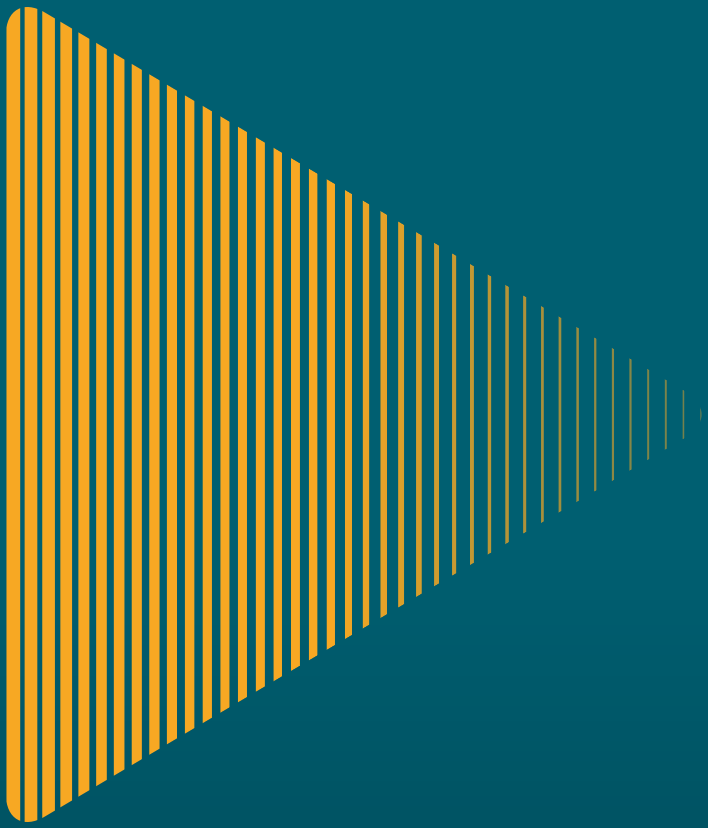
DRONE WARS

CENTER FOR
WAR
STUDIES

SDU
University of
Southern Denmark



MEANING-*LESS* HUMAN CONTROL

Lessons from air defence systems on meaningful human control for the debate on AWS

Ingvild Bode and Tom Watts

Published **February 2021**

Written by **Dr Ingvild Bode**, Centre for War Studies, University of Southern Denmark
and **Dr Tom Watts**, Department of Politics and International Relations, Hertfordshire University

Catalogue of air defence systems available on: https://doi.org/10.5281/zenodo.4485695

Published by the Centre for War Studies, University of Southern Denmark in collaboration with
Drone Wars UK, Peace House, 19 Paradise Street, Oxford OX1 1LD, www.dronewars.net

Acknowledgements

Research for this report was supported by a grant from the Joseph Rowntree Charitable Trust and by funding from the European Union's Horizon 2020 research and innovation programme (under grant agreement No. 852123). We want to express our thanks to the following readers for providing invaluable feedback for previous drafts of this report: Maya Brehm, Justin Bronk, Peter Burt, Chris Cole, Aditi Gupta, Hendrik Huelss, Camilla Molyneux, Richard Moyes, Max Mutschler, and Maaike Verbruggen. We would also like to thank Julia Barandun for her excellent early research assistance.

About the authors

Dr Ingvild Bode is Associate Professor at the Centre for War Studies, University of Southern Denmark and a Senior Research Fellow at the Conflict Analysis Research Centre, University of Kent. She is also the Principal Investigator of an ERC-funded project examining autonomous weapons systems and use of force norms. Her research focuses on processes of normative change, especially in relation to the use of force, weaponised Artificial Intelligence, United Nations peacekeeping, and United Nations Security Council dynamics.

Dr Tom Watts is a Lecturer in Politics and International Relations at Hertfordshire University. His research focuses on American foreign and security policy, remote warfare, and lethal autonomous weapons systems.

CONTENTS

ABBREVIATIONS

ADS	Air defence system
AWS	Autonomous weapons systems
CIC	Combat information centre
CIWS	Close-in Weapons System
GGE	Group of Governmental Experts
ICRC	International Committee of the Red Cross
IFF	Identification, friend or foe
IR655	Iran Air Flight 655
LAWS	Lethal autonomous weapons systems
MH17	Malaysia Airlines Flight 17
NGO	Non-governmental organization
PS752	Ukraine International Airlines Flight 752
ROE	Rules of engagement
SIPRI	Stockholm International Peace Research Institute
TBMs	Tactical ballistic missiles
TN	Track number
UN	United Nations
UN-CCW	1980 United Nations Convention on Certain Conventional Weapons
UNIDIR	UN Institute of Disarmament Research

EXECUTIVE SUMMARY

Autonomous weapons systems (AWS) are the subject of growing debate. At the heart of this debate is whether such systems reduce meaningful human control over the use of force. Much of the current debate on AWS focuses on future technological developments. Yet, a closer examination of how automated and autonomous features have already been integrated into the critical functions of air defence systems highlights that, in some situations, human control has become effectively meaningless. This report argues that air defence systems, whose importance has been neglected in the discussion of AWS, have set important and problematic precedents for the development and regulation of AWS.

This report documents how the development, testing, and operation of air defence systems with automated and autonomous features has, in specific targeting decisions, eroded the substance of meaningful human control over the use of force. In our assessment, it has made human control increasingly *meaningless*. We understand it to be 'meaningless' in two ways. First, in terms of the inability of human agents to exercise deliberative control over air defence systems because of the speeds at which these systems operate, the complexity of the tasks they perform, and the demands human agents are placed under (i.e. human control over the use of force lacks significance). Second, as the cumulative effect that the incremental processes of machine delegation have had on reducing the range and substance of *meaningful human control* in specific targeting decisions (i.e. human control has come to mean less over time).

Air defence systems are operated using controlled weapon parameters and environments. Nevertheless, in terms of human-machine interaction, they demonstrate significant problems regarding the meaningful exercise of human control. As a broader range of tasks have been delegated to machines, the human operators' role in the operation of air defence systems has changed from active control to passive supervision. The cumulative outcome of this shift has been to relegate human agents to minimal but impossibly complex roles. This outcome is acknowledged to be deeply problematic by experts in human-machine interaction, leading to some tactical changes in operating practices. Yet, it has not prompted a more fundamental reassessment of whether the ongoing integration of more and more automated and autonomous features in air defence systems is desirable or appropriate.

This indicates that decades of developing and operating air defence systems with automated and autonomous features have contributed toward the tacit acceptance of an *unspoken* norm. Norms are understandings of appropriateness that guide behaviour. They do not necessarily point to what is universally appropriate, but often to what a particular group of actors deems as suitable in a particular context. Air defence systems have contributed toward an emerging norm of what constitutes 'appropriate' human-machine interaction, a vital element of meaningful human control. This emerging norm is *implicitly* understood rather than *explicitly* stated by policymakers. It precedes and runs parallel to the international debate at the United Nations Convention on Certain Conventional Weapons that has been ongoing since 2014. It is deeply problematic because, in our reading, it normalises a reduced role for human operators in targeting decisions as acceptable and 'appropriate'. It thus runs counter to and undercuts the ongoing, deliberative attempts at codifying the meaningful human control as a general obligation as part of potential new international law on AWS.

