraft, submarines and radar. The way the software recognizes them varies depending on the nature of the targets, but generally it uses simple criteria: tanks are often recognized based on their shape and height, missiles are typically detected based on velocity, radio-frequency emission or both, while submarines are usually identified based on their acoustic signature. Robotic sentry weapons are the only type of weapon system to use ATR software to detect human targets. The actual recognition capability is very crude and the software can only recognize that

---

[8] For an analysis of the entire targeting process and how it may be carried out at the tactical, operational and strategic level see Ekelhof, M., 'Human control in the targeting process', ed. R. Geiß, *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law and Security* (German Federal Foreign Office: Berlin, 2016), pp. 66–75.

[9] Roff, H., 'Sensor-fused munitions, missiles and loitering munitions: speaker's summary', Autonomous Weapon Systems: Implication of Increasing Autonomy in the Critical Functions of Weapons, Expert Meeting, Versoix, Switzerland, 15–16 Mar. 2016, pp. 33–34.

[10] For an analysis of the difference between reactive and deliberative systems see chapter 2 of this report.

the target is a human. It does not have the ability to distinguish whether the human is a civilian or a soldier (this type of weapon system is further discussed in section XI).

The performance of ATR systems in general is also highly sensitive to variations in the environment. One report found that ATR performs reliably under favourable weather conditions and when the target is located in an uncluttered background.[11] As soon as the weather conditions deteriorate, and the target background becomes cluttered, the false detection rate of ATR software increases significantly. This means that weapon systems using this technology cannot be used safely in all circumstances.

ATR systems can only *recognize* predefined target types. They are unable themselves to make the evaluations necessary to ensure an attack complies with the rules and principles of international law in the conduct of hostilities, namely the obligations of distinction, proportionality and precaution (these principles are discussed in more detail in chapter 4). It is perhaps worth noting that systems *can* in fact apply the principle of distinction, but only in a very crude manner: they simply ignore everything that does not match the predefined target. For example, they are unable to evaluate whether the detected target has surrendered or is *hors de combat* for another reason. The only known exception is Samsung's SGR-A1, a sentry guard robot (now retired), which could detect surrender motions (arms held high to indicate surrender).[12] Existing systems are also unable to detect whether the target is surrounded by civilians and civilian objects, which would be a fundamental requirement to the application of the principles of proportionality and precaution.[13]

Slow progress of ATR technology

The aforementioned limitations of ATR technology are not due to a lack of progress in sensor technology; rather, they are the consequence of two recurrent problems associated with the development of ATR algorithms.

The first is the lack of training and test data. Target recognition algorithms need to be trained and tested on a large sample of data that is related to the mission scenario so as to expose the algorithm to all the variables that it will be expected to handle. This means that the dataset needs to include appropriate data about the target, but also conceivable variations owing to possible changes in operating environment (e.g. different backgrounds or weather conditions). For many target types (notably humans) and operational situations, finding data remains a fundamental challenge. This problem has been compounded by the fact that these datasets are often considered to be classified information that cannot be circulated among the community of industry, governmental and academic experts that are involved in the development of ATR technology.[14]

The second problem is that machine-learning techniques, such as deep learning, which could significantly facilitate the programming of ATR algorithms—notably by making ATR systems capable of learning by themselves the difference between military target objects and civilian objects (e.g. a tank versus a school bus)—raise concerns with regard to predictability. As previously discussed in chapter 2, learning systems operate like 'black boxes'. Humans have trouble understanding how they learn: the data and sensory input and the data output of the system are observable but the process leading from input to output is unknown or difficult to comprehend. This creates some uncertainty as to how the system might react to input data that is very different

---

[11] Ratches, J., 'Review of current target recognition systems', *Optical Engineering*, vol. 50, no. 5 (2011), pp. 1–7.

[12] 'Samsung Techwin SGR-A1 Sentry Guard Robot', Global Security, [n.d.].

[13] Translating the requirements of proportionality and precaution into an algorithmic form remains challenging, there is an open debate among experts as to whether it will ever be possible. For further discussion see chapter 4 of this report.

[14] Ratches (note 11), pp. 1–7.

Box 3.1. Typology of the human–weapon command-and-control relationship according to Human Rights Watch

Human Rights Watch describes the typology of the human–weapon command-and-control relationship as follows.

    1. *Human-in-the-loop weapons*: robots that can select and deliver force only with a human command.

    2. *Human-on-the-loop weapons*: robots that can select and deliver force under the oversight of a human operator who can override the robot's actions.

    3. *Human-out-of-the-loop weapons*: robots that are capable of selecting targets and delivering force without any human input or interaction.

Source: Docherty, B., *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch/International Human Rights Clinic: Washington, DC, 2012), p. 2.

from that used during the training phase (i.e. the phase during which the system is presented with numerous examples, such as pictures of trucks and pictures of tanks, in order to improve its performance). As a result, the use of machine learning for the development of ATR software has so far been restricted to experimental research.[15]

Human–machine command-and-control relationship

For many systems, the limitations of ATR technology are not fundamentally problematic because the weapon systems are intended to operate as a decision aid and in an operational context where the presence of civilians and civilian objects is unlikely.

Human-in-the-loop: ATR as a decision aid

Nearly one-third of the systems (50 of 154) identified by the SIPRI dataset as having autonomous targeting capabilities use ATR as a 'decision aid' for human operators. ATR software is mainly used in cases when a target is beyond the visual range of a human operator or moving too fast for a human to identify and track. Such ATR systems may be capable of detecting, tracking, prioritizing and selecting targets autonomously but human operators retain the decision to engage the target. In the terminology that Human Rights Watch (HRW) has developed to describe categories of autonomous weapon systems, these could be designated as 'human-in-the-loop' weapons (see box 3.1).[16] An additional 31 systems are known to use ATR as a decision aid, but it is unclear whether they engage autonomously. This category mostly includes air defence systems.

ATR for human-on-the loop and human-out-of-the-loop systems

Around one-third of the systems (49 out of 154) identified by the SIPRI dataset as having autonomous targeting capabilities have the capacity to engage with targets without the direct involvement of a human operator.[17] Primarily, these are weapon systems that are intended to protect ships, ground installations or vehicles against incoming projectiles. They are generally operated under human supervision ('human-on-the-loop' weapons in the HRW terminology) and have different modes of engagement. They use the autonomous mode only in situations where the time of engagement would be too short for humans to be able to respond.

---

[15] Warwick, G. and DiMasco, J., 'Machine learning key to automatic target recognition', *Aviation Week & Space Technology*, 26 May 2016.

[16] Docherty, B., *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch/International Human Rights Clinic: Washington, DC, 2012).

[17] It was impossible to determine, using open sources, whether these ATR were coupled with an automated fire control, which would allow the sytems to attack targets without the direct approval of an operator.

ATR in unarmed systems

ATR technology is not the sole prerogative of weapon systems. It can also be found in unarmed military systems. These are typically unarmed systems used for ISR missions with the aim of feeding target information to another weapon system or to a command-and-control chain. The SIPRI dataset identified 24 systems (23 unarmed systems and 1 whose armed status is unknown) of that sort. One notable example is the aforementioned MQ-4C Triton, the Northrop Grumman surveillance UAS. The MQ-4C can detect and classify targets using advanced image and radar return recognition software. The MQ-4C can be pre-programmed to zoom in on particular target types and relay images that might be of specific interest to human operators.[18] The MQ-4C can also be used to mark targets to assist with the strike and for making post-strike assessments.

V. Autonomy for intelligence

A third important application area of autonomy in weapon systems is intelligence. SIPRI found that autonomy is used in at least 56 military systems to collect and process various types of information that might not be directly related to targeting but that might be of critical relevance from a command-and-control perspective.

Functions and capabilities

Automated detection of objects and events

The types of information that existing weapon systems can handle remain relatively simple. In most cases, the information processing takes the form of automated detection of simple objects or events that match specific predefined criteria. Examples of this process include (but are not limited to) the following.

1. *Detection of explosive devices.* This capability is mainly found in robotic weapon systems that are specifically designed for bomb ordnance disposal. They use various kinds of sensors depending on the type of explosive device they are supposed to detect (e.g. landmines, sea mines or improvised explosive devices, IEDs). These systems are usually managed by human operators, who are generally in charge of the actual disposal of the explosive device by remote control. However, some recent systems, such as the Counter IED and Mine Suite (CIMS) developed by IAI (Israel), are now capable of executing the entire process autonomously from detection to destruction.[19]

2. *Detecting perimeter intrusion.* This is a relatively unsophisticated function that is typically found in robotic platforms that are intended to secure known perimeters, such as military bases, borders or warehouses. The detection process is straightforward: the systems are programmed to detect movement or the presence of unauthorized living forms using a suite of sensors. One notable example is the Mobile Detection Assessment and Response System (MDARS), a UGS developed jointly by General Dynamics for the US Army and the US Office for Naval Research for autonomous patrol of storage sites and warehouses. It can autonomously detect intrusions using forward-looking infrared (FLIR), radio-frequency identification (RFID), radar, and light detection and ranging (LADAR) sensors. It can also control inventories through special radio-frequency transponder tags.[20]

[18] Rogoway (note 5).
[19] Eshel, T., 'Israeli smart multi-sensor counters IEDs', Defense Update, 7 Oct. 2014.
[20] Hudson, C., Nguyen, H. and Mailey, C., 'Unmanned systems research and development at SPAWAR Systems Center Pacific', Space and Naval Warfare Systems Center, Feb. 2008, p. 10.

3. *Detecting the location of gunfire or other weapon fire.* Another autonomous feature found on military robots (but also possibly on law enforcement robots) is the autonomous detection of gunfire or other weapon fire. This feature is intended to improve protection of human forces on the ground. The RedOWL optional sensor of the 510 Packbot—a ground robot developed by US-based iRobot (now part of Endeavor Robotics)—is used to locate snipers and mortars; however, the system does not attack these targets, it simply conveys information about the direction and range to forces on the ground. It has a reported accuracy rate of 94 per cent.[21]

4. *Detection of objects of interest in ISR missions.* Most unmanned systems that are currently in use in ISR missions have no on-board ability to analyse the intelligence information they collect; all the data that is captured must be monitored and assessed by human analysts off-board—a set-up that is labour-intensive and requires a robust and reliable communication broadband.[22] Hence, an emerging feature among new-generation unmanned systems that are specifically aimed at ISR missions is the inclusion of image data processing software that permits systems to autonomously find information of interest and relay that information to human analysts for disambiguation.[23] As in the case of targeting, the actual perceptual intelligence of these systems remains relatively rudimentary. It is limited to the detection of large objects. A case in point is the ScanEagle, a small UAS developed by Boeing, which is reportedly able to detect autonomously objects of interest on the sea surface. The system relies on a Visual Detection and Ranging (ViDAR) sensor that can only differentiate between water and non-water; the system can only detect non-aqueous objects and cannot discriminate between such objects.[24] There is a clear interest in improving the object recognition ability of such systems, notably using recent advances in machine learning but, as previously discussed, progress in this area has been limited to the R&D level.[25]

It should be stressed that, as of October 2017, no deployed system is capable of producing an advanced situation analysis that would, for instance, enable the detection of suspect human behaviour on the battlefield. Computer vision technology has made great strides in biometrics and object recognition, but it still struggles to infer abstract meanings from images, video footage or real-life situations.[26] Cutting-edge computer vision systems can recognize some simple human actions such as walking, running and hand waving, but they are unable to determine the intentions behind these actions (e.g. why a person might be running). Making computers capable of understanding complex actions and goal-oriented activity continues to be a fundamental research problem. In other words, it remains challenging using the currently available technology to develop autonomous image processing systems able to detect potential human enemy targets based on the behaviour or actions of those targets. There are, however, a number of research projects that aim to equip weapon systems and unarmed unmanned military systems with such capabilities. One notable illustration of this is the US Office for Naval Research's project entitled Automated Image Understanding Thrust, which is attempting to develop techniques to infer intentions and threats in surveillance imagery.[27]

---

[21] Dunnigan, J., 'A chip on the shoulders that kills snipers', Strategy Page, 13 Dec. 2010.

[22] Tucker, P., 'Robots won't be taking these military jobs anytime soon', *Defense One*, 22 June 2015.

[23] US Department of Defense (DOD), Defense Science Board, *Task Force Report: Role of Autonomy in DOD Systems* (DOD: Washington, DC, 2012); and Scheidt, D., 'Organic persistent intelligence, surveillance and reconnaissance', *John Hopkins APL Technical Digest*, vol. 31, no. 2 (2012).

[24] 'Watch: Insitu launches ScanEagle upgrades', iHLS, 1 June 2016, <http://i-hls.com/archives/70040>.

[25] Stevenson, B., 'Boeing to test high levels of unmanned autonomy', FlightGlobal, 24 Oct. 2016.

[26] Karpathy, A., 'The state of computer vision and AI: we are really, really far away', Andrej Karpathy Blog, 22 Oct. 2012.

[27] US Office of Naval Research, 'Computational methods for decision making program', [n.d.].

Intelligence data generation

A second category of autonomous features worth mentioning in the context of intelligence relates to the ability of systems to collect data and generate intelligence information. Three specific functions could be highlighted.

1. *Map generation.* One feature that is particularly common for underwater systems and is emerging in the latest generation of surveillance and reconnaissance aerial systems is the ability to autonomously generate details about the environment. One notable example is Shield AI, a tactical UAS currently under development by Shield AI (USA). The Shield AI can generate three-dimensional (3-D) maps using cameras, lasers, and inertial and ultrasonic sensors. It requires no human piloting or GPS.[28]

2. *Threat assessment.* Another function that is generally found in defensive systems is automated threat assessment. In this case, the system is programmed to evaluate the level of risk based on predefined criteria. Israel's Iron Dome missile defence system, for instance, can assess where an incoming missile will detonate, and suggest countermeasures accordingly—if the incoming missile is not threatening particular military or civilian assets, the system may suggest not to attack it to save munitions.[29] The intrusion detection assessment system of the aforementioned MDARS also includes an algorithm that allows it to generate a 'threat score' each time an intrusion is detected.

3. *Big data analytics.* One other development in autonomy for intelligence worth noting—although it does not take place on-board weapon systems due to the high demand in computing power—is the use of big data analytics for pattern recognition in intelligence data. Advances in machine-learning algorithms allow the military command to find correlations in large and potentially heterogeneous sets of intelligence data.[30] One recent illustration of this capability is the alleged use of machine-learning algorithms by the USA to search the Global System for Mobile (GSM) communication metadata of 55 million mobile phone users in Pakistan. The algorithm was trained to track down couriers carrying messages between al-Qaeda members.[31] It reportedly eventually helped the USA to locate the residence of Osama bin Laden.

Human–machine command-and-control relationship

The actual nature of the command-and-control relationship between the systems that include the functions presented above depends very much on the nature of the functions, the systems and the mission circumstances. The majority of the functions are not safety critical, hence they generally do not require direct supervision. The two notable exceptions are the detection of explosive devices and threat assessment in defensive systems, as both of these functions are related to the use of kinetic force.

VI. Autonomy for interoperability

A fourth notable application area of autonomy is interoperability, defined here as the ability of military equipment and troops to operate in conjunction with each other. The SIPRI dataset includes 55 military systems capable of executing tasks or a mission in cooperation with other systems (machine–machine teaming) or combat troops (human–machine teaming).

---

[28] Tucker, P., 'Special operators are getting a new autonomous tactical drone', *Defense One*, 11 Sep. 2016.

[29] Raytheon, 'Iron Dome weapon system: defence against rockets, artillery and mortars', [n.d.].

[30] US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016).

[31] Robbin, M., 'Has a rampaging AI algorithm really killed thousands in Pakistan?', *The Guardian*, 18 Feb. 2016.

Machine–machine teaming

Machine–machine teaming can take different forms. The most basic expression of that capability is information sharing: systems are connected and can communicate with each other to share sensor or intelligence information, including sometimes target information, but each pursues its own goals. Collaborative autonomy is a more advanced model, where multiple systems are capable of coordinating their actions to achieve a common goal. This requires a software architecture that commands and controls the actions of the 'collective system' or the 'system of systems' as a whole. The system of systems can be composed of heterogeneous systems—for example a mix of UASs and unmanned surface systems—or a 'swarm' of identical, and generally relatively small and low-cost systems, which will then operate as a coherent entity. In the former case, the software architecture predetermines the specific role of each system within the larger group. In the latter case, the software architecture is designed to govern a collective behaviour, and achieve effects that each system could not achieve individually.[32]

Functions and capabilities

Machine–machine teaming is still a nascent capability. In already deployed systems, it takes only a primitive form: it is limited to basic exchange or relaying of sensory data and target information. Collaborative autonomy is a capability that is actively researched but it has not yet reached the point where it can be turned into a viable operational capability. Systems that the SIPRI dataset identified as capable of operating as part of a swarm or another collective system are all under development or still in a demonstration phase. However, many experts foresee that the technological advances are such that collaborative autonomy could become an operational reality in the coming years.[33] The types of collaborative operations that are feasible at the R&D level with the current state of technology include the following.

1. *Coordinated mobility*. The most fundamental and technologically mature form of collaborative operation is coordinated mobility. Making air, land or maritime systems autonomously move in formation is a relatively simple and well-understood operation. Systems simply need to keep a predetermined distance from each other. Similar to autonomous navigation, the main technical difficulty is the nature of the environment. Coordinated mobility is much easier to achieve in the air and sea domains than the land domain because these are inherently less complex. There are an increasing number of systems under development that feature this capability. One notable example is the UTAP-22, a UAS developed by Kratos (USA). The systems, which reached operational demonstration phase in 2016, showed that in test flights they could fly in formation with other systems in several different scenarios and configurations, including as a formation lead, lead–follow or wingman.[34]

2. *Coordinated ISR operation over a large geographical area*. One of the most anticipated near-term applications of collaborative autonomy is the deployment of small, low-cost and 'disposable' UASs for ISR missions. There are numerous R&D projects

---

[32] Tan, Y. and Zheng, Z., 'Research advance in swarm robotics', *Defense Technology*, vol. 9, no. 1 (Mar. 2013), pp. 18–39.

[33] The most anticipated near-term application of that capability includes the deployment of micro and small UASs for ISR missions in cluttered environments, and swarms of unmanned systems for vehicle protection and anti-access, and anti-access and area denial. Arquilla, J. and Ronfeldt, R., *Swarming and the Future of Conflict* (RAND Corporation: Santa Monica, CA, 2005); Scharre, P., *Robotics on the Battlefield Part II: The Coming Swarm* (Centre for a New American Security: Washington, DC, Oct. 2014); Golson, J., 'The Navy's developing little autonomous boats to defend its ships', *Wired*, 10 June 2014; and 'US military's new swarm of mini drones', *Defense News*, 17 May 2015.

[34] Kratos, 'Kratos' third UTAP-22 flight exceeds objectives, successfully performing all primary and alternate test points', Press release, 21 Dec. 2015.

investigating that possibility, some of which have already reached demonstration phase. One that is worth mentioning is the Perdix autonomous swarm project that the Strategic Capability Office of the US DOD conducted in partnership with the US Naval Air System Command, and which culminated in 2016 with the successful testing of a surveillance operation involving a swarm of 103 micro-UASs (Perdix drones, manufactured by the Lincoln Laboratory of the Massachusetts Institute of Technology, MIT). The test received significant media coverage, partly because it was one of the largest decentralized swarm operations ever launched in an open environment, but also because it demonstrated that low-cost and disposable micro-UASs could be safely launched from high altitude and at high speed by combat aircraft (on this occasion they were launched at a speed of Mach 0.6 from three F/A-18 Super Hornets).[35]

3. *Anti-access/area-denial (A2/AD) manoeuvres.* Another foreseeable application of collaborative autonomy is for perimeter surveillance and protection. The CARACaS (Control Architecture for Robotic Agent Command and Sensing) project led by the US Office of Naval Research provides a good illustration of the progress that has been achieved towards that specific end. The CARACaS project aims to develop a control architecture that would allow a boat swarm fleet to conduct complex surveillance and security manoeuvres autonomously, including (*a*) encircling surface vessels for control or access denial; and (*b*) detecting enemy vessels that present potential threats based on their behaviour, and autonomously assigning drones to trail or track them.[36] The achievements of the project were successively showcased in real-life demonstration exercises in 2014 and 2016. During an exercise in 2016, the swarm showed that it could collaborate to identify, surround and harass an enemy vessel with little human supervision—the 13-boat fleet was supervised by only one human operator.[37]

4. *Distributed attacks.* One other model of collaborative autonomy that is being investigated at the R&D level is the development of a control architecture in which weapon systems could autonomously distribute targets among themselves. One development project worth mentioning here is the Enhanced Awareness and Forward Operating Capability (EA Focus) for the Unmanned Aerial System Project sponsored by the Defence Science and Technology Laboratory (UK) and jointly carried out by Blue Bear Systems Research (UK) and Deep Vision (Canada). Its objective is to create multi-layer systems of UASs, whereby a higher-level UAS, which would act as the central authority, would identify a target and then hand it over to a lower-level UAS such as a nano- or micro-UAS.[38]

It should be noted that if the aforementioned R&D projects show that it is possible to develop workable control algorithms for various types of autonomous mission by systems of systems, there are still significant important technical obstacles to their formal adoption by the armed forces. To be deemed viable from an operational perspective, these systems need to demonstrate not only that they carry out the mission they are supposed to well, but also that they are sufficiently sophisticated to work in situations where an intelligent adversary might be capable of defeating them using, for instance, spoofing techniques, decoys, or cyber or electronic attacks. This is a much more complex challenge (this point is discussed further in chapter 4).

---
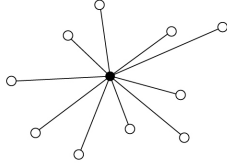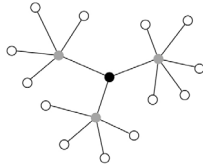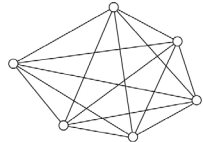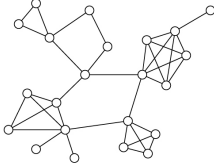
[35] US Department of Defense, 'Department of Defense announces successful micro-drone demonstration', Press Release NR-008-17, 9 Jan. 2017; and Houck, C., 'These swarming drones launch from a fighter jet's flare dispensers', *Defense One*, 9 Sep. 2016.

[36] Tucker, P., 'Inside the Navy's secret swarm robot experiment', *Defense One*, 5 Oct. 2014; and Tucker, P., 'The US Navy's autonomous swarm boats can now decide what to attack', *Defense One*, 14 Dec. 2016.

[37] Tucker, 'The US Navy's autonomous swarm boats can now decide what to attack' (note 36).

[38] Blue Bear, 'Deep Vision and Blue Bear partner on ASUR project', News report, [n.d.].

Table 3.1. Command-and-control structure for collective systems, including swarms

Command-and-control structure	Description	Pros	Cons
Centralized control (centralized)	Individual elements communicate with a centralized planner that coordinates all tasks	Can find an optimal or 'good enough' solution quickly	Require high bandwidth to transmit data to centralized sources and send instructions back to the swarm Vulnerable to communication disruption
Hierarchical coordination (centralized)	Individual elements are controlled by 'squad'-level agents that are, in turn, controlled by a higher-level controller		
Coordination by consensus (decentralized)	All elements of the systems communicate with one another and use 'voting' or auction-based methods to converge to a solution	Can find solutions to complex problems Can work with low bandwidth between the different elements	Finding the optimal solution may take multiple iterations, and, hence, time
Emergent coordination (decentralized)	Coordination arises by individual swarm elements reacting to one another, like an animal swarm	Can work with no direct communication between elements, hence immune to direct communication jamming	

Source: Scharre, P., *Robotics on the Battlefield Part II: The Coming Swarm* (Centre for a New American Security: Washington, DC, Oct. 2014), p. 39.

Human–machine command-and-control relationship

How to exercise meaningful command-and-control over an autonomous system of systems—and a swarm in particular—remains a nascent area of research. Systems of systems can be controlled in multiple ways. The systems can be controlled directly by a human operator who sends commands to (*a*) one system which then distributes them to the rest of the network (centralized control); or (*b*) the system of systems as a whole (decentralized control).[39] Current R&D projects that investigate the operational benefits and capabilities of collective systems have assessed both options. Each has pros and cons depending on the type of systems of systems and the type of mission (see table 3.1 for a more detailed presentation of command-and-control structures for collective systems).

In the specific case of swarming technology, military planners seem to have a preference for decentralized control because it allows the collective systems to be more

---

[39] Scheidt, D. and Schultz, K., 'On optimizing command and control structures', Paper presented at the 16th International Command and Control Research and Technology Symposium (ICCRTS 2011), Quebec, Canada, 21–23 June 2011.

scalable and resilient to the loss of individual units.[40] It is also assumed that as the number of elements in the collective system increases, it becomes more practical for humans to control the system as a whole than manage individual elements.[41]

All the R&D projects that were identified involve the presence of human operators that supervise the operation of (parts of the) collective systems. Details about how the human–machine command-and-control relationship took form in practice were generally very sparse. It was clear, however, that the human–machine ratio (i.e. number of human operators per number of systems), as well as the way human operators interact with the collective systems, would vary depending on the type of systems and type of missions. In a report on the potential of swarming, Scharre outlines some different possibilities that might be examined by the military research community. These included the following.[42]

1. The human operator develops a detailed plan of action that the collective system implements, allowing the system to adapt to changing circumstances; the operator might also intervene to micro-manage the behaviour of elements of the system of systems.

2. The human operator defines a high-level task (such as finding an enemy target) and then lets the collective system find the most optimal way to perform it, either through centralized or decentralized coordination.

3. Should the behaviour of collective systems be too complex to be viably commanded and controlled by one human operator, several human controllers might have to be assigned different responsibilities, such as monitoring the health of the systems, setting high-level goals or approving high-risk actions.

Human–machine teaming

Human–machine teaming can be interpreted in many different ways. For the SIPRI dataset, it is specifically understood as an ability of weapon systems to work independently alongside humans towards the execution of a specific mission. This excludes traditional approaches to 'teleoperation' or 'remote presence', where humans perceive the world and execute an action through the systems.

Functions and capabilities

Like machine–machine teaming, technological developments in the area of human–machine teaming are still immature. The limitations of AI mean that autonomous systems do not have sufficient situational awareness and decision-making capacity to really work in a peer relationship with humans. The capabilities that SIPRI identified involve only simple manoeuvres. For systems operating in the land domain this would be following human soldiers that wear a radio-emitting device that the robot can track and follow.[43] In the air domain, this would be unmanned systems operating as a loyal wingman for a manned aircraft and executing relatively straightforward pre-programmed manoeuvres, such as fly in formation, target marking or post-strike assessment, as well as delivering weapon payload when ordered to by the pilot of the manned aircraft.[44]

---

[40] Scharre (note 33).
[41] Scharre (note 33).
[42] Scharre (note 33), p. 40.
[43] Eshel, T., 'Amstaff robot expands capabilities as tactical support UGV', Defense Update, 31 Oct. 2011.
[44] United States Air Force (USAF), *RPA Vector: Vision and Enabling Concepts 2013–2038* (USAF: Washington, DC, 2013); and USAF, Deputy Chief of Staff for ISR, *Small Unmanned Aircraft Systems (SUAS) Flight Plan: 2016–2036: Bridging the Gap Between Tactical and Strategic* (USAF: Washington, DC, 2016).

It should be noted that human–machine teaming remains very much an experimental capability. There are, for instance, no unmanned systems currently in use that are capable of acting as a loyal wingman. Many of the R&D projects that might deliver realistic options—including the 'autonomous reliable teammate technology' (ART2) project of the US Air Force Research Laboratory, which aims to retrofit old F-16 combat aircraft with autonomous flight capacity to turn them into autonomous unmanned loyal wingmen—have not yet reached the stage of operational demonstration. In the case of the ART2 project, the proof-of-concept demonstration is planned for 2022.[45]

Human–machine and control relationship

One notable technical obstacle to human–machine teaming is the limitation of existing human–machine communication.

The ultimate model of human–machine teaming for many military planners would be a situation where operators could describe and give directions—before and during operations—using natural language, and where robots or autonomous systems could report on their actions or ask for additional input or assistance when they met an unexpected situation. This model is not yet achievable with current technology. Speech-interface technology has developed enormously in recent years (notably thanks to the standardization of voice-commanded digital assistants in smartphones), but it still falls short of what would be expected for a peer-to-peer human–machine communication. State of the art speech interfaces are steadily improving at speech recognition (recognizing words being said) but they still have major difficulties with understanding speech (recognizing what is being discussed).[46] For now, they can only handle simple queries and the fault rate remains fairly high.[47] The technology is yet to reach the point where systems can (*a*) comprehend complex spoken phrases; and (*b*) maintain an understanding of what is being discussed at an abstract level. These are two fundamental requirements for effective communication with humans.

This is the reason why most communication between humans and existing weapon systems continues to occur through visual interfaces such as personal computers or tactile displays. This mode of interaction places a high cognitive load on human vision and is impractical in many operational situations where humans and weapon systems would have to collaborate. Voice command-and-control is used in some existing systems, notably manned combat aircraft and some robotic systems, but due to the limitations of speech recognition technology, it is only used to activate non-critical functions or order very basic actions such as stop or follow-me.[48]

VII. Autonomy for the health management of systems

A fifth and less common application area of autonomy in weapon systems is the health management of systems. SIPRI identified 42 systems which include functions that allow them to manage some aspects of their functioning or survival.[49]

---

[45] Pomerleau, M., 'Loyal wingman program seeks to realize benefits of advancements in autonomy', C4ISR, 19 Oct. 2016.

[46] Knight, W., '10 breakthrough technologies 2016: conversational interfaces', *MIT Technology Review*, vol. 119, no. 2 (Mar./Apr. 2016); and Tuttle, T., 'The future of voice: what's next after Siri, Alexa and Ok Google', *Recode*, 27 Oct. 2015.

[47] Guo, J., 'Google's new artificial intelligence can't understand these sentences. Can you?', *Washington Post*, 18 May 2016.

[48] Voice recognition for command-and-control can be found in the most recent generations of combat aircraft such as the F-16 Vista and F-35 Lightning (Lockheed Martin), the JAS 39 Gripen (Saab), the Mirage (Dassault) and the Eurofighter Typhoon (Airbus). However, it is only used to operate non-critical functions. Schutte, J., 'Researchers fine-tune F-35 pilot-aircraft speech system', Air Force Link, 15 Oct. 2007; Englund, C., 'Speech recognition in the JAS 39 Gripen aircraft: adaptation to speech at different G-loads', Master's Thesis in Speech Technology, Royal Institute of Technology, Stockholm, 11 Mar. 2004; and Eshel (note 43).

[49] It is unclear whether this figure is representative. It is likely that autonomous health management capabilities go largely unreported in newspaper articles or sales brochures dedicated to systems.

Functions and capabilities

Systems have been able to monitor their own status for a long time. Health monitoring has its roots in the discovery of feedback loops more than a century ago.[50] The extent to which existing weapon systems can act upon health status remains limited. The systems that SIPRI identified can only undertake self-recharging/-refuelling and detect and diagnose system faults and failures. Self-maintenance and self-repair remain, at this stage, only experimental capabilities.

Self-recharging/-refuelling

Power management is increasingly a standard feature on robotic systems, be they for civilian or military use. As for other functional application areas, the extent to which systems can execute the recharging and refuelling procedure autonomously is very much determined by the operational conditions. Making systems capable of recharging or refuelling from a fixed dock station is, from an engineering perspective, relatively simple. Many inexpensive domestic robots can do this. By contrast, making a UAS capable of aerial refuelling is much more challenging. Such a procedure remains, for now, reserved to high-end technology demonstrators like Northrop Grumman's X-47B (USA).[51]

Fault detection and diagnosis

There is generally little information on whether or not deployed systems and systems in development include mechanisms for fault detection and identification. One report by the Defense Science Board of the US DOD noted in that regard that such capability is not a focus area in the development of robotics and unmanned vehicles but that the technology exists.[52] Existing systems seem able to detect only very basic problems or problems that are external to the systems themselves. For example, the SW-4 RUAS, an optionally piloted aircraft developed by Leonardo-Finmeccanica (Italy) and Pzl Swidnik (Poland), can only detect problems such as engine failure, vortex ring state and loss of datalink.[53]

Self-repair

Self-repair requires both the ability to self-modify and the availability of new parts or resources to fix broken parts. Existing physical systems still lack these properties. Modular robotics is one area of robotics research that is experimenting with such capability. Modular robots consist of identical robotic modules that can autonomously and dynamically change their aggregate geometric structure to suit different loco-motion, manipulation and sensing tasks. They can self-repair by detecting the failure of a module, ejecting the bad module and replacing it with one of the extra modules.[54]

Human–machine command-and-control relationship

Systems that include health management capabilities are all remotely controlled or supervised by human operators. The use of autonomy for health management is primarily aimed at easing the task of human operators or reducing the workload of maintenance personnel.

[50] Rid, T., *Rise of the Machines: A Cybernetic History* (W. W. Norton and Company: New York, 2016), p. 32.

[51] 'Fueled in flight: X-47 B first to complete autonomous aerial refuelling', *Navair News*, 22 Apr. 2015.

[52] US Department of Defense (DOD), Defense Science Board (note 30), p. 13.

[53] Donald, D. and Dubois, T., 'Pilot is optional for certain missions', AIN Online, 9 Nov. 2015.

[54] Fitch, R., Rus, D. and Vona, M., 'A basis for self-repair robots, using reconfiguring crystal module', Institute of Electrical and Electronics Engineers (IEEE)/Robotics Society of Japan (RSJ) International Conference on Intelligent Robots and Systems 2000 (IROS 2000), Takamatsu, Japan, 30 Oct.–5 Nov. 2000.
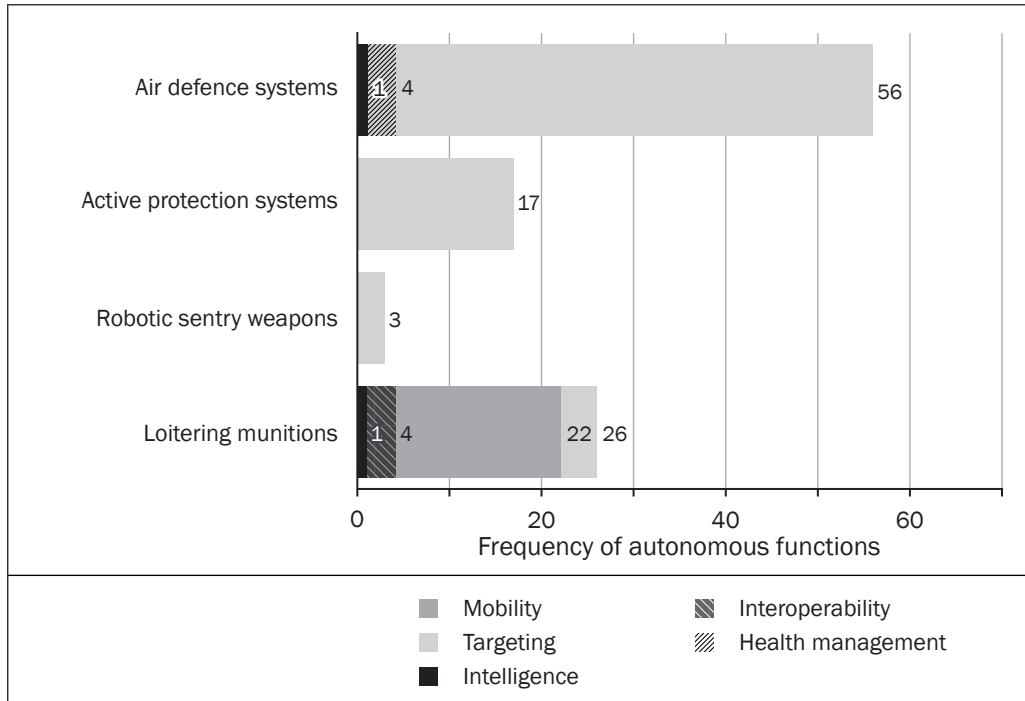
Figure 3.3. Autonomy in 'semi-autonomous' and 'autonomous' weapon systems

Source: SIPRI dataset on autonomy in weapon systems.

VIII. Mapping existing 'semi-autonomous' and 'autonomous' weapon systems

Weapon systems that, once deployed, can independently detect, identify, track, select and potentially attack targets without human involvement do not belong to a distant future; in fact, some have been used for decades. These include some types of (*a*) air defence systems; (*b*) active protection systems; (*c*) robotic sentry weapons; (*d*) guided munitions; and (*e*) loitering weapons (see figure 3.3). The following sections will discuss each category in turn, addressing the same set of questions: What are they? When and where were they developed? Where are they used? How 'autonomous' (from a technical standpoint and from the perspective of the human–machine command-and-control relationship) are they?

IX. Air defence systems

Background

Definition and characteristics

Air defence systems are weapon systems that are specifically designed to nullify or reduce the effectiveness of hostile air action. These can be parsed out in different categories depending on their end use—for example, missile defence systems, anti-aircraft systems and close-in weapon systems (CIWSs). All these systems operate in the same way: they use a radar to detect and track incoming threats (missiles, rockets or enemy aircraft), and a computer-controlled fire system that can prioritize, select and potentially autonomously attack these threats. They can be differentiated based on the following criteria.

1. *The range of engagement.* CIWSs such as the GoalKeeper (Netherlands) or the Phalanx (USA) are designed to defend a limited geographical zone such as the area

around a ship or military base (point defence). Missile defence systems such as the Iron Dome (Israel) can provide protection over a large geographic area such as a border, city or military formation manoeuvre area (area defence) (see figures 3.4 and 3.5).

2. *The types of targets they can engage.* Target types include missiles, rockets or enemy aircraft. There can be differences between target types even within the same subcategory (e.g. CIWS): the Centurion C-RAM, which is a land-based CIWS, can only engage with incoming air projectiles, while the Phalanx, which operates on ships, can also defend against surface vehicles, notably fast attack craft. In this case, the difference seems primarily related to the risks of collateral damage and wrongful target engagements, which are higher on land than at sea.

3. *The type of countermeasures.* The majority of air defence systems use 'hard-kill' measures to defeat incoming threats. They fire missiles or bullets (and in the future laser) at the incoming target. Some systems can also use 'soft-kill' measures whereby electronic countermeasures change the electromagnetic, acoustic or other signature of the targeted system thereby altering the tracking and sensing behaviour of the incoming threat.

History and availability

Automatic air defence systems have existed for decades. The very first automatic air defence system, the Mark 56, was in fact invented during World War II by Bell Laboratories and the MIT Radiation Lab.[55] The oldest model that is still in use is the Soviet S-75 Dvina, introduced in 1957.

The automatic air defence system is also a relatively widespread technology. SIPRI found that at least 89 countries have automatic air defence systems in their arsenal, and 63 countries deployed more than one type of air defence system. The number of countries that develop and manufacture such systems is much smaller, however (see figure 3.6). Countries that have produced the largest variety of automatic air defence systems are the USA (11 different systems) and Russia (8 different systems).

Autonomy and human control

Functions and capabilities

Autonomy in air defence systems has no other function than supporting targeting. The aim is to detect, track, prioritize, select and potentially engage incoming air threats more rapidly and more accurately than a human possibly could. Two examples that highlight the performance of such systems are the S-400 Triumf and the Rapier. The S-400 Triumf, a Russian-made air defence system, can reportedly track more than 300 targets and engage with more than 36 targets simultaneously, at a distance of up to 250 kilometres.[56] The Rapier, which is produced by MBDA and is the UK's primary air defence system, takes 6 seconds from target detection to missile launch.[57]

The technology behind air defence systems has not fundamentally changed since the invention of the Mark 56. The performance of radar and fire control systems has certainly improved but the operating principle remains the same.

Target detection/identification. Air defence systems typically use a radar system to detect potential targets. The radar system emits radio-frequency signals and detects targets based on the return of the reflected signals. Incoming threats are generally identified using two simple criteria: trajectory and velocity. For example, to determine

---

[55] Mindell, D., 'Automation's finest hour: radar and systems integration in World War II', eds A. Hughes and T. Hughes, *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After* (MIT University Press: Boston, MA, 2000).

[56] O'Halloran, J. and Foss, C. (eds), *Jane's Land-Based Air Defence 2010–2011* (IHS Jane's: Coulsdon, 2010).