Our research findings are based on a qualitative catalogue of automated and autonomous features in a global selection of twenty-eight air defence systems. This is coupled with a detailed analysis of human-machine interaction in four different air defence systems involved in failures that brought down civilian and military aircraft in friendly fire incidents. Our catalogue orientates the study of air defence systems in the wider debates on AWS away from a focus on their technical capabilities toward their role in 'normalising' the continued integration of autonomous features, and their consequences for what counts as meaningful human control in specific targeting decisions.

By drawing attention to the emergence of meaning*less* human control, this report demonstrates that the further integration of autonomous features into weapons systems is neither as desirable nor as inevitable as is generally assumed. To help facilitate a period of critical reflection, following the suggestions put forward by the Campaign to Stop Killer Robots, we support new international law on AWS based on meaningful human control as a central, positive obligation. To help ensure that such legal safeguards ensure meaningful rather than meaningless human control over the use of force, we make the following recommendations for stakeholders involved in the international debate on LAWS at the UN-CCW:

- Current practices of how states operate weapons systems with automated and autonomous features in specific use of force situations should be brought into the open and scrutinised. As we argue, such operational practices shape what constitutes 'meaningful' human control, especially the quality and type of human-machine interaction.
- Beyond air defence systems, more in-depth studies of the emerging standards set for meaningful human control produced by the use of other existing weapons systems with automated and autonomous features are required. Such studies can provide practical insights into the existing and future challenges to human-machine interaction created by autonomy and automation that, if not explicitly addressed, may shape silent understandings of appropriateness.
- Our study of air defence systems highlights that while all three components of meaningful human control (technological, conditional, and human-machine interaction) are important, control through human-machine interaction is the decisive element in ensuring that control remains meaningful. This is not least because human-machine interaction highlights meaningful human control at the specific point of using a weapons system, rather than the exercise of human control at earlier stages, such as during research and development.
- Control through human-machine interaction should be integral to any codification of meaningful human control in disarmament debates. We identify three prerequisite conditions needed for human agents to exercise meaningful human control: (1) a functional understanding of how the targeting system operates and makes targeting decisions, including its known weaknesses (e.g. track classification issues); (2) sufficient situational

understanding; and (3) the capacity to scrutinise machine targeting decision-making rather than over-trusting the system. Of course, human operators should also have the possibility to abort the use of force.

- These three prerequisite conditions (functional understanding, situational understanding and the capacity to scrutinise machine targeting decision-making) of ensuring meaningful human control in specific targeting situations set hard boundaries for AWS development that should be codified in international law. In our assessment, they represent a technological Rubicon which should not be crossed as going beyond these limits makes human control essentially meaningless. Adhering to these conditions does not only help ensure that human control remains meaningful, it also has the potential of easing the pressure put on human operators of air defence systems who are currently, unintentionally, set up to fail.
- The complexity inherent to human-machine interaction means that there will be limits to exercising meaningful human control in specific targeting decisions. Ensuring the stringent training of human operators is a necessary precondition for maintaining meaningful human control but is not a panacea. This inconvenient truth should be made clear to all relevant stakeholders.

1

INTRODUCTION

The development, testing, and usage of autonomous weapons systems (AWS) is arguably the defining security, legal, and ethical challenge of the twenty-first century. The militarisation of Artificial Intelligence (AI) is presented as the 'Third Revolution' in warfare after the invention of gunpowder and nuclear weapons.[1] The UK Ministry of Defence has identified "the use of autonomy and machine learning" as key in "achieving Defence capabilities".[2] Contrary to its public opposition toward AWS development, the British government has funded the research and development of key enabling technologies through the Defence and Science Technology Laboratory and various contractors.[3] Further afield, the United States, China, Russia, Israel, France, and South Korea are also heavily investing in the weaponisation of AI.[4]

Reflecting the transformative potential of these technologies, the legal, ethical, and strategic implications of AWS development have been scrutinised by a variety of civil society organisations, think tanks, and academics.[5] According to some, unmanned aerial vehicles represent a "halfway house" toward the development of 'fully' autonomous weapons systems.[6] The important precedents created by the decades long use of weapons technologies with automated and autonomous features for the practice and interpretation of meaningful human control have, however, yet to be fully explored.

---

[1] Future of Life Institute, "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," July 28, 2015, https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.

[2] UK Ministry of Defence, "Science and Technology Strategy 2017," 2017, 15, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/655514/Science_and_Technology_Strategy_lowres.pdf.

[3] Drone Wars UK, "Off the Leash: The Development of Autonomous Military Drones in the UK," 2018, 55, https://dronewarsuk.files.wordpress.com/2018/11/dw-leash-web.pdf.

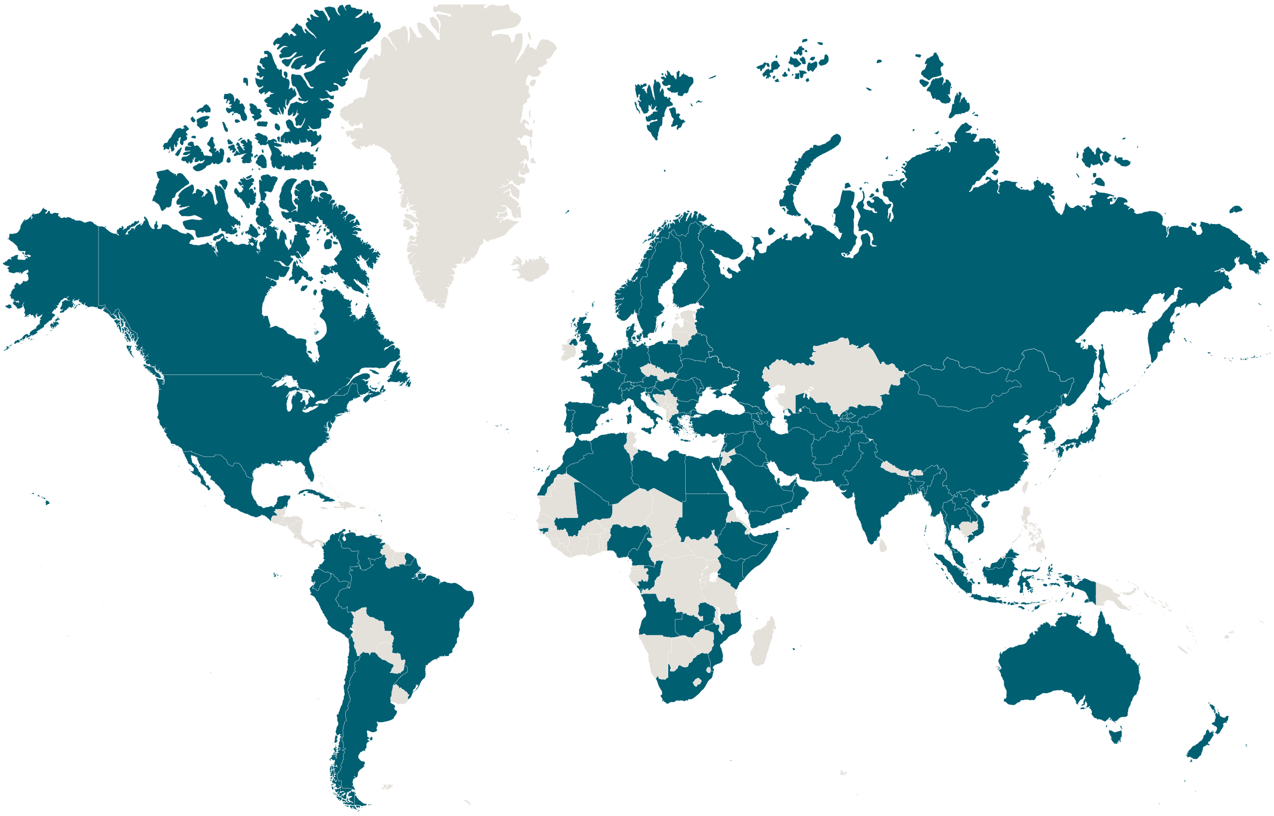
[4] Frank Slijper, Alice Beck, and Daan Kayser, "State of AI. Artificial Intelligence, the Military, and Increasingly Autonomous Weapons" (Utrecht: PAX, April 2019).

[5] Regina Surber, "Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats" (Zurich: ICT for peace foundation, Zurich Hub for Ethics and Technology, February 2018), https://ict4peace.org/wp-content/uploads/2018/02/2018_RSurber_AI-AT-LAWS-Peace-Time-Threats_final.pdf; Article 36, "Structuring Debate on Autonomous Weapons Systems," November 2013, http://www.article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf; IPRAW, "Focus on Technology and Application of Autonomous Weapons," Focus On Report (Berlin: International Panel on the Regulation of Autonomous Weapons, August 2017); Michael C Horowitz and Paul Scharre, "Meaningful Human Control in Weapon Systems: A Primer" (Washington, DC: Center for New American Security, March 2015), https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf; M. L. Cummings, "Artificial Intelligence and the Future of Warfare," in *Artificial Intelligence and International Affairs. Disruption Anticipated*, ed. Chatham House (London: Chatham House, 2018), 7–18; Jürgen Altmann and Frank Sauer, "Autonomous Weapon Systems and Strategic Stability," *Survival* 59, no. 5 (September 3, 2017): 117–42; Slijper, Beck, and Kayser, "State of AI. Artificial Intelligence, the Military, and Increasingly Autonomous Weapons"; Human Rights Watch, "Losing Humanity: The Case against Killer Robots" (Human Rights Watch, 2012); Human Rights Watch, "Heed the Call: A Moral and Legal Imperative to Ban Killer Robots," 2018.

[6] Peter Burt, "Off the Leash. The Development of Autonomous Military Drones in the UK" (Drone Wars UK, 2018), 3, https://dronewars.net/wp-content/uploads/2018/11/dw-leash-web.pdf.

Air defence systems are prime examples of these existing technologies. In use by at least 89 militaries globally,[7] these systems identify, track, and are capable of intercepting airborne threats in order to defend a platform, base, or population from attack. In comparison to 'fully' autonomous weapons systems "[…] that, once activated, can track, identify, and attack targets with violent force without further human interaction",[8] air defence systems are not deemed to be problematic. This is because they are assumed to operate under strict limitations that (supposedly) amount to meaningful human control.

States operating air defence systems[9]



This report challenges this narrative. Closer scrutiny reveals that air defence systems with automated and autonomous functions *are problematic* for two principal reasons. **First, the integration of automation and autonomy into the critical functions of air defence systems has already shaped what is understood to be the 'appropriate' role of human operators in specific use of force decisions**, deeply compromising its 'meaningfulness'. As air defence systems with automated and autonomous functions have proliferated to more and more states, these precedents have shaped a problematic, emerging norm defined here as an understanding of appropriateness that guides behaviour. This emerging norm centres on a less demanding understanding of the appropriate limits to meaningful human control in specific targeting decisions. **Second and relatedly, this norm has consequences for the ongoing deliberative efforts held at international forums to set a standard of meaningful human control. These deliberations risk being undercut by the real-world use of air defence systems with autonomous and automated features if these practices are not openly acknowledged and critically discussed**.

---

[7] Vincent Boulanin and Maaike Verbruggen, "Mapping the Development of Autonomy in Weapons Systems" (Stockholm: Stockholm International Peace Research Institute, 2017), 37, https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf.

[8] Noel Sharkey, "Staying in the Loop: Human Supervisory Control of Weapons," in *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal Bhuta et al. (Cambridge: Cambridge University Press, 2016), 3.

[9] Graphic created from the SIPRI dataset on autonomy in weapons systems. See also Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 40.

Introduced by the civil society organisation Article 36 in 2013,[10] the concept of meaningful human control has become a focal point of the international debate on AWS. At the Group of Governmental Experts (GGE) meetings under the auspices of the UN Convention on Certain Conventional Weapons (UN-CCW), both states parties and civil society actors, led by the Campaign to Stop Killer Robots,[11] are promoting meaningful human control as a new norm. Norms do not necessarily point to what is considered universally appropriate, but often to what a particular group of actors deems as suitable in a specific context.[12] If states parties decided on a negotiation mandate, the meaningful human control norm could be legally codified and provide the basis for a treaty on AWS, which prohibited the development of AWS not meeting this standard.

Group of Governmental Experts on LAWS at the UN-CCW in session, November 2017
Source Ingvild Bode

This report provides a detailed analysis of how meaningful human control is exercised in one existing type of weapons system with automated and autonomous features: air defence systems. It responds to calls for "further analysis of existing and emerging technology [to] help determine which […] components should be codified in a legal instrument as prerequisites for meaningful human control".[13] It shows that only by assessing existing systems can we understand which components of meaningful human control should be deemed crucial as their absence generates deeply problematic consequences.

The report is structured into five sections. Section 1 introduces the report and its key issues. Section 2 summarises the debate on autonomous features in the critical functions of weapons systems and introduces our working definitions of autonomy and automation. It also provides a detailed discussion of meaningful human control as being composed of three elements – a technological element, a conditional element, and a decision-making element[14] – and distinguishes between different levels of human control in specific targeting decisions.

---

[10] Article 36, "Killer Robots: UK Government Policy on Fully Autonomous Weapons," April 2013, http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf.

[11] Campaign to Stop Killer Robots, "Who Wants to Ban Fully Autonomous Weapons?," 2020, https://www.stopkillerrobots.org.

[12] Ingvild Bode and Hendrik Huelss, "Autonomous Weapons Systems and Changing Norms in International Relations," *Review of International Studies* 44, no. 3 (2018): 393–413.

[13] Campaign to Stop Killer Robots, "Key Elements of a Treaty on Fully Autonomous Weapons," November 2019, 4, https://www.stopkillerrobots.org/wp-content/uploads/2020/04/Key-Elements-of-a-Treaty-on-Fully-Autonomous-WeaponsvAccessible.pdf.

[14] Vincent Boulanin et al., "Limits of Autonomy in Weapon Systems. Identifying Practical Elements of Human Control." (Stockholm International Peace Research Institute and International Committee of the Red Cross, June 2020), 27; Campaign to Stop Killer Robots, "Key Elements of a Treaty."

Section 3 analyses the integration of automation and autonomy into the critical functions of air defence systems through a qualitative catalogue of twenty-eight such systems. We start by defining what an air defence system is, describing their history, and explaining how they work. We then summarise existing databases on AWS and outline the novelty of, and contribution made by, our qualitative catalogue of automated and autonomous features in air defence systems. This section also comments on the methodology used to generate this catalogue and the limits of working with open-source data.