[57] 'Rapier low-level ground-to-air defense missile system', Army Recognition, [n.d.].
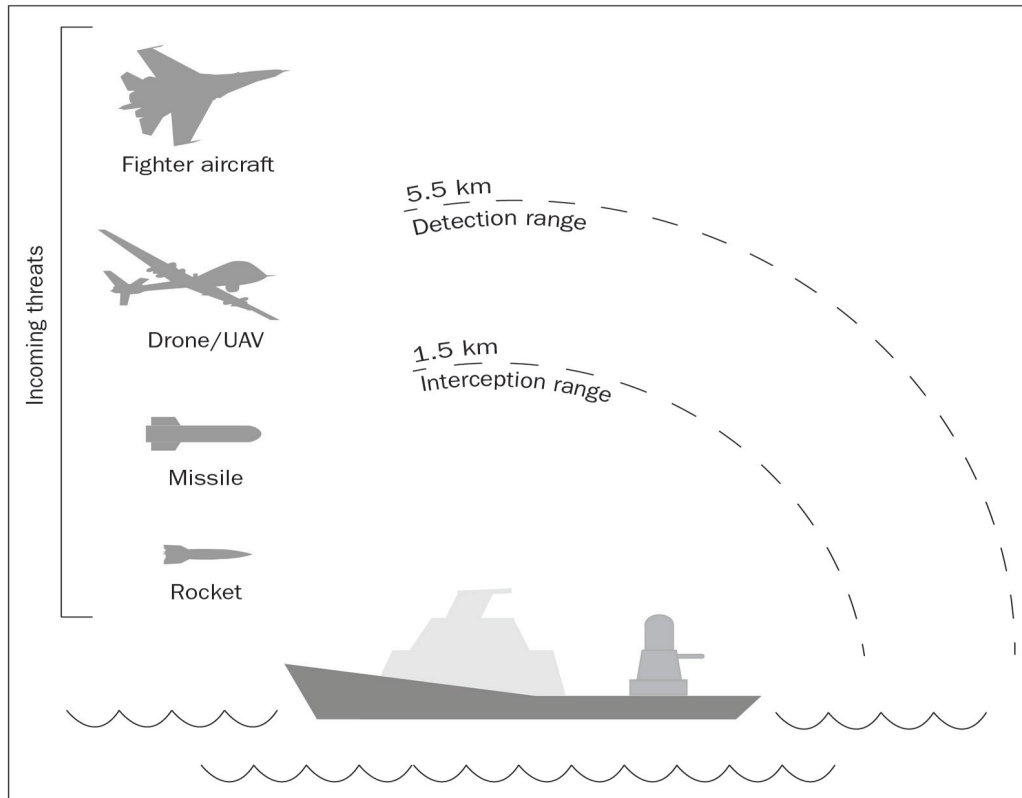
Figure 3.4. Short-range air defence systems: Phalanx close-in weapon system

Source: Friedman, N., *The Naval Institute Guide to World Naval Weapon Systems* (Naval Institute Press: Annapolis, MD, 1989).

whether a detected object represents a threat, the control system of the Phalanx CIWS asks itself the following questions.

1. Is the range of the target increasing or decreasing—that is, is the target approaching the defence point?
2. Will the target hit the relevant defence point if it follows its predicted path?
3. How fast is the target moving?

The Phalanx system is programmed to engage only those targets that are aiming at the defence point and that travel between predefined velocity ranges; thus, the system cannot engage with targets with too high or too low a velocity, although these limits can be adjusted manually.[58]

To avoid the risk of engagement with civilian objects, deployed systems are generally updated during operations with information concerning the flight paths of civilian aircraft and friendly military aircraft, using this data to create engagement zones in which it is safe to carry out an attack. Missile defence systems often include identification, friend or foe (IFF) systems to reduce the risk of friendly fire. The IFF system interrogates the incoming target to determine whether it is a friendly or hostile one (in practice, IFF systems can only positively identify friendly targets).[59] It should be noted that IFF systems are uncommon on CIWSs.

Target prioritization. When several incoming threats are detected, systems typically proceed to a threat assessment to determine which target to engage first. Once again,

[58] Stoner, R., 'R2D2 with attitude: the story of the Phalanx close-in weapon systems (CIWS)', Naval Weapons, 30 Oct. 2009.
[59] 'Identification friend or foe', Global Security, 7 July 2011.
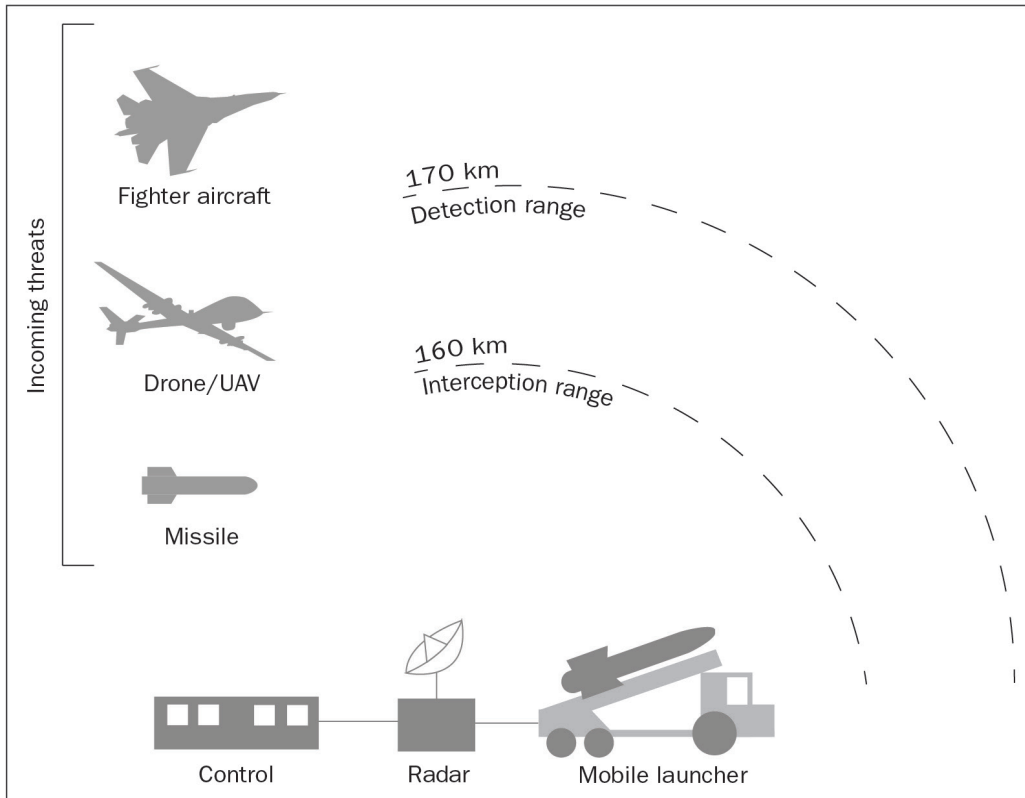
Figure 3.5. Long-range air defence systems: Patriot missile defence system

Source: O'Halloran, J. and Foss, C. (eds), *Jane's Land-Based Air Defence 2010–2011* (IHS Jane's: Coulsdon, 2010).

the assessment is made based on pre-set parameters. In the case of CIWSs, they generally engage the target that represents the most imminent threat to the ship or specific location that they are supposed to protect. For missile defence systems, such as the Iron Dome, the parameter very much depends on the operational scenario, but the assessment works in the same way: the system assesses where the incoming missile or rocket is likely to land and evaluates accordingly whether it is worth deploying countermeasures.[60]

Target engagement. The fire control systems of air defence systems have two modes of engagement: human-in-the-loop and human-on-the-loop. In the human-in-the-loop mode, the operator must always approve the launch, and there are one or several 'decision leverage points' where operators can give input on and control the engagement process. In the human-on-the-loop mode, the system, once activated and within specific parameters, can deploy countermeasures autonomously if it detects a threat. However, the human operator supervises the system's actions and can always abort the attack if necessary.

Human control

Existing systems seem to be governed by different rules of engagement, but information is too scarce to make detailed comparisons between them. It is assumed, however, that CIWSs, because they work as a last line of defence, are likely to operate on a human-on-the-loop mode as standard. In the case of missile defence systems, such as the Patriot system or Iron Dome, selection of the mode of engagement will depend on the nature and imminence of the threats and the operational circumstances. For the Aegis Combat System, which is an integrated combat system that can conduct

[60] Raytheon (note 29).
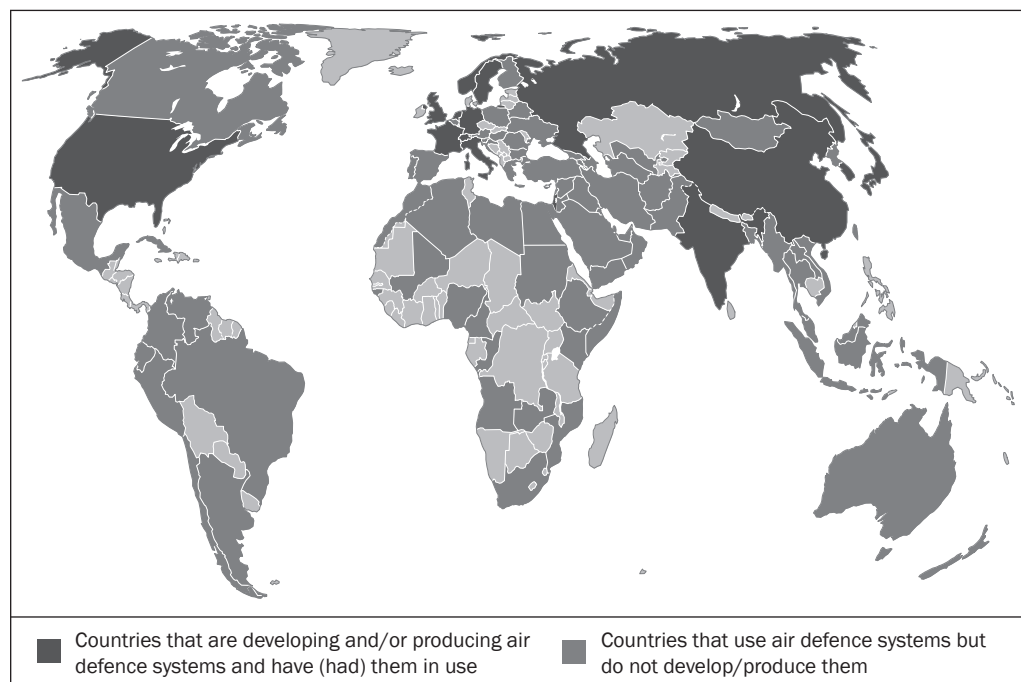
Figure 3.6. Countries with 'automatic' or 'semi-automatic' air defence systems

Source: SIPRI dataset on autonomy in weapon systems.

both defensive and offensive missions, the use of full automatic mode is reserved for self-defence against anti-ship cruise missiles.[61]

All air defence systems are intended to operate under human supervision. The decision to activate the system is retained by a commander who also maintains oversight during the operation and can stop the weapon at any time. However, history has shown that direct human control and supervision is not always a remedy to the problems that emerge with the use of advanced autonomy in the targeting process. One tragic example is the human failure that led to the destruction of a commercial aircraft—Iran Air Flight 655—on 3 July 1988 by the Aegis Combat System on the *USS Vincennes*, a US Navy warship. It was reported that the Aegis Combat System accurately detected Flight 655 and notified the crew that it was emitting signals on a civilian frequency and climbing. However, the crew on the *USS Vincennes* mistook the airliner for an attacking combat aircraft and decided to shoot it down. According to reports, the commanding officers were under stress when assessing the information provided by the Aegis Combat System and had a preconceived notion that the airliner was a combat aircraft descending to attack. As a result, they took the decision to respond, believing that they were defending themselves.[62] This incident illustrates that human supervision is no intrinsic guarantee of reliable use; rather, it may be a source of problems if personnel are not properly trained, or if the information interface provided by the system is too complex for a trained operator to handle in an urgent situation.[63]

---

[61] Ozkan, B. et al., 'Three simulation models of naval air defense', 10th International Command and Control Research and Technology Symposium, McLean, VA, 13–16 June 2005.

[62] Morrison, J. et al., 'Implications of decision making research for decision support and displays', eds J. Cannon-Bowers and E. Salas, *Decision Making Under Stress: Implications for Training and Simulation* (American Psychological Association: Washington, DC, 1998).

[63] United Nations Institute for Disarmament Research (UNIDIR), *Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies* (UNIDIR: Geneva, 2016); Mindell, D., *Our Robots, Ourselves, Robotics and the Myths of Autonomy* (Viking: New York, 2015); and Hawley, J., *Patriot Wars: Automation and the Patriot Air and Missile Defence Systems* (Center for a New American Security: Washington, DC, 2017).
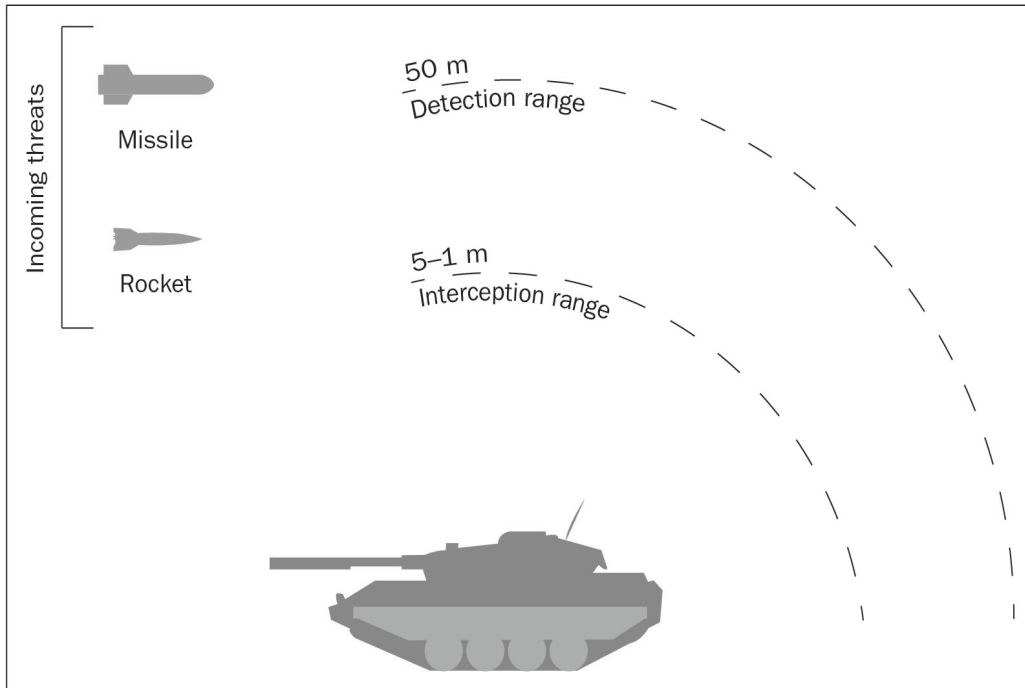
Figure 3.7. Active protection systems: T-80 Arena KAZT

Source: US Department of the Navy (USDN), US Marine Corps, *Antiarmor Operations*, MCWP 3-15.5 (USDN: Washington, DC, 2000).

X. Active protection systems

Background

Definition and characteristics

Active protection systems (APSs) are weapon systems that are designed to protect armoured vehicles against incoming anti-tank missiles or rockets. APSs operate on the same basic principle as air defence systems (see figure 3.7). They combine a sensor system, typically a radar, IR or ultraviolet (UV) detection sensor that detects incoming projectiles, with a fire control system that tracks, evaluates and classifies the incoming threat(s). The systems then launch the appropriate countermeasures (hard-kill or soft-kill) at the optimal location and point in time. Hard-kill countermeasures usually consist of firing rockets or shotgun blasts at the incoming projectiles to (*a*) alter the angle at which they approach the armoured vehicle; (*b*) decrease the chances of penetration; (*c*) trigger a premature or improper initiation of the warhead; or (*d*) destroy the outer shell. Soft-kill measures include using IR jammers, laser spot imitators or radar jammers to prevent the guided munitions from remaining locked onto the vehicle that the APS is meant to protect.

History and availability

Like air defence systems, APSs have been developed, produced and used for several decades. The oldest of the 17 APSs referenced in the SIPRI dataset was introduced in 1978, while research efforts to develop such systems can be traced back to the 1950s. APS technology aims to increase the survivability of armoured vehicle systems and can also contribute to improving their manoeuvrability and deployability, as it reduces the need to equip them with heavier armour.[64]

---

[64] Osborn, K., 'Almost science fiction: new US army tech instantly destroys enemy fire', *National Interest*, 9 May 2016.
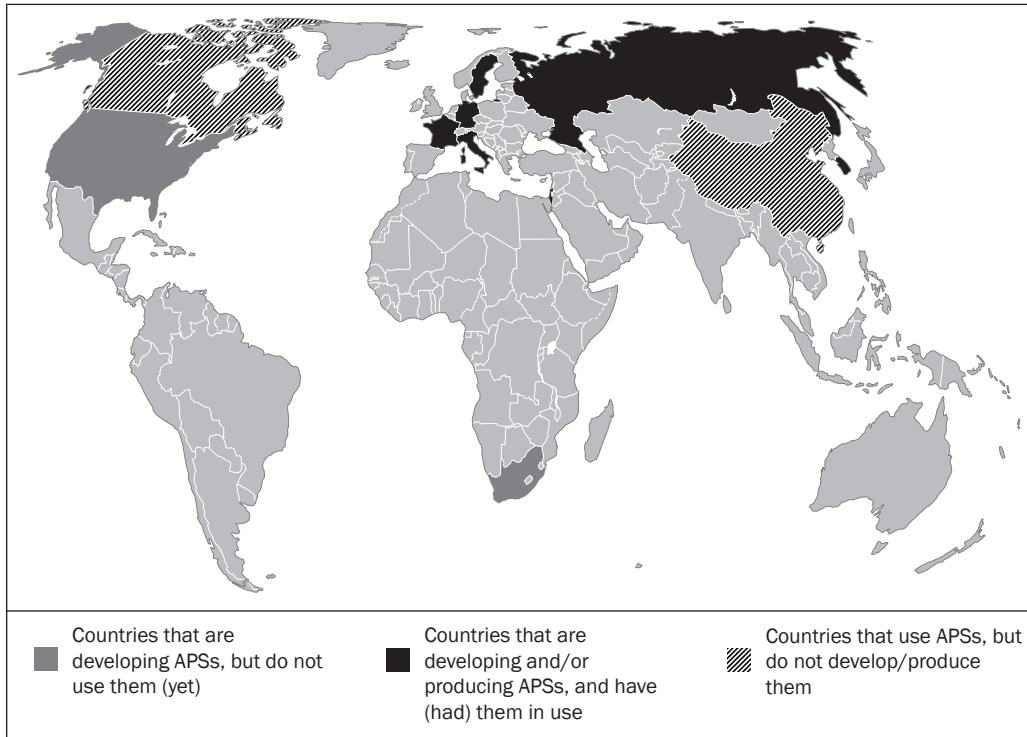
Figure 3.8. Countries with active protection systems (APSs)

Source: SIPRI dataset on autonomy in weapon systems.

There has been growing interest in APSs over the past decade, notably as a result of the proliferation of anti-tank guided missiles (ATGMs) and rocket-propelled grenades (RPGs) to non-state armed groups. The value of APSs against such threats was recently demonstrated during the 2014 Gaza–Israel conflict, during which Israel's armed forces used the Israeli-made Trophy APS on its Merkava tanks. Thanks to the APS, the Israeli armed forces reported far fewer losses of armoured vehicles than during the 2006 Israel–Lebanon war.[65]

Only a limited number of countries develop and produce APSs. The SIPRI dataset includes nine producing countries: France, Germany, Israel, Italy, South Korea, Russia, South Africa, Sweden and the USA (see figure 3.8). Israel and Russia have produced the widest variety of APS models. The number of reported users of APSs is also low. SIPRI identified 10 countries that have developed or acquired off-the-shelf armoured vehicles equipped with APSs (Canada, China, France, Germany, Israel, Italy, Russia, South Korea, Sweden and the USA). There are clear reports indicating that Israel and Russia have used APSs in combat conditions.[66] It is worth noting that the USA, which pioneered the development of APSs, has never formally acquired or fielded such a system. The US military appears to have some concerns as to the extremely short reaction time of APSs, their vulnerability to countermeasures and the potential threat that their use may pose to civilians and friendly forces. Many systems use explosive rounds to defeat incoming projectiles, which creates a high risk of collateral damage and unintended casualties. A recent report noted that the USA is now looking for solutions to overcome these issues and is fast-tracking the acquisition of APSs for combat vehicles; it is likely that this is in response to the rise in accessibility of ATGMs and RPGs, and concerns about Russian military activities in Eastern Europe.[67]

---

[65] Feickert, A., *Army and Marine Corps Active Protection System (APS) Efforts*, Congressional Research Service (CRS) Report for Congress R44598 (US Congress, CRS: Washington, DC, 30 Aug. 2016).

[66] Feickert (note 65).

[67] Feickert (note 65).

Autonomy and human control

Functions and capabilities

The SIPRI dataset identified 17 different APS models (note that only hard-kill APSs were considered): seven are in use, three are still under development, six have been developed but never formally acquired or used, and one has been retired. All these systems operate more or less in the same way, but there are variations in terms of actual capabilities.

Target detection. APSs function by detecting incoming projectiles in various ways (radar, IR or UV detectors). The area that an APS can cover varies. The first operational APS—the Drozd (Russia)—could only cover the forward 60 degrees of a tank's gun turret; the tank crew had to move the turret to change the tank's protective profile. The APS that currently equips Russia's newest tank, the T-14 Armata, reportedly covers only threats that are lateral to the turret, which means that it cannot protect the tank against air-launched guided missiles or projectiles that use a top-attack mode.[68] A newer model, the Afghanit, developed by the KBP Instrument Design Bureau (Russia), which is expected to be mounted on the T-14 Armata in 2017, is reported to include 360 degree active electronically scanned array radar and UV detectors that will provide the vehicles with a hemispheric coverage.[69]

Target identification. The way APSs identify and classify incoming threats is very similar to CIWSs: their sensors evaluate the speed and trajectory of the incoming threats. Some systems, such as Israel's Trophy, include additional advanced features that allow the system to also calculate the shooter's location.[70]

Target prioritization. The ability to simultaneously detect and track multiple targets seems to be a standard feature of APSs. The soon-to-be-deployed Afghanit will supposedly be capable of detecting and tracking up to 40 ground targets and 25 aerial targets.[71] As with CIWSs, the parameters that APSs use to prioritize targets are classified information but are very likely to be a combination of risk variables such as time until impact and nature of the incoming projectiles.

Target engagement. Each of the 17 models of APS identified by SIPRI is designed to execute the entire process of detecting, identifying, tracking and selecting incoming projectiles in complete autonomy. This is due to the fact that APSs are supposed to act within a time frame that is far too short to allow human authorization or supervision of the target engagement. An APS's reaction time is its key performance measure. One recent report noted that an APS with a reaction time of 300 milliseconds would only be able to intercept a typical anti-tank missile if it were launched from at least 400 metres away. By contrast, an APS with a reaction time of 0.5 milliseconds would be able to intercept an anti-tank missile launched from within 10 metres of the vehicle.[72] Finding detailed information about each APS's reaction time has proved difficult. One system that is known to have a reaction time of less than 1 millisecond is the Active Defense System produced by Rheinmetall Defence (Germany), an APS that has been described as a 'fairly mature system'.[73]

[68] De Larrinaga, N., 'Return of the Bear', *Jane's Defence Weekly*, 16 Mar. 2016, p. 27.

[69] Majumbar, D., 'Russia's dangerous T-14 Armata tank: ready for war next year?', *National Interest*, 15 Apr. 2016.

[70] Osborn (note 64).

[71] Eshel, T., 'New Russian armor: first analysis: Armata', Defense Update, 9 May 2015; and 'Russian Armata tank becomes impervious to depleted uranium shells', Sputnik, 22 Sep. 2016.

[72] Feickert (note 65).

[73] Feickert (note 65), p. 16.

Human control

Once activated, APSs are supposed to function in complete autonomy. However, when they are mounted on a manned vehicle, humans inside the vehicle can override or manually shut down the system in case of problems. There is less open-source information available on how the functioning of APSs on unmanned systems is supervised and managed by human operators. Published details on Israel's Trophy APS suggest that when it is mounted on an unmanned system, it will shut down in the event of a communication failure with the remote operator.[74] As APSs have seen only limited use in combat, little is known about the doctrines that govern their use and the effects that their use might have on civilians and friendly forces. Reportedly, the use of the Trophy APS during the 2014 Gaza–Israel conflict did not result in any civilian casualties.[75]

XI. Robotic sentry weapons

Background

Definition and characteristics

Robotic sentry weapons are gun turrets that can automatically detect, track and (potentially) engage targets. They can be used as stationary weapons or be mounted on various types of vehicles. They resemble CIWSs but they use smaller calibre rounds and are usually employed as anti-personnel weapons (see figure 3.9).

History and availability

Robotic sentry weapons remain relatively rare. SIPRI identified only three different models, namely Samsung's SGR-A1 (South Korea), Raphael's Sentry Tech (Israel) and DODAAM's Super aEgis II (South Korea). The development of each of these systems was completed only very recently: 2006 for the SGR-A1, 2007 for the Sentry Tech and 2010 for the Super aEgis II.[76] The Super aEgis II and the Sentry Tech are currently in use; the SGR-A1 is already retired.

Israel and South Korea are the only two countries that currently produce and sell anti-personnel sentry weapons (see figure 3.10). Both countries initiated the development of these systems for border security purposes. Israeli armed forces used the Sentry Tech for protecting Israel's border along the Gaza Strip.[77] South Korea invested in the development of the SGR-A1 and Super aEgis II for potential deployment in the Demilitarized Zone (DMZ)—the buffer zone at the border between North and South Korea. The Korean War Armistice Agreement of 1953 prohibits the deployment of weapons in the zone, so these systems have never been fielded in the DMZ. The South Korean Army has, however, deployed the SGR-A1 on an experimental basis outside South Korea, notably in Afghanistan and Iraq.[78] DODAAM has also reportedly exported its Super aEgis II to a small number of countries, specifically Qatar and the United Arab Emirates (UAE), where it is used to protect air bases and some critical infrastructure.[79]

---

[74] International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems: Implication of Increasing Autonomy in the Critical Functions of Weapons, Expert Meeting, Versoix, Switzerland, 15–16 Mar. 2016*, Meeting report (ICRC: Geneva, Aug. 2016).

[75] International Committee of the Red Cross (note 74).

[76] DODAAM also sells a version of the Super aEgis II, called the 'Athena', which is mounted on an unmanned ground system. DODAAM, 'Combat robot (lethal): Athena', [n.d.].

[77] The South Korean Army has allegedly also used the Athena in combat operations in Iraq, but very little public information exists about this. [Improving the Korean robots], Army Guide, 13 Nov. 2005 (in Russian); and Parkin, S., 'Killer robots: the soldiers that never sleep', BBC, 16 July 2015.

[78] Rabiroff, J., 'Machine gun-toting robots deployed on DMZ', *Stars and Stripes*, 12 July 2010; and Tarantola, A., 'South Korea's auto-turret can kill a man in the dead of night from three clicks', Gizmodo, 29 Oct. 2012.

[79] Parkin (note 77).

Thermal
sensor

Remote reckoning system

12.7 mm calibre
standard

Day colour
camera

Laser range
finder

Azimuth continuous   360°
Elevation   +60° to –20°
Azimuth and elevation rate 60°/sec

Human-sized target detection range
day, 3 km – night, 2.2 km
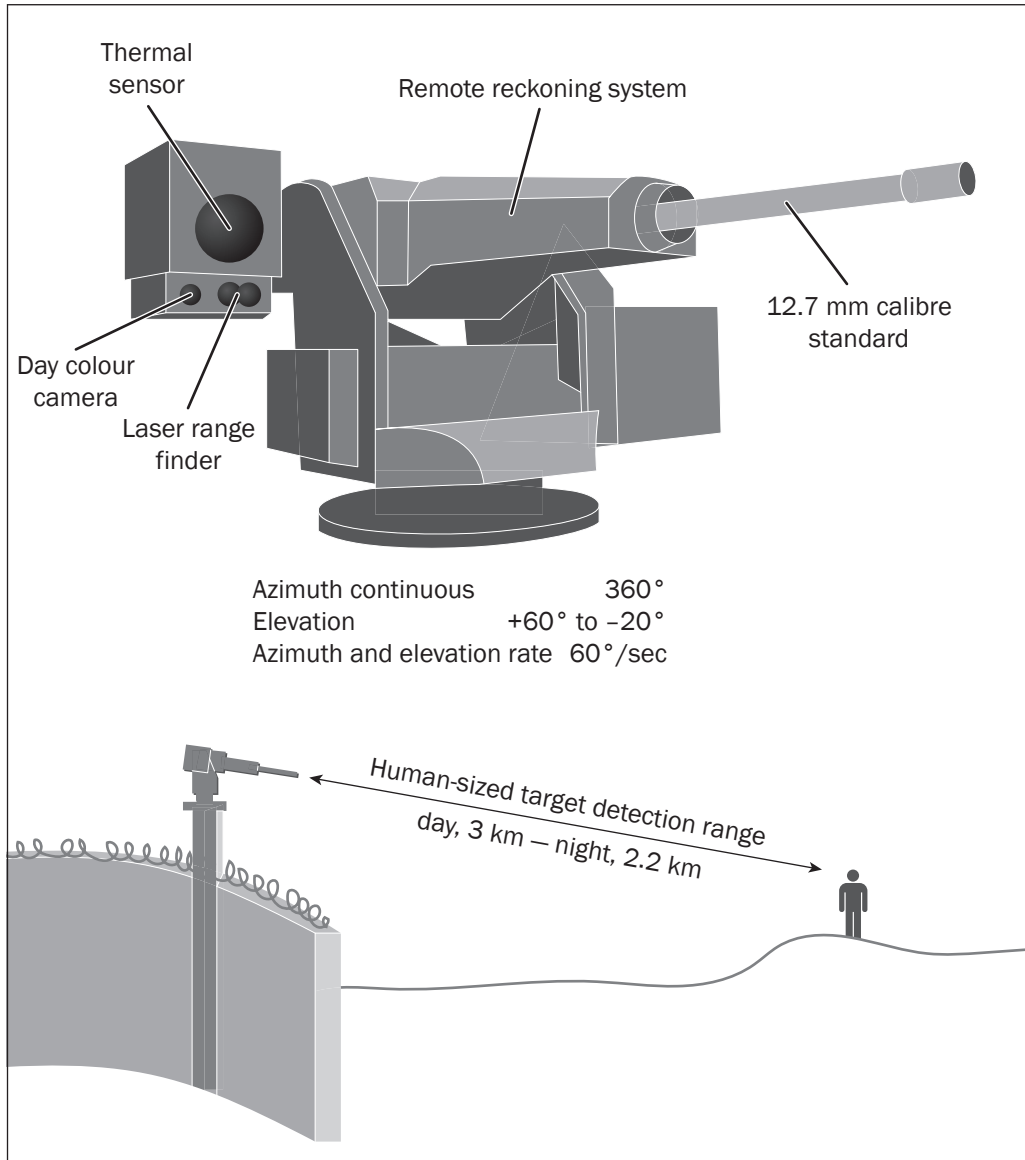
Figure 3.9. Robotic sentry weapons: DODAAM's Super aEgis II

Source: DODAAM, 'Combat robot (lethal): Super aEgis II', [n.d.].

Autonomy and human control

Functions and capabilities

As they are currently employed, robotic sentry weapons might be more accurately described as weaponized autonomous surveillance systems. Autonomy serves primarily to guarantee that they are keeping a sharp and unblinking eye on the perimeters under their protection.

Target detection. All three robotic sentry weapon systems commonly use a combination of digital cameras and IR cameras to detect targets within a relatively large perimeter. The Super aEgis II, for instance, can supposedly detect and lock on to human-sized targets at a distance of up to 2.2 km at night and 3 km in daylight.

Target identification. Robotic sentry weapons recognize targets based chiefly on heat and motion patterns. They are therefore unable to distinguish between 'civilian' and 'military' human targets. They do, however, include some features that allow them to detect more than simple human presence. The SGR-A1 can reportedly recognize surrender motions (arms held high to indicate surrender), while the Super aEgis II can
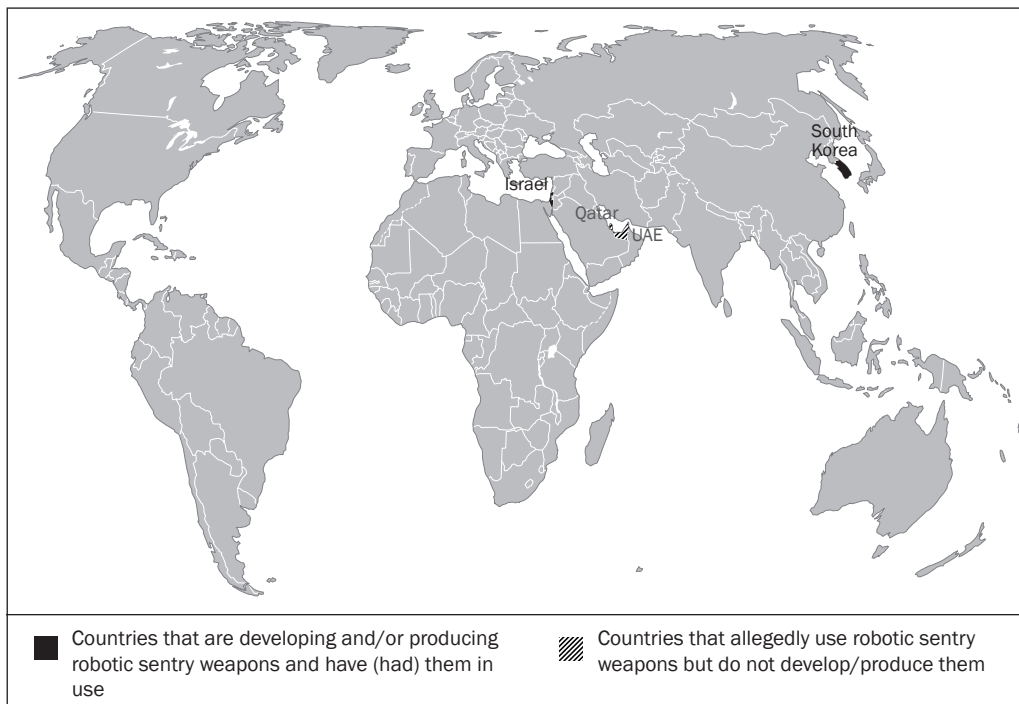
Figure 3.10. Countries with robotic sentry weapons

Source: SIPRI dataset on autonomy in weapon systems.

sense whether a human target is carrying explosives under his or her outer clothing.[80] According to reports, DODAAM is working on a software update that would enable the Super aEgis II to identify whether the target is a 'friend or foe', based on the features of the target's uniform.[81]

Target engagement. The SGR-A1, the Sentry Tech and the Super aEgis II each feature different modes of target engagement. The SGR-A1 and the Sentry Tech reportedly only have the possibility of alerting an operator to the presence of a human in the surveillance zone; at that point, a human operator takes control over the system. The operator then uses the video and audio equipment mounted on the system to establish communication and issue a warning to a person or people that the system has detected. Depending on the target's reaction, the human operator might decide to fire or not to fire the weapon.[82] In its original design, the Super aEgis II was intended to execute all the steps in the process fully autonomously. It was built with a speech interface that allows it to interrogate and warn detected targets. Prospective users of the system reportedly expressed concern that it might make mistakes and requested the introduction of safeguards. DODAAM therefore revised the system to include three modes: human-in-the-loop (the human operator must enter a password to unlock the robot's firing ability and give the manual input that permits the robot to shoot); human-on-the-loop (a human operator supervises and can override the actions of the system); and human-out-of-the-loop (the system is fully autonomous and not supervised in real time by a human operator). According to DODAAM, all the current users have configured the system to human-in-the-loop mode.[83]

---

[80] 'Samsung Techwin SGR-A1 Sentry Guard Robot' (note 12); Parkin (note 77).

[81] Parkin (note 77).

[82] Hughes, R. and Ben-David, A., 'IDF deploys sentry tech on Gaza border', *Jane's Defence Weekly*, 6 June 2007.

[83] Parkin (note 77).

Human control

As they are currently employed, robotic sentry weapons hand over control to a human command-and-control centre once targets are detected. The SGR-A1, for instance, reportedly requires a minimum of two people to operate each robot, one operator and one commander.[84] The question of whether the use of a robotic sentry weapon in a fully autonomous mode would be lawful is still a matter of contention. Some have argued that the system's inability to distinguish between civilian and military targets and make proportionality assessments would make the use of full autonomous mode necessarily unlawful. Others have argued that the legality of the system is very much context-based and that using the system in human-out-of-the-loop mode would not be legally problematic as long as it is deployed in an area where (*a*) it is reasonable to assume there would be no civilian presence; and (*b*) circumstances would make the use of force proportionate (e.g. the DMZ).[85]

XII. Guided munitions

Background

Definition and characteristics

Guided munitions—also called smart bombs or precision-guided munitions—are explosive projectiles that can actively correct for initial-aiming or subsequent errors by homing in on their targets or aim-points after being fired, released or launched.[86] Guided munitions were excluded from the SIPRI mapping exercise primarily for reasons of feasibility. Providing a detailed mapping of existing guided munitions would have been a study in itself.[87] It is also debatable whether guided munitions may be described as autonomous weapon systems, as they are assigned targets in advance by human operators. They only use autonomy to travel to, track or engage the pre-assigned target. However, analysis of their development and use provides some interesting insights into the advance of targeting technology and human control.[88]

Guided munitions encompass a wide range of systems—missiles, torpedoes, sensor-fused munitions and encapsulated torpedo mines—which may feature very different characteristics and properties. With regard to their targeting capacity specifically, they can be categorized depending on the following.

1. *Readjustment capacity.* Some munitions, generally missiles, can be re-targeted or have their flight path adjusted mid-flight. Systems that are not capable of this are called fire-and-forget weapons.

2. *Guidance control system.* Projectiles with *pre-set guidance* have their target and flight path programmed before launch, while those under *remote guidance* receive directions from an operator over wire or radar. Projectiles with *homing guidance* can recalculate their flight path based on calculations on-board. Many projectiles with homing guidance can still receive instructions from the operator over a data link.

---

[84] Rabiroff (note 78).

[85] Brehm, M., *Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems under International Humanitarian Law and Human Rights Law*, Academic Briefing no. 9 (Geneva Academy: Geneva, 2017).

[86] Watts, B. D., *Six Decades of Guided Munitions and Battle Networks* (Centre for Strategic and Budgetary Assessments: Washington, DC, Mar. 2007), pp. ix–18.

[87] An overview of the autonomous capabilities of many guided munitions can be found in Roff, H., 'Dataset: survey of autonomous weapons systems', Arizona State University, Global Security Initiative, Sep. 2016.