Section 4 provides an in-depth analysis of how the use of air defence systems has set precedents circumscribing the exercise of meaningful human control in specific targeting decisions. The section investigates five failures of air defence systems with automated and autonomous features involving civilian and military airplanes: Iran Air Flight 655 (1988), Malaysian Airlines MH 17 (2014), Ukrainian Airlines PS752 (2020), and two instances of fratricide in the Second Gulf War (2003). Coverage of these incidents includes more substantial information about the role of human operators than is generally available for entries into our qualitative catalogue of air defence systems. **Our analysis of faults arising from humans and machines operate together shows that: (1) the role of human operators is changed in human-machine interaction; (2) the role of human operators has been minimised; and (3) at the same time, the role of human operators has been made impossibly complex.**

Finally, section 5 offers a critical conclusion and summarises our policy recommendations. These affirm meaningful human control as a general obligation of new international law on AWS, but urge a prioritisation of the human-machine interaction element in accounting for what makes human control meaningful in specific targeting situations.

The Phalanx system is an example of a highly automated and widely used Close-in Weapons System. This system was developed for the American military during the 1970s and its baseline capabilities have since been upgraded and improved **Source** Official U.S. Navy Page / Flickr

2

DEFINING KEY TERMS: AUTONOMY, AUTOMATION, AND MEANINGFUL HUMAN CONTROL

Artificial Intelligence (AI) – in simple terms, the "attempt to make computers do the kinds of things that humans and animals do"[15] – has evolved from the genesis of computing in the 'Colossus' computers used to support the Allied war effort during the Second World War. In recent decades, advancements in computer processing power have underpinned significant breakthroughs in the complexity and range of military tasks which have been delegated to machines. Advancements in narrow forms of AI, which are designed to enable machines to perform specific tasks, have been integral to AWS development and research.[16]

Colossus 10 in Block H at Bletchley Park in the room now containing the Tunny gallery of The National Museum of Computing **Credit** Good, Jack; Michie, Donald; Timms, Geoffrey (1945) **Source** Wikimedia

---

[15] Margaret Boden, "Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS) Meeting. Presentation to Panel 1 – Technological Dimension" (United Nations Office, Geneva, November 17, 2017).

[16] Austin Wyatt, "Charting Great Power Progress toward a Lethal Autonomous Weapon System Demonstration Point," *Defence Studies* 20, no. 1 (January 2, 2020): 2.

Existing AWS limit the degree of the control exercised by human agents to determining where, when, for how long, and against what types of targets a system can operate.[17] With the possible exception of the Harpy loitering munitions, 'fully' autonomous weapons systems capable of independently moving through their environments, selecting target(s) and modifying their objectives are yet to be developed.[18] Automated and autonomous features have been integrated into a variety of different weapons systems, however. This includes aerial combat vehicles such as the BAE Taranis,[19] stationary sentries such as the Samsung Techwin SGR-A1, and ground vehicles such as the Kalashnikov Concern Uran-9.[20] Automated and autonomous features in these systems fulfil different functions, which include, but are not restricted to, communication, detection, and mobility. It is therefore more accurate to speak of *autonomous features* in weapons systems rather than *autonomous weapons systems*[21] as if this were "a clearly definable category".[22]

Taranis taxiing at Warton, Lancashire **Source** Think Defence / Flickr

The integration of automated and autonomous features into a weapons system's critical functions has been singled out as particularly problematic. Critical functions relate to the selection and engagement of calculated targets without human intervention.[23] In the case of air defence systems, autonomous

---

[17] Boulanin et al., "Limits of Autonomy in Weapon Systems," 18.

[18] Andreas Krieg and Jean-Marc Rickli, *Surrogate Warfare: The Transformation of War in the Twenty-First Century* (Washington, DC: Georgetown University Press, 2019), 105–6; Wyatt, "Charting Great Power Progress toward a Lethal Autonomous Weapon System Demonstration Point," 2.

[19] The BAE Taranis programme is an experimental "advanced prototype autonomous stealth drone", referred to by the manufacturer as a "technology demonstrator". Drone Wars UK, "Off the Leash: The Development of Autonomous Military Drones in the UK," 4; BAE Systems, "Taranis," BAE Systems Products, 2020, https://www.baesystems.com/en/product/taranis.

[20] Vincent Boulanin, "Mapping the Development of Autonomy in Weapon Systems" (Stockholm: Stockholm International Peace Research Institute, December 2016), 9, https://www.sipri.org/publications/2016/other-publications/mapping-development-autonomy-weapon-systems.

[21] Paul Scharre and Michael C Horowitz, "An Introduction to Autonomy in Weapon Systems" (Center for New American Security, February 2015), 8, https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf?mtime=20160906082257; Heather Roff, "Sensor-Fused Munitions, Missiles, and Loitering Munitions," in *Expert Meeting: Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons*, ed. ICRC (Geneva: International Committee of the Red Cross, 2016), 33.

[22] Elvira Rosert and Frank Sauer, "How (Not) to Stop the Killer Robots: A Comparative Analysis of Humanitarian Disarmament Campaign Strategies," *Contemporary Security Policy* online first (2020): 14.

[23] ICRC, "Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems," April 11, 2016, https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system. Despite this focus on the identification and interception of threats, the decision by a machine to use force comes at the end of a long chain of interrelated processes. Autonomous functionality can be integrated at any stage of this process, and is not restricted to the tracking, identification, and interception of threats alone Merel A. C. Ekelhof, "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting," *Naval War College Review* 71, no. 3 (2018): 1–34; Merel Ekelhof, "Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation," *Global Policy* 10, no. 3 (2019): 343–48.

features can chiefly relate to their critical functions, as we will show in section 3. Often, these connect to *lethality* (e.g. a systems capacity to project lethal force), leading to the umbrella term lethal autonomous weapons systems (LAWS). These are defined as "systems that, once activated, can track, identify and attack targets with violent force without further human intervention".[24] AWS can be used to project lethal force. Yet, the problematic consequences of integrating autonomous features[25] apply in general to "acting with the intent to cause physical harm, i.e. violence".[26] We therefore use the more general term AWS throughout the report and only refer to LAWS when speaking to the international debate at the UN-CCW, as the discussion there is specifically focused on *lethal* autonomous weapons systems.

Despite this seemingly straightforward understanding, autonomy has been defined in multiple ways.[27] Contributors to this debate – be they states, institutions, or defence manufacturers – invariably have a stake in defining autonomy in a way which advances their perceived interests. The controversies surrounding the UK's definition of autonomy offer a good illustration of these dynamics (figure 1).

Figure 1 The UK's definition of AWS

The British government has publicly committed itself to "maintaining human control over its weapon systems as a guarantee of oversight and accountability", "not possess[ing] fully autonomous weapon systems" and "ha[ving] no intention of developing them".[28] As one Government spokesperson forcefully put it in 2013: "Let us be absolutely clear that the operation of weapons systems will always be under human control".[29]

Both the Ministry of Defence's *Joint Doctrine Note 2/11 The UK Approach to Unmanned Aircraft Systems* published in March 2011 and the updated *Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems* published in July 2017 define an autonomous system in the following terms:

> *An autonomous system is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.*[30]

---

[24] Sharkey, "Staying in the Loop," 3.

[25] Rosert and Sauer, "How (Not) to Stop the Killer Robots," 13.

[26] Peter Asaro, "Algorithms of Violence: Critical Social Perspectives on Autonomous Weapons," *Social Research: An International Quarterly* 86, no. 2 (2019): 541.

[27] See Article 36, "Shifting Definitions – The UK and Autonomous Weapons Systems," July 2018, http://www.article36.org/wp-content/uploads/2018/07/Shifting-definitions-UK-and-autonomous-weapons-July-2018.pdf; Michael Carl Haas and Sophie-Charlotte Fischer, "The Evolution of Targeted Killing Practices: Autonomous Weapons, Future Conflict, and the International Order," JOUR, *Contemporary Security Policy* 38, no. 2 (2017): 281–306.

[28] UK Ministry of Defence, "Letter in Response to Natalie Samarasinghe, Executive Director of the United Nations Association, UK and Richard Moyes, Managing Director, Article 36.," December 8, 2017, https://una.org.uk/file/12812/download?token=gPhuFZ3V; UK Ministry of Defence, "Joint Doctrine Publication 0-30.2. Unmanned Aircraft Systems," 2017, 14, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf.

[29] Article 36, "Killer Robots: UK Government Policy on Fully Autonomous Weapons," 2.

[30] UK Ministry of Defence, "Joint Doctrine Publication 030.2," 13 emphasis added; UK Ministry of Defence, "Joint Doctrine Note 2/11. The UK Approach to Unmanned Aircraft Systems," 2011, sec. 205, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN_211_UAS_v2U.pdf.

What is controversial about this definition is the explicit reference to "higher-level intent and direction". Whilst a technical qualification, the use of this specific phrase is understood to have political and legal implications. It has been criticised for setting a "futuristic" and "unrealisable" threshold for what qualifies as autonomy,[31] being "out of step" with how autonomy is defined by practically all other states,[32] and for making it difficult to determine the UK's position on the use of less sophisticated AWS nearing development.[33]

A 2017 House of Lords Select Committee report on AI noted that the focus on "higher-level intent and direction" (…) "limits both the extent to which the UK can meaningfully participate in international debates on autonomous weapons" and "hamstrings attempts to arrive at an internationally agreed definition".[34] It consequently recommended that the government "realign" its definition of autonomous weapons systems to be consistent with those used by other countries.[35]

Contributors to the AWS debate may therefore deliberately vary their definition of autonomy because this "affects what technologies or practices they identify as problematic and their orientation toward a potential regulatory response".[36] To illustrate, actors may use the terms 'automated' or 'highly automated' rather than 'autonomous' in referring to a weapons system's critical functions because they imply a higher level of human control. Likewise, they may add stringent requirements to any definition of autonomy in order to avoid criticism of systems which are currently in development.[37] Defence companies often "play up the sophistication and autonomy of their products in marketing, and downplay them when scrutinised by international bodies such as the United Nations".[38] To navigate these ambiguities and contextualise our subsequent analysis, our working definitions of autonomy, automation, and meaningful human control are provided below.[39]

---

[31] Article 36, "The United Kingdom and Lethal Autonomous Weapons Systems," April 2016, 1, https://article36.org/wp-content/uploads/2016/04/UK-and-LAWS.pdf; Article 36, "Shifting Definitions – The UK and Autonomous Weapons Systems."

[32] Sharkey quoted in House of Lords, Select Committee on Artificial Intelligence, "AI in the UK: Ready, Willing and Able?," Report of Session 2017-19 (London: House of Lords, April 16, 2018), 103, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf,.

[33] Hayley Evans, "Too Early for a Ban: The U.S. and U.K. Positions on Lethal Autonomous Weapons Systems," *Lawfare* (blog), April 13, 2018, https://www.lawfareblog.com/too-early-ban-us-and-uk-positions-lethal-autonomous-weapons-systems; Article 36, "The United Kingdom and Lethal Autonomous Weapons Systems," 2.

[34] House of Lords, Select Committee on Artificial Intelligence, "AI in the UK: Ready, Willing and Able?," 105.

[35] House of Lords, Select Committee on Artificial Intelligence, 105.

[36] Maya Brehm, "Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapons Systems under International Humanitarian and Human Rights Law," Academy Briefing No. 9 (Geneva Academy, May 2017), 13.

[37] Richard Moyes, "Target Profiles" (Article 36, August 2019), 2, http://www.article36.org/wp-content/uploads/2019/08/Target-profiles.pdf.

[38] Quoted in House of Lords, Select Committee on Artificial Intelligence, "AI in the UK: Ready, Willing and Able?," Report of Session 2017-19 (London: House of Lords, April 16, 2018), 26, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf,.

[39] These definitions are summarised to contextualise our analysis, but we acknowledge that in practice these distinctions are not always easy to uphold.

2.1 Automation and Autonomy

Figure 2 Automated and autonomous functions

	Definition	Example
Automated functions	These functions are performed by following a chronological sequence of pre-determined actions without direct human control. The system typically has sensors that help sequence it actions, but these are limited by its programming.	The command and control module of an air defence system is capable of completing the full targeting process once switched on, but cannot deviate from the sequence of actions specified by its programming.
Autonomous functions	The system's sensors are at the centre of 'decision-making' in these functions. The system acts within a pre-defined/pre-programmed range of actions [current state of systems] or relies on machine learning [potential future systems]. It can choose between multiple options for action.	The command and control module of an air defence system is capable of completing the full targeting process independently. While it works within the limits established by its programming, it can select different courses of action on the basis of sensory input.

Autonomy is, as others have argued, a relative concept with multiple interpretations.[40] It can broadly be defined as the "ability of a machine to perform a task without human input".[41] On this basis, an 'autonomous' system is one "that, once activated, can perform some tasks or functions on its own".[42] Automation is a term that overlaps with these understandings and is often used synonymously with it. The difference between the two is not always clear.[43]

Basic definitions from the field of robotics offer some distinctions. According to robot ethics professor Alan Winfield, automation means "running through a fixed pre-programmed sequence of action", while autonomy means that "actions are determined by its sensory inputs, rather than where it is in a preprogramed sequence".[44] With this in mind, much like a robot vacuum cleaner, "a robot can be autonomous but not very smart".[45] Autonomy does not *necessarily* imply a high level of intelligence.[46] Robotics professor Noel Sharkey defines an automatic robot as "carr[ying] out a pre-programmed sequence of operations or mov[ing] in a structured environment".[47] An autonomous robot, on the other hand, "[…] operates in open and unstructured environments". As Sharkey explains, "[t]he robot is still controlled by a program but now receives information from its sensors that enable it to adjust the speed and direction of its motors (and actuators) as specified by the program".[48]

---

[40] Krieg and Rickli, *Surrogate Warfare*, 105.

[41] Scharre and Horowitz, "An Introduction to Autonomy in Weapon Systems," 5.

[42] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 5.

[43] Martin Hagström, "Characteristics of Autonomous Weapon Systems," in *Expert Meeting: Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons*, ed. ICRC (Geneva: International Committee of the Red Cross, 2016), 23.