[88] Gubrud, M., 'Killer robots and laser-guided bombs: a reply to Horowitz & Scharre', Mark Gubrud's blog, 4 Dec. 2014; and Horowitz, M. and Scharre, P., 'Do killer robots save lives?', *Politico*, 19 Nov. 2014.
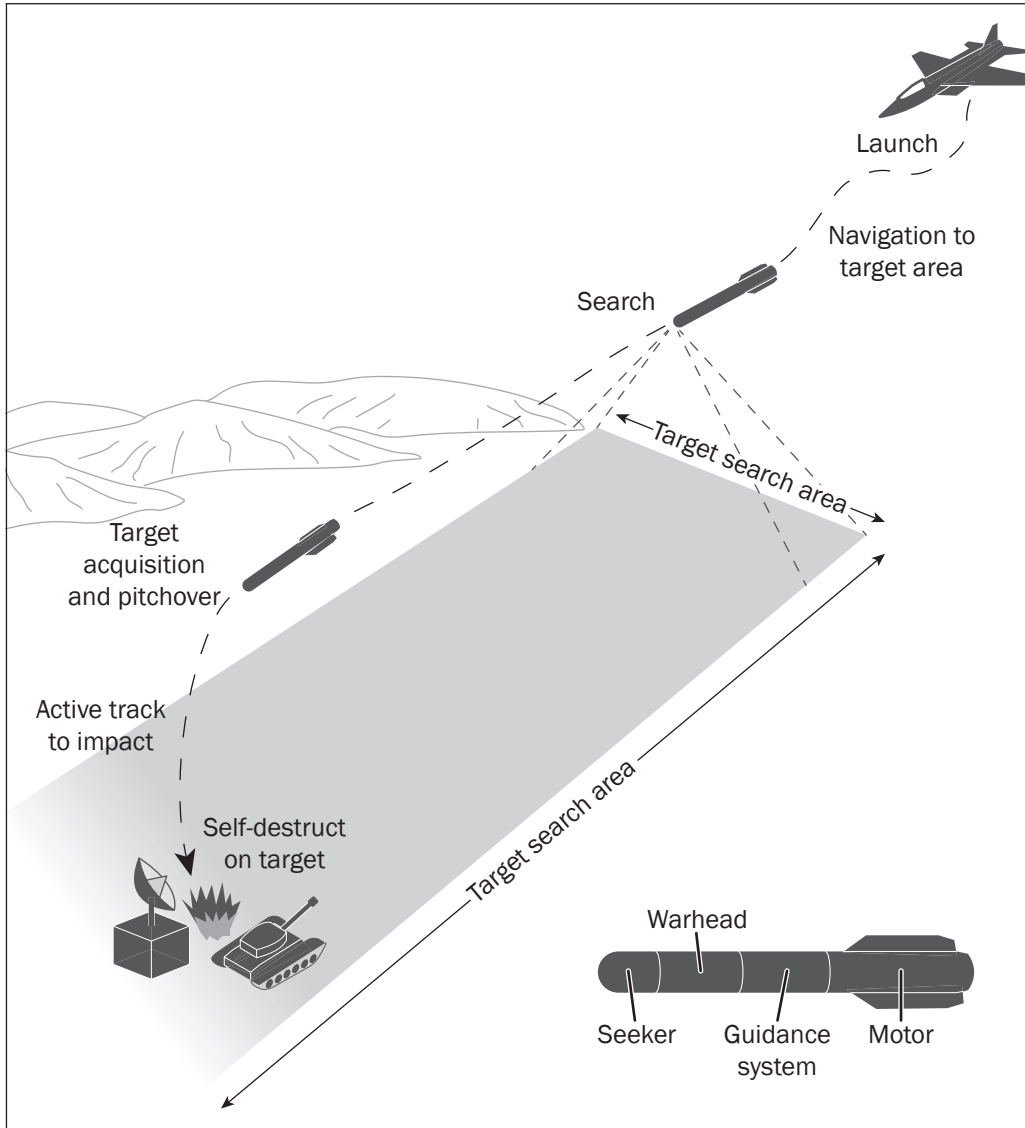
Figure 3.11. Guided munitions: Dual-Mode Brimstone

Source: British Royal Air Force (RAF), 'Brimstone', *Royal Air Force: Aircraft and Weapons* (RAF: 2003), p. 87.

3. *Seekers.* The seekers of guided munitions can be *passive*, searching for the reflection of signals such as noise or heat (IR), or *active*, actively sending out signals to search for (often radar). Active seekers are more accurate, but often have a shorter range, weigh more and are easier to detect by enemies. *Semi-active radar homing* is when a radar dish on the ground sends out a radar wave to the target, and the munition follows the reflection of those waves from the target. Some projectiles only use active seekers in the last leg of the flight shortly before impact; this is known as *terminal guidance*.

4. *Target designation style.* Projectiles ordered to *go-onto-location-in-space* hit the particular geographic location where a target is situated through specific GPS coordinates or by pointing a laser at the target. Projectiles ordered to *go-onto-target* hit a particular target based on its signature, often through IR (heat) or radar. They can lock in on a target during launch or when the target is within line-of-sight. In addition, they can be sent in the direction of the target and lock in after launch when the target is beyond-line-of-sight, if the range of the seeker is insufficient.[89]

---

[89] Horowitz, M. and Scharre, P., *An Introduction to Autonomy in Weapon Systems*, Center for a New American Security (CNAS) Working Paper (CNAS: Washington, DC, 2015), p. 8.

History and availability

Guided munitions have been around for a long time, and the oldest models were introduced during World War II. In the early years, guided munitions were mostly used in the air and underwater, as the capacity to hit moving targets is especially useful when the target can evade in three dimensions. Their use was expanded in the 1970s with the introduction of laser-guided bombs, which were developed to hit very precise locations on the ground. These were used for the first time during the Viet Nam War. The precision and accuracy of guided munitions has significantly improved in recent decades, most notably with the introduction of satellite communication and GPS.[90]

The technology has also become available to a growing number of countries. Many countries now produce short-range guided munitions and, due to the fact that they are relatively affordable, they are used widely, including by non-state actors.[91] Long-range missiles have also proliferated, but to a lesser extent, as they are expensive and require significant infrastructure networks. They are only accessible to major military powers.[92]

Autonomy and human control

As previously mentioned, the vast majority of guided munitions use autonomy only to find, track and hit targets or target locations that have been pre-assigned by humans. In that sense, autonomy does not support the target selection process; it solely supports the execution of the attack.[93]

The few guided munitions with some target selection autonomy include the Long-Range Anti-Ship Missile (LRASM) (USA), Dual-Mode Brimstone (UK) and the Naval Strike Missile/Joint Strike Missile (NSM/JSM) (Norway). These are all missile systems. Only the latter two are further discussed here, as details are scarce on the LRASM.

Functions and capabilities

In contrast to regular guided missiles, the Dual-Mode Brimstone and the NSM/JSM are not assigned a specific target; rather, they are assigned a target area, where they will have the task of finding targets that match a predefined target type (see figure 3.11). They function in this respect very much like loitering weapons (loitering weapons are discussed in more detail in section XIII of this chapter).

Mobility. Before launch, the missiles are assigned a specific area they are allowed to engage. The operator has to assess whether within that area there is a risk of hitting friendly forces or civilians or civilian objects, and program the systems accordingly.[94] The operator also sets parameters such as altitude and minimum time of flight.

Target detection/identification. The missiles are programmed to search for specific target types, automatically rejecting targets that do not fit their assigned signature. The Dual-Mode Brimstone only targets armoured vehicles and reportedly can

---

[90] Their use has increased significantly. While 6% of all munitions in the 1990–91 Gulf War were precision-guided, in 2014 it was stated that 96% of all US strikes in Syria were precision-guided and 100% of all NATO strikes on Libya. Watts (note 86); Thompson, M., 'These are the weapons the U.S. is using to attack ISIS', *Time*, 23 Sep. 2014; and Mueller, K., *Precision and Purpose: Airpower in the Libyan Civil War* (Rand Corporation: Santa Monica, CA, 2015),

[91] Huiss, R., *Proliferation of Precision Strike: Issues for Congress*, Congressional Research Service (CRS) Report for Congress R42539 (US Congress, CRS: Washington, DC, 14 May 2012).

[92] Watts, B., *The Evolution of Precision Strike* (Center for Strategic and Budgetary Assessment: Washington, DC, 2013).

[93] Both fire-and-forget systems and systems with on-board homing guidance are sometimes described as being autonomous. However, 'autonomous' in missile terminology is not the same as 'autonomous' in the context of LAWS. In these missile systems, 'autonomy' refers only to the fact that they find the pre-assigned target themselves.

[94] Kongsberg Gruppen, 'Kongsberg naval and joint strike missiles update: precision strike annual review (PSAR-14)', 13 Mar. 2014, p. 21.

identify buses, cars and buildings as invalid targets through high-resolution radar images.[95] The NSM/JSM targets ships. Kongsberg, the company that develops the system, reports that operators use photographs of ships to semi-automatically create silhouettes and 3-D IR models, which are stored in the NSM/JSM Target Library System, providing a database of potential targets. The NSM/JSM then uses an imaging IR seeker to identify particular features of a ship, allowing it to determine the ship's class. Kongsberg states that there is close to zero probability of the NSM/JSM inadvertently attacking civilian ships.[96]

Target prioritization/engagement. These systems are likely to include parameters that allow them to prioritize between targets, but this is classified information. With regard to target engagement, the NSM/JSM reportedly can decide whether to engage on the 'tactical situation/scene data' following criteria assigned a priori by the operator. These criteria include the zones it can fly and engage in, altitude, minimum time of flight, 'target approach heading' and 'minimum detection by target'. The criteria are reprogrammable to meet different rules of engagement.[97]

The NSM/JSM can also decide how to engage the target, as it can optimize the point of impact on the target and select the warhead fuze profile accordingly.[98] The Dual-Mode Brimstone, when fired in a salvo, can also coordinate with each other to optimize impact, reduce the likelihood that multiple missiles hit the same target if that is undesirable or assign a staggered order to hit targets.[99]

Human control

The Dual-Mode Brimstone is the only guided munition featuring target selection autonomy that is currently operational. It works like a fire-and-forget missile. Once launched, the missile operates in full autonomy; it does not include a human-in-the-loop mode.[100] However, it can be optionally guided with an external laser as well, providing control to the operator if needed. The precise nature of the human–systems command-and-control relationships used by the NSM/JSM and the LRASM remains unclear.

XIII. Loitering weapons

Background

Definition and characteristics

Loitering weapons—also labelled 'loitering munitions' or 'suicide drones'—are a hybrid type of weapon system, which fits a niche between guided munitions and unmanned combat aerial systems (UCASs). Loitering weapons combine the purpose and attack mode of guided munitions (loitering weapons dive-bomb their targets) with the manoeuvrability of UCASs. They can loiter for an extended time to find and strike targets on the ground (see figure 3.12).[101] Their operational utility lies in the fact that they (*a*) are not aimed at a predefined target but rather a target area (in contrast to guided munitions); and (*b*) are disposable. They can conduct offensive and defensive missions that might be deemed dangerous or risky for other types of unmanned systems or

---

[95] 'Brimstone', Think Defence, [n.d].

[96] Kongsberg Gruppen (note 94), pp. 18–22.

[97] Kongsberg Gruppen (note 94), pp. 18–22.

[98] Kongsberg Gruppen (note 94), pp. 18–22.

[99] 'Brimstone', Army Technology, [n.d.].

[100] 'Brimstone' (note 95).

[101] Gilli, M. and Gilli, A., 'The diffusion of drone warfare? Industrial, organizational and infrastructural constraints', *Security Studies*, vol. 25, no. 1 (2016), p. 67.
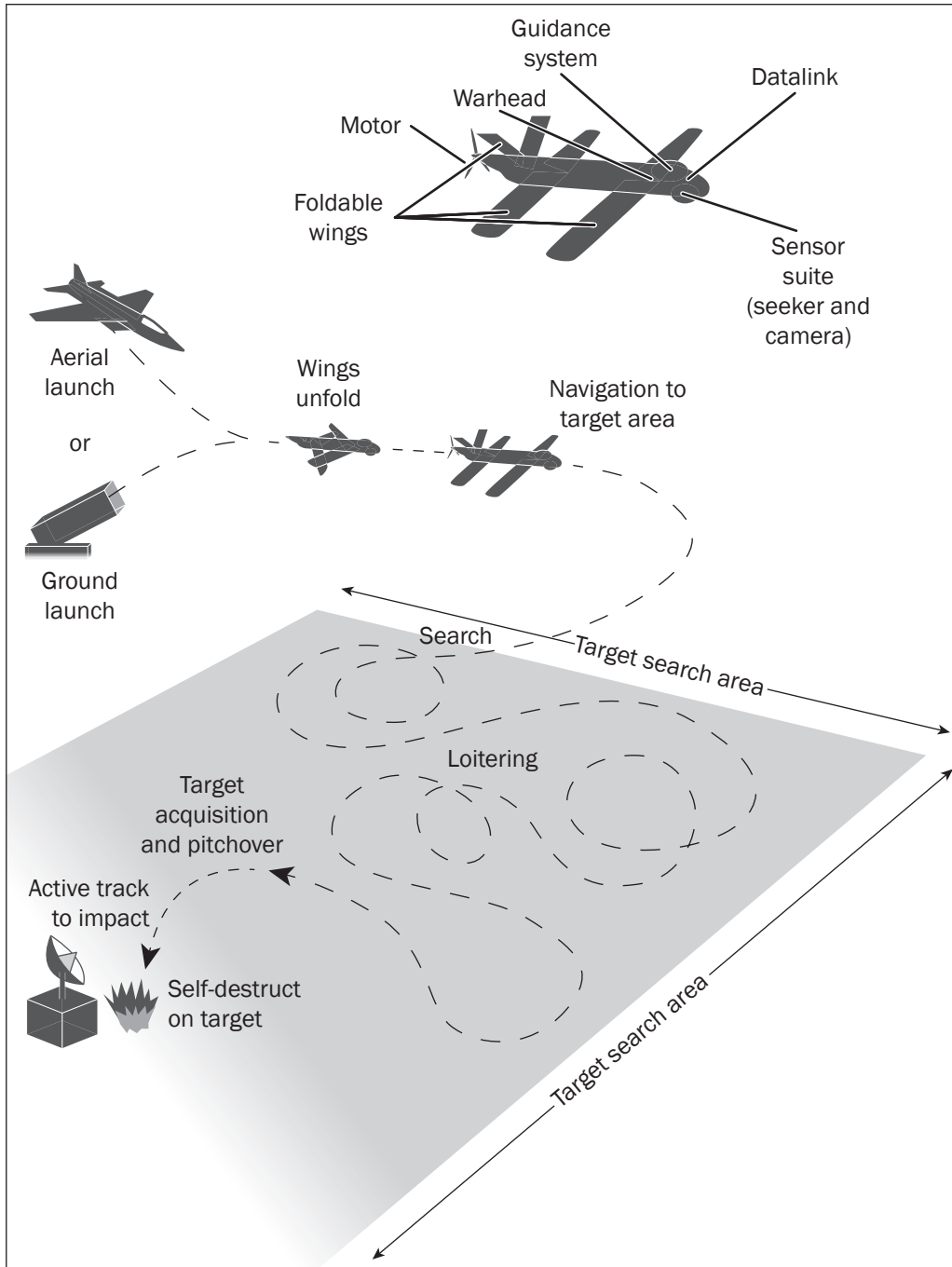
Figure 3.12. Loitering weapons

manned systems, such as suppression of enemy air defences (SEAD), support of artillery, and A2/AD.[102]

Extant loitering weapons come in all sizes and shapes. Variables that fundamentally differentiate them include the following.

---

[102] Examples of these modern guided munitions are the Brimstone and the Long-Range Anti-Ship Missile, while the Tomahawk Anti-Ship Missile operated in this manner in the 1980s. Markoff, J., 'Fearing bombs that can pick whom to kill', *New York Times*, 11 Nov. 2014; British Royal Air Force (RAF), 'Brimstone', *Royal Air Force: Aircraft and Weapons* (RAF: 2003), p. 87; and Scharre, P., 'The opportunity and challenge of autonomous systems', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO: Norfolk, VA, 2015), p. 12.

Countries that are/were developing loitering munitions with autonomous engagement but do not use them (yet)

Countries that are developing and/or producing loitering munitions with autonomous engagement and have (had) them in use

Countries that use loitering munitions with autonomous engagement but do not develop/produce them

Countries that are/were developing loitering munitions without autonomous engagement but do not use them (yet)

Countries that are developing and/or producing loitering munitions without autonomous engagement and have (had) them in use

Countries that use loitering munitions without autonomous engagement but do not develop/produce them

Figure 3.13. Countries with loitering weapons *with* and *without* autonomous engagement

Source: SIPRI dataset on autonomy in weapon systems.

1. *Loitering time.* Loitering weapons vary greatly in their loitering times. Aerovision's Switchblade (USA) can loiter for 10 minutes, while IAI's Harpy Next Generation (Harpy NG) (Israel) can loiter for 9 hours.[103]

2. *Payload/size.* Systems for counter-insurgency are small and have light payloads and short loitering times. Some can be folded and carried in a backpack by troops. The larger systems (up to 4 metres) are the size of missiles, with a payload up to 32 kg.[104] Many are folded into tubes or canisters and launched like a missile; they then unfold their wings mid-flight.

3. *Nature of human–machine command-and-control relationship.* The majority of loitering weapons are remotely operated, but some systems, notably those that are used for SEAD, can work in complete autonomy after launch.

4. *Recoverability.* Some systems, notably larger systems, potentially have the ability to return to base if they do not find any relevant targets or if the mission has to be aborted. The majority of existing models, however, are non-recoverable. They self-destruct if they do not find any relevant targets.

History and availability

The first generation of loitering weapons was developed in the late 1980s and early 1990s. They were typically relatively large systems designed to conduct SEAD and other types of long-range stand-off engagement behind enemy lines. Israel pioneered the development of this technology. SEAD had played an important role during Israel's war with Lebanon in 1982 and Israel had been highly successful in its SEAD operations by using a combination of anti-radiation missiles and decoy UASs.[105] Based on that experience, it developed the Harpy, an anti-radiation loitering weapon, which was capable of finding and attacking radar installations autonomously. In the early 1990s Germany, the UK and the USA started to develop similar systems. Development was later cancelled in all three countries, which allowed Israel to dominate the international market for this specific type of system.[106] Israel has exported the Harpy and similar concepts to numerous countries.

In the late 2000s and early 2010s a new generation of loitering weapons emerged, following the USA's military intervention in Iraq and Afghanistan. Most usually, they are small- or medium-sized (up to 1.5 metres) models for use in counter-insurgency missions as artillery or close-air support. The smallest model, also known as squad-level loitering weapons, can be carried in a soldier's backpack and deployed very quickly.[107] The USA pioneered the development of these systems and—together with Israel—dominates the market. However, similar systems are now also being developed by a number of countries with mid-sized defence industries such as Iran, Poland and Turkey (see figure 3.13).[108]

Autonomy and human control

The large majority of loitering weapons operate under remote control. The SIPRI dataset identified only four operational systems that can find, track and attack targets in complete autonomy once launched: the Orbiter 1K 'Kingfisher', the Harpy, the Harop

---

[103] Systems considered for the dataset all have a loitering time above 30 minutes. This excludes systems commonly considered loitering weapons such as the Switchblade.

[104] Gettinger, D. and Michel, A. H., 'Loitering munitions in focus', Center for the Study of the Drone at Bard College, 10 Feb. 2017.

[105] Sanders, R., 'UAVs: an Israeli military innovation', *Joint Forces Quarterly* (Winter 2003), p. 115.

[106] Gilli and Gilli (note 101), pp. 50–84.

[107] Rapaport, A., 'Loitering munitions alter the battlefield', Israel Defence, 30 June 2016.

[108] Gettinger and Michel (note 104).

and the Harpy NG (all from Israel).[109] As previously noted, Germany, the UK and the USA all started development on loitering weapons with a fully autonomous engagement mode. Examples of these systems include (*a*) the Low Cost Autonomous Attack System (LOCAAS) (USA); (*b*) the Non-Line-of-Sight Launch System (NLOS-LS) (USA); (*c*) the Taifun/TARES (Germany); and (*d*) the Battlefield Loitering Artillery Direct Effect (BLADE) (UK).[110]

None of these systems went beyond the R&D phase. Besides technical and cost issues, a key reason for the cancellation of these programmes was the controversy around the use of autonomy for targeting. The US Air Force, for instance, was reportedly reluctant to have a weapon system that it could not control at all times.[111]

Functions, capabilities and human control

The Harpy, which is the oldest system, operates in complete autonomy. The Harop and the Harpy NG, which are upgrades of the Harpy, as well as the Orbiter 1K, include both a human-in-the-loop and fully autonomous mode.[112] However, the fully autonomous mode seems to be reserved only for SEAD missions.[113] In such circumstances, they operate very much like an anti-radiation missile.

Once launched, the loitering weapon flies to the predetermined geographical area in which it is allowed to engage, using GPS coordinates or pre-programmed flight routes. Upon arrival, it activates its anti-radar seeker to search for and locate potential targets. It may use pre-programmed rules to prioritize between targets (the Harpy NG, for instance, can operate with multiple pre-programmed scenarios). If it cannot find the prioritized targets, it is supposed to move on to, and engage with, secondary targets.[114]

The human-in-the-loop mode seems to be preferred for operations against high-value targets such as armoured vehicles. In such cases, loitering weapons use optical and IR sensors to search for, locate and monitor predefined target types. A human operator supervises the system and retains the ability to abort the attack up until a few seconds before impact. In those circumstances, the operator may order the system to self-destruct, return to a loitering mode or, in some cases, return to base.

The Harop was recently used in armed conflict. Azerbaijan's armed forces used the system in Nagorno–Karabakh in April 2016 to hit six Armenian military targets, including a bus full of volunteers, artillery systems, air defence systems and a military runway. Azerbaijan's armed forces reportedly used the human-in-the-loop mode.[115]

XIV. Conclusions

The main conclusion from this extensive review of the state of autonomy in weapon systems is that autonomy is already a reality of weapon system development and use. There are many concrete examples that CCW delegates could use to bring more focus to the debate on the legal and ethical challenges posed by autonomy. Most of the

---

[109] Scharre (note 102), p. 12.

[110] Van Blyenburgh, P., 'UAVs: current situation and considerations for the way forward', Defense Technical Information Centre, Compilation Part Notice ADP010752 (2000); Watts (note 86) p. 212, p. 282; 'TARES unmanned combat air vehicle (UCAV), Germany', Army Technology, [n.d.]; Biass, E., 'Ground stinging drones', *Armada International*, no. 6 (Dec./Jan. 2012); 'Loitering munition capability demonstration: LMCD', Defense Update, 20 Nov. 2015; and Scharre (note 102), p. 12.

[111] Watts (note 86), p. 282.

[112] 'Loitering munitions against disappearing targets', Israel Defense, 13 July 2016; 'Harpy air defense suppression system', Defense Update, [n.d.]; and 'HAROP', Global Security, [n.d.].

[113] Arkin, D., 'Loiter, lock onto a target and launch!', Israel Defense, 29 Nov. 2015; and 'Aeronautics has unveiled a new loitering system', iHLS, 21 May 2015, <http://i-hls.com/archives/62237>.

[114] Zitun, Y., 'The missile that looks like a UAV', Ynet News, 17 Feb. 2016.

[115] 'Harop made 6 precise shots in Karabakh', Azeri Defense, 14 Apr. 2016.

systems and capabilities described above are likely to be deemed unproblematic to the large majority of states and civil society experts. However, it would be interesting if delegates could engage in a general discussion about (*a*) why these weapon systems or capabilities are unproblematic as they are currently used; and (*b*) in what operational circumstances (both in time and space) their use would become problematic from a legal, ethical and operational perspective. Engaging in a scenario exercise of this type could help the CCW community to discuss more openly the legality and acceptability of autonomy in weapon systems, and could also facilitate a discussion on meaningful human control. Possible case studies could include loitering weapons (for existing systems) and swarms of small UASs (for more futuristic systems).

The case of loitering weapons would be interesting to explore because they are offensive weapons that are currently in use. It would be instructive for the CCW discussion to clarify in what circumstances fielding such weapon systems might be deemed (il)legal and morally (un)acceptable. Variables that would need to be discussed in the course of the scenario exercise include (*a*) the nature and complexity of the area of deployment; (*b*) the loitering time; and (*c*) the human–machine command-and-control relationship during the loitering phase. Loitering munitions might also provide some useful historical insight. Some of the models of loitering munitions that were developed in the 1990s—notably the LOCAAS (1990–2007)—were eventually cancelled. Reportedly, this was partly because they were set to operate fully autonomously. It would, therefore, be interesting to learn more about the debate that surrounded their development and cancellation.

The case of a swarm of small UASs would be more thought-provoking because (*a*) it is an emerging capability that has been tested through various R&D projects, but not yet formally deployed in operations; (*b*) it could be used for a variety of missions; and, more importantly, (*c*) it might require a new paradigm in terms of human–machine command-and-control relationships. Thus, not only would the scenario exercise have to review the legality and acceptability of these systems for different types of missions and operational circumstances, but it would also have to take into consideration variations in models of command-and-control for swarm operations.

4. What are the drivers of, and obstacles to, the development of autonomy in weapon systems?

I. Introduction

Making predictions about what types of systems, operational capabilities and concepts might emerge in the near-, middle- and long-term future is not easy. The one certainty is that there are many variables that need be taken into consideration. Technological progress is a central variable but not the only one. Weapon systems are not created in a vacuum. The fact that the feasibility of a technology is demonstrated at the R&D level does not necessarily mean that the military will want or be able to adopt it quickly or easily as there are numerous political, economic, operational, ethical and cultural factors that also come into play.[1]

This chapter aims to (re)connect the discussion on the emergence of LAWS with the current reality of weapon system development. It maps the spectrum of factors that are currently driving, but also limiting, the development and adoption of autonomy in military systems in general and weapon systems in particular. The objective is simple: to help CCW delegates to obtain a more concrete sense of the current speed and trajectory of the development and adoption of autonomy in weapon systems.

The chapter addresses two fundamental questions.

1. To what extent and why are major arms-producing countries interested in further developing autonomy in weapon systems?
2. What are the different factors that could slow down or limit the further increase of autonomy in weapon systems?

The chapter is made up of two main sections. Section II briefly discusses the extent to which major military powers have articulated a strategic reflection on the development of autonomy in weapon systems and maps out the spectrum of arguments that are commonly mobilized to justify the development of autonomy within weapon systems. Section III maps out the variety of technical, political and economic obstacles to further increasing autonomy in weapon systems. A concluding section (section IV) discusses how the key points raised can be filtered into CCW discussions.

II. Mapping the drivers: to what extent and why is the military interested in autonomy?

Autonomy as a part of future strategic capability calculations

Robotics first, autonomy second

A first important observation is that, from what is visible at the policy level, there is not (yet) a declared 'arms race' on autonomy. SIPRI reviewed the most recent (up to March 2017) official defence strategy publications of the 10 largest arms-producing countries—namely the USA, the UK, France, Russia, Italy, Japan, Israel, Germany, South Korea and India—and China.[2] It found that the USA is the only country that has

---

[1] Smith, R., *Military Economics: The Interaction of Power and Money* (Palgrave Macmillan, Basingstoke, 2009), p. 132.
[2] Based on the share of arms sales of companies listed in the SIPRI Top 100 for 2014. The SIPRI Top 100 lists the world's 100 largest arms-producing and military services companies (excluding those based in China). These are ranked by volume of arms sales. While not covered by the SIPRI Top 100 due to the lack of data on arms sales, China is also considered as one of the largest arms-producing countries. SIPRI considers that at least 9 of the 10 major state-owned conglomerates under which the Chinese industry is organized would be listed in the Top 100 if official data

officially specified that the advance of autonomy is a central component of its future strategic capability calculations. While it is publicly known that the development of autonomy is also a major component of Israel's defence strategy, this has not been officially articulated in a strategy document. Other countries, such as the UK, France, Russia, Japan, South Korea and China, have also expressed interest in the topic in official publications to varying degrees. The references to autonomy by these countries are usually in the context of discussions on UASs, but many are now beginning to widen their focus to include other systems.[3]

Each of the arms-producing countries included in the study maintains that robotic technologies (alongside cyber-warfare technologies) will contribute to shaping the future of warfare and that the acquisition of such technologies, particularly UASs, should be a priority. Most countries are still in the early stages of the adoption of robotic technologies. Their views on autonomy in their military doctrines are therefore likely to develop and mature as they increasingly integrate robotic technologies into their arsenals.

The USA currently sets the benchmark for autonomy in the military sphere. It has pushed the boundaries of what is technically feasible over the past decade—thanks to an unmatched level of investment in relevant R&D areas. It is also shaping, through numerous reports and publications, the way experts and other countries are thinking about the potential and limitations of the military application of autonomy, especially among its North Atlantic Treaty Association (NATO) allies. Much of the current debate on autonomy in weapon systems and robotic systems is influenced by US technological achievements and doctrinal visions that have been elaborated and promoted by US publications. This chapter is no exception. The analysis is primarily based on narratives developed by the US DOD or expert commentary on those narratives. Before reviewing in detail the arguments that are typically invoked to motivate the adoption of autonomy, it is useful to put the USA's current interest in autonomy in a historical context and discuss how it is understood by its allies and peer competitors.

The USA's Third Offset Strategy: setting the benchmark for the future of autonomy in the military sphere

Autonomy has been part of US strategic calculations for a long time.[4] The US DOD was already working on the possible applications of AI technology and the development of new doctrinal concepts for the 'automated battlefield' in the 1980s. In 1983 the USA established the Strategic Computing Initiative under the auspices of its Defense Advanced Research Projects Agency (DARPA), a project worth $1 billion that aimed to develop high-performance machine intelligence for military applications.[5] The initiative was terminated in 1993, partly because it appeared that it would not succeed in creating artificial general intelligence (AGI) as originally planned (AGI is discussed in more detail in chapter 5).[6] The USA nonetheless continued to carry out some R&D

---

was available. Fleurant, A. et al., 'The SIPRI Top 100 arms-producing and military services companies, 2014', SIPRI Fact Sheet, Dec. 2015.

[3] Kania, E., 'Chinese advances in unmanned systems and the military applications of artificial intelligence—the PLA's trajectory towards unmanned, "intelligentized" warfare', Testimony before the US–China Economic and Security Review Commission, 23 Feb. 2017; Le Drian, J., 'L'intelligence artificielle: un enjeu de souveraineté nationale' [Artificial intelligence: an issue of national sovereignty], *L'intelligence artificielle: des libertés individuelles à la sécurité nationale* [Artificial intelligence: individual freedoms to national security] (Eurogroup Consulting: Paris, 2017), pp. 11–24; British Ministry of Defence (MOD), *Advantage through Innovation: The Defence Innovation Initiative* (MOD: London, Sep. 2016); and Kashin, V. and Raska, M., *Countering the US Third Offset Strategy: Russian Perspectives, Responses and Challenges*, S. Rajaratnam School of International Studies (RSIS) Policy Report (RSIS: Singapore, Jan. 2017).

[4] Rid, T., *Rise of the Machines: A Cybernetics History* (W. W. Norton & Company: New York, 2016).

[5] Åkersten, I., 'The strategic computing programme', ed. A. M. Din, *Arms and Artificial Intelligence: Weapons and Arms Control Applications of Advanced Computing* (SIPRI/Oxford University Press: Oxford, 1987), p. 93.

[6] The initiative was replaced by the Accelerated Strategic Computing Initiative, which solely focused on

on specialized AI, notably on applications such as pilot assistants and autonomous vehicles.[7] These efforts fitted into the wider ambition of the US DOD in the 1990s to exploit the advance of information and communications technology (ICT) to lead a so-called Revolution in Military Affairs.[8] The US DOD's interest in autonomy continued to grow during the 2000s and early 2010s, mainly driven by its increasing use of robotic systems, especially UASs and ground robots, in US interventions in Iraq and Afghanistan.[9] In the past few years, the US DOD and each branch of the US military have commissioned studies on the potential and limitations of autonomy in unmanned systems, which have been widely commented on in the USA and globally.[10]

In 2015 the USA elevated the issue of autonomy to the highest strategic level with the publication of its new 'Defense Innovation Initiative', which is also referred to as the 'Third Offset Strategy'.[11] Like the first two offset strategies that were introduced during the cold war, the Third Offset Strategy is based on the idea that the USA should seek to leverage emerging and disruptive technologies in innovative ways to offset the advantages of potential adversaries and maintain its strategic superiority.[12] Each of the previous offset strategies had a specific 'technological sauce' as Robert Work, US Deputy Secretary of Defense, puts it.[13] For the First Offset Strategy, it was the miniaturization of nuclear components, which enabled the adoption of tactical nuclear weapons for conventional deterrence. For the Second Offset Strategy, it was the development of digital microprocessors, information technologies, new sensors and stealth, which enabled the USA to develop precision-guided weapons and achieve dominance in conventional warfare. For the Third Offset Strategy, a key component is going to be AI (particularly machine learning) and autonomy.[14] The reasons for this are manifold (and will be presented in the next subsection), but the main rationale is that AI and autonomy could leverage many operational benefits that could allow the US military to improve the strength and cost-effectiveness of its forces. In this way, it would continue to outmatch Russia, China, Iran or North Korea, even if those countries were to catch up with the USA in the development of high-end weapon technologies, such as precision-guided munitions, robotic technology, cyber and electronic warfare capabilities and A2/AD denial technologies.[15]

supercomputing and did not include research on AI. Roland, A. and Shiman, P., *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983-1993* (MIT Press: Cambridge, MA, 2002).

[7] McCorduck, P., *Machines Who Think* (A K Peters/CRC Press: Natick, MA, 2004), pp. 430–31.

[8] Rasmussen, M. V., *Risk Society at War* (Cambridge University Press: Cambridge, 2007), pp. 43–90.

[9] Between 2003 and 2009 it invested $20 billion under the Future Combat Systems programme to develop heterogeneous teams of networked UASs, UGSs and manned vehicles. The programme was cancelled prematurely due to cost overruns. Freedberg, S., 'Total cost to close out cancelled army FCS could top $1 billion', Breaking Defense, 19 June 2012; and The Dwight D. Eisenhower School for National Security and Resource Strategy, *Spring 2014 Industry Study: Final Report: Robotics and Autonomous Systems Industry* (National Defense University: Washington, DC, 2014), pp. 5–6.

[10] These studies include: US Department of Defense (DOD), Defense Science Board, *The Role of Autonomy in DOD Systems: Task Force Report* (US DOD: Washington, DC, 2012); United States Air Force (USAF), *RPA Vector: Vision and Enabling Concepts 2013–2038* (USAF: Washington, DC, 2013); Endsley, M. R., *Autonomous Horizons: System Autonomy in the Air Force: A Path to the Future, Volume 1: Human-Autonomy Teaming* (United States Air Force, Office of the Chief Scientist: Washington, DC, 2015); and US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016).

[11] The Third Offset Strategy is the brainchild of Robert Work, US Deputy Secretary of Defense, who is also the former chief executive officer of the Center for a New American Security, a US think-tank that has been promoting greater adoption of robotic technologies and autonomy by the USA's military for several years.

[12] The first strategy in the 1950s relied on tactical nuclear superiority to neutralize the Soviet Union's numerical advantages in conventional forces. As the Soviet Union achieved parity with the USA in the nuclear realm in the 1960s, a second strategy was adopted in the 1970s. The second strategy was centred on the development of high-tech conventional weapons, including precision-guided weapons and stealth aircraft, that could more accurately strike conventional forces.

[13] US Department of Defense (US DOD), 'Remarks by Deputy Secretary Work on Third Offset Strategy, as delivered by Deputy Secretary of Defense Bob Work', Brussels, Apr. 2016.

[14] US Department of Defense (note 13).

[15] 'A2/AD technologies' is an umbrella term that covers a wide spectrum of kinetic and non-kinetic technologies that serve to protect specific geographic areas against unmanned and manned systems. Freedberg, S., 'People, not tech: DepSecDef work on 3rd Offset, JICSPOC', Breaking Defense, 9 Feb. 2016; and Sadler, B., 'Fast followers, learning

Many commentators have noted that it is too early to assess the impact of the Third Offset Strategy and, in fact, it remains unclear whether it will survive President Donald Trump's administration.[16] Nevertheless, the publication of the Third Offset Strategy has provoked reactions from the USA's allies and also its direct peer competitors (i.e. Russia and China).

The academic and policy conversations on the implications of the Third Offset Strategy for Europe and NATO are still in their very early stages. However, some observers have noted that there is some concern that the Third Offset Strategy will widen the technology and capability gap between the USA and the other members of NATO and disrupt interoperability and industrial competition.[17] It is still unclear whether European military powers, notably France, Germany and the UK, will completely accept the USA's strategic calculation about AI and autonomy. Daniel Fiott, a defence analyst at the European Institute for Security Studies, has noted that these countries are unlikely to fully support the concept, for the same reason that they did not immediately go along with the concept of Revolution in Military Affairs in the 1990s: they have different strategic priorities from the USA, hence different military–technological ambitions.[18] The USA is in the business of projecting power on a global basis and has the strategic imperative to leap ahead of its rivals in East Asia, Eastern Europe and the Middle East. By contrast, European states on the whole are chiefly concerned with Europe's immediate neighbourhood and do not see the need, given the current security landscape, to prioritize high-tech, US-style capabilities. In other words, European states have different strategic requirements and do not have the same incentives as the USA to use the potential of AI and autonomy strategically. It is evident, however, that some capabilities promoted by the Third Offset Strategy would be considered beneficial in the European context, notably the use of autonomous systems for ISR, data processing, cyber-defence or offensive operations in A2/AD exclusion zones or 'bubbles' (this will be further discussed in the next subsection).[19] Furthermore, European states might also decide to invest in autonomy-related R&D to stay competitive in the global arms market.[20]

Neither Russia nor China appear, at this stage, to have made any official comments about the Third Offset Strategy. Key studies based on Russian and Chinese expert literature found that the military in Russia and China have paid great attention to the Third Offset Strategy and have started to formulate possible reactions. One of the studies on Russia reports that it intends to respond in two ways. First, Russia aims to counter the Third Offset Strategy by using the main principle from the USA's own First Offset Strategy—that is, it intends to offset US dominance in conventional warfare through the development of a wide array of strategic and tactical nuclear weapons. Second, Russia aims to develop similar indigenous R&D programmes, although these would be more narrowly focused and on a smaller scale.[21] This matches another analysis of Russian military capabilities, which states that Russia's main priority areas

machines and the Third Offset Strategy', *Joint Force Quarterly*, vol. 83 (Oct. 2016).

[16] Metha, A., 'Pentagon no. 2: how to keep the third offset going in the next administration', *Defense News*, 2 May 2016; and Johnson, T., 'Will the Department of Defense invest in people or technology?', *The Atlantic*, 29 Nov. 2016.

[17] Quencez, M., *The Impossible Transatlantic Discussion on the U.S. Third Offset Strategy*, The German Marshall Fund of the United States (GMF), Policy Brief no. 41 (GMF: Washington, DC, Oct. 2016); Fiott, D., 'A revolution too far? US defence innovation, Europe and NATO's military–technological gap', *Journal of Strategic Studies*, vol. 40 (June 2016); Fiott, D., 'Europe and the Pentagon's Third Offset Strategy', *RUSI Journal*, vol. 161, no. 1 (2016), pp. 26–31; and Simón, L., 'The "Third" US Offset Strategy and Europe's "anti-access" challenge', *Journal of Strategic Studies*, vol. 39, no. 3 (2016), pp. 417–45.

[18] Fiott, 'A revolution too far? US defence innovation, Europe and NATO's military–technological gap' (note 17), p. 3.

[19] Sadler (note 15).

[20] Fiott, 'A revolution too far? US defence innovation, Europe and NATO's military–technological gap' (note 17), p. 16.

[21] Kashin and Raska (note 3), p. 14.

are strategic nuclear deterrence and strategic aerospace defence, after which come military robotics and UASs.[22]

A review of the debate in China on the Third Offset Strategy concludes that Chinese experts think that the strategy is (*a*) a trap to drag China and Russia into harmful technological competition; (*b*) a hoax to cover the USA's weaknesses; or (*c*) a competitive strategy to strengthen US dominance.[23] Some Chinese analysts have started to assess the strategic choice that China should make in response. Their suggestions include improving the management of China's defence industry—the main challenge being that the Chinese defence industry should be able to match leaps in innovation by the USA—and developing key elements of unmanned systems, AI for military applications, and countermeasures to US technologies (e.g. soft-kill measures against unmanned systems).[24] It should be noted that China has been heavily investing in robotics over the past few decades and is believed to be working on a number of weapon systems concepts that resemble those that the US military R&D agencies have developed. However, China reportedly still lags behind the USA in many of the more fundamental technology areas, especially engines for UASs, data links and sensors.[25]

Benefits of autonomy

What is the benefit of increasing autonomy in weapon systems? For US military planners, increasing autonomy offers many advantages that can help the military to overcome a number of operational and economic challenges associated with manned weapon systems.[26]

Operational benefits

The operational benefits of autonomy are well known. In a nutshell, as previously pointed out by Scharre, autonomy provides the possibility of fielding forces with greater speed, agility, accuracy, persistence, reach, coordination and mass.

Speed. One of the most important advantages of autonomy is speed. Autonomy can make weapon systems execute the so-called observe, orient, decide, act (OODA) loop much faster than any human ever could, which explains why autonomy is deemed particularly attractive for time-critical missions or tasks such as air defence (detecting and targeting high-velocity projectiles) and air-to-air combat (between combat aircraft). It is also well suited to cyber-defence (discovering and neutralizing a cyber-attack) and electronic warfare (analysing and countering new enemy signals).[27] In addition, autonomy provides opportunities at the higher command-and-control level as it enables systems to collect, collate and analyse data in ways human operators cannot (be that in terms of complexity, volume or speed). This drastically improves the quality and speed of decision cycles.

Agility. A second and correlated benefit is agility. Autonomy can make weapon systems far more agile from a command-and-control perspective as it reduces the need

[22] Persson, G. et al., *Russian Military Capability in a Ten-year Perspective 2016*, FOI-R-4326-SE (Swedish Defence Research Agency: Stockholm, Dec. 2016), p. 152.

[23] Fan, G., 'A Chinese perspective on the US Third Offset Strategy and possible Chinese responses', Study of Innovation and Technology in China (SITC) Research Brief, 3 Jan. 2017.

[24] Ray, J. et al., *China's Industrial and Military Robotics Development*, Report for the US–China Economic and Security Review Commission (Center for Intelligence Research and Analysis, Defense Group Inc: Vienna, VA, Oct. 2016), pp. 55–56.

[25] Kania (note 3).

[26] Sadler (note 15); US Department of Defense, Defense Science Board (note 10); The Dwight D. Eisenhower School for National Security and Resource Strategy, *Spring 2016 Industry Study: Final Report: Robotics and Autonomous Systems* (National Defense University: Washington, DC, 2016), p. 3; and Krishnan, A., *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate: Burlington, VT, 2009).

[27] Byrnes, M., 'Nightfall: machine autonomy in air-to-air combat', *Air and Space Power Journal*, vol. 28, no. 3 (2014).

Table 4.1. Possible missions for autonomous (weapon) systems according to US strategic documents

Key advantages of autonomy	Type of mission
Speed: speed of light implementation of the OODA loop	Air defence Cyber-defence Electronic warfare
Agility: reduced reliance on command-and-control	ISR Cyber-warfare Electronic warfare Submarine and mine-hunting Logistics operations
Persistence: constant performance of unmanned systems for 'dull, dirty and dangerous' missions	Air defence Long ISR Countermine operations Casualty evacuation Logistics operations in enemy territory
Reach: access to GPS and communication-denied environments	ISR in A2/AD environments Submarine and mine-hunting Casualty evacuation Logistics operations in A2/AD environments Strikes in A2/AD environments
Coordination: ability to coordinate large groups of weapon systems in a structured and strategic way	Force protection Combat operations in A2/AD environments ISR in complex and cluttered environments

A2/AD = anti-access/area-denial; GPS = Global Positioning System; ISR = intelligence, surveillance and reconnaissance; OODA = observe, orient, decide, act.

Source: US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016).

to be in constant contact with human operators. On-board autonomous sensing could, for instance, mean that a weapon system deployed on a reconnaissance mission would only have to report on key information. This, in turn, would (*a*) reduce the need to maintain a constant communication link between the system and the military command; and (*b*) allow the military to scale down on the number of human operators and analysts required to oversee the system and process information.

Accuracy. A third benefit is that autonomy can improve the accuracy of weapon systems, which provides the opportunity to apply force in a more effective and discriminate way. Advances in accuracy have reduced the need to employ weapon payloads with a large destruction radius, which, in turn, has reduced the risk of collateral damage. Progress in sensor technology has also increased the ability of the military to discriminate between targets, notably between what may be deemed as lawful and unlawful targets.[28]

Persistence. A fourth benefit is that autonomy improves a weapon system's persistence, meaning that its performance remains unaltered over time. The performance level of a weapon system that is destined for so-called dull, dirty or dangerous missions (sometimes referred to as 3D tasks), such as air defence, long surveillance missions, countermine operations or logistics operations in enemy territory, might deteriorate over time due to the human operator's cognitive and physical limitations (e.g. fatigue, boredom, hunger or fear). Autonomy removes these limitations.

Reach. A fifth benefit is that autonomy can give weapon systems greater reach. It grants access to operational theatres that were previously inaccessible to remotely controlled unmanned systems or too risky for soldiers or manned systems. Such theatres include A2/AD bubbles and areas where there are harsh operating environments for

[28] Freedberg, S., 'Naval drones "swarms", but who pulls the trigger', Breaking Defense, 5 Oct. 2014.

humans (and where communication is limited), such as deep water, the Arctic and, potentially, outer space.

Coordination and mass. Finally, autonomy also provides new opportunities for collaborative operations as it permits weapon systems to operate in large groups, or 'swarms', in a much more coordinated, structured and strategic way than if they were individually controlled by a human operator. Some military planners have argued that progress in collaborative autonomy, combined with advances in low-cost robotic platforms, has the potential to reintroduce a capability that the modern military has progressively lost due to the rising cost of weapon systems: mass.[29] Swarming technology could prove useful for many different types of missions, including (*a*) force protection (autonomous swarms could act as agile mines to protect perimeters around military assets); (*b*) force penetration (large numbers of autonomous systems could be used—for instance—to confuse, deceive or wear down enemy defences); and (*c*) ISR in cluttered and adversarial environments that represent a high lethality risk for combat troops (autonomous swarms could be used to explore buildings and locate enemy combatants or civilians).

In short, autonomy provides multiple operational benefits across a very large spectrum of missions, from defensive, ISR and logistics missions to combat missions (see table 4.1).

Economic benefits

Another important driver of autonomy is the promise that it could reduce the manpower requirements for military operations and thereby potentially provide some cost-saving opportunities.

Nearly all weapon systems that are in use today are operated by one operator or more, regardless of whether the system is manned (inhabited) or unmanned (uninhabited). Embedding greater autonomy in the systems could arguably help to achieve greater manpower efficiency, which could in turn translate into personnel cost savings.[30] One study by the US Air Force found that shifting to a model where one pilot would control several unmanned aerial vehicles (UAVs) at the same time instead of just one could allow personnel reductions of 50 per cent or greater.[31]

Reducing the number of pilots and operators could have a virtuous circle effect on operating costs. This is an argument that is generally used to support the transition from manned aircraft to unmanned autonomous aircraft. It is established that one of the most important cost factors associated with the use of manned combat aircraft is the need for continuous practice. Combat aircraft pilots must fly in real conditions to be properly trained, and have to fly between 10 and 20 hours a month to maintain their skill set. More frequent flights mean higher running costs (e.g. more fuel) and higher maintenance costs.[32] Unmanned autonomous aircraft, on the other hand, can 'sit on a shelf' for extended periods of time without losing any of their operational capability. These systems might be more expensive to develop and acquire, but they would be comparatively more cost-effective over their lifetime.[33]

---

[29] Arquilla, J. and Ronfeldt, R., *Swarming and the Future of Conflict* (RAND Corporation: Santa Monica, CA, 2005); and Scharre, P., *Robotics on the Battlefield Part II: The Coming Swarm* (Centre for a New American Security: Washington, DC, Oct. 2014).

[30] Manpower is, however, an issue for all countries. Manpower is foremost a problem for countries that have professional armed forces and a stagnant or declining military budget or a decreasing (e.g. Russia) or aging (e.g. Japan) population. US Department of Defense, Defense Science Board (note 10); and Scharre (note 29).

[31] United States Air Force (note 10).

[32] One study reported that maintaining a UAV in a long and continuous loiter surveillance mission required a rotating crew of 10 pilots, 10 sensor operators, 10 mission controllers and another 80 people to manage and process the data. Scharre (note 29).

[33] Byrnes (note 27), p. 57.

How much autonomy should weapon systems have?

The current state of the CCW discussions on LAWS demonstrates that determining the level of autonomy weapon systems should have is far from straightforward. The official position in the USA is that, while there are many benefits to increasing the autonomy of weapon systems, there should always remain some level of human supervision. An autonomous system that can operate entirely independently of human direction would, arguably, be useless from a military standpoint. The military wants and needs to exert control over a system's actions.[34] In other words, full autonomy is not, and cannot be, the objective.

To explain the US approach to the development of autonomy to the public and tackle potential concerns associated with the development of autonomous weapon systems, Work (the lead architect of the US Third Offset Strategy) often invokes the following comparison based on characters from popular science fiction:

When people hear me talk about this [autonomy], they immediately start to think of Skynet and Terminator, I think more in terms of Iron Man ... A machine to assist a human, where a human is still in control in all matters, but the machine makes the human much more powerful and more capable.[35]

According to this narrative, the purpose of pursuing advances in autonomy is not to remove humans from unmanned systems altogether, but to change their role by creating new models of human–machine collaborations in which the capabilities of both humans and machines can more effectively complement each other.[36] Autonomous systems would replace or support human operators in the execution of tasks for which their cognitive capabilities are not required or for which they are too limited. Human operators, on the other hand, would use their cognitive capabilities to make qualitative judgements that machines are not capable of.

This vision does not necessarily mean that autonomous systems would be under direct and constant human supervision. A report from the US Defense Science Board explains that the relationship between the human and the system could be dynamic and alternate between two teamwork styles: 'remote presence' (the human works through systems to perceive and act in real time at a distance, either in an active or supervisory role); and 'taskable agency' (the system is delegated sole responsibility for the task or mission while the human attends to other tasks or missions).[37] The choice of team style would depend on the nature of the overall mission, the type of tasks to be executed during the mission and the evolution of the mission circumstances. The control of some of the system's functions or the weapon system in its entirety would need, in some cases, to be passed back and forth between the human operator and the machine.[38]

As to the question of whether taskable agency would be reasonable for combat missions, the official narrative of the US administration is that it very much depends on the situation. In an interview about autonomy, Work maintained that autonomy is perfectly justified for defensive missions such as air defence or cyber-defence, but the use of autonomy in offensive operations is much more problematic.[39] The current approach, which was formally set into a policy in 2012, is that a human should exert

---

[34] US Department of Defense, Defense Science Board (note 10), pp. 23–24.

[35] Freedberg, S., 'Iron Man, not Terminator: the Pentagon's sci-fi inspirations', Breaking Defence, 3 May 2016.

[36] Endsley (note 10).

[37] US Department of Defense, Defense Science Board (note 10), p. 45.

[38] Endsley (note 10); and US Department of Defense, Defense Science Board (note 10), pp. 23–24.

[39] Work, R., 'Ending keynote: art, narrative, and the Third Offset', Atlantic Council Global Strategy Forum 2016, 2 May 2016.

an appropriate level of human judgement over the decision to attack. An autonomous system might only be authorized to engage targets that have been previously identified by human military command.

The US emphasis on human–machine teaming is, however, only one approach. Work admits that other countries, notably authoritarian regimes, might approach the governance of autonomy differently, and embrace a more 'Skynet' approach to the development of autonomous systems:[40]

Authoritarian regimes who believe people are weaknesses in the machine will naturally gravitate toward totally automated solutions ... Why do I know that? Because that's exactly the way the Soviets conceived their reconnaissance-strike complex: It was going to be completely automated.[41]

One author argues that states or non-state groups, if they have little consideration for international law and would not care about collateral damage or other consequences associated with the use of unsupervised autonomy, would also be likely to embrace operational concepts where humans would have little control over the use of weapons.[42]

III. Mapping the obstacles to further incorporation of autonomy in weapon systems

While there are manifold reasons for increasing autonomy in weapon systems, there are also many technical, institutional, legal, moral and economic factors that slow, and in some cases block, the further incorporation of autonomous capabilities into weapon systems.

Technical hurdles: issues of performance and safety

Technology is trailing behind expectations

The most obvious obstacles to the further development and use of autonomy in weapon systems are the limitations of the technology itself.[43] As discussed in the previous chapter, major progress has been achieved in many capability areas. The state of the art, while impressive, still trails by a wide margin the cultural perception of what advanced autonomous weapon systems ought to be able to do in a military context, namely operate safely and reliably in complex, uncertain and adversarial environments.

Autonomous capabilities that are showcased in deployed weapon systems and in systems under development still lack a certain flexibility.[44] They function only in situations that the programmers could foresee and plan for at the design phase, and they have no ability to generalize from previous experience and adapt to novel situations.[45] A human would need to remain in a supervisory role to intervene and handle all cases and situations that the systems were not programmed to address. This poses a fundamental problem to their deployment in battlefield situations that are complex, dynamic and involve an adversary that could seek to defeat the system by using decoy and deception tactics such as spoofing and cyber and electronic attacks. Unsupervised

---

[40] 'Skynet' is a term taken from different science-fiction franchises that refers to a dystopian vision of an army of robots capable of individually making intelligent decisions, sharing information and receiving orders from a super-computer. Freedberg, S., 'Robot wars: centaurs, Skynet, and swarms', Breaking Defense, 31 Dec. 2015.

[41] Freedberg (note 40).

[42] Freedberg (note 40).

[43] Versprille, A., 'Army still determining the best use for driverless vehicles', *National Defense* (June 2015).

[44] Endsley (note 10), p. 5.

[45] Cummings, M., *Artificial Intelligence and the Future of Warfare*, Research Report (Chatham House: London, 2017).

autonomy would only be a safe and reliable option in situations that were predictable to the programmers and non-adversarial, for example, aerial attack against predetermined targets on the territory of more technologically inferior forces.[46] To be truly valuable from an operational standpoint, existing autonomous capabilities should be able to be used safely and reliably in a much greater number of situations. But for that, they would have to be more adaptive and be able to deal with a much higher level of environmental uncertainty than is currently possible.

A related technical challenge concerns the limits of machine perception. Today's AI technology gives weapon systems only a very crude ability to make sense of their operating environment. This means that weapon systems that rely on this technology for targeting are certainly not able to execute the type of qualitative evaluations that are necessary to comply with the cardinal principles of international law: distinction, proportionality and precaution in attack (this is further discussed in the following subsection on institutional, legal and normative obstacles). Thus, at least for the foreseeable future, humans will have to continue to play the crucial role of receiver and arbitrator of tactical information on the battlefield.

A final technical issue is that, in some cases, there remain some small hardware problems that limit the military viability of autonomous weapon systems concepts. A recent illustration of this was the decision by the US military to shelve the development of Boston Dynamics' Pack Mule due to the limitations in its propulsion systems. The Pack Mule, which currently uses a gas engine, was deemed too noisy, while an alternative electric power version could not carry enough equipment (40 pounds instead of 400) to meet the demands required of the system.[47] The problem of power management is common to all robotic systems, but is particularly challenging in the case of small robotic systems. The viability of swarm operations using small low-cost UASs is fundamentally limited by the current state of battery technology, which prevents them from being used over long distances or long periods. Also, the smaller the systems are, the less computer power they can take on-board, which in turn affects their ability to execute the type of complex calculations that would permit them to operate advanced autonomous operations. These problems are not insurmountable or exclusive to autonomous systems, but they certainly negatively impact the cost/benefit analysis that is inherent to the procurement process within the military.

Finding the right human–machine ratio is difficult

If human operators can be expected to continue to play a key role in the supervision of weapon systems for the reasons mentioned above, maintaining a safe and meaningful interaction between the human operator(s) and weapon systems becomes increasingly challenging as the systems' level of autonomy rises.[48] Risks associated with human supervision of advanced automation are well known and include automation complacency, under-trust and out-of-the-loop problems.

Automation complacency (also known as 'automation bias') is a phenomenon whereby humans overly rely on a system. Research has shown that the more reliable human operators perceive the system to be, and the more their cognitive resources are mobilized elsewhere (which can happen, for example, when operators are expected to multitask or control multiple systems at once), the less likely they are to monitor it properly.[49] Over-reliance on autonomy in the context of weapon systems processes can

---

[46] 'Predictable' here means that the designers would be able to foresee possible scenarios that could be modelled in mathematical terms.

[47] Vincent, J., 'US military says robotic pack mules are too noisy to use', The Verge, 29 Dec. 2015.

[48] Murphy, R. and Burke, J., 'The safe human–robot ratio', eds M. Barnes and F. Jentsch, *Human–Robot Interaction in Future Military Operations* (CRC Press: Boca Raton, FL, 2010).

[49] Parasuraman, R., Molloy, R. and Singh, I. L., 'Performance consequences of automation-induced "complacency"',

have dramatic consequences—the most dramatic being a situation in which human operators using an ATR system as a decision aid would accept the computer recommendation without seeking any disconfirming evidence and would end up engaging a friendly or unlawful target.[50]

Under-trust is the opposite situation. In this case, human operators place insufficient reliance on the automated process. This generally happens when operators deal with systems that are known for producing false-positives or system errors, or systems that have user interfaces that are prone to inducing misinterpretation. A typical consequence of under-trust is that human operators sometimes ignore relevant information provided by the system or override its actions without justification. Several incidents involving automated air defence systems have been caused in this way. A famous example is the destruction of a commercial aircraft—Iran Air Flight 655—on 3 July 1988 by an Aegis Combat System stationed on the *USS Vincennes*, a US Navy warship.[51]

The more autonomous a process is, the harder it is for human operators to react to a problem correctly and in a timely manner. A number of empirical studies have demonstrated that when human operators shift from an active role of 'controller' to a passive role of 'supervisor', they lose some situational awareness. Maintaining constant vigilance is extremely difficult for humans—decrement of vigilance can occur after as little as 30 minutes.[52] The out-of-the-loop control problem happens when emergency or critical situations occur (e.g. system failure situations that can be addressed by a human operator) and the human operator is unable to regain sufficient situational awareness to react appropriately and in time.[53] The out-of-the-loop problem is well known within the commercial aircraft industry, and many aviation accidents have occurred because of a sudden transfer of control from the autopilot to the human pilot.[54] It is also a problem that has been associated with the use of missile defence systems.

An illustrative anecdote is provided by John Hawley of the US Army Research Laboratory in relation to two friendly fire incidents involving the US Patriot missile defence system during Operation Iraqi Freedom in 2003. According to Hawley, the reaction of the commanding Major General in charge of the Patriot system at the time was as follows: 'How do you establish vigilance at the proper time? 23 [hours] and 59 [minutes] of boredom followed by one minute of panic.'[55]

This statement sums up one of the most fundamental problems posed by the rise of autonomy in weapon systems: how to calibrate human control over increasingly autonomous systems and ensure that it remains adequate and effective. This problem has legal and ethical implications, which will be discussed in the next subsection. Providing a technical solution to this problem is challenging. On this topic, David Mindell, Professor of Aeronautics and Astronautics at the MIT, notes that:

It takes more sophisticated technology to keep the humans in the loop than it does to automate them out … On a commonly used scale of levels of autonomy, level one is fully manual control and level 10 is full autonomy … history and experience show that the most difficult, challenging and worthwhile problem is not full autonomy but the perfect five—a mix of human and machine and the

International Journal of Aviation Psychology, vol. 3, no.1 (1993); and Murphy and Burke (note 48).

[50] Sharkey, N., 'Staying in the loop: human supervisory control of weapons', eds N. Bhuta et al., *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press: Cambridge, 2016), p. 36.

[51] For further detail see chapter 3 of this report.

[52] Endsley (note 10), p. 6.

[53] Murphy and Burke (note 48), p. 45.

[54] One of the most recent cases was the crash of Air France Flight 447 in 2009. See Mindell, D., *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking: New York, 2015), pp. 1–2.

[55] Hawley, J. K., *Automation and the Patriot Air and Missile Defense System* (Center for a New American Security: Washington, DC, 2017), p. 6.

optimal amount of automation to offer trusted, transparent collaboration, situated within human environments.[56]

What constitutes a safe human–machine ratio, a 'perfect five' as Mindell says, is context-based and depends on a number of variables, including (*a*) the type and number of tasks to be executed; (*b*) the nature and complexity of the operating environment; (*c*) the sophistication of the systems; and (*d*) the cognitive workload of the human operators.

However, field research on command-and-control of unmanned systems in military operations could establish that, with the current state of technology, it would be premature, and potentially unsafe, to move away from a paradigm where a single unmanned system is operated and supervised by multiple controllers—the typical ratio for aerial systems is *many:1* while the ratio for ground systems is *2:1*—and shift to a paradigm where one operator alone would be in charge of a single platform (*1:1*) and potentially multiple platforms at the same time (*1:many*), as some military planners have suggested.[57] Robin Murphy and Jenny Burke, human–robot interaction (HRI) researchers at the University of South Florida, note that 'expecting a soldier to fly/ drive a robot, interpret sensor data, and remain sufficiently aware of the surroundings is unrealistic' and that 'autonomous navigation may not be sufficient to safely reduce the ratio due to the human out-the-loop problem'.[58]

The need to maintain a safe and reliable human–machine ratio can represent, in this respect, a significant obstacle to the further incorporation of autonomous capabilities in weapon systems. The fact that some autonomous capabilities, such as swarming, may now be technically mature does not mean that they can easily be adopted and fielded by the military. Increases in autonomy necessitate a rethink of the human control element leading to corresponding changes at the techno-organizational level. These changes could include (*a*) developing and placing new training requirements on personnel; (*b*) creating iteratively adequate human–machine teaming processes; and (*c*) potentially developing new human–machine interfaces to ensure that human control remains effective.[59] It is no surprise that the US Third Offset Strategy places a strong emphasis on investment in technologies that should enable a greater and swifter integration of human and machine cognitive capabilities.[60]

New methods of validation and verification are needed

A third important technical hurdle to the further incorporation of autonomy in weapon systems is the limitation of existing validation and verification (V&V) procedures. V&V procedures are parallel but different test and evaluation processes that are intended to provide the assurance that a newly developed system meets the user's needs and works correctly.[61] These procedures, particularly verification, are crucial to determine whether a new weapon system can be used safely, consistently and legally, and therefore whether it can be certified for use.[62]

---

[56] Mindell, D., 'Driverless cars and the myths of autonomy', Huffington Post, 14 Oct. 2015; Reese, H., 'Why robots still need us: David A. Mindell debunks theory of autonomy', Tech Republic, 13 Oct. 2015; and Mindell (note 54).

[57] Murphy and Burke (note 48), p. 48.

[58] Murphy and Burke (note 48), p. 48.

[59] Hawley (note 55); and Hoffman, R. R., Cullen, T. M. and Hawley, J. K., 'The myths and costs of autonomous weapon systems', *Bulletin of the Atomic Scientists*, vol. 72, no. 4 (2016).

[60] Freedberg, S., 'Centaur army: Bob Work, robotics, and the Third Offset Strategy', Breaking Defense, 9 Nov. 2015; and Pomerleau, M., 'Man-machine combo key to future defence innovation', *GCN*, 13 Nov. 2015.

[61] Russell, S., Dewey, D. and Tegmark, M., *Research Priorities for Robust and Beneficial Artificial Intelligence* (Future of Life Institute: Boston, MA, 2015).

[62] That includes ensuring that the system is capable of compliance with international law, as required by Article 36 of 1977 Additional Protocol I to the 1949 Geneva Conventions. Boulanin, V., 'Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems', SIPRI Insight on Peace and Security, no. 2015/1, Nov. 2015.

The progress of autonomy represents, however, a growing challenge for the community of experts that develop and implement V&V procedures. Traditional methods, which typically attempt to verify the correctness of systems in all possible conditions, become progressively more inadequate as autonomous systems grow in complexity.[63] As autonomous systems become more intelligent, interactive and capable of adapting to complex and dynamic environments, it becomes, practically and financially, infeasible to continue to test all ranges of inputs to, and possible states of, the system (this is further explained in box 4.1).[64] It should be stressed that this problem is not exclusive to autonomous weapon systems, but is applicable to all autonomous systems. However, militaries have very high standards for V&V, as safety and reliability are paramount concerns in combat situations.

The current circumstances are such that existing V&V methods can, for now, only be viably implemented for autonomous systems that are static (i.e. that do not incorporate a learning behaviour) and destined to operate in predictable conditions, typically a controlled or semi-controlled environment that features limited or well-understood environmental changes. They do not, or only with difficulty, permit testing of the performance, safety and reliability of autonomous systems that (*a*) incorporate online learning capabilities; (*b*) operate in highly open and non-controlled, hence potentially unpredictable, environments; and (*c*) interact with other autonomous systems or humans. It is commonly agreed within the test and evaluation community that new V&V methods are needed to address these limitations.[65] Until these methods are developed and adopted, the limitations of existing V&V procedures will remain a bottleneck to the incorporation of a higher level of autonomy in weapon systems. This was, at least, the understanding of the Office of the US Air Force Chief Scientist, which concluded in a report dated 2010, that 'It is possible to develop systems having high levels of autonomy, but it is the lack of suitable V&V methods that prevents all but relatively low levels of autonomy from being certified for use'.[66]

It could be said, however, that in the recent past some military systems have by-passed the formal test process due to pressing operational demands. This was the case, for instance, for the MQ1-Predator UAS, which was deployed by the US Air Force, despite the fact that it had failed the operational test and evaluations.[67]

Institutional, legal and normative obstacles

The aforementioned technical limitations feed into another series of obstacles that are more of a political nature, including (*a*) the ambivalent relationship of military organizations with new technologies and autonomy; (*b*) obligations to comply with international law and domestic law; and (*c*) increasing normative pressure from within civil society about the importance of maintaining meaningful human control over weapon systems.

[63] Autonomy Community of Interest Test and Evaluation, Verification and Validation Working Group, *Technology Investment Strategy 2015–2018* (Office of the Assistant Secretary of Defense for Research & Engineering: Washington, DC, 2015).

[64] Endsley (note 10), p. 23; and Autonomy Community of Interest Test and Evaluation, Verification and Validation Working Group (note 63), p. 5.

[65] Autonomy Community of Interest Test and Evaluation, Verification and Validation Working Group (note 63), p. 5.

[66] Office of the US Air Force Chief Scientist, *Technology Horizons: A Vision for Air Force Science and Technology 2010–30*, vol. 1, AF/ST-TR-10-01 (Air University Press/Air Force Research Institute: Maxwell Air Force Base, AL, Sep. 2011), p. *xx*.

[67] Macias, F., 'The test and evaluation of unmanned and autonomous systems', *ITEA Journal*, vol. 29 (2008), pp. 388–95.

Box 4.1. Validation and verification: existing methods and their limitations

Definition

Validation procedures seek to provide the assurance that the system meets its specified requirements, is fit for purpose and does not have unwanted behaviours or consequences (did the engineers build the right system?). *Verification* aims to ensure that the system satisfies formal properties derived from the requirements and design specifications (did the engineers build the system correctly?).

Methods

Validation and verification (V&V) of systems can be realized using *formal methods* and *testing*. There are differences between these two approaches, which are, all in all, complementary. *Formal methods* are a 'deductive' approach that consists of finding a mathematical proof of the correctness of the system. The main advantage of formal methods is that, if they can provide a strong proof, they demonstrate that the system will work correctly in all cases. The disadvantage is that implementing formal methods is a complex endeavour, which requires translating the properties of the system into a formal mathematical language. The more complex the system is, the more computer power-intensive and time-consuming the process becomes. *Testing*, by contrast, is an 'inductive' approach to verification that infers the correctness of the system based on a representative sample of test cases. Unlike formal methods, testing can only provide a weak proof of correctness. Testing provides the opportunity, however, to see how the system performs in real or (close to real) conditions, through field testing or through computer modelling and simulation.

Limitations

There are no established standards for testing autonomous systems. Advances in autonomy also pose a number of unique challenges, including the following.

  1. *State–space explosion problem*. Advances in autonomy commonly allow a system to react to more environmental stimuli and to have a larger decision space (i.e. range of options at its disposal). This means that the number of possible inputs and system states grows accordingly, which makes exhaustive verification, practically and financially, increasingly unfeasible. The state–space explosion problem is particularly exacerbated in cases where autonomous systems are to be used in unpredictable environments or in interaction with humans (who are by definition unpredictable) and other autonomous systems. Exhaustive V&V becomes impossible, as it is infeasible for humans to foresee all the possible combinations of events that could lead to a system failure. There will always be corner cases. The challenge then is to determine what constitutes a representative and realistic set of scenarios that could be explored through formal methods and computer simulations.
  2. *Learning*. Online learning has the potential to boost significantly the intelligence and adaptiveness of systems. The problem from a V&V perspective is that online learning entails an automatic reparameterization and partial reprogramming of the system. Each time the system learns something new, its performance and correctness need to be re-tested and re-evaluated. There is currently no reliable methodology for testing and evaluating systems capable of online learning.

Sources: Background interviews with experts conducted by the author in October 2016; and Autonomy Community of Interest Test and Evaluation, Verification and Validation Working Group, *Technology Investment Strategy 2015–2018* (Office of the Assistant Secretary of Defense for Research and Engineering: Washington, DC, 2015).

Institutional frictions: the ambivalent relationship of military organizations with new technologies and autonomy

Military organizations are traditionally slow to embrace technological change, particularly when it might profoundly alter the way wars are fought.[68] Autonomy is no exception. As it stands, while the potential of autonomy for air defence and non-combat dull, dirty or dangerous tasks, such as ISR, bomb ordnance disposal, and search and rescue, is now widely recognized, there is still some resistance (and in some cases opposition) to increasing reliance on autonomous functions in weapon systems, notably for combat operations. This hesitance is often bemoaned by those military

---

[68] Kaldor, M., *The Baroque Arsenal* (Hill and Hang: New York, 1981); Dunne, P., 'Economics of arms production', ed. L. Kurtz, *Encyclopedia of Violence, Peace and Conflicts* (Elsevier: Oxford, 2009); Toffler, A. and Toffler H., *War and Anti-War: Making Sense of Today's Global Chaos* (Warner Books: London, 1995); and Jungdahl, A. M. and MacDonald, J. M., 'Innovation inhibitors in war: overcoming obstacles in the pursuit of military effectiveness', *Journal of Strategic Studies*, vol. 38, no. 4 (2014), pp. 1–33.

planners and experts who consider autonomy and robotics as the main components of the future of warfare. In a recent report, Scharre expresses his regret that:

Just as the Navy initially resisted the transition from sail to steam-powered ships and elements of the Army dismissed air power and fought against the shift from horses to tanks, some parts of the military continue to resist the expansion of uninhabited systems into traditional combat roles. As a result, the DOD is failing to invest in game-changing technology.[69]

The caution (some would say conservatism) of some military organizations towards new technologies is a well-documented historical phenomenon. The reasons of the past behind the opposition to a shift to air power, steam-powered ships and tanks are not that different from those that explain the scepticism and resistance towards autonomy today. These reasons include (*a*) a lack of trust in unproven technology; (*b*) organizational stasis and cultural resistance within military services; (*c*) a lack of coherent vision between military services; and (*d*) a strong focus on current strategic and operational priorities.

Lack of trust. Military personnel do not have a natural aversion to change. However, as Michel Goya, a historian and former Colonel with the French Army, pointed out during an interview, 'every change in a military context is risk'.[70] A new technology might not work well at first, and its adoption, in any case, requires a learning and adaptation period, which during a conflict may be a source of vulnerability or of reduced efficiency. This is why military personnel often hold on to old technologies, despite the fact that newer and potentially more effective technologies are available. They would rather rely on a technology they know how to use and that they can trust than a technology they fear might fail at a crucial moment.[71] From this standpoint, the challenges of autonomy are manifold. First, due to the technological problems mentioned above, there is still a widespread lack of trust that advanced autonomy can perform as intended in all situations.[72] Second, the applications of autonomy that are often presented as game-changing (e.g. swarming or multi-vehicle control) remain early concepts that have only been demonstrated by R&D projects. Third, autonomy is not a concrete and visible 'object', but a diffuse capability that is hidden within the system (Goya argued that military personnel not only have a preference for what they know, but also for 'what they can see').[73] It is little surprise, in this context, that articles, opinion pieces and reports that promote autonomy place such importance on test and evaluation procedures as well as the research on HRI, and encourage an incremental approach to the adoption of autonomy: slow and progressive adoption of new autonomous capabilities is seen as a way to build trust and overcome these concerns.[74]

Organizational stasis and cultural resistance. Military organizations are also historically characterized as hierarchical and rigid in nature. The adoption of autonomy offers great potential, but it would require, in some cases, significant organizational and personnel changes. The academic literature has shown that military innovations rarely go through unless they are supported by advocates in positions of power within the institutions.[75] The rigid military hierarchy also means that some locally empowered

---

[69] Scharre, P., *Robotics on the Battlefield Part I: Range, Persistence and Daring* (Center for a New American Security: Washington, DC, 2014), p. 5.

[70] Goya, M., Interview with author, 15 Nov. 2016.

[71] Scharre (note 69), p. 5.

[72] Wheeler, S., *Trusted Autonomy: Conceptual Developments in Technology Foresight*, Defence Science and Technology Group Report, DST-Group-TR-3153 (Australian Government, Department of Defence: Victoria, 2015).

[73] The Dwight D. Eisenhower School for National Security and Resource Strategy (note 26).

[74] Scharre (note 69) p. 35.

[75] It could be said that, in the USA, the nomination of Robert Work—who in his previous position of Executive Director of the Center for a New American Security devoted much of his attention to the potential of robotic technologies and autonomy—as Deputy Secretary of Defense has played a key role in the rise of AI and autonomy on the R&D and procurement agendas of the US Department of Defense in recent years. Rosen, S., 'New ways of war:

individuals or groups within the military services, based on prior beliefs about the nature and means of warfare or specific interests, 'can use their unique influence over training and procurement', which inhibits innovation and prevents new technologies from reaching the battlefield.[76] This is very well illustrated by the fact that many combat aircraft pilots, who constitute the elite of the air force, continue to oppose the adoption of UAVs (or remotely piloted aircraft as they have been rebranded) in some countries.[77] A survey of the attitude of US Air Force pilots towards UAVs found that many pilots had wrapped up their professional identity so tightly around the act of flying that they would rather leave the service than fly a remotely piloted aircraft (one-third of the participants).[78] Needless to say, these pilots are also strongly sceptical about the potential of autonomy for multi-vehicle control, which could further disrupt the operational paradigms they are used to.[79] Similar views are held by many other groups of military service personnel. Within each branch of the military, there are those who see the emergence of unmanned systems and autonomy as a threat (or insult depending on the case) to their professional identity, as it would, in their view, progressively make their profession less 'noble' or eventually render their core skills and responsibilities obsolete. It is very likely that the cultural resistance is a generational phenomenon, and it will start to fade as the next generation of personnel come to occupy leadership positions. The process is bound to take time, however.

Lack of coherent vision between military services. A related problem is the lack of coherent vision between military services. Because they face different operational realities, the various branches of the military services can hold differing beliefs about which technologies are important for the future of warfare. This can lead to a situation where one service might embrace and push for the adoption of a new technology, while another might resist it. The relationship of the US military to swarming and multi-vehicle control provides a case in point: the US Air Force remains sceptical about, and is not actively pursuing, technological developments in that area, while the US Navy and, to a lesser extent, the US Army are enthusiastic about the operational possibilities that these technological developments could create.[80] The lack of coherent vision may make it harder for the procurement agencies and arms industry to plan for future acquisition, as it inherently leads to budget competition and mixed signals as to what should be the focus of R&D efforts.

Strategic and operational priorities. Another, and perhaps the most important, factor that shapes the military's relationship with new technologies is the strategic and operational reality itself. Military services typically give priority to technologies they deem the most useful in the current paradigm. This is particularly true in a context where budgetary resources for acquisition are limited. Military services generally would rather invest in readily available technology that they know will be useful in the theatre of operations where they are likely to be deployed than in new futuristic concepts that might not deliver capabilities for years or decades, and might be useful only in a limited number of war scenarios. This is one of the reasons why, as discussed in section II of this chapter, different countries and different military services show varying levels of interest in, and approaches to, autonomy. It also explains why many

---

understanding military innovation', *International Security*, vol. 13, no.1 (1988), pp. 134–68; and Goldman, E., 'Cultural foundations of military diffusion', *Review of International Studies*, vol. 32, no. 1 (2006), pp. 69–91.

[76] Jungdahl and MacDonald (note 68), p. 2.

[77] Mindell (note 54), pp. 113–58. Note that this is not true in all countries. In Israel, for instance, the Air Force has been quite positive towards unmanned systems and the development of autonomy.

[78] Cantwell, H., *Beyond Butterflies: Predator and the Evolution of Unmanned Aerial Vehicles in Air Force Culture* (School of Advanced Air and Space Studies: Maxwell Air Force Base, AL, 2007), pp. 81–85. See also Byrnes (note 27), pp. 48–75.

[79] Scharre (note 29), p. 37.

[80] Goya (note 70).

of these countries and military services might not currently see an immediate value in exploring the futuristic concept of autonomous combat operations (e.g. swarms of micro and small unmanned systems for A2/AD or UCASs for deep-strike). These capabilities would be highly valuable in an interstate conflict scenario between two modern militaries. However, for many countries such a scenario is not credible in the current strategic context.[81]

In sum, considering the military's prudent approach to new technology, the incorporation of autonomy in weapon systems is likely to take place at a slow pace, primarily through incremental changes—at least for the foreseeable future. The only event that could significantly speed up the process, and make military services overcome some of their current cultural barriers (notably with regard to the use of autonomy for combat operations), would be a drastic change in the global security landscape and the outbreak of a major armed conflict between two or more modern militaries. Here again, history has shown that wars can accelerate the introduction and adoption of many inventions by the military, the most recent examples being the wars in Iraq and Afghanistan, which accelerated the adoption of ground robots and UASs by the US military.[82]

Legal restrictions on the use of autonomy for the use of force

Should any military organization, for whatever reasons, decide to accelerate the incorporation of autonomy in weapon systems, and explore the potential of fully autonomous weapon systems for combat operations, they would still have to make sure that the weapon can be used in compliance with international law.

Article 36 of Additional Protocol I of the 1949 Geneva Conventions requires states to employ a mechanism, sometimes referred to as a 'weapon review', a 'legal review' or an 'Article 36 review', which can determine the lawfulness of any weapon, means or method of warfare before it is used in an armed conflict.[83] This mechanism is aimed at preventing the use of weapons that violate international law. From a legal perspective, autonomy does not raise fundamentally new issues, unless it is used to support the targeting process.[84] The targeting process requires a complex assessment to ensure that an attack takes place according to the fundamental rules and principles of IHL in the conduct of hostilities (also known as the laws of targeting): distinction, proportionality and precaution in attack. A truly autonomous weapon system would have to be capable of following each of these three rules to be considered lawful.

The rule of distinction requires a determination as to whether the target is lawful and hence not a civilian, civilian object or a person *hors de combat*.[85] As discussed in the previous chapter, some systems can already comply with the principle of distinction but only in a very crude manner. They can only recognize large military targets (e.g. tanks, radar or missiles) but are unable to appreciate the surrounding context of the

---

[81] The USA, Russia and China aim to project power on a global basis and they do so in the form of large-scale, long-range, non-nuclear air and missile strikes at great distances from their territory. European states do not on the whole feel the need to project their power globally so they have not prioritized concomitant capabilities. See also Fiott, 'A revolution too far? US defence innovation, Europe and NATO's military–technological gap' (note 17), p. 2.

[82] Rosen, S., *Winning the Next War* (Cornell University Press: Ithaca, NY, 1991).

[83] Arguably, this obligation applies to all states regardless of whether they are parties to Additional Protocol I. International Committee of the Red Cross (ICRC), *A Guide to the Legal Review of Weapons, Means and Methods of Warfare* (ICRC: Geneva, 2006), p. 4. The vast majority of states do not have laws or processes to review the legality of new weapons. Out of the 174 states parties to Additional Protocol I, only a limited number of states (fewer than 20) are known to have a weapon review mechanism in place. These are predominantly countries that have modern defence industries. The USA, which is not party to Additional Protocol I, also conducts weapon reviews. See also Boulanin (note 62).

[84] For a detailed analysis see Boulanin (note 62).

[85] On the rule of distinction see Articles 41, 48, 50 and 52 of 1977 Additional Protocol I to the 1949 Geneva Conventions. International Committee of the Red Cross (ICRC), *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II)*, opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978.

target—the Harpy, for instance, can recognize a radar, but cannot appreciate whether the radar is surrounded by civilians or civilian objects. Their use could be deemed lawful when they are employed in remote areas generally devoid of civilians or civilian objects. To be used against human targets and in more cluttered, dynamic and populated areas, they would need to include much more sophisticated perception and decision-making capabilities. Nathalie Weizmann from the Columbia Law School notes that for a system to determine whether the individual is a legitimate target, it would need to:

Be able to evaluate a person's membership in the state's armed forces (e.g. distinct from a police officer) and his or her membership in an armed group (with or without a continuous combat function), whether or not he or she is directly participating in hostilities, and whether or not he or she is *hors de combat* ... [It] would also need to be able to, first, recognize situations of doubt that would cause a human to hesitate before attacking and, second, refrain from attacking objects and persons in those circumstances.[86]

This is far beyond what is feasible today with AI technology, but this might not be impossible to engineer.[87] Some roboticists (e.g. Ronald Arkin) argue that, eventually, robots will be better than humans at applying the distinction principle.[88]

The rule of proportionality poses a much greater technical challenge. It prohibits attacks that may be expected to cause incidental loss of 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.[89] A proportionality assessment requires, in other words, evaluation of the military advantage balanced against expected collateral damage. Evaluating the expected amount of collateral damage is not an insurmountable technical problem, as it is essentially a quantitative determination of the number of civilians that may be harmed incidentally as the result of an attack. According to Michael Schmitt, Chairman of the Stockton Center for the Study of International Law at the US Naval War College, major militaries have already mechanized this process with a system called Collateral Damage Estimation Methodology, which relies on scientific data and objective standards.[90] Evaluating what constitutes a military advantage, however, is more difficult as it is an assessment that is (*a*) subjective—there are no objective metrics or parameters of what constitutes a military advantage; and (*b*) contextual—it must be based on the reliable information available at the time of the attack on the target and its surroundings; it is, in other words, a case-by-case determination.[91]

There seems to be general agreement among legal experts who write on this topic that it is unlikely that an autonomous system will ever be capable of independent value judgements as required by the proportionality rule.[92] However, Jeffrey Thurnher,

---

[86] Weizmann, N., *Autonomous Weapons Systems under International Law*, Academy Briefing no. 8 (Geneva Academy: Geneva, 2014), p. 14.