[44] A. F. T. Winfield, *Robotics: A Very Short Introduction*, Very Short Introductions 330 (Oxford: Oxford University Press, 2012), 12.

[45] Winfield, 13.

[46] Noel Sharkey, "Saying 'No!' To Lethal Autonomous Targeting," *Journal of Military Ethics* 9, no. 4 (December 2010): 376, https://doi.org/10.1080/15027570.2010.537903.

[47] Noel Sharkey, "Automating Warfare: Lessons Learned from the Drones," *Journal of Law, Information and Science* 21, no. 2 (2012): 141.

[48] Noel Sharkey, 141.

Autonomy does not necessarily imply a high level of intelligence: robot vacuum in action
Source Wikimedia commons

Following these distinctions, automation implies less sophisticated forms of action vis-à-vis autonomy because automated systems follow a sequenced form of action. This makes them potentially more predictable, understood as the possibility to "reasonably foresee how a weapon will function in any given circumstances of use and the effect that will result".[49] *But, when it comes to challenges inherent to human-machine interaction, both automated and autonomous features trigger similar problematic consequences because they increase the complexity of the system (see section 4).*

Figure 3 illustrates the growing inclusion of automated and autonomous features into weapons systems along a spectrum of autonomy.[50]

Figure 3 Spectrum of autonomy

remote controlled	automated features	autonomous features	fully autonomous
	complex human-machine interaction		

On the one side of the autonomy spectrum, are remote controlled systems such as medium-altitude long-endurance drones like the MQ-9 Reaper in which human agents remain in manual control of targeting functions. On the other side of the spectrum, are fully autonomous systems, where humans are no longer involved in specific use of force decisions that are instead administered by the system, which operates completely on its own. Systems with both automated and autonomous features can be found at different stages in the middle of this spectrum, a zone that we coin complex human-machine interaction. Here, systems operate under the supervision of a human but this supervision differs in quality depending on the range and type of tasks 'performed' via automated and autonomous features – something we explore in more detail in the next section.

---

[49] Boulanin et al., "Limits of Autonomy in Weapon Systems," 7.
[50] See also Burt, "Off the Leash," 12.

A member of 16th Regiment Royal Artillery operating the tracking system of a Rapier Field Standard C system **Source** defenceimagery.mod.uk

2.2 Meaningful Human Control

The politicisation of autonomy's definition, coupled with the rapid rate of technological advances, has pushed the international debate on AWS toward a focus on human-machine interaction. This is epitomised in the concept of meaningful human control,[51] originally coined by the non-governmental organisation (NGO) Article 36 in 2013[52] but since developed by other stakeholders. Although it was not framed in this way, the international community has been wrestling with the issue of human control since the debate on landmines. The conceptual focus of meaningful human control that comes out of the debate on AWS thereby opens up novel perspectives on existing weapons systems.[53]

Actors at the Group of Governmental Experts (GGE) on LAWS meetings at the UN-CCW in Geneva, the principal forum of intergovernmental debate on the issue, display an emerging consensus around the unacceptability of machines exercising lethal force without human supervision.[54] According to UN Secretary-General António Guterres, such an outcome would be "politically unacceptable, morally repugnant and should be prohibited by international law".[55] Advocates now promote the codification of an obligation to maintain meaningful human control as part of a regulatory framework that would also prohibit the development and usage of LAWS not meeting this requirement.[56]

---

[51] Contributors to the debate also use other labels such as appropriate levels of human judgement or sufficient human control. Merel Ekelhof, "Autonomous Weapons: Operationalizing Meaningful Human Control," *ICRC Humanitarian Law & Policy* (blog), August 15, 2018, https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/.

[52] Article 36, "Killer Robots: UK Government Policy on Fully Autonomous Weapons."

[53] We want to thank Richard Moyes for drawing our attention to these two points.

[54] See the inclusion of human-machine interaction into the Guiding Principles UN-CCW, "Draft Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. UN Document No. CCW/GGE.1/2019/CRP.1/Rev.2," August 21, 2019, 3.

[55] UN Secretary-General, "Secretary-General's Message to Meeting of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems," United Nations, March 25, 2019, https://www.un.org/sg/en/content/sg/statement/2019-03-25/secretary-generals-message-meeting-of-the-group-of-governmental-experts-emerging-technologies-the-area-of-lethal-autonomous-weapons-systems.

[56] Rosert and Sauer, "How (Not) to Stop the Killer Robots."

Group of Governmental Experts on LAWS at the UN-CCW in session, August 2018 **Credit** Ingvild Bode

One complication here are the multiple understandings of what meaningful human control implies, and the extent to which these are shared among states parties to the UN-CCW. Such differences in opinion are not surprising. Article 36 proposed the concept of meaningful human control to open up "a space for discussion and negotiation" in policy discourse around how human control should be understood.[57]

At a minimum, the concept of meaningful human control indicates "[t]hat a machine applying force and operating without any human control whatsoever is broadly unacceptable".[58] Thus "[…] a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense".[59] A *meaningful* exercise of human control requires "humans to deliberate about a target before initiating any and every attack".[60] Meaningful human control is not therefore limited to a human operator undertaking a single targeting decision. Rather it refers to the input and supervision of multiple human operators at different stages of the targeting process.[61] This thinking is consistent with the wider call to focus less on the technologies that enable AWS, and more on the processes through which they may facilitate the exercise of force, i.e. the identification and selection of calculated targets.[62]

The concept of meaningful human control has been credited with pushing the regulatory agenda on LAWS beyond the gridlocked debate on how to define autonomy. It has opened up the space for a greater focus on the legal and ethical

---

[57] Richard Moyes, "Meaningful Human Control over Individual Attacks," in *Expert Meeting: Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons*, ed. ICRC (Geneva: International Committee of the Red Cross, 2016), 46.

[58] Heather M. Roff and Richard Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons" (Article 36, April 2016), 1, http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf.

[59] Roff and Moyes, 1.

[60] Suchmann quoted in Victoria Brownlee, "Retaining Meaningful Human Control of Weapons Systems," UN Office for Disarmament Affairs, October 18, 2018, https://www.un.org/disarmament/update/retaining-meaningful-human-control-of-weapons-systems/.

[61] IPRAW, "A Path Towards the Regulation of LAWS," IPRAW Briefing (Berlin: International Panel on the Regulation of Autonomous Weapons, May 2020), 3; Ekelhof, "Moving Beyond Semantics on Autonomous Weapons."