[87] The company that is developing anti-personnel sentry guns for the South Korean armed forces is reportedly working on the development of computer vision software that could distinguish between military personnel and civilians based on visual features such as uniforms, but it is unclear whether these systems will be able to assess whether individuals who are *a priori* legitimate targets may be surrendering or *hors de combat*.

[88] Marchant, G. et al., 'International governance of autonomous military robots', *Columbia Science and Technology Law Review*, vol. 12 (2011), pp. 273–315; and Arkin, R., 'Lethal autonomous systems and the plight of the non-combatant', *AISB Quarterly*, no. 137 (2013), pp. 1–9.

[89] Article 51(2) of 1977 Additional Protocol I to the 1949 Geneva Conventions.

[90] Schmitt, M., 'Autonomous weapon systems and international humanitarian law: a reply to the critics', *Harvard National Security Journal*, vol. 73, no. 2003 (2012), pp. 19–20.

[91] Schmitt (note 90), pp. 19–20.

[92] Thurnher, J., 'Means and methods of the future: autonomous systems', eds P. A. L. Ducheine, M. N. Schmitt and F. P. B. Osinga, *Targeting: The Challenges of Modern Warfare* (Asser Press: The Hague, 2016), p. 189; Schmitt (note 90), pp. 19–20; Docherty, B., *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch: Washington, DC, 2012), p. 33; and Sharkey, N., 'Saying "no!" to lethal autonomous targeting', *Journal of Military Ethics* vol. 9, no. 4 (2010), pp. 369–83.

legal adviser to the NATO Rapid Deployable Corps, and Schmitt both argue that it might be possible, albeit technically challenging, to pre-programme acceptable values for the situation that an autonomous system might encounter.[93] Thus, a human would do a proportionality assessment before the launch of the system and predetermine the maximum amount of collateral damage for specific targets. To be in compliance with the principle of proportionality, these values would have to be set conservatively. As is the case with existing systems, the military might have to restrict the action of the system in time because the assessment would have to remain valid throughout the weapon's deployment. The military might also need to apply geographical restrictions. Compliance with the principle of proportionality would certainly be less problematic to achieve in a remote, open and unpopulated area than in dynamic and populated areas where the notion of what constitutes a military advantage might change quickly.

The rule of precaution is closely linked to the obligation of proportionality as it stipulates that those who plan or decide upon an attack shall (*a*) do everything feasible to verify that the objectives to be attacked are military in nature and are not civilians, civilian objects or subject to special protections; and (*b*) take all feasible precautions in the choice of means and methods of attack with a view to avoiding or minimizing injury to civilians or civilian objects.[94] As with the principle of proportionality, the type of assessments required to comply with the principle of precaution are highly complex and very difficult to translate into an algorithmic form. It might be possible to pre-programme some 'ethical rules' within a system (e.g. not engaging a target when there might be some uncertainty about its nature and instead requesting human approval), but these might, in some situations, affect the actual military or operational value of the system, as it would potentially provide an opportunity to the enemy to defeat the system. According to Thurnher, the principle of precaution also means that the use of autonomous weapon systems could be deemed unlawful if it could be reasonably established that it was feasible to use another system (e.g. a manned or remotely controlled system) that would provide better protection of civilian objects without sacrificing military advantage. In other words, the use of autonomous weapon systems would be lawful only when the systems would be the best choice for the situation.[95]

In short, IHL requirements on the use of weapons mean that, based on the current level of technology, the military is bound to keep the development and use of autonomous targeting capabilities 'on a tight leash', to borrow an expression recently used by Scharre and Michael Horowitz.[96] The use of weapons that can select and engage targets outside direct human supervision can only be lawful in a limited number of circumstances, typically against predetermined targets in low-clutter and static environments, where complex assessments are rarely required. Complex operational contexts and dynamic environments would require humans to remain as receivers and arbitrators of tactical information, as the nature of the qualitative assessments that such situations demand to ensure compliance with targeting law cannot reasonably—and might never be—conducted by a weapon system.

To sum up, it could be technically possible to remove human decisions from all parts of the control chain of a weapon system. However, there are some legal limitations that restrict the development and use of autonomous targeting capabilities and require human operators to maintain, in most circumstances, some form of human control or oversight over the weapon system's behaviour.

---

[93] Thurnher (note 92), p. 189; and Schmitt (note 90), p. 20.

[94] Article 57(2) of 1977 Additional Protocol I to the 1949 Geneva Conventions.

[95] Thurnher (note 92), p. 190.

[96] Scharre, P. and Horowitz, M., 'Keeping killer robots on a tight leash', *Defense One*, 14 Apr. 2014.

Normative pressure from within civil society on maintaining meaningful human control over weapon systems

In addition to the aforementioned legal limitations, there is also a growing normative opposition within civil society to the development of autonomous weapon systems, which makes the development and use of such systems potentially politically sensitive for the military.

The issue of autonomy in weapon systems has attracted growing attention in recent years from the general public, notably thanks to the significant advocacy work of the Campaign to Stop Killer Robots, a coalition of 61 international, regional and national NGOs that calls for a pre-emptive ban on the development, production, use and trade of 'killer robots'.[97]

The campaign has been successful at creating political momentum on the issue of autonomy in weapon systems. It can certainly take credit for the fact that there is now a formal intergovernmental discussion within the CCW framework and that the concept of meaningful human control has been identified as a possible basis for regulation or control of autonomy in weapon systems.[98] The campaign has also been successful at mobilizing an opposition within the expert community, both on the humanities side (lawyers, ethicists and philosophers) and on the engineering side (AI researchers and roboticists). In 2015 the Future of Life Institute released an open letter that calls for a ban on 'offensive weapons beyond human control'. The letter was signed by 3105 AI and robotics researchers and other leading figures from academia and the private sector, such as Stephen Hawking, Noam Chomsky, Elon Musk (Chief Executive Officer of Tesla), Steve Wozniak (co-founder of Apple), Peter Norvig (Research Director at Google), and another 17 701 individuals (as of February 2017).[99] In 2016 the Institute of Electrical and Electronic Engineering, which is the world's largest technical professional organization with over 400 000 members in 160 countries, included in its very first report of its Global Initiative for Ethical Consideration in Artificial Intelligence and Autonomous Systems a recommendation to technical organizations to accept that 'meaningful human control of weapon systems is beneficial to society'.[100]

The extent to which the Campaign to Stop Killer Robots reflects the view of the general public on the use of autonomous weapon systems remains unclear, as there have only been a handful of public opinion surveys conducted so far. The survey with the largest geographic coverage was conducted in 2015 by the Open Robotics Initiative and involved 1002 participants from 49 countries. It concluded that participants in the survey were largely opposed to the development and use of LAWS by the military.[101]

Another study focused on the US public reached similar conclusions but also found that public opinion opposing autonomous weapons was highly contextual and could rise or fall depending on circumstances. It found that the fear of other countries or non-state actors developing LAWS made the survey participants significantly more supportive of the USA developing them. Moreover, it appeared that the survey

[97] The campaign defines 'killer robots' as fully autonomous robots that would be able to select and fire on targets without human intervention. Campaign to Stop Killer Robots, <http://www.stopkillerrobots.org/the-problem>. For extensive coverage of the arguments in favour of a ban see Docherty (note 92).

[98] United Nations Institute for Disarmament Research (UNIDIR), *The Weaponization of Increasingly Autonomous Technologies, Considering Ethics and Social Values*, UNIDIR Resources no. 3 (UNIDIR: Geneva, 2015).

[99] Future of Life Institute, 'Autonomous weapon systems: an open letter from AI and robotics researchers', July 2015.

[100] Institute of Electrical and Electronic Engineering (IEEE) Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, Version 1 (IEEE: 2016).

[101] A total of 67% considered that all types of LAWS should be banned; 56% considered that LAWS should not be developed; 85% considered that LAWS should not be used for offensive purposes; 71% considered that the military should use remotely controlled weapons rather than LAWS. Moon, A. and Nuttall, R., *The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll* (Open Robotics Initiative: Vancouver, 2015).

participants became more willing to support the use of autonomous weapons when their use reduced risk to the national armed forces.[102]

More studies will be needed to draw the conclusion that there is widespread public opposition to the development of autonomous weapon systems or to argue that these systems would violate the public conscience provision of the so-called Martens Clause in Additional Protocol I.[103] It could also be argued that public opinion on the use of autonomous systems might change over time, as autonomous systems become more integrated into civilian life.

Economic obstacles

Finally, in addition to technical and political issues, there are also a number of economic factors that come into play in the adoption of autonomy in weapon systems. The most obvious one is cost management. There are limits to what states can afford, which makes the question of developing autonomy in weapon systems dependent on the strategic context.

Financial constraints: limitations to what can be afforded

One of the key arguments that is often flagged in support of incorporating greater autonomous capabilities in weapon systems is that autonomy might provide cost-saving opportunities. However, in some cases, the cost reductions might be offset by the increase of, or creation of new, cost items associated with the development of autonomy. The systems would be more expensive to develop and acquire, and while they might need fewer human operators, they might require more engineers to develop and maintain them. They might also necessitate the creation of new and more expensive training routines for the people who will use them. Development and testing costs might be higher than expected, which is often the case, and training and maintenance might take longer and be more complex than for other types of system.[104] There are, in any case, many switching costs associated with the adoption of autonomous capabilities in weapon systems. These can include (*a*) acquiring new or upgrading enabling technologies (in some cases it might require developing a new class of weapon systems); (*b*) (re)creating adequate doctrines and concepts of use; (*c*) developing new or modifying existing training and maintenance procedures; (*d*) adapting the logistical chain and existing support infrastructure; and (*e*) developing new V&V procedures.

It is not feasible here to provide a proper estimation of these costs and determine the extent to which they are prohibitive (or not) for a given country, as there are too many variables to consider. However, as pointed out by Ben Fitzgerald and Kelley Sayler in a recent report on the global defence industry, 'the rate of diffusion and the type of adopters of a given technology are both likely to vary as a function of the financial intensity and organizational capital required for adoption'.[105] It can be assumed, in this context, that the economic barriers to the adoption of autonomy in weapon sys-

---

[102] Horowitz, M., 'Public opinion and the politics of the killer robots debate', *Research and Politics* (Jan.–Mar. 2016), pp. 1–19.

[103] The Martens Clause is a legal principle in Article 1(2) of 1977 Protocol I Additional to the Geneva Conventions, which states as follows: 'In cases not covered by this Protocol or by other international agreements, civilians and combatants 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.' On the Martens Clause see Ticehurst, R., 'The Martens Clause and the laws of armed conflict', *International Review of the Red Cross*, no. 317 (30 Apr. 1997); and Asaro, P., 'Jus nascendi, robotic weapons and the Martens Clause', eds R. Calo et al., *Robot Law* (Edward Elgar Publishing: Cheltenham, 2016).

[104] Vautravers, A., 'Economic drivers: are the assumptions correct?', United Nations Institute for Disarmament Research Conference on Considering the Drivers of the Weaponization of Increasingly Autonomous Technologies, Geneva, Switzerland, 11 Nov. 2015.

[105] Fitzgerald, B. and Sayler, K., *Technology, Strategy and the Future of the Global Defense Industry* (Center for a New American Security: Washington, DC, 2014), p. 10.
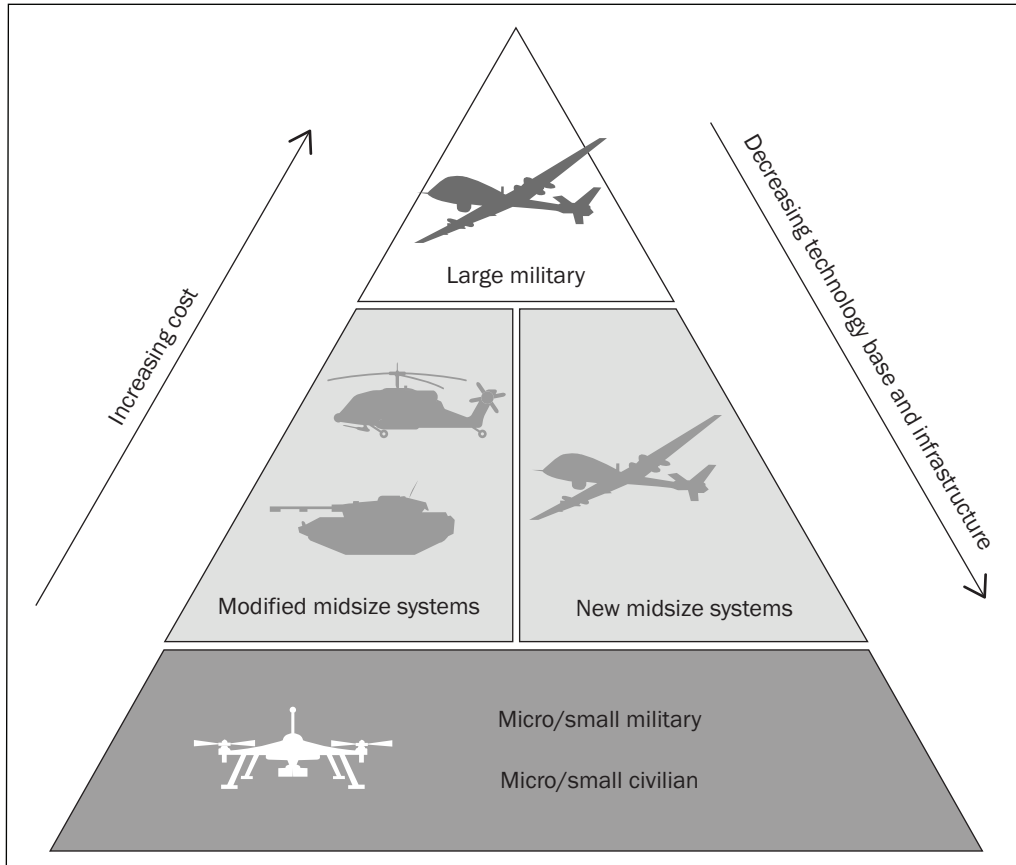
Figure 4.1. Weapon systems accessibility based on financial and organizational capital required for adoption

Source: Sayler, K. A., *World of Proliferated Drones* (Center for a New American Security: Washington, DC, 2015). Modified by the author.

tems depend upon the type of weapon systems. The rule of thumb is that the larger and more sophisticated weapon systems are, the more costly they are to acquire and the more infrastructure and skills they require to be operated. The extent to which their design is military specific or dual-use (i.e. a design that can be used both for military purposes and for civilian purposes) is also an important factor of cost and accessibility. Weapon systems that are heavily based on civilian technologies, notably off-the-shelf platforms, are much more affordable and accessible than systems that are highly military specific in their design.

Weapon systems can be divided into three categories depending on the entry barriers to their adoption by given actors (state or non-state). These are *high* for large advanced combat weapon systems, *medium* for middle-sized unmanned systems and modified legacy weapon systems and *low* for micro and small unmanned systems (see figure 4.1).

High entry barriers. Large weapon systems, such as large stealth drones, hypersonic glide vehicles and medium altitude and long-endurance UAVs, are highly military specific in their design. Only an exclusive group of states has the resources required to develop and operate them: states with the most mature defence and technology industrial bases and the highest military expenditures. These include the USA, China, Russia, the UK, France and Germany (see table 4.2). Some countries, such as Saudi Arabia, would theoretically have the financial resources to buy them off-the-shelf from another country, but they might not be able to do so due to export control policies of the producing countries and strategic considerations. These systems typically include

high-end components that are covered by strategic export control regulations, which often prevent their diffusion to all but a limited number of countries. In some cases, these regulations also limit the ability of the importing countries to use these systems in operations. French armed forces, for instance, have to ask for authorization from the US Congress before they deploy the Reaper drone in an operation.[106]

Medium entry barriers. Middle-sized unmanned systems are potentially accessible to a wider and growing number of countries, mainly thanks to the fact that many such systems are dual-use in their design and may be acquired off-the-shelf from the civilian sector. Major militaries usually acquire systems that are military specific in their design. Middle-sized unmanned systems still require significant infrastructure to be operated, which means that they can only be employed by major militaries. There is also the opportunity to modify legacy systems, including manned systems (e.g. helicopters, fixed-wing aircraft and transport vehicles) into unmanned systems using various types of add-on kits. This latter option has the advantage of providing new capabilities without having to engage in an entirely new development and production cycle (which usually takes up to 25 years).

Low entry barriers. Micro and small unmanned systems fall within the most accessible category of systems. They generally do not need a large support infrastructure, and can be developed using relatively inexpensive civilian off-the-shelf components or directly acquired as pre-assembled platforms from the civilian sector. The market for small robotic platforms, notably recreational and professional drones, has expanded dramatically in recent years. It is now possible to buy small remotely piloted UASs equipped with features such as GPS waypoint navigation and high-definition video camera sense-and-avoid capabilities for less than a few thousand US dollars.[107] These systems are not only accessible to nearly any state's military, but are also available to non-state actors and individuals.

In short, the accessibility of weapon systems with autonomous capabilities may vary widely: small and micro weapon systems could be adopted and used by virtually anyone, including lone-wolf terrorists, while large and highly capable autonomous weapon systems, such as stealth combat UASs, will remain the prerogative of a very small number of states.

It should also be noted that some of the few states that have a capability to develop, produce and operate very high-end weapon systems are currently dealing with some budgetary issues, which constrain their ability to engage, in the near term, in the acquisition of new, next-generation weapon systems.

The USA, the UK, France, Germany, Italy and Israel have seen their level of military expenditure stagnate or decline over the past decade, mostly as a result of austerity policies. At the same time, the unit production cost of weapon systems has continued to rise. This has led to a situation in which some states can no longer afford the development and acquisition of the latest or best designs of advanced weapon systems (e.g. combat aircraft) unless they engage in cooperation projects with other states (an issue that remains politically complicated). And even when states can afford to develop or acquire such systems, they are tending to procure steadily decreasing numbers of units (this phenomenon is commonly known as Augustine's Law in defence economics;

---

[106] Sayler, K. et al., *Global Perspectives: A Drone Saturated Future* (Center for a New American Security: Washington, DC, 2014).

[107] Sayler, K., *A World of Proliferated Drones* (Center for a New American Security: Washington, DC, 2015), p. 11.

Table 4.2. Countries with the highest military expenditure, 2006–15.

Rank	Country	Spending, 2015 ($ b.)	Change 2006–15 (%)	World share 2015 (%)	Spending as share of GDP (%)	
					2015	2006
1	USA	596	−3.9	36	3.3	3.8
2	China	[215]	132	[13]	[1.9]	[2.0]
3	Saudi Arabia	87.2	97	5.2	13.7	7.8
4	Russia	66.4	91	4.0	5.4	3.5
5	UK	55.5	−7.2	3.3	2.0	2.2
6	India	51.3	43	3.1	2.3	2.5
7	France	50.9	−5.9	3.0	2.1	2.3
8	Japan	40.9	−0.5	2.4	1.0	1.0
9	Germany	39.4	2.8	2.4	1.2	1.3
10	South Korea	36.4	37	2.2	2.6	2.5
11	Brazil	24.6	38	1.5	1.4	1.5
12	Italy	23.8	−30	1.4	1.3	1.7
13	Australia	23.6	32	1.4	1.9	1.8
14	UAE	[22.8]	136	[1.4]	[5.7]	[3.2]
15	Israel	16.1	2.6	1.0	5.4	7.5
Total top 15		**1 350**		81		
World total		**1 676**	19	100	2.3	2.3

[ ] = SIPRI estimate; GDP = gross domestic product; b. = billions; UAE = United Arab Emirates.
Source: Perlo-Freeman, S. et al., 'Trends in military expenditure, 2015', SIPRI Fact Sheet, Apr. 2016.

see box 4.2).[108] Another limiting factor is that the cost of legacy procurement projects still absorbs the vast bulk of these states' procurement budgets.[109]

Thus, although states like the USA, the UK and France have invested in technology demonstrators that could replace current legacy systems, notably manned combat aircraft (e.g. Dassault's nEUROn, BAE System's Taranis and Northrop Grumman's X-47B), it remains very unclear in the current budgetary situation whether and when they might lead to official acquisition programmes. A noteworthy development in this respect was the recent decision of the US Navy to abandon the first-of-a-kind semi-autonomous UCAS acquisition programme (known as the UCLASS programme: unmanned carrier-launched strike and surveillance aircraft) and retire the X-47B. To motivate its decision, the US Navy invoked budget constraints and the fact that it deemed it more valuable to reallocate the budget of the UCLASS to the acquisition of a larger number of manned aircraft (F/A-18E/F Super Hornet and F-35C Joint Strike Fighter).[110] Given the current context, the incorporation of autonomy in high-end weapon systems, such as combat aircraft, armoured vehicles and combat ships, is more likely to come through upgrades and modifications of legacy platforms undergoing a lifetime-extension programme than the creation of new platforms, at least in the coming two decades.

---

[108] That is, large weapon systems whose development started during the cold war and that are coming to maturity only now. Kirkpatrick, D., 'Trends in the costs of weapon systems and the consequences', *Defence and Peace Economics,* vol. 13, no. 3 (2004), p. 270.

[109] In the US case, that would include the F-35C Joint Strike Fighter and the C-130 Hercules tactical transport airlift aircraft. US Department of Defense (DOD), Office of the Under Secretary of Defense (Controller), *Program Acquisition Cost by Weapon System: United States Department of Defense Fiscal Year 2016 Budget Request* (DOD: Washington, DC, Feb. 2015).

[110] Freedberg, S., 'Good-bye, UCLASS: hello, unmanned tanker, more F-35Cs in 2017 budget', Breaking Defense, 1 Feb. 2016.

Bureaucratic barriers: inadequate acquisition processes

In addition to the issue of budgetary resources, there are also some practical issues related to the way the defence acquisition process works in most arms-producing countries.

That process, which was developed in the post-World War II era and which has seen relatively few changes since the end of the cold war, is reportedly ill-suited to the development of autonomous capabilities.[111] Its main weakness is that it is highly 'vehicle-centric'—that is, it revolves around the development of identifiable weapon platforms (e.g. aircraft, missiles, trucks and submarines)—and places significant emphasis on the hardware parts of systems; whereas the critical capabilities provided by autonomy are embedded in system software.[112] Hardware and software require, moreover, entirely different acquisition procedures. Typically, the acquisition process for hardware stretches over decades, while the development of software necessitates rapid acquisition cycles and regular updates. The model in place in most countries makes it such that software parts of new systems can, in some cases, be already obsolete by the time the systems enter into service. A related problem is that the acquisition programmes are frequently led by a single defence contractor. Generally, this contractor acts as a systems integrator and may often use proprietary standards that limit the ability to incorporate new software functionality or to upgrade the system using new software from different vendors.[113]

Defence acquisition processes also experience difficulty in quickly adopting new technologies from the commercial sector, which is clearly leading innovation in the fields of AI and robotics.[114] Most defence procurement agencies are still looking for a reliable method to assess which commercial innovations have military potential, and how to integrate them rapidly (i.e. how to test and refine them for military use).[115] At the same time, defence procurement processes continue to be characterized as excessively bureaucratic, which puts off most non-defence companies.[116] The long production cycles and the restriction on intellectual property rights are additional factors discouraging civilian companies, especially commercially successful companies, from developing solutions for the military sector.[117] There is also reluctance by some actors in the commercial sector to see the military benefiting from innovations in the fields of robotics and AI. Some companies openly refuse to sign contracts with the military for ideological reasons.[118] This is the case with Google DeepMind, for instance.[119] Other companies are believed to be concerned about the perception of consumers and the risk of bad publicity.[120]

---

[111] US Department of Defense, Defense Science Board (note 10), p. 10.

[112] US Department of Defense, Defense Science Board (note 10), p. 22.

[113] Gonzales, D. and Harting, S., *Designing Unmanned Systems with Greater Autonomy* (RAND Europe: Brussels, 2014), p. 24; and Fitzgerald, B. et al., *Open Source Software and the Department of Defence* (Center for a New American Security: Washington, DC, 2016).

[114] Fiott, 'A revolution too far? US defence innovation, Europe and NATO's military–technological gap' (note 17), pp. 9–10.

[115] On this topic see the speech by Chuck Hagel, US Secretary of Defense, at the Defence Innovation Days, Newport, Rhode Island, 3 Sep. 2014. The USA created in 2016 a Defense Innovation Unit whose mission is to facilitate the import of commercial innovation to the defence sector.

[116] Tama, J., *There's No App for That: Disrupting the Military–Industrial Complex* (Brookings: Washington, DC, 2015).

[117] Dyer, G., 'Robot soldiers', *FT Magazine*, 17 July 2015; and Tucker, P., 'As Pentagon dwindles, Silicon Valley sells its newest tech abroad', *Defense One*, 22 Apr. 2016.

[118] In 2015 the Future of Life Institute published an open letter against autonomous weapon systems. As of Feb. 2017, the letter had been signed by thousands of AI and robotics experts and researchers, including many of the leading figures from academia and the private sector. See Future of Life Institute (note 99).

[119] D'Onfro, J., 'Google's robot group struggles to fill leadership vacuum as it shoots for ambitious launch before 2020', *Business Insider*, 8 Nov. 2015.

[120] Mulrine, A., 'Pentagon cybersecurity strategy comes with olive branch to Silicon Valley', *Christian Science Monitor*, 23 Apr. 2015.

Box 4.2. The limits of affordability: Augustine's Law and the escalating cost of weapon systems

Augustine's Law is a famous observation made by Norm Augustine before he became President of Lockheed Martin. It refers to the fact that the unit cost of certain high-technology equipment increases at an exponential rate with time, forcing cuts in production numbers. Finding that the unit cost of combat aircraft was rising by a factor of four every 10 years, while military budgets were increasing only linearly, he humorously warned that: 'In the year 2054 the entire defense budget will purchase just one tactical aircraft. This aircraft will have to be shared by the Air Force and Navy 3½ days each per week except for leap year, when it will be made available to the Marines for the extra day.'

  Augustine's Law is still valid today. An analysis of some 30 classes of weapon systems has shown that their unit production costs have grown in most cases at 5–10 per cent each year since the end of World War II. For mature and less complex systems, such as rifles and machine guns, the rate of growth has been lower; for systems including high-end electronics (e.g. anti-tank helicopters) the rate of growth has been higher. The fundamental consequence of the rise in unit cost is that it pushes down procurement quantities. Defence economist David Kirkpatrick reported that the total number of aircraft in the British Royal Air Force declined by 80 per cent between 1954 and 1993. However, these reductions are deemed by some to be offset by improved performance.

Sources: Augustine, N., *Augustine's Law and Major System Development Programs*, revised and enlarged 2nd edn (American Institute of Aeronautics and Astronautics: New York, 1983), p. 50; Kirkpatrick, D., 'Trends in the costs of weapon systems and the consequences', *Defence and Peace Economics*, vol. 13, no. 3 (2004), pp. 259–73; and Hartley, K., 'The case for defence', *Defence and Peace Economics*, vol. 21, nos 5–6 (2010), pp. 409–26.

   None of the issues represents an insurmountable obstacle to the further incorporation of autonomous capabilities in weapon systems, but they clearly slow down the process. Defence acquisition bureaucracies are famous for their inertia and this is unlikely to change in the near future. A notable illustration of this is the difficulty that the USA is facing in reforming its acquisition systems, despite the determination of the US DOD's leadership. The reforms that Work initiated as part of the Third Offset Strategy to facilitate innovation, and specifically to address the aforementioned limitations, have reportedly had very little success so far.[121]

IV. Conclusions

The military has multiple reasons to accelerate the incorporation of autonomy into weapon systems. Autonomy has the potential to make weapon systems achieve greater speed, accuracy, persistence, reach, coordination and mass on the battlefield. This will not necessarily result in a linear development towards full autonomy, with the 'man being taken out of the unmanned' altogether, so to speak. There are a number of technical, institutional, legal, normative and economic challenges associated with the development of autonomy which mean that the military will, at least in the near and middle term, continue to want and need to exert control over the weapon systems used by armed forces.

   The main conclusion with regard to the CCW discussion is that the focus on 'fully autonomous weapon systems' is problematic, as it does not reflect the reality of how the military is envisioning the future of autonomy in weapon systems, nor does it allow for discussion of the full spectrum of challenges raised by the progress of autonomy in weapon systems in the near term. Autonomy will transform the way humans interact with weapon systems and make decisions on the battlefield, but will not eliminate their role. Therefore, it might be more helpful if the CCW debate moves away from the question of whether fully autonomous weapon systems are possible and potentially legal or acceptable, and starts investigating the following fundamental questions.

---

[121] Fitzgerald, B. and Dejonge Schulman, L., *12 Months In—8 Months Left: An Update on Secretary Carter's Innovation Agenda* (Center for a New American Security: Washington, DC, 2016).

1. How are the advances in autonomy changing the nature, location and timing of human decision making and action in warfare and why does this matter?

2. What control do we expect humans to maintain over the weapon systems they use?

3. How can we ensure that human control remains adequate or meaningful as weapon systems possess increasingly complex and autonomous capabilities?

5. Where are the relevant innovations taking place?

I. Introduction

A central issue of concern in the CCW discussions is the impact that a potential regulation or ban on LAWS would have on innovation, notably in the civilian sphere, considering that much of the technology on which these systems might be based could be dual-use (i.e. capable of being used for both civilian and military purposes).[1] The empirical foundations of these concerns are the focus of this chapter. It sheds light on the 'innovation ecosystem' that is driving the development of autonomy in weapon systems. Specifically, it maps out where relevant innovations are taking place from three different perspectives (*a*) a science and technology perspective (the field of R&D); (*b*) a geographical perspective (the location of key R&D institutions); and (*c*) a sector perspective (whether innovation is driven by civil or military research). It should be emphasized that the purpose here is neither to advocate, nor argue against, the development of a new protocol; rather, the purpose is to set an unbiased baseline for future discussions on the feasibility and impact of a protocol dedicated to LAWS.

Section II provides a brief background on innovation and discusses the challenges associated with mapping innovation in machine autonomy. Sections III, IV and V discuss relevant developments within academia, state-funded R&D and the private sector, respectively. The concluding section (section VI) summarizes the key findings and presents some conclusions for future CCW discussions.

II. What is innovation and why is it difficult to track in the context of machine autonomy?

Innovation in the context of autonomy is a complex issue. Thus, it is useful as a first step to clarify some of the essentials of the R&D process through which new technologies, particularly military technologies, are generally created. This section also reviews the methodological challenges associated with mapping innovation in the area of machine autonomy.

The essentials of innovation

Types of R&D efforts

Innovation typically results from formal R&D efforts, which can broadly be divided into three categories: basic research, applied research and experimental development. These three categories can be summarized as follows.

1. *Basic research* is about advancing the state of science and knowledge at the fundamental level, through theoretical or experimental inquiry.
2. *Applied research* is about researching new methods and techniques to address concrete socio-technical problems. In contrast to basic research, applied research pursues a concrete objective.
3. *Experimental development* builds on the findings of basic research and applied research to improve existing, or develop new, technology—be it new materials, products, systems or services.

---

[1] The concept of 'innovation' refers both to the process, and the result of, technological development that leads to the creation of new methods, ideas or products.

The last phase is crucial as far as the development of marketable technologies is concerned—be they civilian or military technologies. It is during this phase that new methods, ideas or products possibly take their final shape. Thus, from a regulatory point of view, it is the phase that is the most important to monitor and control. However, the importance of basic and applied research should not be underestimated, as major technological breakthroughs cannot happen without basic and applied research.

The innovation ecosystem: key players and their relationship

R&D efforts leading to innovation can be conducted within any of the following three types of institutions: (*a*) university research laboratories; (*b*) government civilian and military funding agencies and research laboratories; and (*c*) private sector laboratories.

University research laboratories are usually more concerned with research than development. Their primary role is to advance science and technology through fundamental research. In contrast to the research labs of private companies, their research efforts do not necessarily need to translate into innovations that may be monetized. Nonetheless, universities in many countries are increasingly involved in various forms of collaboration with industry to advance technologies with potentially marketable applications—including military applications. In addition, in some countries, notably in the USA and China, it is not uncommon that university labs receive military funding to work on R&D projects that are of interest to the armed forces. Their involvement, however, rarely goes beyond the applied research phase. Some universities have internal rules that specifically limit their ability to participate in weapons development. MIT, for instance, allows its researchers to receive military funding only for basic and applied research. Stanford University has a policy stating that research professors may hire students of any nationality to work on research projects. Because there are likely to be additional nationality restrictions on persons working on sensitive military R&D projects, research teams from Stanford University may be prevented from participating if those teams include students from certain states.

The restrictions on research conducted by universities is one of the reasons why academic researchers sometimes establish businesses to run alongside their academic activities. By forming a private company, researchers can receive further funding for experimental development or directly exploit commercially the findings of their academic research. One of the most notable examples is Boston Dynamics, which was founded in 1992 by Marc Raibert as a spin-off of the MIT Leg-Lab—a research group then headed by Raibert, focused on the development of self-balancing legged robots. Since then, Boston Dynamics has received multiple R&D contracts with the US DOD, which led to the development of advanced and widely discussed prototypes of four- and two-legged robots.[2]

Thus, the contribution of university labs to innovation is not limited to delivering research that tackles generic socio-technical problems; such labs also serve as incubators for talents and ideas that can then grow in the private sector.

Government civilian and military funding agencies and research laboratories focus usually on basic and applied research, but in some cases they can also conduct experimental development projects.[3] The types of applied research that they commonly fund or conduct can be subdivided into two categories: strategic applied research and specific applied research. Strategic applied research projects have a purely prospective

---

[2] It should be noted that Boston Dynamics was acquired by Google in 2013, which indicated following the acquisition that Boston Dynamics would complete its contractual engagement with the US Department of Defense but would not seek new military contracts.

[3] In Europe, the CNRS (France's National Center for Scientific Research), the Max Planck Institutes and others are typically basic research driven. France's CEA (Alternative Energies and Atomic Energy Commission) and Germany's Fraunhofer Institute are more application driven.

nature and are meant to explore technological developments over very long time frames (e.g. 10–50 years). Specific applied research projects are directed towards near-time and specific innovation. Typically, they test and demonstrate the usefulness of new technologies in the light of short-run requirements (e.g. 5–10 years).

Historically, government funding and research institutions, notably military funding agencies and military research laboratories, have played a key role in the development of game-changing technology, for the simple reason that they were able to invest in R&D projects that neither academia nor private companies were willing or able to support or articulate alone. DARPA in the USA is perhaps the best known of these institutions. DARPA has long been a leader of innovation in many areas of science and technology, including AI and robotics—the two fields of science and technology that are essential to the development of autonomy (see section III of this chapter). A number of DARPA projects have not only resulted in entirely new techniques or ways of doing things, but have also resulted in disruptive innovations (i.e. innovations that create new markets or disrupt existing markets) such as stealth technology, GPS and the Internet. One of DARPA's distinguishing features is that it is somewhere between a funding agency and a research agency. It designs, funds and oversees projects, but it 'outsources' the actual research process to academic institutions and private companies. This model has been particularly effective at facilitating the deployment of innovation to the marketplace. A number of countries, including Russia, Japan and—more recently—China, have attempted to reproduce the DARPA model.[4]

Private sector laboratories may conduct all types of R&D efforts, but their focus is mostly on the development part of R&D. When they conduct basic and applied research, it is usually with the intention of developing marketable products and services. Here, it is worth underlining that there are notable differences in the ways in which commercial companies and defence companies research and develop new technologies.[5] These differences generally derive from the fact that defence companies, at least companies that are specialized in arms production, and civilian companies operate in fundamentally different market conditions. Civilian companies have to invest heavily in R&D to remain competitive. Innovation allows them to attract customers and gain or retain market share. Defence companies, on the other hand, operate in a market characterized by monopsony (i.e. a market where there is only one customer: the state).[6] Their R&D efforts are therefore largely determined by the evolution of government demand. They need to adapt their research agenda to priorities that are set, and volume resources made available, by the military. In general, they do not invest in the development of new military products or services without the guarantee that they will be able to sell them.[7] This is not to say that some defence companies do not make significant self-funded R&D efforts, but these are usually of a smaller scale than those of their commercial counterparts, and they often reflect a conservative approach towards developing new technology.[8] An appreciation of these differences is essential to an understanding of why defence companies may appear less proactive

[4] Xin, H., 'China to create its own DARPA', *Science Magazine*, 11 Mar. 2016; Beckhusen, R., 'Putin wants a DARPA of his own', *Wired*, 25 June 2012; and Reuters, 'Japan to tap technology for military use, in another step away from pacifism', *Financial Express*, 14 Nov. 2013.

[5] Tama, J., *There's No App for That: Disrupting the Military–Industrial Complex* (Brookings: Washington, DC, 2015), p. 27.

[6] It is very rare for defence companies to develop new and capital-intensive projects for governments other than the one in the country in which they primarily operate.

[7] Defence companies usually make states cover most R&D costs associated with the production of new military technologies. Sköns, E., *The Globalization of the Arms Industry*, PhD Dissertation (Bradford University: Bradford, 2009), p. 45.

[8] According to the research firm Capital Alpha Partners, the combined R&D budgets of 5 of the largest US defence contractors (about $4 billion) amounts to less than half of what companies such as Microsoft or Toyota spend on R&D in a single year. Lynn III, W. J., 'The end of the military–industrial complex', *Foreign Affairs*, Nov./Dec. 2014.

than their commercial counterparts in the area of machine autonomy—something that will be discussed further in section V of this chapter.

Innovation ecosystem. Together, university labs, government research agencies and private sector laboratories form what is often referred to as an 'innovation ecosystem'. Some economic studies have shown that a state's ability to deliver high-quality innovation is concomitant with its ability to create or facilitate interaction between these different institutions, be that in terms of exchange of fundamental knowledge, personnel or funding.[9]

The relationship between civilian and military innovation

When discussing the relationship between civilian and military innovation, it is important to note that 'innovation' has two meanings: it may refer both to the process of, and the result of, technological development.

Innovation that results from military R&D (i.e. R&D that is funded or conducted by military research institutions) can find applications in the civilian sphere and vice versa.[10] What determines whether the result of innovation is both a military and civilian (i.e. dual-use) technology is its end use.[11]

In many areas of science of technology there is nothing fundamental at the basic and applied research level to determine whether a certain area is civil or defence oriented.[12] This is particularly the case for most enabling technology areas such as electronics, computer programming and advanced materials. The divergence between civilian and military innovation generally emerges during the development stage of the R&D cycle, as it is during that stage that the end-user requirements are factored in. One well-established difference between the military and the civilian sectors is the fact that the military end user often places greater emphasis on performance, survivability and reliability of the technology than on aesthetics and cost, while the civilian end user might focus on cost-limitation, user-friendliness or aesthetics.[13] This is especially true of final systems such as vehicles or ICT.

Thus, contemporary military technologies, even weapon systems, rarely originate only from 'pure' military research efforts; rather, they result from developments in both civilian and military R&D that have synthesized in military applications.[14] This trend is not new, but has been increasing rapidly over the past 20 years thanks to the growing role played by electronics and ICT in the design of military systems. Electronics and ICT are prime examples of dual-use technology, whose development has been chiefly driven by the commercial sector for decades.

The key conclusion from this brief introduction is that to understand the dynamics of innovation in military technology, it is useful to consider the entire R&D cycle, not just the phase of experimental development that primarily takes place within the industry. The remaining sections of this chapter aim to map out and analyse the relevant R&D efforts carried out within academia, government research agencies and the

[9] Wadhwa, V., 'Silicon Valley can't be copied', *MIT Technology Review*, 3 July 2013; and Dutta, S., Lanvin, B. and Wunsch-Vincent, S. (eds), *Global Innovation Index 2016* (Cornell University, INSEAD and World Intellectual Property Organization: Ithaca, NY, Fontainebleau and Geneva, 2016).

[10] There is a large volume of literature discussing, in depth, the relationship between civilian, military and dual-use innovation. Most of this literature was published around the end of the cold war, when a number of experts explored the possibility of diverting military resources to the civilian sector. Since then, a relatively limited number of academic studies have been published on the topic. Carter, A. et al., *Beyond Spin-Off: Military and Commercial Technologies in a Changing World* (Harvard Business School: Boston, MA, 1992).

[11] Cowan, R. and Foray, D., 'Quandaries in the economics of dual technologies and spill-overs from military to civilian research and development', *Research Policy*, vol. 24, no. 6 (1995), p. 851.

[12] Davis, I., *Military R&D in Europe, Collaboration Without Control?* (Oxford Research Group: Oxford, 1992), p. 11.

[13] Kaldor, M., *The Baroque Arsenal* (Hill and Hang: New York, 1981); and Dunne, P., 'Defense industrial base', eds K. Hartley and T. Sandler, *Handbook of Defense Economics*, vol. 1 (Elsevier: Amsterdam, 1995).

[14] Davis (note 12), p. 11.

private sector. Before continuing with this mapping exercise, it is first necessary to discuss the extent to which it is actually feasible to map innovation in the context of autonomy.

Innovation and autonomy

Mapping innovation in machine autonomy poses a major challenge from a methodological standpoint. Autonomy has no established definition. It is not a specific technology area with well-defined boundaries, or a dedicated academic discipline or distinct market sector.[15] Autonomy is not even technology per se; rather, it is a property that can be attached to very different types of technology.

Moreover, as explained in chapter 2, while machine autonomy is always made possible by the integration of the same types of enabling technologies, the characteristics of these enabling technologies vary significantly depending on their relevance to the applications and capabilities of interest. This means that, even in the context of military weapon systems, the underlying technological architecture may vary within and between systems, depending for instance on the nature of the tasks that are executed, the weapon system's mission and the nature of the operating environment. Therefore, it is not feasible to capture and discuss in a single study all the technological developments that may be relevant to advances of autonomy in weapon systems.[16]

To make the scope of this study more manageable, an emphasis has been placed on the development of software technologies that allow autonomous weapon systems or subsystems to feature greater perception and decision-making capabilities. Some developments related to hardware components, such as sensor technology and computer processor technology, will be discussed briefly because they are in some cases directly relevant to the performance of software technologies.[17]

III. A science and technology perspective: autonomy and academia

This section maps out the networks of research disciplines and research issues that are involved directly (or in some cases indirectly) in the development of autonomous capabilities in weapon systems. It also provides an overview of the global academic landscape in this area and identifies the locations of the world's leading academic research institutions in this field.

Core disciplines

At the basic science and technology level, advances in machine autonomy derive primarily from research efforts in three disciplines: AI, robotics and control theory (see figure 5.1).

Artificial intelligence

According to John McCarthy, who coined the concept in 1955, AI can be broadly defined as the 'science and engineering of making intelligent machines'.[18]

---

[15] 'Autonomy' is defined here as the ability of a technology to execute a task, or tasks, without human input, using interaction of computer programming with the environment. This definition is based on one previously proposed by Andrew Williams. Williams, A., 'Defining autonomy in systems: challenges and solutions', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO: Norfolk, VA, 2015).

[16] US Department of Defense (DOD), Office of Technical Intelligence, Office of the Assistant Secretary of Defense for Research and Engineering, *Technical Assessment: Autonomy* (DOD: Washington, DC, Feb. 2015), p. 24.

[17] For a more detailed discussion on what autonomy is and how it is created see chapter 2 of this report.

[18] Dale, R., 'An introduction to artificial intelligence', ed. A. Din, *Arms and Artificial Intelligence* (SIPRI/Oxford University Press: Oxford, 1987), p. 33.
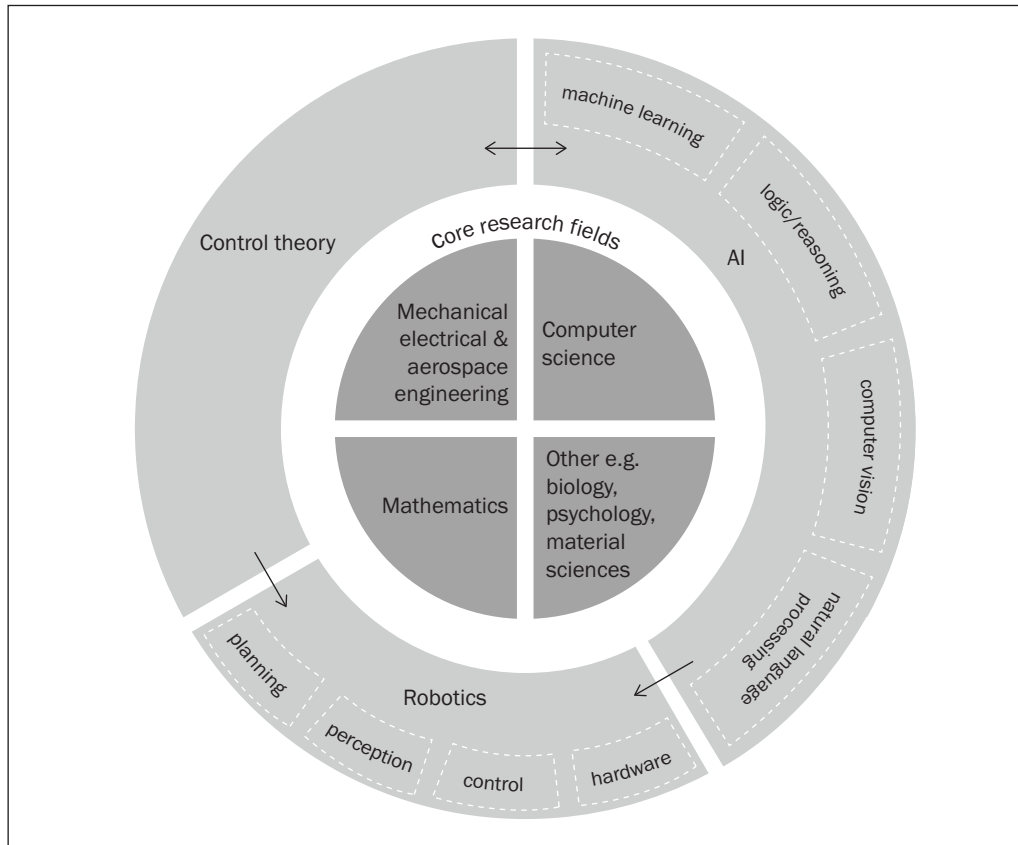
Figure 5.1. Major research fields in machine autonomy

As an academic discipline, AI mainly falls within computer science. The centre of gravity of AI research is difficult to delineate satisfactorily, partly because the concept of AI means different things to different people, and partly because its subject matter, intelligence, is hard to define. Historically, core AI research has focused on problem solving through logic and reasoning. Many researchers and engineers continue to think of AI in those terms. Others see it as an umbrella term that covers all the research issues associated with making machines do tasks that humans label as intelligent (e.g. observing the world through vision, learning and natural language processing).[19]

One distinction worth mentioning here is the difference between specialized AI (weak AI) and AGI (strong AI) (see box 5.1). Most current research relates to the development of specialized/weak AI. AGI has always fascinated AI researchers, but it remains a fundamental technical challenge. There are, in fact, strong disagreements as to whether it would even be possible to design AGI computer programs.[20]

Robotics

Robotics is a field of science and engineering that is dedicated to the development of robots (i.e. computer-enabled machines that can sense and purposefully act on or in their environment).[21] As an academic discipline, robotics is at the crossroads between mechanical engineering, electrical engineering and computer science.

---

[19] See e.g. Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Pearson Education: Harlow, 2014).

[20] Dileep, G., 'Killer robots? Superintelligence? Let's not get ahead of ourselves', *Washington Post*, 4 Nov. 2015; and Adams, T., 'AI: we are like small children playing with a bomb', *The Guardian*, 12 June 2016.

[21] Winfield, A., *Robotics: A Very Short Introduction* (Oxford University Press: Oxford, 2012); and 'Why is it so difficult to define "robot?"', Robohub, 29 Apr. 2016.

Broadly speaking, R&D in robotics falls into one of two generic categories. The first category consists of R&D efforts that mainly focus on development and integration of the hardware parts of robots, notably the actuators and the end-effectors.[22] These efforts aim to improve, for instance, the agility, endurance, flexibility, hardiness, size or velocity of robots. This category includes specific subfields of research such as soft robotics (which covers the construction of robots that are made from soft and transformable material) or nano-robotics (which covers robots that range in size from 0.1 to 10 micrometres and are constructed of nanoscale or molecular components).

R&D efforts in the second category mainly focus on the development of the software parts that control the robot's behaviour. These can be further divided into two subcategories: (*a*) those that seek to improve the ability of humans to remotely control the behaviour of the robot (e.g. through haptic control); and (*b*) those that seek to develop robots capable of governing their own behaviour. The latter subcategory is a fundamental part of research in the area of machine autonomy and is where the robotics and AI disciplines directly overlap. The terms 'AI robotics', 'cognitive robotics' or 'autonomous robots research' are sometimes used in, or to refer to, this area of robotics research.[23] Basic research areas that are shared by the AI and robotics research community and that are of key importance to the development of autonomy include the following.

1. *Computer vision*. The development of computers and robots capable of acquiring, processing, analysing and understanding visual data.

2. *Natural language processing*. The development of computers and robots capable of acquiring, processing, analysing and generating human language.

3. *Machine learning*. The development of computers and robots capable of adapting to their environment and improving performance based on past experiences and training rather than a pre-programmed model of the world.

4. *Search and planning*. The development of computers and robots capable of developing or adapting plans of action to achieve desired goals.

5. *Logical and symbolic reasoning*. The development of computers and robots capable of reasoning and drawing inferences from a database of facts and logical rules.

6. *Human–machine interaction*. The development of improvements to the way in which humans and machines (either computers or robots) work together.

7. *Manipulation*. The development of robotic systems capable of manipulating physical devices in a precise way (e.g. so that the correct level of pressure is applied to an object when grasped).

8. *Locomotion*. The development of legged robotic systems capable of autonomous motion control.

9. *Human–machine interaction*. The development of improvements to the way in which humans and machines (either computers or robots) work together.

10. *Collaborative intelligence*. The development of several individual machines capable of completing a task collectively (e.g. as a swarm).

11. *Validation and verification*. The development of methods to ensure that intelligent systems satisfy certain desired formal properties and meet formal requirements (i.e. that they do not have unwanted behaviours or consequences).

[22] End-effectors are the physical devices that assert physical force on the environment: wheels, legs and wings for locomotion, as well as grippers and, of course, weapons. Actuators are the 'muscles' that enable the end-effectors to exert force and include things such as electric motors, hydraulic cylinders and pneumatic cylinders.

[23] Khamassi, M. and Doncieux, S., 'Nouvelles approches en robotique cognitive' [New approaches in cognitive robotics], *Intellectica*, vol. 1 (2016); and Murphy, R., *Introduction to AI Robotics* (MIT Press: Cambridge, MA, 2000).

Box 5.1. Artificial general intelligence versus specialized artificial intelligence

In the artificial intelligence (AI) community, the concepts of *artificial general intelligence* (AGI) and *strong AI* refer to a general-purpose AI that would be as intelligent as, or even more intelligent than, humans. A system with AGI would be able to make sense of the world itself and develop its own meaning for the environment it encounters. AGI does not currently exist and remains for now in the realm of science fiction. Currently, *specialized AI* or *weak AI* is the only type of AI technology in existence. A system with specialized AI can make complex decisions based on reasoning and past sets of data, but needs to be trained and pre-programmed for specific applications. Such systems have no capability to think beyond the limits of their programming.

Each of these research areas constitutes a separate subfield of academic research. Most of them have dedicated academic conferences, as well as research teams with university labs. It should be noted that members of the research community vary significantly depending on the subfield under consideration. V&V, for instance, is a topic that is covered by a relatively small community of scholars when compared with topics such as computer vision or manipulation. In the subfield of machine learning—an area that is of growing importance to the development of intelligent systems—and to a lesser extent in the subfield of robotics, academics are being lured away from universities to private companies, many of which have a vested interest in the development of machine learning.[24] Reportedly, the most renowned scholars in machine learning are now primarily affiliated with private companies rather than universities.[25]

Control theory

A third academic discipline of central importance to the development of autonomy is control theory. Control theory is an interdisciplinary branch of engineering and mathematics that deals with the behaviour of dynamical systems. It is, in this respect, relevant to nearly all fields of mechanical and electrical engineering. Robots, cars, aircraft engines, submarines and assembly lines all fundamentally rely on control systems—hence control theory—to function properly.[26]

Control theory provides some of the theoretical foundations to the development of automation and autonomy, and robotics technologies more largely. Its most important contribution is the principle of closed-loop feedback control. Closed-loop feedback control means that the systems possess a monitoring feedback that allows them to continuously correct their output. The feedback is created by a sensor that measures the system's actual output and a controller that calculates adjustments to keep the measured variable within a desired set range. All autonomous systems or autonomous functions that execute physical force in their physical operating environment (e.g. self-driving vehicles and autopilots in aircraft) use closed-loop feedback control. The control strategies may vary, however, from one system to another. The main control techniques in control theory include hierarchical control, adaptive control, intelligence control and optimal control. Key research topics in control theory include stability, controllability and observability, control specification, and model identification.

The relationship with other academic disciplines at the basic and applied research levels

Academic research in the fields of AI and robotics is essentially an interdisciplinary pursuit. This is particularly notable at the basic and applied research levels, where

[24] Hernandez, D. and King, R., 'Universities' AI talent poach by tech giants', *Wall Street Journal*, 24 Nov. 2016; and Ramsey, M., 'Carnegie Mellon reel after Uber lure away researchers', *Wall Street Journal*, 31 May 2015.

[25] Levy, S., 'How Google is remaking itself as "machine learning first company"', *Backchannel*, 22 June 2016.

[26] Yamamoto, Y., 'Control systems are ubiquitous', IEEE Control Systems Society, [n.d.].

researchers often attempt to connect with, and learn from, many scientific disciplines, including biology, psychology and linguistics.

Biology

Interest in biology derives from the fact that the natural world has been, and continues to be, a central source of inspiration for AI and robotics scholars. In the field of AI, many researchers are seeking to draw upon recent discoveries in neuroscience about the structure and functions of the human brain. AI researchers are particularly interested in exploiting that knowledge to generate advances in machine cognition, notably with regard to learning and decision making.[27] The connection with biology is even more palpable in the field of robotics. The shape and behaviour of robots are often inspired by the shape and behaviour of natural bodies. Iconic examples include the legged robots developed by Boston Dynamics. These were developed based on research on animal locomotion conducted by the company's founder while he was head of the MIT's Leg-Lab.[28] The current development of swarm robotics also builds heavily on biological research into swarm intelligence in the animal world.[29]

Psychology

The AI and robotics research community works in close cooperation with researchers in psychology. Human psychology and cognition provide important benchmarks for AI and robotics researchers who are modelling the behaviour and cognitive abilities of intelligent computers and robots.

Linguistics

The AI and robotics research community's interest in linguistics is driven by two key factors: (*a*) improving the ability of machines to process natural language; and (*b*) gaining an understanding of how language is structured, which could help to unravel some of the more complex aspects of human brain function as communication through language is one of the most complex of all human activities.[30] Basic and applied research in AI and robotics that bridge with linguistics can, for instance, aim to improve knowledge representation and reasoning within computers and robots.

Leading university laboratories

Currently, there are no worldwide university rankings focusing on both AI and robotics that enable an assessment of the leading university labs in these research areas. SIPRI used a simple indicator, the volume of affiliated publications in relevant subject matters, to obtain a broad impression of the global academic landscape and the locations of key academic research institutions. This is certainly an imperfect benchmark, as the number of publications neither reflects the quality nor the impact of the research. Also, it tends to give greater importance to large universities. Nevertheless, it gives some idea of where productive universities are located.

   To this end, the Microsoft Academic Search Index (MASI) has proved to be a useful tool, as it references research publications (it also includes labs operated by private companies) on a number of key AI-related topics—AI in general, machine learning,

---

[27] Potter, S., 'What can AI get from neuroscience', eds M. Lungarella et al., *Fifty Years of AI* (Springer Verlag: Berlin/ Heidelberg, 2007); van der Velde, F., 'Where artificial intelligence and neuroscience meet: the search for grounded architectures of cognition', *Advances in Artificial Intelligence* (2010), pp. 1–18; and Khamassi and Doncieux (note 23).

[28] Knight, W., 'Robots running this way', *MIT Technology Review*, 3 June 2014.

[29] Tan, T. and Zheng, Z-Y., 'Research advances in swarm robotics', *Defence Technology*, vol. 9, no. 1 (Mar. 2013), pp. 18–39.

[30] Rosenberg, R., 'Artificial intelligence and linguistics: a brief history of a one-way relationship', *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (1975), pp. 379–92.

human–machine interaction, natural language processing and computer vision—and uses that data to rank the top 10 universities in each of these topic areas. The rankings are listed in the appendix of this report. Unfortunately, MASI does not provide similar rankings for robotics and autonomous systems. The 15 research institutions that were the most often referenced in MASI's publications database for the period 2000–16, using the keywords 'autonomous systems', 'robotics' and 'mobile robots', are listed in the appendix.[31]

The key lesson learned from these rankings is that, in each topic area, the academic landscape is largely dominated by US universities, most notably Carnegie Mellon University, Stanford University, MIT and the University of California, Berkeley. Outside of the USA, universities that are the most productive on these topics are based in Western Europe, South Korea and China.

IV. A geographical perspective: state-funded R&D

Assessing R&D efforts of the largest arms-producing countries

SIPRI attempted to map autonomy-related state-funded R&D of the largest arms-producing countries—namely the USA, the UK, Russia, France, Italy, Japan, Israel, South Korea, Germany, India and China (based on SIPRI data on defence companies' arms sales and national levels of military expenditure; see table 5.1).[32] This has proved to be challenging for two reasons: (*a*) open source information about national military R&D is very often scarce (the USA and countries in the European Union, EU, being notable exceptions); (*b*) the science and technology foundations of autonomy are, as previously discussed, very diffuse. Therefore, the following review discusses each state's R&D efforts on AI and robotics generally. It presents countries' strategies, policies and budgetary efforts on AI and robotics (up to March 2017), in both the civilian and military spheres, and considers, when possible, R&D development related to autonomy.

The United States

The USA pioneered investment in AI and robotics R&D in the 1950s. From the beginning, the US DOD has played a key role in setting priorities and channelling state R&D funding in these areas of technology.[33] As discussed in the previous chapter, the DOD's interest in AI and robotics has waxed and waned over time but it has never stopped investing in R&D in these areas, including autonomy-related applications.[34] Following the publication of Third Offset Strategy research, the DOD released for the first time an articulated roadmap and a consolidated budget for R&D in autonomy. The DOD roadmap identified four areas of priority: (*a*) improving human–autonomous systems interaction and collaboration; (*b*) advancing machine perception, reasoning and intelligence; (*c*) developing scalable teaming of autonomous systems; and (*d*) creating new test, evaluation and V&V procedures for adaptive autonomous systems. The

---

[31] Microsoft Academic Search Index, accessed 9 Dec. 2016, <https://academic.microsoft.com>.

[32] Based on the share of arms sales of companies listed in the SIPRI Top 100 for 2014. The SIPRI Top 100 lists the world's 100 largest arms-producing companies and military services companies (excluding those based in China). These are ranked by volume of arms sales. While not covered by the SIPRI Top 100 due to the lack of data on arms sales, China is also considered as one of the largest arms-producing countries. SIPRI considers that at least 9 of the 10 major state-owned conglomerates under which the Chinese industry is organized would be listed in the Top 100 if official data was available. Fleurant, A. et al., 'The SIPRI Top 100 arms-producing companies and military services companies, 2014', SIPRI Fact Sheet, Dec. 2015.

[33] The US DOD is a major source of state funding into R&D in general, as it controls half of the federal R&D budget. It funds everything from basic research (especially by universities) and applied research (especially by industry) to complete system development.

[34] McCorduck, P., *Machines Who Think* (A. K. Peters: Natick, MA, 2004), pp. 430–31.

Table 5.1. Government research and development (R&D) spending in the 10 largest arms-producing countries and China

Country	Arms production data (SIPRI)	Military expenditure data (SIPRI)	Government budgets on R&D (OECD)		
	Share of arms sales in the SIPRI Top 100 for 2014 (%)	Military expenditure 2014 $ b.	Military R&D 2014 $ b. (constant 2010)	Total R&D $ b. (constant 2010)	Share of military R&D in total R&D (%)
USA	*54.4*	596.0	64.4	126.8	*50.8*
UK	*10.4*	55.5	2.3	13.7	*16.9*
Russia	*10.2*	66.4	. .	19.6	. .
France	*5.6*	50.9	1.1	16.7	*6.6*
Italy	*3.0*	23.8	0.1	10.3	*0.9*
Japan	*2.3*	40.9	1.5	33.3[c]	*4.4*
Israel	*1.9*	16.1	. .	1.6[d]	. .
South Korea	*1.7*	36.4	2.7	20.3	*13.5*
Germany	*1.6*	39.4	1.2	29.9	*3.8*
India	*1.2*	51.3	. .	. .	. .
China	. .[a]	(215.0)[b]	. .	. .	. .

. . = not available or not applicable; b. = billions; OECD = Organisation for Economic Co-operation and Development.

[a] Chinese companies are not covered by the SIPRI Top 100 due to the lack of data on which to make a reasonable estimate of arms sales for most companies. Nonetheless, some information is available on the 10 major state-owned conglomerates under which most of the Chinese arms industry is organized. Based on the overall industry picture and on limited information on individual companies, at least 9 of these 10 companies would almost certainly be in the Top 100 if figures for arms sales were available. Of these, 4–6 would probably be in the top 20, and 2—the aircraft producer AVIC and the land systems producer Norinco—may be in the top 10.

[b] SIPRI estimate.

[c] Military figure based on underestimated data.

[d] Figure does not include military R&D.

Sources: Fleurant, A. et al., 'The SIPRI Top 100 arms-producing companies and military services companies, 2014', SIPRI Fact Sheet, Dec. 2015; Perlo-Freeman, S. et al., 'Trends in military expenditure, 2015', SIPRI Fact Sheet, Apr. 2016; and Organisation for Economic Co-operation and Development (OECD), Statistics Database on Research and Development, <http://www.oecd.org/innovation/inno/researchanddevelopmentstatisticsrds.htm>.

DOD spent $149 million on these priority areas in 2015 (see figure 5.2) and a total of $18 billion was earmarked for continued investment in autonomy for 2016–20.[35]

The DOD's internal research agencies were the primary recipients of these funds. DARPA received the largest share (29 per cent) followed by the Office of Naval Research (25 per cent), the US Army Research Laboratory (13 per cent) and the Air Force Research Laboratory (5 per cent). The remaining 28 per cent was allocated directly to R&D within universities and private companies.[36] The research agendas of DARPA, the Office of Naval Research, the Army Research Laboratory and the Air Force Research Laboratory are somewhat different. DARPA traditionally focuses on fundamental research and so-called moonshot developmental projects (i.e. ambitious exploratory projects undertaken without any expectation of near-term benefit). Its autonomy-related R&D projects are all primarily intended to explore middle- and long-term capabilities. Many of them, if successful, could deliver important advances in autonomy in weapon systems. Notable projects include the following.

[35] Bornstein, J., 'DOD autonomy roadmap: autonomy community of interest', National Defense Industrial Association 16th Annual Science and Engineering Conference/Defense Tech Exposition, Springfield, VA, 24–26 Mar. 2015; and Hunter, A. et al., *Defense Acquisition Trends, 2016: The End of the Contracting Drawdown* (Center for Strategic and International Studies: Washington, DC, Apr. 2017), p. 11.

[36] Bornstein (note 35).
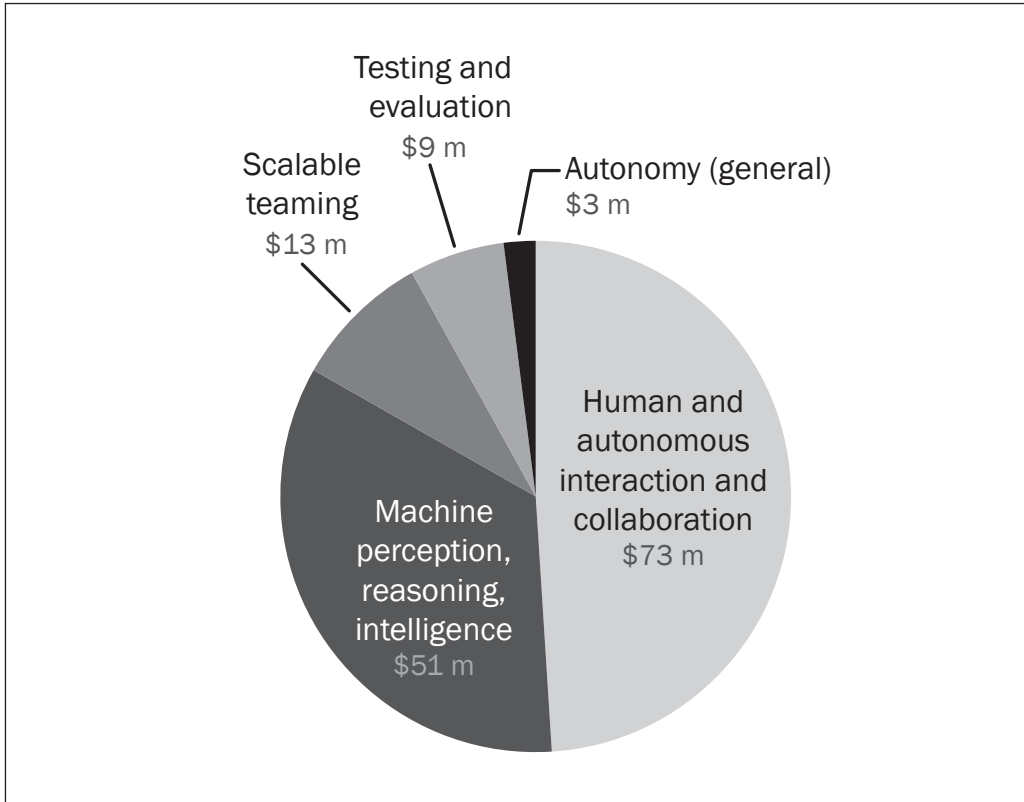
Figure 5.2. US Department of Defense (US DOD) funding distribution on applied research and advanced technology development on autonomy in US fiscal year 2015 in millions of US dollars

Note: 'Advanced technology development' in US DOD terminology refers to experimental development (i.e. all efforts that have moved into the development and integration of hardware for field experiments and tests).

Source: Bornstein, J., 'DOD autonomy roadmap: autonomy community of interest', National Defense Industrial Association 16th Annual Science and Engineering Conference/Defense Tech Exposition, Springfield, VA, 24–26 Mar. 2015.

1. *TRACE (Target Recognition and Adaption in Contested Environments)*. A project that aims to use the most recent advances in machine learning to improve the performance of automatic recognition systems.

2. *CwC (Communicating with Computers)*. A project that seeks to improve human trust in autonomous systems and facilitate interaction between human and autonomous systems by making computers capable of symmetric communication with humans.

3. *CODE (Collaborative Operations in Denied Environment)*. A project that seeks to make it possible for a group of UASs to conduct a coordinated attack in a denied environment under one person's supervisory control.

The research laboratories of each military branch have traditionally had a more short-term focus. They explore capabilities that meet a specific operational demand. However, this does not mean that they do not conduct fundamental research. Notable projects include the following.

1. *CARACaS (Control Architecture for Robotic Agent Command and Sensing)*. An Office of Naval Research project that seeks to develop a control architecture for swarm operations.

2. *SMET (Squad Multipurpose Equipment Transport)*. An Army Research Laboratory project that seeks to develop a semi-autonomous vehicle for logistical transport.

3. *'Trust in Autonomy for Human Machine Teaming'*. An Air Force Research Laboratory project that seeks to understand the trust dynamic between humans (e.g. pilots, ISR operators and analysts) and robots, and develop advanced human–robot teaming.

It is worth noting that the US Government has also invested heavily in autonomy-related projects outside the DOD. One initiative worth mentioning is the National Robotic Initiative, which has funded projects on collaborative robotics for various government agencies since 2011. Its budget for 2017 amounted to $221 million, of which slightly less than half ($103 million) was allocated to the DOD and the rest to civilian research.[37] In October 2016 the US Government proposed the creation of a similar initiative on AI, which would eventually double or triple the current government funding on AI ($1.1 billion on unclassified research in 2015). The priority areas identified are big data, computer vision, theoretical understanding, general AI, scalability, human-like AI, reliability and enabling hardware.[38]

The United Kingdom

The UK is the European country that invests the most in military R&D. There are no figures on the current level of funding that the UK allocates to R&D on military applications of AI and robotics technologies. Most of the relevant research is conducted either by Qinetiq, a former public defence research agency that was privatized in 2001, or the Defence Science and Technology Laboratory (DSTL). Qinetiq regularly performs collaborative research with private companies and universities. The DSTL conducts R&D internally (40 per cent of its budget) and also funds research with universities and companies (60 per cent of its budget).[39] Since 2010 the DSTL has funded a number of autonomy-related R&D projects (for a total of at least £16 million, which is approximately $21 million), addressing issues such as computer vision, sensor processing, swarming and autonomous navigation for unmanned systems (a list of these projects is available at the SIPRI website).

On the civilian side, the development of robotics and autonomous systems is one of the 10 priority areas of the UK's 2017 Modern Industrial Strategy, a plan that, in part, seeks to offset the impact of the UK's withdrawal from the EU (commonly referred to as 'Brexit') on academic and private R&D in the UK.[40]

Russia

Russia's public R&D efforts on robotics are predominantly defence oriented. This is typical of Russia's approach to innovation.[41] The Russian Government prioritizes military technology on the basis that innovation will eventually spin off into the civilian sphere and deliver benefits to society as a whole.

---

[37] National Science Foundation, 'National Robotics Initiative 2.0: ubiquitous collaborative robots (NRI-2.0)', NSF 17-518, [n.d.]; and Sargent, J. et al., *Federal Research and Development Funding: FY2017*, Congressional Research Service (CRS) Report for Congress R44516 (US Congress, CRS: Washington, DC, 27 Jan. 2017).

[38] National Science and Technology Council (NSTC), Networking and Information Technology Research and Development Subcommittee, *The National Artificial Intelligence Research and Development Strategic Plan* (NSTC: Washington, DC, Oct. 2016); and Executive Office of the President, National Science and Technology Council (NSTC), Committee on Technology, *Preparing for the Future of Artificial Intelligence* (NSTC: Washington, DC, Oct. 2016).

[39] Marcum, M., 'Assessing high-risk, high-benefit research organizations: the "DARPA effect"', Study of Innovation and Technology in China (SITC) Policy Brief, 2 Jan. 2014.

[40] British Government, *Building our Industrial Strategy*, Green Paper, (British Government: Jan. 2017).

[41] Adamsky, D., 'Defense innovation in Russia: the current state and prospects for revival', IGCC Defense Innovation Briefs, Jan. 2014, pp. 5–7.

Military robotics is a priority area for Russia's new rearmament programme, which has a budget of 20 trillion roubles ($346 billion) for the period 2016–25.[42] Reportedly, Russia aims to catch up with the progress that the USA has achieved in this field.[43] Key recent developments were the launch of the Robotics 2025 Programme and the creation of the Skolkovo Robotics Centre (SRC) in 2014. Detailed figures on how much Russia has invested in these initiatives have not been disclosed.

The Robotics 2025 Programme is implemented by the Ministry of Defence and the Foundation for Advanced Studies (FPI)—a new military research centre modelled on DARPA.[44] The FPI's total budget in 2015–16 was 4.5 billion roubles ($78 million), of which a significant part was reportedly allocated to R&D in robotics.[45] With this programme, Russia aims to foster the development of all types of robotics, from unmanned vehicles to bio-, micro- and nano-robots.

Autonomy also seems to be an important topic of interest. Autonomous navigation, command-and-control for collaborative operations and swarming, and autonomy for decision-making support have been presented as priority R&D areas.[46]

The SRC was founded to improve the synergy between state research institutes, universities and companies working on robotics, especially civilian robotics. It connects 35 state research institutes from industry and academia, which together represent 20–25 per cent of all Russian entities conducting R&D on civilian robotics.[47] The SRC reportedly collaborates extensively not only with the FPI but also with research institutions in other countries, particularly those in China. Autonomy is clearly a key focus area for the SRC, as it includes among its R&D priorities computer vision, navigation, control in dynamic and unstructured environments, human–machine interactions and human augmentation systems.[48]

France

France generally provides few details on military R&D activities. However, in February 2017 Jean-Yves Le Drian, the French Minister of Defence, announced that AI will play an increasingly important role in developing new military technologies to ensure that France does not fall behind its allies, specifically the UK and the USA. One of the ways this will be accomplished is to improve the use of state-funded academic research for military (and industrial) applications. According to the announcement, investment in AI research will be essential to support France's strategic autonomy. Some of the most important topic areas mentioned were intelligent sensor processing and real-time scene analysis, collaborative autonomy, war gaming and simulations, and countering cyber-attacks in real time.[49]

In January 2017 the Ministry of Higher Education and Research released a report on AI that made a series of recommendations to help France to maintain its progress

[42] 'Russia to focus on robotic weaponry in arms procurement', Sputnik, 11 Dec. 2013.

[43] Kashin, V. and Raska, M., *Countering the US Third Offset Strategy: Russian Perspectives, Responses and Challenges*, S. Rajaratnam School of International Studies (RSIS) Policy Report (RSIS: Singapore, Jan. 2017); and Roffey, R., 'Russian science and technology is still having problems: implications for defense research', *Journal of Slavic Military Studies*, vol. 26, no. 2 (2013), p. 165.

[44] This includes the Robotics R&D and Experimental Centre and the Main Directorate of Scientific Studies and Engineering Support of Advanced Technologies (Innovative Research).

[45] Ria Novosti, [Russia will increase investment into the development of new weapons, including robots], 15 Oct. 2014 (in Russian).

[46] Kozyulim, V. and Efimov, A., [The new James Bond: a machine with a licence to kill], *Security Index*, vol. 22, no. 1 (116) (2016) (in Russian).

[47] Efimov, A., 'From Russia with robots...', The Disruptory, 29 Feb. 2016.

[48] Skolkovo Robotics, 'Skolkovo Robotics Center: Russia's leading commercialization hub for civilian robotics', [n.d.].

[49] Le Drian, J., 'L'intelligence artificielle: un enjeu de souveraineté nationale' [Artificial intelligence: an issue of national sovereignty], *L'intelligence artificielle: des libertés individuelles à la sécurité nationale* [Artificial intelligence: individual freedoms to national security] (Eurogroup Consulting: Paris, 2017), pp. 11–24.

in the field of AI more generally. These included the creation of initIA: a €550 million ($639.5 million) research fund aimed at stopping the 'brain drain' of France's leading research experts to (foreign) industry. The report recommended strengthening research on computer vision; human–machine collaborations; big data; making AI reasoning understandable; problem solving; collaborative intelligence; general AI; and the ethical and social concerns of AI, including the protection of privacy.[50] As yet, the government has not released concrete plans, but has promised to fund infrastructure and research, and to include AI systematically in its innovation strategy.[51]

Italy

There is no information available on the extent to which Italy is making military R&D efforts on AI or robotics. It is established, however, that AI and robotics are elements of its army's modernization plan.[52] Italy has made significant investments in civilian applications of AI and robotics in recent years. In 2016 the Italian Government launched Industria 4.0 (2017–20), a €30 billion ($34.8 billion) initiative designed to support the development of smart manufacturing, smart energy and smart services through nine different technologies, including big data, collaborative machines, and autonomous cooperative robotics and sensors.[53]

Japan

Japan is, together with the USA, one of the countries that has historically pioneered the development of robotics. In contrast to the USA, Japan's efforts have primarily been focused on advancing robotics for civilian uses. Japan notably has a long history of researching and developing industrial robots and humanoid service robots (for further discussion on industrial and service robots see section V of this chapter).

In 2015 Japan launched the New Robot Strategy, an initiative that is aimed at maintaining Japan's competitive advantage as China, South Korea and the USA make advances in this field.[54] The New Robot Strategy is primarily focused on the development of intelligent industrial robots and service robots for elderly care (to assist Japan's aging population). The R&D priorities in this respect include the development of sensing and recognition technology, mechatronics (i.e. technology combining electronics and mechanical engineering), and actuator and control technology.[55] The exact amounts that Japan would invest in R&D as part of this initiative were not officially disclosed but were estimated by one source to be $350 million.[56] Japan has also launched a number of initiatives in recent years that are believed to channel important investments in R&D on robotics as well as AI.[57] In addition, it launched

---

[50] French Government, *Rapport de Synthèse: France Intelligence Artificielle* [Synthesis Report: France Artificial Intelligence] (French Government: Jan. 2017), pp. 7–14.

[51] French Ministry of Higher Education and Research, *La stratégie IA en France* [The IA strategy in France] (French Ministry of Higher Education and Research: 21 Mar. 2017), p. 11.

[52] Nones, M. and Marrone, A. (eds), *The Transformation of the Armed Forces: The Forza NEC Program*, IAI Research Paper (Istituto Affari Internazionali: Rome, 2012).

[53] Italian Ministry of Economic Development, 'Italy's plan: Industria 4.0', Jan. 2017.

[54] Inagaki, K., 'Google and IBM overshadow Japanese tech groups in global AI race', *Financial Times*, 4 Feb. 2016; and Water, R. and Muyayama, K., 'Are Japanese robots losing their edge to Silicon Valley', *Financial Times*, 11 Jan. 2016.

[55] Headquarters for Japan's Economic Revitalization (HJER), *New Robot Strategy: Japan's Robot Strategy: Vision, Strategy, Action Plan* (HJER: 2 Oct. 2015).

[56] Kurata, K., *Overview on the Current Policy Trends in Robotics and AI in Japan* (Embassy of Japan in the UK: London, 18 Feb. 2016).

[57] These include the following: Society 5.0/Promotion of a Super Smart Society (26 trillion yen/$228 billion); Productivity Revolution by Investment for the Future (a total of 34.81 billion yen/$333 million) for AI, robotics, Internet of Things (IoT) and big data in 2016 by the Ministry of Economy, Trade and Industry; Strategic Innovation Promotion Program on Innovative Design and Manufacturing Technologies (2.55 billion yen/$22 million) and Automated Driving Systems (2.32 billion yen/$20 million); the construction of a deep learning supercomputer

three research centres in 2016 dedicated to AI and big data. These were to receive $1 billion over 10 years.[58]

Japan also has an interest in the military application of robotics. The 2016 Defense Technology Strategy listed research into fields related to unmanned systems as one of its four key priorities.[59] The 2017 defence budget notably earmarked 900 million yen ($7.8 million) for research into 'autonomous surveillance technology' for unmanned underwater vehicles.[60]

It is worth noting that Japan's R&D on military robotics has for a long time been constrained by the Japanese Constitution, which prohibited universities from conducting military R&D.[61] Military R&D could only be conducted by military research institutes or private companies. Defence companies had very little incentive themselves to engage in military R&D because the Constitution also prohibited military exports. In 2014 the Japanese Government changed the Constitution to lift these restrictions and facilitate innovation in military technology.[62] Cooperation between military research institutes and civilian research institutes developing robotics has since intensified. The Japanese Government initiated a programme in 2015 under which the New Energy and Industrial Technology Development Organization, a civilian R&D agency that is a driving force in the development of robotic technologies, will conduct dual-use advanced R&D projects on the same model as the USA's DARPA.[63]

Israel

There is little official information available on Israel's efforts in military R&D. Its strategy, budget and ongoing activities remain confidential.[64] It is common knowledge, however, that military robotics is a key technology area for the Israel Defense Forces (IDF). Israel and the USA pioneered the development and adoption of armed unmanned systems, and Israel is one of the leading exporters of unmanned systems. It supplies unmanned systems for all domains: aerial, maritime and ground. Many of them feature remarkable autonomous capabilities.

Israel's interest in military robotics and autonomy is reportedly driven by two fundamental considerations: to reduce the risk to its military personnel and to give the IDF, which are numerically small, a qualitative edge on Israel's enemies in the region.