[62] Andree-Anne Melancon, "What's Wrong with Drones? Automatization and Target Selection," *Small Wars & Insurgencies* 31, no. 4 (May 18, 2020): 801–21.

implications of machines assuming a greater role in use of force decisions.[63] As Article 36 explain, "[t]he defining feature of lethal autonomous weapons systems is that they would be systems that operate without meaningful human control".[64]

Other key civil society stakeholders have further developed the concept of meaningful human control. Two 2020 publications are noteworthy: (1) *Limits of Autonomy in Weapons Systems: Identifying Practical Elements of Human Control* jointly authored by SIPRI and the ICRC;[65] and (2) *Key Elements of a Treaty on Fully Autonomous Weapons Systems* authored by The Campaign to Stop Killer Robots.[66] Three dimensions of **meaningful human control** can be inferred from these publications:

- **A technological element** that enables human agents to exercise control over the use of force through the design of weapon parameters. This includes, for example, "limits on target type",[67] "predictability and reliability of the system",[68] or "fail-safe requirements"; [69]
- **A conditional element** that sets operational limits to how weapons systems can be used in order to enhance human control. This includes restrictions on *where* or *when* they can be used, the "duration of the system's operation",[70] or the creation of "exclusion zones, physical barriers, warnings";[71]
- **A decision-making element** that defines acceptable forms of human-machine interaction through ensuring *appropriate* levels of human supervision. This includes, for example, ensuring that the human decision-maker *understands* how the weapons systems functions and can "deactivate" the system if necessary.[72]

Both studies agree that retaining meaningful human control requires a focus on *all three elements*.[73] As such a narrow focus on the technological component is rejected as being sufficient for ensuring "meaningful" human control.[74] But, the reports otherwise do not weigh the relative importance of the three elements. As the Campaign to Stop Killer Robots note, "[w]hile none of these components are independently sufficient to amount to meaningful human control, all have the potential to enhance control in some way. In addition, the components often work in tandem".[75]

This interim conclusion has significant consequences for assessing human control in air defence systems. Both publications capture air defence systems in their remit – the Campaign to Stop Killer Robots, for example, proposes regulating "*all* weapons systems that select and engage targets on the basis of sensor inputs".[76] Despite this focus, air defence systems are not considered problematic from a meaningful human control perspective.[77] Instead, both sets

---

[63] Boulanin et al., "Limits of Autonomy in Weapon Systems," 3.

[64] Article 36, "The United Kingdom and Lethal Autonomous Weapons Systems," 1.

[65] Boulanin et al., "Limits of Autonomy in Weapon Systems."

[66] Campaign to Stop Killer Robots, "Key Elements of a Treaty," 6.

[67] Boulanin et al., "Limits of Autonomy in Weapon Systems," 27.

[68] Campaign to Stop Killer Robots, "Key Elements of a Treaty," 4.

[69] Boulanin et al., "Limits of Autonomy in Weapon Systems," 27.

[70] Campaign to Stop Killer Robots, "Key Elements of a Treaty," 4.

[71] Boulanin et al., "Limits of Autonomy in Weapon Systems," 27.

[72] Boulanin et al., 27.

[73] Boulanin et al., 33; Campaign to Stop Killer Robots, "Key Elements of a Treaty," 4.

[74] Boulanin et al., "Limits of Autonomy in Weapon Systems," 9.

[75] Campaign to Stop Killer Robots, "Key Elements of a Treaty," 4.

[76] Campaign to Stop Killer Robots, "Key Elements of a Treaty on Fully Autonomous Weapons. Frequently Asked Questions," February 2020, own emphasis.

[77] This exclusion is explained as a consequence of two factors: first, that air defence systems such as the Iron Dome and the Phalanx "operate within tight parameters in relatively controlled environments and target munitions rather than people"; and second, that humans remain able to "override" the system's decisions. Campaign to Stop Killer Robots, 3.

of stakeholders very clearly orient their definition of meaningful human control around emerging technologies rather than systems already in use.

This focus on the future is understandable and highly relevant. *But, as our report demonstrates, it is also important to carefully assess and interrogate the precedents for what counts as meaningful human control set by existing technologies.* Our analysis of air defence systems suggests that some elements may be more important than others in defining *meaningful* human control in a substantive way. To be considered *meaningful*, the decision-making element should not be outweighed by the conditional or the technological elements.

In order to more precisely capture the human-machine relationship shaping the use of air defence systems, we distinguish between five different levels of human control that human operators may exercise. These range from (a) humans deliberating about specific targets before initiating an attack, (b) humans choosing from a list of targets suggested by a program, (c) programs selecting the targets and needing human approval before attack, (d) programs selecting targets and allocating humans a time-restricted veto at the lowest level, to (e) programs selecting calculated targets and initiating attacks without human involvement.[78]

Figure 4 Levels of human control (based on Sharkey 2016)

(a) humans deliberate about specific targets before initiating an attack	in-the-loop
(b) humans choose from a list of targets suggested by a program	in-the-loop
(c) programs select the calculated targets and needs human approval before attack	on-the-loop
(d) programs select calculated targets and allocate humans a time-restricted veto before attack	on-the-loop
(e) programs select calculated targets and initiate attacks **without human involvement**	

The image of the control loop, typically referred to as the OODA loop (orient, observe, decide, act),[79] helps to visualise the relationship between the human and the system in specific situations when targets are selected and engaged rather than in earlier phases of the targeting process, e.g. strategic planning.[80] Levels (a) and (b) are classified as systems with human operators 'in the loop' because human agents *actively* participate in the selection of specific targets and the decision to use force.[81] Levels (c) and (d) situate humans 'on-the-loop' as "the operator sets goals, monitors system actions, and intervenes [only] when necessary". Consequently, they *react* to the specific targets suggested by the machine.[82] As documented in section four, due to the time constraints generally involved in the use of air defence systems, the distinction between (c) and (d) can in practice collapse in on-the-loop scenarios: while human approval before attack may be needed (level c), this becomes a *de facto* time-restricted veto (level d) when the air defence system is operating at machine speed. In the final category (e), weapons can operate without any human control and humans are therefore out of the loop.

---

[78] Based on Sharkey, "Staying in the Loop," 34–37.

[79] The OODA loop was developed by US Air Force Lieutenant General John Boyd, see Wendy R. Anderson, Amir Husain, and Marla Rosner, "The OODA Loop: Why Timing Is Everything," *Cognitive Times*, December 2017, 28–29; Taylor Pearson, "The Ultimate Guide to the OODA Loop," Taylor Pearson, 2017, https://taylorpearson.me/ooda-loop/.

[80] Burt, "Off the Leash," 11.

[81] John K. Hawley, "Patriot Wars. Automation and the Patriot Air and Missile Defense System," Voices from the Field (Center for New American Security, January 2017), 3.

[82] Hawley, 3.

To conclude, efforts to operationalise meaningful human control in the context of deliberations at the UN-CCW are ongoing. This matters for our analysis: what is actually ‘meaningful’ about meaningful human control remains unresolved. What is yet to be acknowledged is how this debate is heavily influenced by the practices of developing and operating air defence systems. Over time these have shaped perceptions of what constitutes ‘appropriate’ human-machine interaction – the decision-making element to meaningful human control – without this having been (publicly) discussed or deliberated upon. In fact, as we discuss in section 4, the level of human control in specific targeting decision-making situations is on the low end of Sharkey’s spectrum cited above, signalling very limited human control. We refer to this throughout the report as the human operator’s substantially reduced, essentially meaningless, role in specific targeting decisions.

3

AIR DEFENCE SYSTEMS, AUTONOMY AND AUTOMATION

Air defence systems can be defined as "weapons systems that are specifically designed to nullify or reduce the effectiveness of hostile air action".[83] They are used to identify, track, and if necessary, engage airborne threats to protect platforms, military installations, and people from aerial attack. Air defence systems perform different military functions. They can be designed to engage a variety of targets including different types of missiles, manned and unmanned aircraft, helicopters and rocket, artillery and mortar attacks. Close in Weapons Systems (CIWS) such as the AK-630M, the Goalkeeper and the Phalanx which are installed on warships are also able to engage small surface craft and, in the case of the upgraded AK-630M-2, "open enemy manpower and firing points on the shore".[84] The majority of air defence systems rely on a 'hard kill' projectile, usually a missile or bullet, to kinetically destroy an incoming target. Some air defence systems also process 'soft kill' features such as electronic countermeasures which can disable targets without necessarily destroying them.[85]

Air defence systems have proliferated globally. On SIPRI estimates, 89 states operate air defence systems (figure 6).[86] These include global military powers such as all five permanent members of the United Nations Security Council – the US, the UK, France, Russia, and China – and regional powers including Brazil, Egypt, India, Japan, and Turkey. The most widely used type of air defence system is reported to be shipborne CIWS. On some estimates, 2,000 of these systems are operated by over 45 states.[87]

The development of air defence systems is tied to advances in airpower. Their modern history can be traced to the interwar period. As faster and more manoeuvrable aircraft entered service with air forces across the world, the time human operators had to calculate firing solutions and respond to aerial attack shrunk.[88] During the Second World War, the US Navy invested heavily in improving its anti-aircraft capabilities. This included the development of computers to help calculate the distance and trajectory of attacking aircraft.[89] Despite these advancements, human operators retained meaningful control

---

[83] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 37.

[84] Rosoboronexport, "AK-630M-2," Rosoboronexport Naval Systems, 2020, http://roe.ru/eng/catalog/naval-systems/shipborne-weapons/ak-630m-2/.

[85] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 37.

[86] Boulanin and Verbruggen, 37.

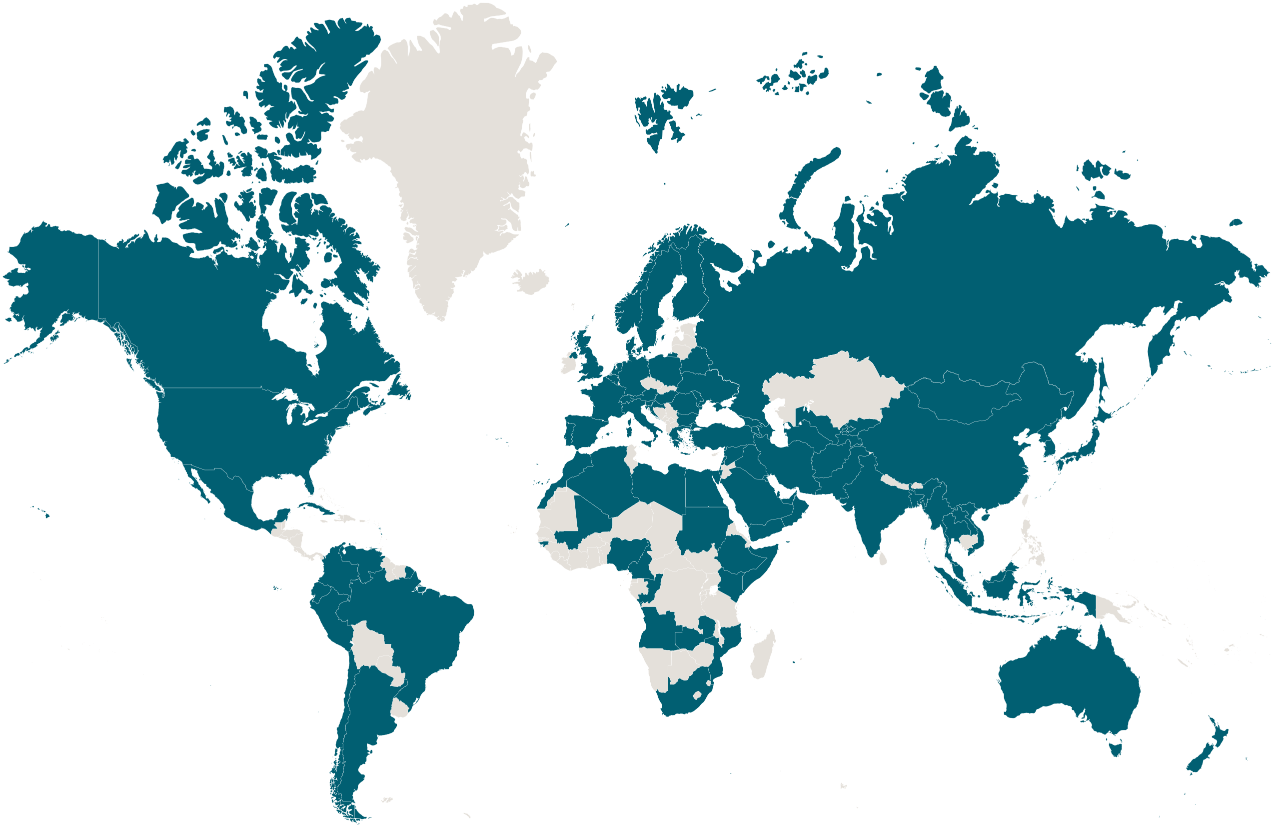
[87] James Farrant and Christopher M. Ford, "Autonomous Weapons and Weapon Reviews: The UK Second International Weapon Review Forum," *International Law Studies* 93 (2017): 395.

[88] Antoine J. Bousquet, *The Eye of War: Military Perception from the Telescope to the Drone* (Minneapolis, MI: University of Minnesota Press, 2018), 55.

[89] Bousquet, 56.

over the use of force. Gunners on stationary gun platforms manually tracked enemy warplanes to keep them within their weapon crosshairs even if, as with the US Navy's Mark 14 Gunsight introduced during the Second World War, these automatically corrected for the lag between where a human operator was aiming and where the target would be by the time they had fired.[90] Whilst the invention of radar increased the range at which enemy aircraft could be detected, as others have argued, it was "simply a tool that provides information which is useful only if procedures are developed to allow for a successful interception of an approaching aircraft".[91] During the Battle of Britain, for example, radar was integrated into a centralised command and control structure which was dependent upon the expertise of the staff at operational headquarters to 'filter', plot and pass on the information received from radar sites to squadrons of intercepting aircraft.[92]

Figure 5 States operating air defence systems[93]



Meaningful human control over air defence systems began to weaken with the development of supersonic aircraft and more advanced missiles. These increased the complexity of air defence whilst further reducing the time human operators had to react to aerial threats. In the UK, the 1957 Defence White Paper "suggested that ballistic missiles would make the UK's extant air defences all-but obsolete as no defence against them existed".[94] To better defend Britain's nuclear deterrent against possible Soviet attack, Defence Minister Sandy proposed that fighter aircraft be steadily phased out and "replaced by a

---

[90] Bousquet, 55–59.

[91] David Zimmerman, "Information and the Air Defence Revolution, 1917–40," *Journal of Strategic Studies* 27, no. 2 (June 2004): 370, https://doi.org/10.1080/0140239042000255968.

[92] Zimmerman, 384–87.

[93] Graphic created from the SIPRI dataset on autonomy in weapons systems. See also Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 40.

[94] David Jordan, "Britain's Air Defences: Inventing the Future" (Freeman Air & Space Institute, King's College London, 2020), 5, https://www.kcl.ac.uk/security-studies/assets/david-jordan-air-defence.pdf.

ground-to-air guided missile system".[95] In collaboration with the US, research into "defence even against bombardment by ballistic rockets" would also be "intensified".[96]