(19.5 billion yen/$166 million); Economic Revitalization (estimated at 10 trillion yen/$88 billion); and many smaller projects. Kurata (note 56); Japanese Ministry of Finance, 'Highlights of the draft FY2017 budget', 22 Dec. 2016; Ministry of Economy, Trade and Industry, 'Key points of the METI-related FY 2016 budget', [n.d.]; Japanese Cabinet Office, Council for Science, Technology and Innovation (CSTI), *Comprehensive Strategy on Science, Technology and Innovation 2016* (CSTI: 24 May 2016); and Yoshida, R., 'Abe orders drafting of new stimulus package to breathe life into Japan's economy', *Japan Times*, 12 July 2016.

[58] Matsuoka, S., 'FLOPS to BYTES: accelerating beyond Moore's Law is enabled from data', Convergence with Data Science: A New Beginning for HPC, SOS 21 Workshop, Swiss National Supercomputing Centre, Davos, Switzerland, 22 Mar. 2017.

[59] Japanese Ministry of Finance (note 57), p. 26.

[60] Japanese Ministry of Finance (note 57), p. 32.

[61] That rule actually prevented two researchers from the University of Tokyo from participating in the USA's DARPA Robotic Challenge in 2014. To participate, the researchers had to quit their positions at the university and found a robotics company (Schaft Inc). Schoff, J., 'Robotics diplomacy and the US–Japan alliance', *The Diplomat*, 15 Mar. 2016.

[62] Hokazono, H., 'The role of science and technology for Japan's self defense', Acquisition, Technology and Logistics Agency, 19 Dec. 2016; Sugai, H., *Japan's Future Defence Equipment Policy* (Brookings: New York, 2016), pp. 33–34; and Japanese Ministry of Defence (MOD), *Strategy on Defense Production and Technological Bases: Toward Strengthening the Bases to Support Defense Forces and 'Proactive Contribution to Peace'* (MOD: June 2014), pp. 29–30.

[63] Xinhua, 'Japan's leading science body expresses concerns over gov't sponsored military research', 8 Mar. 2017; and Kelly, T. and Kubo, N., 'Japanese civilian R&D agency to get military role to spur arms innovation', Reuters, 19 Mar. 2015.

[64] The exact numbers are not made public, but were estimated in 2006 to be 30% of total R&D. Brozska, M., 'Trends in global military and civilian research and development (R&D) and their changing interface', *Proceedings of the International Seminar on Defence Finance and Economics*, vol. 13 (2006), p. 286.

Commentators have argued that Israel's success in the field of robotics, and high-technology more generally, has been facilitated by the efforts of the Israeli Government to foster not only a very close interaction between the IDF, the defence industry (which is mostly government owned) and universities, but also an ecosystem that favours technology spin-off (defence to civilian) and spin-in (civilian to defence).[65]

South Korea

Like Japan, South Korea's R&D efforts in the area of AI and robotics have been oriented towards the development of civilian applications. The South Korean Government has issued numerous investment plans in robotics in recent years, which were all primarily dedicated to the development of industrial robots—a business that is essential to South Korea's economy and in which it is a market leader.[66] These include the Joint Robot Industry Development Initiative, the second Basic Plan for Intelligent Robot Development (2014–18), the Seven Robot Fusion Business Strategies Roadmap, the Smart Robot Basic Plan and the Robot Future Strategic Vision 2022 (2012).[67] In 2016 the South Korean Government decided to invest $840 million between 2016 and 2020 to boost R&D in AI.[68] The government will fund the establishment of civilian-led research by six conglomerates (chaebol), which together will invest $2.6 million.[69] This AI plan identified five priority areas: advancing natural language processing, image recognition, robot locomotion, HRI and semantic understanding.[70]

There is limited open-source information on South Korea's current R&D efforts in military robotics. It is known that South Korea is investing significantly in military research and it is attaching value to the development of unmanned systems, notably for missions such as ISR, anti-submarine surveillance and land warfare.[71] As discussed in chapter 3, South Korea pioneered, along with Israel, the development of robotic sentry weapons.[72]

Germany

There is very little public information available on Germany's military R&D activities.[73] The 2015 strategy paper to strengthen the German defence industry makes

[65] IAI, IMI and Rafael are government-owned defence companies. Elbit is privately owned. Breznitz, D., *The Military as a Public Space: The Role of the IDF in the Israeli Software Innovation System* (MIT Press: Cambridge, MA, 2002); Swed, O. and Butler, J., 'Military capital in the Israeli hi-tech industry', *Armed Forces and Society*, vol. 41, no. 1 (2015), p. 127; The Dwight D. Eisenhower School for National Security and Resource Strategy, *Spring 2015 Industry Study: Final Report: Robotics and Autonomous Systems* (National Defense University: Washington, DC, 2015), p. 16; and Honig, B., Lerner, M. and Raban, Y., 'Social capital and the linkages of high-tech companies to the military defense system: is there a signaling mechanism?', *Small Business Economics*, vol. 27, no. 4 (2006), pp. 419–37.

[66] Investment in industrial robotics is a way for South Korea to stay competitive in comparison with China. Temperton, J., 'Samsung developing robots to replace cheap Chinese labour', *Wired*, 19 Oct. 2015.

[67] These plans and their budgets include more than just R&D spending, which cannot be isolated. Hong, J., 'South Korea to boost its robot industry with a new development initiative', Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland), Feb. 2017; and Ren, F. and Sun, X., 'Current situation and development of intelligence robots', *ZTE Communications*, vol. 14, no. S1 (Dec. 2016).

[68] This investment was already in progress, but the investment plan was accelerated in reaction to the decision of Lee Sodol, a South Korean professional Go player, to assist in the development of Google DeepMind's AlphaGo, which is an AI computer program developed to play Go. Chi-dong, L., 'Government to invest 1 tln won in artificial intelligence', Yonhap News Agency, 17 Mar. 2016; and Zastrow, M., 'South Korea trumpets $860-million AI fund after AlphaGo "shock"', *Nature*, 23 Mar. 2016.

[69] The conglomerates are Hyundai Motor, KT, LG Electronics, Naver, Samsung and SK Telecom.

[70] Ha, J., 'Artificial intelligence industry in South Korea', Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland), 22 Mar. 2016.

[71] In 2016 South Korea also announced defence reforms, which include strengthening defence R&D capabilities by $17 billion. 'Korea: defence industry equipment', Export.gov, 18 Aug. 2016; and Park, J., 'South Korean private–public partnership to invest $2.6B in robot industry by 2018', Robohub, 11 Aug. 2014.

[72] Weitz, R., 'South Korea's defence industry: increasing domestic capabilities and global opportunities', Korea Economic Institute of America, Academic Paper Series, 7 Nov. 2013, p. 4.

[73] Paillaid, N. and Butler, C., *Today's Technological Innovations for Tomorrow's Defence*, Armament Industry

no mention of AI or robotics.[74] Germany is known, however, for its investment in industrial robotics—Germany is the second largest producer of industrial robotics in the world.[75] The future of industrial robotics was a key focus area of Germany's New High-Tech Strategy of 2014, which set aside €17 billion ($19 billion) for R&D investment for the period 2014–17.[76]

India

India did not have a dedicated national plan for AI or robotics until July 2017. A defence technology roadmap published in 2013 indicates that India's armed forces have an interest in UGSs for logistics and ISR, image-based target identification and classification, and expert systems for managing the health of sophisticated weapon systems.[77]

India has one military research institute specifically dedicated to AI and robotics: the Centre for Artificial Intelligence and Robotics. It seems to work mostly on tactical communication, communication secrecy and information security.[78]

India is reasonably transparent about its defence R&D programmes. SIPRI found a handful of programmes related to AI and robotics (a list is available at the SIPRI website). Two are directly related to autonomy: one is dedicated to the development of an autonomous UCAS and the other covers the development of an unmanned ground vehicle (UGV) capable of autonomous navigation in a semi-structured environment.

China

AI and robotics are important technology areas for the Chinese Government. The basis of China's R&D efforts in AI and robotics was established by the 863 Programme (also known as the State High-Level Development Plan), which was established in 1986 with the aim of developing China's innovation capacity in a variety of advanced fields of technology, especially information technology and automation.[79] In 2014 the Chinese Government officially named AI and robotics as priority technology areas.[80] Since then, they have been incorporated in a wide range of central economic and scientific plans, including the 13th Five-year Plan for Economic and Social Development of the People's Republic of China (2016–20), the Medium- to Long-term Plan for the Development of Science and Technology, the 13th Defense Science and Technology and Industry Five-year Plan, and the 2025 Defense Science and Technology Plan.

In 2016 China released two plans specifically dedicated to AI and robotics: the Robotics Industry Development Plan (2016–25) and the Three-year Guidance for Internet Plus and Artificial Intelligence Plan (2016–18).[81] The latter paved the way for the creation of three joint research centres, headed by Baidu, China's leading internet service provider and often thought of as China's equivalent to Google, on deep

European Research Group (ARES) Policy Paper no. 10 (ARES: Paris, Dec. 2016), p. 11.

[74] German Ministry for Economic Affairs and Energy, 'Strategiepaper der Bundesregierung zur Stärkung der Verteidigungsindustrie in Deutschland' [Strategy paper of the German Government to strengthen the defence industry in Germany], 8 July 2015.

[75] International Federation of Robotics, 'Executive summary world robotics 2016 industrial robots', 2016.

[76] German Ministry of Education and Research, *The New High-Tech Strategy: Innovations for Germany* (German Ministry of Education and Research: Berlin, 2014); and Cologne Institute for Economic Research, 'The industrial sector: a pillar of the German economy', [n.d.].

[77] Indian Ministry of Defence (MOD), Headquarters Integrated Defence Staff, *Technology Perspective and Capability Roadmap* (MOD: Apr. 2013), p. 7, p. 19.

[78] Indian Ministry of Defence, Defence Research and Development Organisation, Centre for Artificial Intelligence and Robotics, 'Area of work', [n.d.].

[79] Cheung, T., Anderson, E. and Yang, F., 'Chinese defense industry reforms and their implications for US–China military technological competition', Study of Innovation and Technology in China (SITC) Research Brief, 4 Jan. 2017, p. 2.

[80] Cheung, Anderson and Yang (note 79), p. 2.

[81] He, H., 'China's five-year plan to transform its robotics industry', *South China Morning Post*, 6 Apr. 2016.

learning, big data and brain-like AI.[82] China's Academy of Science is currently drafting the Artificial Intelligence 2.0 Plan, which will cover big data, intelligent sensing, cognitive computing, machine learning and swarm intelligence. The Chinese Government is also working on a long-term vision on the development of AI through to 2030.[83]

There are no official figures available on how much China has actually spent or intends to spend on R&D through these various plans. China's investments in these fields are believed to be significant and part of the wider strategy to establish China at the forefront of innovation in ICT.[84] The extent to which the Chinese Government intends to derive military applications from these investments in AI and robotics is also difficult to determine. However, what is known is that China is pursuing a policy of 'civil–military fusion', whereby the military, academia and industry jointly develop and share technology.[85] The People's Liberation Army (PLA) is therefore believed to have been closely involved in, and to have benefited from, the implementation of these plans.[86]

The PLA is not usually transparent with regard to its investment in military R&D. However, the PLA is believed to have identified robotics and unmanned systems as fundamental components of future warfare and made considerable investments in R&D in these areas.[87] China has reportedly achieved major advances in the development of UASs in recent years but still lags behind the USA in the development of autonomous capabilities.[88] There is disagreement among experts about the importance of AI for the PLA, but it is reported that the PLA is also conducting work on intelligent systems.[89]

The European Union

It is worth including the EU in this discussion because it is a major source of R&D funding for many European research institutions, be they public or private. The EU is highly transparent about the R&D projects it funds. Between 2007 and 2016 the EU spent €675 million ($787 million) on AI and robotics under the ICT call (the most important source of funding on AI and robotics) of the Seventh Framework Programme (FP7).[90] In 2014 it announced the start of SPARC, a public–private partnership (PPP) through which €2.8 billion ($3.26 billion) will be invested in robotics between 2014 and 2020.

---

[82] Vena, D., 'China is helping Baidu's AI efforts, and it couldn't come at a better time', Motley Fool, 27 Mar. 2017; and Gershgorn, D., 'China is funding Baidu to take on the US in deep-learning research', Quartz, 22 Feb. 2017.

[83] Other notable plans include the following: the Made in China 2025 Advanced Manufacturing Program; the 12th Five-year Plan for Intelligent Manufacturing; the Guidelines on Promoting the Development of Industrial Robots; and the 12th Five-year Special Plan for Service Robot Technology Development (2012). Cheung, T., 'The Chinese defense economy's long march from imitation to innovation', *Journal of Strategic Studies*, vol. 34, no. 3 (2011), p. 334; and Kania, E., 'China may soon surpass America on the artificial intelligence battlefield', *National Interest*, 21 Feb. 2017.

[84] According to experts, China was historically less capable of developing innovative and cutting-edge technology, so it has been heavily investing in R&D in order to become more competitive globally. Cheung (note 83), p. 334.

[85] Xinhua, 'Xi to head central commission for integrated military, civilian development', 22 Jan. 2017; Xin (note 4); and Grevatt, J., 'China opens military R&D to the private sector', *IHS Jane's Defence Industry*, 8 Aug. 2016.

[86] Kania, E., 'Chinese advances in unmanned systems and the military applications of artificial intelligence—the PLA's trajectory towards unmanned, "intelligentized" warfare', Testimony before the US–China Economic and Security Review Commission, 23 Feb. 2017.

[87] Ray, J. et al., *China's Industrial and Military Robotics Development*, Report for the US–China Economic and Security Review Commission (Center for Intelligence Research and Analysis, Defense Group Inc: Vienna, VA, Oct. 2016).

[88] Ray et al. (note 87), p. 63.

[89] See e.g. Ray (note 87); Kania (note 86); Singh, A., 'Is China really building missiles with artificial intelligence', *The Diplomat*, 21 Sep. 2016; Chase, M. et al., 'Emerging trends in China's development of unmanned systems', Rand Corporation Report (2015); Fan, G., 'A Chinese perspective on the US Third Offset Strategy and possible Chinese responses', Study of Innovation and Technology in China (SITC) Research Brief, 3 Jan. 2017; and Markoff, J. and Rosenberg, M., 'China's intelligent weaponry gets smarter', *New York Times*, 3 Feb. 2017.

[90] European Commission, Community Research and Development Informative Service, 'Cognitive systems and robotics: call 10', [n.d.]; European Commission, 'New robotics projects from 2015 announced', 7 Jan. 2016; and European Commission, 'New Horizon 2020 robotics projects from 2016', 16 Dec. 2016.

The EU committed to invest €700 million ($818 million) and the other €2.1 billion ($2.45 billion) will be supplied by industry and national governments.[91] The project will be funded under the auspices of the ICT call of the Horizon 2020 Programme (H2020), which replaced FP7. Beyond the ICT call, the EU has also funded R&D on AI and robotics in transport, security, future and emerging technologies, and societal challenges calls.

R&D that the EU has funded through FP7 and currently funds through H2020 is for civilian or dual-use purposes (pure defence research was de facto excluded from these funding schemes).[92] It is worth mentioning that the EU placed great emphasis through FP7 on fundamental and applied research projects, particularly those relating to cognitive robotics. With H2020, the EU has sharpened its focus on applied research and technology development. The new PPP funding model, which favours collaborations between academia and industry, is intended to encourage the development of marketable innovation.[93]

In 2016 the EU articulated for the first time a European Defence Research Programme (EDRP), which aimed to encourage joint military R&D efforts in the EU and thereby reduce the replication of similar military R&D efforts at the national level.[94] A total of €25 million ($29 million) was allocated to the programme for 2016–17; €65 million ($75 million) has been requested for 2018–20, and officials hope for another €500 million ($582 million) between 2021 and 2027. Robotics technologies and autonomy are expected to be major focus areas of the EDRP. In fact, the three projects that have so far received funding are all related to autonomy and robotics.[95] These projects are as follows.

1. *EuroSWARM*. A project intended to demonstrate the feasibility of heterogeneous swarms of sensor platforms.

2. *TRAWA*. A project aimed at standardizing detect-and-avoid systems on remotely operated UASs. This project is intended to ensure that multiple UASs can coexist in a single airspace.

3. *SPIDER*. A project aimed at providing the proof of concept of a sensor and surveillance system for awareness and navigation inside buildings during urban warfare.

Conclusions

The majority of the countries in the top 10 of the largest arms-producing countries (and China) have identified AI and robotics as important R&D areas. While there is clear evidence that most are investing significant resources in civilian and dual-use applications, finding details about their equivalent military R&D activities in open sources has proved difficult in many cases. The USA is the country for which most

---

[91] SPARC, *Robotics 2020 Multi-annual Roadmap for Robotics in Europe: Horizon 2020 Call ICT-2016 (ICT-25 & ICT-26)* (SPARC: 3 Dec. 2015).

[92] European Commission, 'Explanatory note on "exclusive focus on civil applications"', [n.d.]. A notable example of a dual-use project was TALOS (Transportable Autonomous Patrol for Land Order Surveillance). This was a Polish-led research project conducted during FP7, which aimed to develop and field test the innovative concept of a mobile, modular, scalable, autonomous and adaptive system for protecting European borders. European Commission, Community Research and Development Informative Service, 'Transportable Autonomous Patrol for Land Order Surveillance', Projects and results, [n.d.].

[93] SPARC (note 91); Huet, C., 'Robotics in H2020: latest developments: robotics PPP* and beyond', Robotics: Science and Systems Conference, Berlin, 24–28 June 2013; and EURobotics, 'Implementation: how SPARC is used', [n.d.].

[94] The European Defence Agency coordinated some defence research, but it was paid for by EU member states, who joined on an ad hoc basis, and carried out by defence companies in those countries. A total of €500 million ($584 million) has been spent on R&D projects since 2004. Most of the projects dealing with autonomy included prototypes of unmanned systems. European Defence Agency, 'Research and technology', [n.d.]. A list of these projects is available at the SIPRI website.

[95] Fiott, D., 'EU defence research in development', European Union Institute for Security Studies, Issue Alert, no. 43 (2016).

public information is available. It is also the only country that has an articulated and identifiable military R&D strategy on autonomy. The USA funds a wide spectrum of R&D projects through DARPA and its military labs. Many of these projects could serve as building blocks for the development of autonomy in weapon systems. It is notable that the USA's current top priority on autonomy, at least in budgetary terms, is the improvement of human–autonomous systems interaction and collaboration. A review of expert commentaries and the policy statements and various R&D projects of other major arms-producing countries indicates that most of these countries are paying a lot of attention to how the US DOD funds R&D in AI and robotics. They seem to be prioritizing the same types of capabilities as those that have been developed or are being developed by the USA: robust navigation for unmanned vehicles, collaborative autonomy and human–autonomous systems interaction.

V. An industry perspective

In which areas of industry do relevant R&D efforts take place? For the reasons presented in section II of this chapter, providing a comprehensive mapping of the private sector landscape with regard to innovations that could shape the future of autonomy in weapon systems is not feasible within the scope of this report. Mapping the robotics industry itself is challenging (see box 5.2). It should be noted that when it comes to innovations in the area of machine autonomy there are two major lines of division within the industry. The first line of division is between the civilian and the defence industries, and the second is between companies with a background in the ICT sector and traditional manufacturing companies (see figure 5.3).

Innovation leadership: civilian versus military industry

Why the civilian industry is driving innovation in the area of autonomous technologies

When considering the industrial landscape, it is striking that it is the civilian industry, not the defence industry, that is currently driving the development and adoption of autonomous technologies. There are three main reasons for this.

The first and fundamental reason is that technologies on which autonomy is created are more often than not dual-use and belong to a realm where the civilian sector has been leading innovation for decades. This is particularly true for computer processing technologies. One trend worthy of note in this respect is the impact that the boom in smartphones has had on the availability, performance, size and cost of computer chips, batteries and sensor technologies, from vision-based sensors (video cameras) and tactile sensors (touch screens) to motion sensors, such as inertial measurement units (IMU). The introduction of the Kinect in 2011, a sensor system developed by Microsoft for its video game platform the X Box, was also remarkable as it provided the robotics community with a very low-cost and efficient 3-D scanner system.[96] Before the X Box-led innovations, 3-D scanner systems were either very expensive or unreliable.[97] The decreasing cost and increasing availability of sensor technologies have made robotic platforms much more affordable to develop and acquire. This trend has notably fuelled the emergence in recent years of low-cost robotic platforms, such as hobbyist drones. The improving performance of civilian sensor technologies is also significant as it allows the military sector to increasingly rely on 'off-the-shelf' components for the development of military robotic platforms. In this respect, the

---

[96] 3-D scanner systems are 3-D perception sensors that enable robots to map out their environment and detect and manipulate obstacles, and also recognize motions, objects and faces.

[97] The future development of driverless vehicles, which will rely on 3-D perception systems for autonomous navigation, is expected to further improve the efficiency, and more importantly, the availability of 3-D perception sensors.
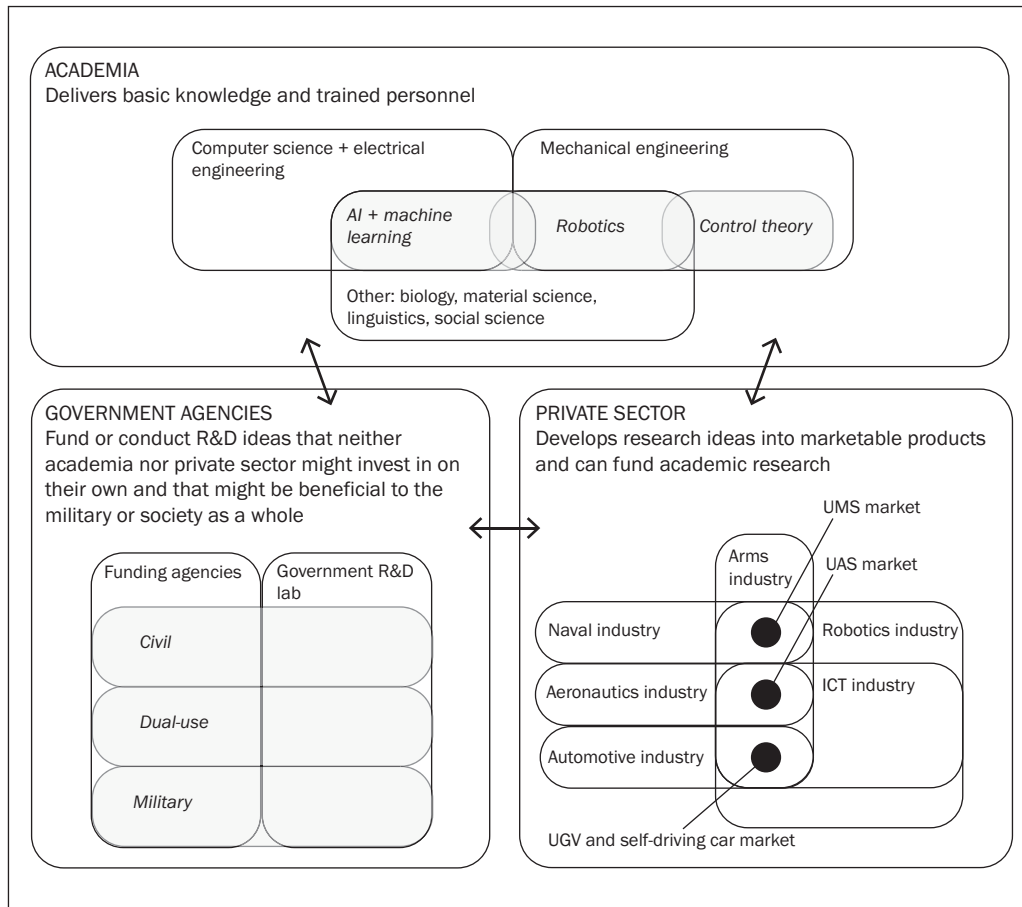
Figure 5.3. Robotic and autonomous systems industry

AI = artificial intelligence; ICT = information and communications technology; R&D = research and development; UAS = unmanned aerial system; UGV = unmanned ground vehicle; UMS = unmanned maritime system.

emergence of the self-driving car industry could help to reduce the cost of high-end sensors that are used in military UGVs.

The second reason, which has been partially discussed in section II of this chapter, is that due to a difference in how the supply and demand relationship is structured in the civilian and defence markets, the civilian industry has had, and has been able, to invest far more than the defence industry in relevant R&D. While a number of large commercial corporations, such as Google, Facebook, Amazon and Toyota Motors, have been making huge investments in the development of autonomous technologies in recent years, large defence companies have had to embrace a 'wait and see' position, due to budget uncertainties and mixed signals from military leaders. Military planners have been stressing the potential of autonomous technologies for decades, but autonomy only emerged as a key strategic priority very recently.

The third reason is more technical, and derives from the fact that autonomous capabilities are generally easier to engineer within civilian systems than military systems. Autonomous ground robots provide a case in point here. Most civilian ground robots (e.g. hotel robots and hospital robots) are intended to operate in environments that are non-adversarial, known and (relatively) predictable. In these conditions, the engineers who design them do not have to take into consideration that intelligent adversaries will actively try to defeat the technology, for example, through spoofing or cyber-attacks.[98] In addition, they can use several engineering tricks to palliate the

---

[98] US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on*

Box 5.2. The robotics industry: an amorphous industry

The robotics industry might be considered, for obvious reasons, as the most central industrial sector with regard to innovations in the area of machine autonomy. Autonomy is not, however, a priority for all stakeholders within this industry. The main dividing line is between 'industrial' robotics and 'interactive' or 'service' robotics.[a] Industrial robotics, which actually represents the largest segment of the robotics industry, has historically had little to no interest in autonomy, as the robots developed in this segment are designed to perform repetitive, pre-programmed tasks in a very controlled environment.[b] Service robots, on the other hand, because they are intended to assist humans in various tasks and possibly evolve in dynamic conditions, usually need to include some level of autonomy in their functioning. Thus, research and development efforts that are relevant to this discussion are primarily related to these types of robots.

Service robots come in all shapes and sizes—from small robotic insects to large unmanned ground systems—and have very different types of military and commercial applications. The service/interactive robot industry is therefore not really a distinct industry; rather, it is an amorphous agglomeration of companies that have diverse backgrounds and that pursue different technological and business objectives (see figure 5.3). These range from large industrial companies venturing into automated systems to support the development of their own industry and market—this includes the carmakers who are moving into the development of self-driving vehicles, logistics companies investing in the business of delivery robots, and aerospace companies moving into the development of unmanned systems—to specialized robotic companies and start-ups that are geared towards a specific market niche. The extent to which service robots need to, and can, work autonomously varies significantly, and depends on the intended end use and the nature of the working environment. Robot applications that will push innovation in the area of autonomy are robots, either civilian or military, that are intended to operate in open and semi-structured environments and perhaps interact with humans. These include the following.

1. *Logistical robots* that will have to operate in urban areas.
2. *Companion robots* that are intended to interact with humans (e.g. elderly care robots).
3. *Military robots* for intelligence, surveillance and reconnaissance missions, demining, and combat.
4. *Aerial robots* for professional use (e.g. photography, agriculture and surveillance).
5. *Self-driving vehicles*.
6. *Underwater vehicles* used for oceanography and underwater maintenance operations.

[a] An 'industrial robot' is defined by the International Federation of Robotics as 'an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation applications'. By contrast, a 'service/interactive robot' is 'a robot that performs useful tasks for humans or equipment excluding industrial automation application', International Federation of Robotics, <http://www.ifr.org/>.

[b] The emergence of so-called collaborative robots has begun to change the dynamic in the industrial robotics segment. Collaborative robots are industrial manipulators that are intended to work safely alongside humans. Tobe, F., 'Why co-bot will be a huge innovation and growth driver for robotics industry', IEEE Spectrum, 30 Dec. 2015.

---

limitations of, or reduce the technical requirements on, sensing and control algorithms. For instance, they can thoroughly pre-map the environment in which the robot will operate so that the robot does not need to identify everything with its sensors and make too many decisions.[99] Obviously, this developmental path to autonomy does not work in the case of military robots. They need to be designed with much better perception and decision-making capabilities because they must be able to cope with unstructured and dynamic terrains as well as the deployment of deception, assault and counter-autonomy technologies by adversaries. Hence, if civilian companies seem to have the edge regarding the development of autonomy in robotics, it is—perhaps primarily—because they face fewer engineering difficulties than their defence counterparts.

Autonomy (DOD: Washington, DC, June 2016), p. 13.

[99] A good illustration of this is the Google car. It is often presented as the apotheosis of autonomous systems development, while in reality the system cannot even recognize a traffic light signal by itself. US Department of Defense, Office of Technical Intelligence, Office of the Assistant Secretary of Defence for Research and Engineering (note 16), p. 12.

Box 5.3. Challenges related to the commercialization of self-driving vehicles

Engineering

Driving is not just about navigation and obstacle avoidance, it is also about social interaction: driving in an urban environment requires frequent interaction with humans—other drivers, cyclists and pedestrians. Machines remain poor at understanding and predicting the behaviour of human road users. The state of the art in computer vision permits self-driving cars to recognize only basic behaviours (i.e. walking, running and looking away). The limitations in perception and communication represent an obstacle to the use of self-driving cars in densely populated environments like city centres. As vehicle situational awareness improves, carmakers are working on challenging engineering problems similar to those faced by the military in terms of autonomous capabilities for targeting and surveillance (i.e. recognizing human behaviour).

Human control

Carmakers also have radically different approaches to how self-driving vehicles should be developed, and how much autonomy they should have. Companies with a background in the information and communications technology sector, such as Google or Uber, are aiming to develop fully autonomous vehicles that might not even include a steering wheel. Traditional carmakers, on the other hand, have a more conservative approach, and favour a 'shared control' model where autonomy would allow vehicles to work in collaboration with human drivers, rather than replacing them.[a] It is unclear for now which model will prevail, as experts have radically different views about which model will guarantee maximum safety.

Ethical

The most salient problem is the so-called car crash dilemma: how should the vehicle deal with a situation where it has to choose between making a manoeuvre that will keep its passenger safe but put a pedestrian or another car driver at risk, and making another manoeuvre that will keep the pedestrian or the other car driver safe but put its passenger at risk? How carmakers, transport regulators and insurance companies resolve this dilemma will be instrumental in determining how societies approach the ethical governance of autonomous systems in the future.

Legal

The development of the self-driving car industry will also contribute to the resolution of some of the legal questions that autonomy poses, notably in terms of liability. Self-driving will only be widely used once insurance companies, transport regulators and carmakers have agreed on who is to blame when a self-driving car is involved in an accident. In this respect, the legal concerns associated with the use of self-driving cars might also push the car industry to develop common standards for testing and evaluation procedures. As it stands, the autonomous systems community still lacks a proper methodology to test complex autonomous control systems. Considering the vested interest that the carmakers have in demonstrating the safety and reliability of their vehicles, it is likely that they will play a crucial role in the development of standards for the validation and verification of autonomous systems.

[a] Toyota's main research project on autonomous vehicles is called 'human-centred artificial intelligence for future intelligent vehicles and beyond'. Some carmakers, including Toyota, have even created research centres directly within universities. Toyota's centres are based at MIT and the University of Stanford, both of which are in the USA. Toyota Global Newsroom, 'Toyota will establish new artificial intelligence research and development company', 6 Nov. 2015.

Key technological developments in the civilian industry

There are a number of civil innovations that deserve attention, either (*a*) because they have remarkable autonomous capabilities and could be adopted off-the-shelf by the military; or (*b*) because the R&D efforts that are invested in their development contribute to helping to solve fundamental socio-technical problems that limit the further advance and adoption of autonomous capabilities in the military sphere. These include small recreational and professional drones; autonomous underwater vehicles used for oceanography and underwater maintenance operations; image and video analysis software for internet referencing and video surveillance; biometric recognition technologies; open source robotics software (e.g. ROS, the Robot Operative System managed by the Open Source Robotics Foundation); and speech interfaces. However, if there is one that deserves extra scrutiny it is the development of self-driving cars.

Self-driving cars remain an emerging technology. In fact, the companies that develop them still disagree about when they might be commercialized on a large scale. The Tesla Group estimates 2018, while Toyota and Ford suggest 2020.[100] In any event, efforts that are put into the development of self-driving cars are significant for at least three reasons.

First, they have generated an important funding stream and a clear research agenda for the community of AI and robotics researchers. One notable illustration of this is the recent creation by Toyota of research centres located in two of the most prestigious US universities in the fields of AI and robotics: Stanford University and MIT.[101] The company has committed $50 million in funding to these universities.[102]

Second, many of the engineering, ethical and legal problems that these companies are seeking to solve, in order to commercialize self-driving cars on a large scale, are common to many other types of civilian, but also military, autonomous systems (see box 5.3). These problems include the following.

1. Using autonomy in complex and populated environments.
2. Finding methods to ensure a trustworthy and reliable interaction and collaboration between humans and autonomous systems.
3. Finding methods to test, evaluate, verify and validate the capability, reliability, suitability and safety of autonomous systems intended to operate in complex and populated environments.
4. Determining how to programme ethical rules into the command-and-control of autonomous systems.
5. Clarifying issues related to liability in the case of accidents presented by autonomous cars (the self-driving car is a test-bed technology for the wider adoption and acceptance of autonomous systems).
6. Finding methods to ensure the integrity of autonomous systems against the threats of spoofing and cyber-attacks.

Third, the volume of production of the commercial car industry usually generates major economies of scale in the production of hardware and software components. Hence, the growth of the driverless car market holds the promise of bringing down the cost of sensors and computer chip technologies for large robotic platforms, including military platforms.[103]

Dual-use in the industry and why the defence industry will continue to play a key role

Although civil companies have clearly taken the lead with regard to the development and adoption of autonomy in robotic systems, traditional arms producers continue to play a key role, for the simple reason that commercial autonomous technologies can rarely be adopted by the military without modifications. The military domain is much more stringent in terms of technical requirements than the civilian sector. Autonomous systems that are intended to operate in battlefield conditions or enemy territory, such as self-driving logistical vehicles, need to include far more advanced

[100] 'When will self-driving cars be available to consumers', Quora, 29 Jan. 2016; Caddy, B., 'Toyota to launch first driverless car in 2020', *Wired*, 8 Oct. 2015; and Lambert, F., 'BMW will launch the electric and autonomous INext in 2021, new i8 in 2018 and not much in-between', Electrek, 12 May 2016.

[101] Toyota Global Newsroom, 'Toyota will establish new artificial intelligence research and development company', 6 Nov. 2015.

[102] With reference to the level of investment in AI by carmakers, some carmakers (such as Toyota) have established themselves in the top 20 of the world's leading software developers. Markoff, J., 'Toyota invest $1 billion in artificial intelligence in U.S.', *New York Times*, 6 Nov. 2016.

[103] Office of Technical Intelligence, Office of the Assistant Secretary of Defence for Research and Engineering (note 16), p. 11.

perception and decision-making algorithms than their civilian counterparts to cope with potential attack or deception attempts from an intelligent enemy. While there are certainly civilian companies that might be prepared to fulfil a defence contract, commercial companies, especially larger companies, have little economic incentive to work with military customers. The bureaucratic process is cumbersome, the value of the contract might be limited in comparison to what can be gained on the civilian market and the contractual requirements in terms of proprietary rights are too stringent.

Defence companies are, therefore, bound to play a central role in delivering autonomous technologies to the military. Innovation in the defence market is demand led, not supply led. Defence companies are inherently dependent on the direction they are given by the military. Companies are unlikely to engage in the development of new autonomous capabilities or new autonomous systems without the assurance that what they do corresponds to specific demands, and that they will be able to obtain a return on their investment. This does not mean that they do not conduct relevant research nor that they do not have the right competences. They need to maintain some ongoing research activities to prepare for future contractual opportunities. However, it is known that the sums major contractors invest in self-funded R&D are minimal in comparison to what major groups from the civilian sector invest. According to the research firm Capital Alpha Partners, the combined R&D budgets of five of the largest US defence contractors (about $4 billion) amounts to less than half of what companies such as Microsoft or Toyota spend on R&D in a single year.[104] While it is very difficult to know what major defence contractors are researching outside advertised R&D state-funded contracts, it is notable that a number of major defence contractors in the USA, and to a lesser extent in Europe, have made efforts in recent years to form closer relationships with the academic institutions working on AI and robotics. The most recent and relevant example is the multi-year collaboration agreement that Lockheed Martin (USA) signed with MIT's Department of Aeronautics and Astronautics (in collaboration with the Computer Science and Artificial Intelligence Laboratory) to work on robotics and autonomous systems.[105]

Approach to autonomy: ICT versus traditional automotive and aerospace companies

The companies that are deemed most likely to play a critical role in shaping the future of AI and robotics are typically (*a*) companies with a background in ICT, such as Apple, IBM, Intel and Microsoft; (*b*) internet services giants, such as Alphabet (Google), Amazon, Baidu and Facebook; or (*c*) large and established corporations from the aerospace and automotive sectors, such as Airbus, Boeing, Toyota and Volvo. Their influence on the fields of AI and robotics (and, therefore, autonomy) takes many forms. To begin with, they have reportedly been luring the most talented individuals in AI, especially those involved in machine learning and robotics research, away from universities.[106] They have large financial resources at their disposal, which have allowed many of them to acquire, in recent years, some of the companies that are considered to be at the forefront of innovation in AI and robotics.[107] Examples include Google's acquisition of Boston Dynamics, perhaps the world's most famous robotics company,

---

[104] Lynn (note 8).

[105] 'MIT, Lockheed Martin launch long-term research collaboration', *MIT News*, 16 May 2016.

[106] Levy (note 25); and Hernandez and King (note 24).

[107] Apple and Google have nearly $180 billion and $60 billion in cash respectively, dwarfing the amount held by any company within the defence sector. In fact, Google has sufficient cash to buy out any of the major defence contractors, which illustrates the size and power both of the company and of the growing global technology sector. Lynn (note 8).

in 2013 and DeepMind, a specialist in AI technology, in 2014, and Intel's acquisition in 2017 of Mobileye, a leader in computer vision for self-driving vehicles.[108] Companies like Google are also able to pour vast cash resources into internal R&D, including basic research.[109]

It is worth emphasizing that ICT companies and traditional aerospace and automotive companies tend to have different approaches to innovation in the field of autonomy: the aerospace and automotive industries tend to have a more conservative and prudent approach to the development of autonomy. This is clearly visible in the case of self-driving vehicles. ICT companies announced when they entered this business that they would aim to develop fully autonomous systems; whereas traditional carmakers, such as BMW, Toyota or Volvo, have favoured a 'shared control' model where autonomy would allow vehicles to work in collaboration with human drivers rather than replace them. This difference in approach derives from the fact that the aerospace and automotive industries have long experience of working with automation in safety-critical systems. They know that they need to put a lot of effort into reliability and security because of the consequences of a failure of their products. ICT companies operate in a paradigm where a systems failure does not matter as much (i.e. the consequences of a systems failure in a mobile phone are likely to have less of an impact than a systems failure in a car). They are therefore more willing to test and adopt cutting-edge technologies and design.

VI. Conclusions

The innovation ecosystem that is driving the advance of autonomy in weapon systems is diffuse, chiefly because the technologies, academic disciplines and industry sectors involved in the development of autonomous capabilities may vary greatly depending on the type of application and systems at issue. Nevertheless, three general observations can be made.

First, much of the fundamental research in the fields of AI and robotics that could feed the advance of autonomy in weapon systems is dual-use. The divergence between civilian and military innovation generally emerges towards the development end of the R&D cycle because civilian and military products often need to meet different performance criteria. Hence, should CCW delegates eventually engage in a formal discussion on the monitoring or regulation of R&D efforts that could lead to the development and production of LAWS, they should focus on the development end of the R&D cycle, as this is where the actual capabilities of LAWS will be definitively created. Attempting to monitor and control R&D at the more basic research level would be challenging from a practical perspective and possibly problematic as it could threaten civilian innovation.

Second, the barriers to entry to the development of robotic systems are very low. Nearly all hardware components that might serve the development of autonomous robots are commercially available. It is even possible to acquire off-the-shelf low-cost robotic systems that feature advanced autonomous capabilities. These may be adopted, modified and weaponized by states but also, and more worryingly, by non-state actors seeking, for instance, to conduct terrorist operations. This scenario has not yet received great attention within the CCW discussions on LAWS, despite the fact it represents an

[108] Gibbs, S., 'What is Boston Dynamics and why does Google want robots', *The Guardian*, 17 Dec. 2013; Smith, R., 'Google is selling Boston Dynamics: but who's buying?', Motley Fool, 25 Apr. 2016; Reuters, 'Google to buy artificial intelligence company DeepMind', 26 Jan. 2014; 'The last AI breakthrough DeepMind made before Google bought for $400m', Physics arXiv Blog, 29 Jan. 2014; and Intel, 'Intel to acquire Mobileye', News release, 14 Mar. 2017.

[109] Levy (note 25); Hernandez and King (note 24); and 'Google creates new European research group to focus on machine learning', Venture Beat, 16 June 2016.

imminent humanitarian risk. While it falls outside the traditional scope of the CCW, it would be prudent for the GGE to allocate some time to this issue in 2018. It could start a discussion on the options that are offered outside the CCW to control or limit the diffusion and malevolent use of key technologies. This could include discussing the possibilities offered by export control mechanisms and self-control by the industry.

Third, future discussions on the development and control of autonomy in weapon systems could usefully benefit from further exchanges of experience with the civilian sector, especially companies developing safety-critical applications of autonomy (e.g. aerospace companies, carmakers and medical robot companies), considering that a number of issues that are central to discussion on LAWS have already been, or are currently being, actively addressed within the civilian sphere. These issues include the following.

1. *How to define and measure autonomy?* This question has been the concern of standardization and regulatory bodies for a long time. The International Organization for Standardization (ISO) and the International Electrotechnical Commission have had multiple projects aimed at generating an official definition of robot-associated terms, including 'autonomy' (ISO 8373:2012). In addition, the US National Highway Traffic Safety Association has adopted the Society of Automotive Engineers' levels of autonomy for self-driving systems, ranging from complete driver control to full autonomy.[110]

2. *How to operationalize meaningful human control?* Civilian industry sectors that produce safety-critical systems (e.g. aerospace, automotive and medical robotics) are facing the same human control dilemmas as the defence sector. They too are dedicating their efforts to finding a model of the human–machine command-and-control relationship that will maximize safety.

3. *How to test the safety and predictability of autonomous technologies?* The commercial aerospace sector has procedures to test and verify advanced automated systems. Inviting experts from this community to talk about existing procedures would help to identify possible best practices for testing and evaluating weapons with advanced autonomous capabilities in the context of Article 36 reviews.

---

[110] Reese, H., 'Updated: autonomous driving levels 0 to 5: understanding the differences', Tech Republic, 20 Jan. 2016.

6. Conclusions

This final chapter summarizes the analysis contained in this report. It focuses on the conclusions that are of particular relevance for diplomats, scholars and advocacy groups that are involved in discussions on LAWS within the framework of the CCW. It begins with a synthesis of the key findings of the mapping exercise.

I. Key findings

This section summarizes the answers the report has provided to the following questions.

 1. What are the technological foundations of autonomy?
 2. What is the state of autonomy in weapon systems?
 3. What are the drivers and obstacles to the development of autonomy in weapon systems?
 4. Where are relevant innovations taking place?

What are the technological foundations of autonomy? Mapping the conceptual and technical foundations of autonomy

Defining autonomy

In simple terms 'autonomy' can be defined as the ability of a machine to execute a task, or tasks, without human input, using interactions of computer programming with the environment. An autonomous system is, by extension, usually understood as a system—whether hardware or software—that, once activated, can perform some tasks or functions on its own. However, autonomy is a relative notion: within and across relevant disciplines, be it engineering, robotics or computer science, experts have a different understanding of when a system or a system's function may or may not be deemed autonomous. A very common approach for assessing autonomy relates to human–machine command-and-control relationships—that is, the extent to which humans are involved in the execution of the task carried out by the machine. With this approach, the systems can be classified into three categories: semi-autonomous, human-supervised autonomous or fully autonomous. A more technical approach to autonomy relates to the sophistication of a system's decision-making capability— that is, the actual ability of a system to exercise control over its own behaviour (self-governance) and deal with uncertainties in its operating environment. From this standpoint, systems are often sorted into three major categories: automatic, automated and autonomous systems. However, the definitions of, and boundaries between, these three categories are contested within and between the expert communities. A third dimension to consider focuses on the types of decisions or functions that are made autonomous within a system. Referring to autonomy as a general attribute of systems is imprecise, if not meaningless, as the nature of human–machine command-and-control relationships and the sophistication of a machine's decision-making capability may vary from one function to another. Discussions on the advance of autonomy in the area of weapon systems need to be situated firmly in their appropriate contexts.

Unravelling the machinery: how does it work?

At the fundamental level, autonomy is always enabled by the same technological architecture: sensors that allow the system to gather data about the world, a suite of computer hardware and software that permits the system to turn data perceived from

the environment into purposeful plans of action, and actuators and end-effectors that allow the system to execute the actions in its operating environment. The actual characteristics of these underlying technologies then differ depending on the nature of the task and the operating environment.

Advances in autonomy in weapon systems are dependent on technological progress in multiple areas. Advances in sensor technologies are certainly crucial as they determine the accuracy of the data that systems can collect on their operating environments. Likewise, advances in computer processing technologies play an important role as they determine the speed at which the software part of a system can 'think' as well as the volume of data that it can efficiently handle. The design of the actuators and end-effectors will also affect the hardiness, endurance and cost of the systems.

The technologies that are deemed the most critical to autonomy, however, are the software elements. It is the complexity of sensing software and control software that actually determines the performance of autonomy in a system. Thus, advances in autonomy fundamentally depend on the ingenuity of human programmers to find a way to break down complex problems into mathematical rules and instructions that a computer can understand.

Creating autonomy: how difficult is it to achieve autonomy?

Achieving autonomy is, by definition, not actually that difficult. It is chiefly based on the extent to which the type of intended task can be modelled mathematically. The more abstract or ill-defined the task specifications, the harder it is to formulate it in terms of a mathematical problem and a solution. Task executions that require qualitative judgements are often problematic because the outcome cannot be assessed in objective terms. Likewise, tasks that require complex interactions (be they adversarial or collaborative) with humans are also difficult to engineer because human behaviour is often unpredictable, and hence hard to model. The extent to which the operating environment is predictable is also essential. The less predictable it is, the harder it is to model in advance, and the more perceptual and decision-making intelligence the system needs to have.

When tasks and operating environments are too complex for a human to model completely, software developers increasingly rely on 'machine learning'. Machine learning is an approach to software development through which systems learn tasks and improve their performance through experience. Machine learning has been around for decades but has made great strides in recent years, which has created important opportunities for the development of autonomy in weapon systems. Machine learning could, for instance, improve the capabilities of the subsystems that rely on pattern recognition mechanisms, such as ATR software, vision-based guidance systems, malware detection software, and anti-jamming systems. Machine learning, however, also poses a number of practical challenges, especially with regard to predictability. In the systems that use deep learning, algorithms operate like black boxes. It is particularly difficult for humans to understand what such systems have learned and how they might react to the input of data that is different from that which was used during the training phase. In the context of weapon systems, this unpredictability could have dramatic consequences. This is one of the reasons why the use of machine learning in the context of weapon systems remains, for now, limited to experimental research.

What is the state of autonomy in weapon systems? Mapping the functions and capabilities of autonomy in weapon systems

Autonomy already supports multiple functions in weapon systems

Extensive research on existing weapon systems clearly shows that autonomy is already used in a wide variety of tasks in weapon systems, including many connected to the use of force. Mobility-related functions are by far the most common autonomous function within weapon systems. When it comes to the use of force, autonomy is already employed to support many steps of the targeting process, from target identification, tracking, prioritization and selection to target engagement in certain cases. The newer systems sometimes include autonomous capabilities for health management (typically refuelling, power management and fault detection), interoperability (i.e. systems can autonomously collaborate with other systems) or battlefield intelligence (i.e. mapping of 2-D and 3-D environments, explosive ordnance disposal etc.).

Systems that include 'automated' targeting capabilities have been produced and used for decades and are primarily employed for defensive purposes

ATR systems, the technology that enables weapon systems to detect targets autonomously, have existed since the 1970s. ATR is used in guided munitions, loitering weapons, air defence systems and within human-operated weapon systems (e.g. manned systems or remote-controlled unmanned systems), usually to aid human operators to locate and engage with enemy targets that are beyond the visual range of human operators or too fast for them to track.

Systems that can acquire and engage targets autonomously are predominantly designed for defensive purposes

Weapon systems that can acquire and engage targets autonomously are predominantly delegated to defensive uses, for example, to protect ships, ground installations or vehicles against incoming projectiles. They are operated under human supervision and are intended to fire autonomously only in situations where the time of engagement is deemed too short for humans to be able to respond. At least 89 countries have deployed or are developing such systems.

Systems that can acquire and engage targets autonomously in 'offensive' missions are constrained as to the types of targets they can fire upon

Loitering weapons are the only 'offensive' weapon system type that is known to be capable of acquiring and engaging targets autonomously. The loitering time and geographical areas of deployment, as well as the category of targets they can attack, are determined in advance by humans. The only operational loitering weapons that are known to be capable of operating in a full autonomous mode are the Harpy, Harop, Harpy NG and the Orbiter 1K 'Kingfisher' (all produced by Israel).

The autonomy of 'deployed' weapon systems remains rudimentary from a technical standpoint

From a technical perspective, the autonomy of most systems remains rudimentary. Autonomous navigation in UASs and unmanned marine systems (UMSs) typically relies on waypoint navigation and collision avoidance technology. Autonomous navigation in UGSs is generally only possible in low-complexity environments that can be pre-mapped in advance, such as borders and perimeters. With regard to the use of force, it should be emphasized that current systems have very basic perception and decision-making capabilities: they can only detect and engage with large material

targets that match predefined criteria. The few systems that are able to detect, prioritize and engage targets autonomously also do so under constrained parameters. The performance of their ATR systems also rapidly deteriorates as the operating environment becomes more cluttered and weather conditions deteriorate.

What are the drivers of, and obstacles to, the development of autonomy in weapon systems? Mapping the factors shaping the future of autonomy in weapon systems

Autonomy provides numerous operational benefits

Autonomy offers substantial benefits from an operational perspective. It provides opportunities for faster and more reliable task executions. It also has the potential to discharge humans from performing or supervising dull, dirty or dangerous tasks, and could thereby contribute—albeit this is debatable—to reducing the manpower burden that heavily impacts military budgets in many countries. In addition, autonomy holds the promise of enabling new operational capabilities: it can permit access to environments that are unreachable for remote-controlled technologies and facilitate swarming operations, which have the potential to provide militaries with greater mass on the battlefield.

'Full autonomy' might not necessarily be the objective

The military has multiple reasons to accelerate the incorporation of autonomy into weapon systems. This does not necessarily mean that there will be a linear development towards full autonomy, with the 'man being taken out of the unmanned' altogether, so to speak. The current narrative of the US DOD—which is the only actor in the world with an identifiable and articulate strategy and policy on autonomy—is that the objective is not to achieve 'full autonomy'; rather, the model is one where autonomy does not replace human decision but adequately complements it. With regard to the use of force, the dominating narrative is that humans should continue to exert control over engagement decisions.

There remain numerous technical challenges to achieving the popular vision of what autonomous weapon systems should be capable of

There are a number of technical issues that limit the further incorporation of autonomous capabilities in weapon systems. Advances in computer vision and machine learning will be needed to fulfil the military vision of what autonomous weapon systems should be able to do: operate safely and reliably in complex, dynamic and adversarial environments. Further research in HRI will be needed to ensure that weapon systems can remain under meaningful human control. Finally, future advances of autonomy in weapon systems are also highly dependent on the development of new V&V procedures. Existing V&V methods only allow for low levels of autonomy to be certified for use.

Institutional, legal, normative and economic factors hinder the incorporation of autonomous capabilities

The development of military technology does not take place in a vacuum. There are a number of institutional, legal, normative and economic factors that make the development of autonomy in weapon systems a controversial topic for military organizations. The adoption of autonomous capabilities remains a sensitive issue within the military. Ensuring military personnel's trust in the safety and reliability of autonomous capabilities is perhaps the key challenge. Autonomy also represents a threat to the very

ethos of some military communities. Some members of the military, across the different services, may see the development of certain autonomous capabilities as a direct threat to their livelihoods. Alternatively, they may view such autonomous capabilities as inadequate when compared with the operational paradigms they are used to.

The military is also constrained by international law—and in some cases national law—which imposes restrictions on the development and use of autonomous targeting capabilities, and requires military command to maintain, in most circumstances, some form of human control or oversight over the weapon system's behaviour. In addition, there is growing normative pressure from within civil society against the use of autonomy for targeting decisions, which potentially makes the development of autonomous weapon systems a politically sensitive issue for the military.

Finally, certain economic variables need to be recognized. There are limits to what can be afforded by national armed forces. Considering the cost of weapon systems, advances in autonomy in weapon systems are bound to take place at different speeds depending on the type of system and country of development. However, it can be assumed that, due to their affordability and accessibility, micro and small weapon platforms will drive the adoption of autonomous capabilities in weapon systems in the short term.

Where are relevant innovations taking place?

Mapping innovation in machine autonomy is challenging from a methodological standpoint as autonomy is neither a specific technology area with well-defined boundaries, nor a dedicated academic discipline or a distinct market sector.[1]

A science and technology perspective

At the basic science and technology level, advances in machine autonomy derive primarily from research efforts in three disciplines: AI, robotics and control theory. The AI and robotics fields potentially overlap. Furthermore, in addition to sharing a number of research issues, they have in common that they are interdisciplinary and the contact point for many other fields of science and technology, including biology, psychology, linguistics and mathematics. There are no worldwide university rankings on both AI and robotics, but there are indications that the academic landscape in these fields is dominated by North American, West European and East Asian universities.

A geographical perspective

SIPRI has attempted to map the governmental R&D efforts in autonomy of the 10 largest arms-producing countries—the USA, the UK, Russia, France, Italy, Japan, Israel, South Korea, Germany and India—and China. The USA is the only country that has released a standalone military R&D strategy on autonomy, to which it also attached a distinct budget line. The largest shares of the funds that the US DOD allocated to applied research and experimental development in the area of autonomy in 2015 were dedicated to (*a*) human and autonomous systems interaction and collaboration; (*b*) machine perception, reasoning and intelligence; (*c*) teaming of autonomous systems; and (*d*) testing and evaluations. The USA also funds numerous civilian or dual-use R&D projects on AI and robotics that could serve as building blocks for the development of autonomous capabilities in weapon systems. The US DOD plays a major role in this process.

---

[1] 'Autonomy' is defined here as the ability of a technology to execute a task, or tasks, without human input, using interaction of computer programming with the environment. This definition is based on one previously proposed by Andrew Williams. Williams, A., 'Defining autonomy in systems: challenges and solutions', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO: Norfolk, VA, 2015).

Most of the other largest arms-producing countries are also known to have identified AI and robotics as important R&D areas. While there is clear evidence that most are investing significant resources in civilian and dual-use applications, finding details about their equivalent military R&D activities in open sources has proved difficult in many cases. A review of expert commentaries and the policy statements and various R&D projects of these countries indicates that most of them are paying a lot of attention to how the US DOD funds R&D in AI and robotics. To some extent they are enacting similar initiatives, but they are generally on a smaller scale or more specific in nature.

An industry perspective

It is established that the civilian industry leads innovation in autonomous technologies. The most influential players are major ICT companies such as Alphabet (Google), Amazon and Baidu, and large companies in the automotive industry, such as Toyota, that have moved into the self-driving car business. Their role is significant because they are developing a range of AI applications and autonomous robots with military potential (including autonomous delivery UAVs, computer vision systems for video analysis, self-driving vehicles and speech recognition interfaces) and also because they dedicate substantial resources to basic R&D relating to autonomy. It is notable that companies with a background in the ICT sector tend to have a more radical approach to the development of autonomy than 'traditional' companies (e.g. the aerospace and automotive industries), as the latter usually place greater emphasis on reliability and safety.

Arms producers are certainly involved in the development of autonomous technologies but the amount of resources that these companies allocate to R&D is far less than that mobilized by large corporations in the civilian sector. However, the role of defence companies remains crucial, because commercial autonomous technologies can rarely be adopted by the military without modifications. The military domain is much more stringent in terms of technical requirements than the civilian sector. Autonomous systems that are intended to operate in battlefield conditions may need far more advanced perceptual and decision-making intelligence than those that operate in civilian contexts.

II. Recommendations for future CCW discussions on LAWS

The key findings of this report have a number of concrete implications for future CCW discussions. These can be crystallized into eight recommendations that are intended to help the newly established GGE to foster a constructive basis for discussion and achieve tangible progress on some of the crucial aspects under debate.

1. Discuss the development of 'autonomy in weapon systems' rather than autonomous weapons or LAWS as a general category

A focus on autonomy as a general attribute of a weapon system is imprecise and potentially misleading. Autonomy may serve very different capabilities in different weapon systems. For each of these capabilities, the parameters of autonomy, whether in terms of the human–machine command-and-control relationship or the sophistication of the decision-making process, may vary greatly, including over the duration of a mission. In this regard, the continued reference to the concept of LAWS in the framework of the CCW is problematic. It traps states and experts into a complex and contentious discussion about the level at which a system might be deemed autonomous, while in reality the concerns—be they from a legal, ethical or operational standpoint—need to

be articulated on the use of autonomy for specific functions or tasks. Future CCW discussions could, therefore, benefit from a conceptual reframing and a shift from a platform- or system-centric approach to a functional approach to autonomy. Focusing on the concept of 'autonomy in weapon systems' rather than the concept of 'LAWS' could foster a much more consensual and constructive basis for discussion.

2. Shift the focus away from 'full' autonomy and explore instead how autonomy transforms human control

Thus far, the debate on LAWS has been platform-centric and also very much focused on the development of 'full autonomy'. The focus on full autonomous systems is somewhat problematic as it does not reflect the reality of how the military is envisioning the future of autonomy in weapon systems, nor does it allow for tackling the spectrum of challenges raised by the progress of autonomy in weapon systems in the short term. Autonomy is bound to transform the way humans interact with weapon systems and make decisions on the battlefield, but will not eliminate their role. Weapon systems will never be 'fully' autonomous in the sense that their freedom of action will always be controlled by humans at some level and their programming will always be the product of human plans and intentions. Hence, when exploring the advance of autonomy, the fundamental issues that should be addressed by the CCW community are those of human control: How is the progress of autonomy changing the nature, location and timing of human decision making and action in warfare? What control should humans maintain over the weapon systems they use and what can be done to ensure that such control remains adequate or meaningful as weapon systems' capabilities become increasingly complex and autonomous?

3. Open the scope of investigation beyond the issue of targeting to take into consideration the use of autonomy for collaborative operations and intelligence processing

There is growing agreement among CCW delegates that autonomy raises issues primarily in the context of targeting processes, whether from a legal, ethical or security standpoint. However, advances in autonomy in other functional areas should remain under scrutiny for a number of reasons. First, some 'non-critical' autonomous functions may act as force multipliers on the offensive capability of weapon systems (example capabilities include navigation, swarming and self-repair). Second, they raise certain concerns in terms of safety and human control: what are the parameters of human control, for instance, when weapon systems operate in a large swarm? Third, the technological developments that fuel advances in some functional areas, such as navigation, may also serve to improve autonomous targeting. The progress of image processing software for vision-guided navigation may, for example, be beneficial to the improvement of target recognition software. To foresee possible advances of autonomy in the area of targeting, it is important to monitor the overall progress of autonomy in weapon systems.

4. Demystify the current advances and possible implications of machine learning on the control of autonomy

If there is one technological development that future GGE discussions should address it is machine learning. Learning is often described as an increasingly important, if not defining, property of the future of autonomy in weapon systems. Among the community of CCW delegates, there seems to be a limited understanding of what machine learning actually is, how it works and to what extent it could unlock significant advances in autonomy in weapon systems. In the light of this, the GGE could assess the potential for machine learning to further advance autonomy in weapon systems and

also examine its limitations. Clarifications about the difference between 'offline' and 'online' learning—whether in terms of potential, limitations or risks—would be particularly welcome. In addition, one other near-term development deserves extra scrutiny: the use of deep-learning algorithms for the training of ATR systems. It would be useful to know what the implications of such a development would be, as they could be key to an assessment of the legality of a system under IHL when conducting Article 36 weapon reviews.

5. Use case studies to reconnect the discussion on legality, ethics and meaningful human control with the reality of weapon systems development and weapon use

To tackle the challenges posed by autonomy and define the possible parameters of human control, it could be useful to engage in a scenario exercise, using as case studies weapon system concepts that are in use or in development and that have been described in the literature. Using concrete examples could help to anchor the legal and ethical discussions on firm ground and allow the international community to deal with the challenges posed by autonomy in the near term. Possible case studies could include loitering weapons (for existing systems) and swarms of small UASs (for more futuristic systems).

Loitering weapons would make for an interesting case study because they are offensive weapons that are already in use. It would be instructive for the CCW discussion to clarify in what circumstances fielding such weapon systems might be deemed (il)legal and morally (un)acceptable. Variables that would need to be discussed in the course of the scenario exercise include (*a*) the nature and complexity of the area of deployment; (*b*) the loitering time; and (*c*) the human–machine command-and-control relationship during the loitering phase.

The case of a swarm of small UASs would be more thought-provoking because (*a*) it is an emerging capability that has been tested through various R&D projects, but not yet formally deployed in operations; (*b*) it could be used for a variety of missions; and (*c*) it might require a new paradigm in terms of human–machine command-and-control relationships. Therefore, not only would the scenario exercise have to review the legality and acceptability of these systems for different types of missions and operational circumstances, but it would also have to take into consideration variations in models of command-and-control for swarm operations.

6. Facilitate an exchange of experience with the civilian sector, especially the aerospace, automotive and civilian robotics industries

Future discussions on the development and control of autonomy in weapon systems could benefit from further exchanges of experience with the civilian sector, considering that a number of issues that are central to discussions on LAWS have already been, or are currently being, actively addressed within the civilian sphere. These issues include the following.

1. *How to define and measure autonomy?* This question has been the concern of standardization and regulatory bodies for a long time. The ISO and the International Electrotechnical Commission have had multiple projects aimed at generating an official definition of robot-associated terms, including 'autonomy' (ISO 8373:2012). In addition, the US National Highway Traffic Safety Association has adopted the Society of Automotive Engineers' levels of autonomy for self-driving systems, ranging from complete driver control to full autonomy.

2. *How to operationalize meaningful human control?* Civilian industry sectors that produce safety-critical systems (e.g. aerospace, automotive and medical robotics) are

facing the same human control dilemmas as the defence sector. They too are dedicating their efforts to finding the model of human–machine command-and-control relationship that will maximize safety.

3. *How to test the safety and predictability of autonomous technologies?* The commercial aerospace sector has procedures to test and verify advanced automated systems. Inviting experts from this community to talk about existing procedures would help to identify possible best practices for testing and evaluating weapons with advanced autonomous capabilities in the context of Article 36 reviews.

7. Investigate options to ensure that future efforts to monitor and potentially control the development of lethal applications of autonomy will not inhibit civilian innovation

Fundamental innovations in the fields of AI and robotics are often dual-use. The divergence between civilian and military innovation generally emerges towards the development end of the R&D cycle because civilian and military products often need to meet different performance criteria. Should CCW delegates eventually engage in a formal discussion on the monitoring or regulation of R&D efforts that could lead to the development and production of LAWS, they should focus on the development end of the R&D cycle, as this is where the actual capabilities of LAWS will be definitively created. Attempting to monitor and control R&D at the more basic research level would be challenging from a practical perspective and possibly problematic as it could threaten civilian innovation. In this respect, it might be helpful to engage with the civilian AI and robotics research communities to learn more about the existing codes of conduct and frameworks of responsible research and innovation (e.g. the 2013 Rome Declaration on Responsible Research and Innovation).[2]

8. Investigate the options for preventing the risk of weaponization of civilian technologies by non-state actors

Although the scope of possible future regulation on the development and production of autonomy and weapon systems should not be too wide in order to minimize the impact on civilian innovation, the risks of weaponization of civilian technologies by non-state actors should be taken into consideration by the CCW community. The barriers to entry to the development of autonomous systems are very low. Most components that may be used to develop autonomy are widely available in the commercial sector. The main limitation to the creation of autonomy is the ingenuity of human programmers. The risk of terrorists or criminal organizations developing low-cost autonomous weapon systems is therefore very real. Thus, an expert presentation on the different options available to the international community to control or limit the diffusion and malevolent use of civilian technologies would be a useful addition to future CCW discussions. This could include discussion of the possibilities offered by export control mechanisms and also the technical solutions that could be applied by companies, such as introducing safeguards into the systems hardware and software that could limit their use or allow for deactivation of the product.

---

[2] European Commission, Science with and for Society, Rome Declaration on Responsible Research and Innovation in Europe, 21 Nov. 2014.

Glossary

Artificial intelligence (AI) (discipline): Broadly defined as the science and engineering of making intelligent machines. In academia, AI is usually described as a branch of computer science that focuses on solving problems through logic and reasoning.

Automatic/automated system: A system—whether hardware or software—that is governed by prescriptive rules and, once activated, can perform some tasks or functions without human involvement. See also: *Autonomy*.

Automated target recognition (ATR): A software program that enables a weapon system to find and track targets based on predefined target signatures.

Autonomous system: A system—whether hardware or software—that, once activated, can perform some tasks or functions without human involvement using interaction of sensors and computer programming with the environment. It differs from an 'automatic system' through its ability to compose and select from different courses of action to accomplish goals based on its knowledge and understanding of the world, itself and the situation. See also: *Automatic/automated system*.

Autonomy: The ability of a system to execute a task, or tasks, without human involvement, using interaction of its sensors and computer programming with the environment.

Critical functions: The functions in a weapon system that allow it to select (i.e. search for, detect, identify or track) and attack (i.e. use force against, neutralize, damage or destroy) targets. See also: *Function (system)* and *Weapon system*.

Function (system): In the context of a system, function refers to the action, task or process that a system performs, which can include, among others, navigation, take-off and landing, fault detection and target identification.

Identification, friend or foe (IFF): A subsystem to determine whether aircraft, vehicles or forces are friendly. An electromagnetic signal is sent by a transponder, and if the receiver responds with a valid reply, it is considered friendly.

Machine learning: An approach to software development that consists of building a system that can learn, and then teaching it what to do using a variety of methods.

Military system: A system of equipment used by the military, which can be, but is not necessarily, armed. A weapon system is a subtype of military system that is, by definition, armed.

Robotics: The field of science and engineering that is dedicated to the development of robots. See also: *Robots*.

Robots: Computer-enabled machines that can sense and purposefully act on, or in, their environment. See also: *Robotics*.

System: A set of components that cooperate as part of a network, forming a unified whole.

Targeting, autonomous: The capability of weapon systems to search for, detect, identify, track, prioritize, select and/or engage with targets. Autonomous targeting can be divided into (*a*) target acquisition (i.e. the ability to search for, detect, identify, track, prioritize and/or select targets to permit the effective employment of weapons); and (*b*) target engagement (i.e. the ability to decide to intercept, use force against, neutralize, damage or destroy targets).

Targeting cycle: The entire process of selecting targets to be attacked, destroyed or taken in warfare. It is a broader process than mere targeting by weapon systems, as it also includes aspects such as setting of strategic goals by high command; design of approved target sets; goal analysis; target development; validation, nomination and prioritization; capability analysis; assignment of capabilities; execution of the mission; and post-strike assessment.

Unmanned systems (UxS): Refers to any type of unmanned system without specifying the domain in which it operates. Includes unmanned aerial systems (UASs), unmanned maritime systems (UMSs) and unmanned ground systems (UGSs).

Waypoint navigation: An aspect of autonomous mobility. The process of automatically following a predetermined path, defined by geodetic coordinates. This can be done through, among other things, GPS, stellar navigation or optical/radar observation of points.

Weapon platform: The platform on which a weapon system is mounted (e.g. a combat aircraft on which missiles are mounted). See also: *Weapon system*.

Weapon system: A system that may consist of multiple physical platforms, including carrier and launch platforms, sensors, fire control systems and communication links needed for a weapon to engage a target. See also: *Weapon platform*.

Appendix

Table A. Air defence systems with autonomous engagement

Name	Country	Company	Tracking range (km)	Status	Year	No. of users
AK-630	Russia	KBP Instrument Design Bureau	6	Development completed	1976	19
Arrow 2	Israel; USA	Elta; IAI; Rafael; Tadiran Electronics	500	Development completed	2000	1
Aspide/Albatros**	Italy	MBDA	45	Development completed	..	16
Aster 30 SAMP/T	France	Eurosam	>100	Development completed	2001	2
Bamse/RBS 23	Sweden	Bofors; Ericsson; Saab	20	Development completed	2002	1
Crotale EDIR	France	Thales	..	Development completed	1978	12
DARDO	Italy	Leonardo-Finmeccanica	3	Development completed	1977	9
High Energy Laser Mobile Demonstrator (HEL-MD)	USA	Boeing; US Army	1.5	Under development	..	0
Kashtan	Russia	KBP Instrument Design Bureau	8	Development completed	1989	5
LD2000	China	Norinco	3 000	Development completed	2005	1
NBS MANTIS	Germany	Rheinmetall	3	Development completed	2011	1
PAAMS/Sea Viper	France; Italy; UK	Eurosam	113	Development completed	2009	2
Pantsir-S1	Russia	KBP Instrument Design Bureau	20	Development completed	2012	10
Patriot	USA	Raytheon	80	Development completed	1984	14
Phalanx	USA	Raytheon	5.5	Development completed	1980	22
RAPID series	France	CTA; Nexter; Thales	..	Development completed	2012	0
SeaRAM	Germany; USA	Diehl; Raytheon	9	Development completed	2014	2
Seawolf	UK	BAE Systems (UK)	6	Development completed	1979	4
Skyguard GDF-005	Switzerland	Rheinmetall	4	Development completed	1985	35

.. = data not available.

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.

* Indicates conflicting information.

** Indicates a lack of sources.

If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table B. Active protection systems with autonomous engagement

Name	Country	Company	Status	Year	No. of users
Afghanit*	Russia	KBP Instrument Design Bureau	Under development	2017	1
AMAP-ADS	Germany	Rheinmetall	Development completed	2009	>2
Arena	Russia	Kolomna-based Engineering Design Bureau	Development completed	1993	1
AWISS	Germany	Diehl	Development completed	2011	0
Drozd	Russia	KBP Instrument Design Bureau	Development completed	1978	>2
Drozd-2	Russia	KBP Instrument Design Bureau	Development completed	1999	0
Galix	France	Giat Industries; Lacroix Defense and Security	Development completed	1992	5
Iron Curtain	USA	BAE Systems (USA)	Development completed	2013	1
Iron Fist	Israel	IMI	Development completed	2006	2
Korean Active Protection System (KAPS)	South Korea	Agency for Defense Development (South Korea); Defense Acquisition Program Administration (South Korea)	Development completed	2012	1
MUltifunctional Self protection System (MUSS)	Germany	EADS (Airbus); Krauss Maffei Wegmann	Development completed	..	1
Quick Kill	USA	Raytheon	Development completed	..	0
Scudo ('Shield')	Italy	Leonardo-Finmeccanica	Development completed	2011	0
Shtora	Russia	Elers Elektron; NII Transmash	Development completed	1988	1
System Hard Kill (Shark)*	France	IBD Deisenroth Engineering; Thales	Development completed	2006	1
Trench Coat**	Israel	Rafael	Under development	..	0
Trophy/ASPRO-A/Windbreaker	Israel	Elta; IAI; Rafael	Development completed	2009	1

.. = data not available.

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.

  * Indicates conflicting information.

  ** Indicates a lack of sources.

  If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table C. Robotic sentry weapons with autonomous engagement

Name	Country	Company	Status	Year	No. of users
Sentry Tech	Israel	Rafael	Development completed	2007	1
SGR-A1	South Korea	Samsung	Development completed	2006	0
Super aEgis II	South Korea	DODAAM Systems	Development completed	..	3

.. = data not available.

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed; (*b*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.
  * Indicates conflicting information.
  ** Indicates a lack of sources.
  If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table D. A non-representative sample of guided munitions

Name	Country	Company	Munition	Target guidance	Status	Year	No. of users
Assigned a specific target or a target type							
Dual-Mode Brimstone	UK	MBDA	Land attack-missile	Radar	Development completed	2008	2
Long-Range Anti-Strike Missile	USA	Lockheed Martin	Anti-ship missile	..	Under development	2018	0
Mark 60 CAPTOR	USA	Goodyear Aerospace	Encapsulated torpedo mine	Sonar	Development completed		0
Naval Strike Missile/Joint Strike Missile	Norway	Kongsberg	Anti-ship missile; Land attack-missile	IR	Development completed	2012	3
PMK-2	Russia	..	Encapsulated torpedo mine	Sonar	Development completed		2
Assigned a specific target							
Black Shark	Italy	WASS	Torpedo	Wire; Sonar	Development completed	2004	6
Perseus	France; UK	MBDA	Cruise missile	LIDAR; Radar; Laser	Under development	..	0
Spike	Israel	Rafael	Anti-personnel missile; Anti-tank missile	Wire; IR; EO; RF	Development completed	1981	26
V-2	Nazi Germany	Peenemünde Army Research Center	Rocket	Gyroscope	Development completed	1944	4

Name	Country	Company	Munition	Target guidance	Status	Year	No. of users
Assigned a target location							
CBU-105 Sensor-Fuzed Weapon	USA	Textron Instruments	Sensor-fuzed weapon	Laser; IR	Development completed	..	8
GBU-39 Small Diameter Bomb	USA	Boeing	Bomb	GPS	Development completed	2007	4
Paveway IV	UK	Raytheon	Bomb	Laser/GPS	Development completed	2008	2
SM Art 155mm	Germany	GIWS	Artillery round	Radar; IR	Development completed	2000	4
Sudarshan	India	Bharat	Bomb	Laser	Development completed	2006	1
Strix	Sweden	Saab Bofors	Mortar round	IR	Development completed	1989	2
Guidance kits to give unguided bombs target guidance							
LT PGB	China	Luoyang Electro-Optics Technology Development Centre (EOTDC)	Guidance kit	Laser	Development completed	2006	1
Joint-Direct Attack Munition	USA	Boeing	Guidance kit	INS; GPS	Development completed	1997	31
Spice	Israel	Rafael	Guidance kit	EO; GPS	Development completed	2003	1

.. = data not available; EO = electro-optical; GPS = Global Positioning System; INS = inertial navigation system; IR = infrared; RF = radio frequency.

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.

* Indicates conflicting information.

** Indicates a lack of sources.

If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table E. Loitering weapons with autonomous engagement

Name	Country	Company	Loitering time (minutes)	Status	Year	No. of users
Battlefield Loitering Artillery Direct Effect (BLADE)	Israel; UK; USA	BAE Systems (UK); Lockheed Martin; Praxis; Raytheon; Uvision	..	Cancelled	..	0
Harop	Israel	IAI	360	Development completed	2005	5
Harpy	Israel	IAI	120	Development completed	≤1994	6
Harpy NG	Israel	IAI	540	Development completed	2016	1
Low Cost Autonomous Attack System (LOCAAS)	USA	Lockheed Martin	30	Cancelled	2007	0

Name	Country	Company	Loitering time (minutes)	Status	Year	No. of users
Non-Line-of-Sight Launch System (NLOS-LS) LS	USA	Lockheed Martin	30	Cancelled	2011	0
Orbiter 1K 'Kingfisher'	Israel	Aeronautics	180	Development completed	2016	1
Tacit Rainbow	USA	Texas Instruments	..	Cancelled	1991	0
TARES/Taifun	Germany	Rheinmetall	240	Cancelled	2006	0

.. = data not available.

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.
 * Indicates conflicting information.
 ** Indicates a lack of sources.
 If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table F. A non-representative sample of unmanned combat systems with autonomous capabilities in their critical functions

Name	Country	Company	Domain	Status	Year	No of users
Autonomous target engagement and target acquisition						
Taranis	UK	BAE systems (UK)	Air	Under development	2030	0
No autonomous target engagement, only autonomous target acquisition						
Amstaff	Israel	Automotive Robotic Industries	Ground	Development completed	2010	2
Sea Hunter	USA	DARPA; Leidos	Maritime	Under development	2018	0

Note: 'Country' may include funding countries. 'Development completed' includes systems with the following developmental statuses: (*a*) developed: acquisition status unclear; in use; in production; and (*b*) developed: not acquired; retired. 'Under development' includes systems with the following developmental statuses: prototype: testing/evaluation; technology demonstrator; under development.
 * Indicates conflicting information.
 ** Indicates a lack of sources.
 If there is no star, the information can be considered reasonably trustworthy.

Source: SIPRI Dataset on autonomy in weapon systems.

Table G. Top 10 research institutions in the field of artificial intelligence based on volume of academic publications in sample of relevant topics, 2011–16

Ranking by publication topic

Rank	Artificial intelligence	Machine learning	Human–machine interaction	Natural language processing	Computer vision
1	Massachusetts Institute of Technology (USA)	Microsoft (USA)	Microsoft (USA)	Microsoft (USA)	Microsoft (USA)
2	Carnegie Mellon University (USA)	Max Planck Society (Germany)	Carnegie Mellon University (USA)	Google (USA)	Massachusetts Institute of Technology (USA)
3	Microsoft (USA)	Carnegie Mellon University (USA)	Massachusetts Institute of Technology (USA)	Max Planck Society (Germany)	Stanford University (USA)
4	Stanford University (USA)	IBM (USA)	University of Washington (USA)	Stanford University (USA)	Chinese Academy of Science (China)
5	University of Toronto (Canada)	Google (USA)	Georgia Institute of Technology (USA)	Carnegie Mellon University (USA)	ETH Zurich (Switzerland)
6	University of California, Berkeley (USA)	Stanford University (USA)	University College London (UK)	IBM (USA)	University of California (USA)
7	IBM (USA)	Chinese Academy of Science (China)	Stanford University (USA)	Massachusetts Institute of Technology (USA)	Carnegie Mellon University (USA)
8	Centre national de la recherche scientifique (France)	University of Toronto (Canada)	University of California, Berkeley (USA)	Centre national de la recherche scientifique (France)	Tsinghua University (China)
9	Nayang Technology University (South Korea)	Massachusetts Institute of Technology (USA)	IBM (USA)	University of Illinois, Urbana-Champaign (USA)	French Institute for Research in Computer Science and Automation (France)
10	University College London (UK)	University of California, Berkeley (USA)	Newcastle University (UK)	University of Edinburgh (UK)	University of California, Berkeley (USA)

Source: Microsoft Academic Search Index, accessed 9 Dec. 2016, <http://academic.research.microsoft.com/>.

Table H. Top 15 research institutions in the field of robotics based on volume of academic publications in sample of relevant topics, 2000–16

Ranking by publication topic

Rank	Autonomous systems	Robotics	Mobile robots
1	Carnegie Mellon University (USA)	Massachusetts Institute of Technology (USA)	Carnegie Mellon University (USA)
2	Massachusetts Institute of Technology (USA)	Carnegie Mellon University (USA)	Massachusetts Institute of Technology (USA)
3	University of Washington (USA)	Stanford University (USA)	University of Southern California (USA)
4	Imperial College London (UK)	John Hopkins University (USA)	Georgia Institute of Technology (USA)
5	University of Toronto (Canada)	University of Southern California (USA)	University of Freiburg (Germany)
6	Georgia Institute of Technology (USA)	Centre national de la recherche scientifique (France)	University of Pennsylvania (USA)
7	University of Michigan (USA)	University of Pennsylvania (USA)	Stanford University (USA)
8	University of California (USA)	Harvard University (USA)	Centre national de la recherche scientifique (France)
9	Stanford University (USA)	French Institute for Research in Computer Science and Automation (France)	ETH Zurich (Switzerland)
10	Centre national de la recherche scientifique (France)	Vattikuti Urology Institute (Finland)	École Polytechnique Fédérale de Lausanne (France)
11	University of Illinois, Urbana-Champaign (USA)	Columbia University (USA)	University of Tokyo (Japan)
12	Ohio State University (USA)	Cleveland Clinic (USA)	Tokyo Institute of Technology (Japan)
13	Beckman Institute for Advanced Science and Technology (USA)	Jet Propulsion Laboratory (USA)	University of Sydney (Australia)
14	National Technical University Athens (Greece)	École Polytechnique Fédérale de Lausanne (France)	University of Michigan (USA)
15	University of California, Berkeley (USA)	University of Illinois, Urbana-Champaign (USA)	Jet Propulsion Laboratory (USA)

Source: Microsoft Academic Search Index, <http://academic.research.microsoft.com/>.

sipri

**STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE**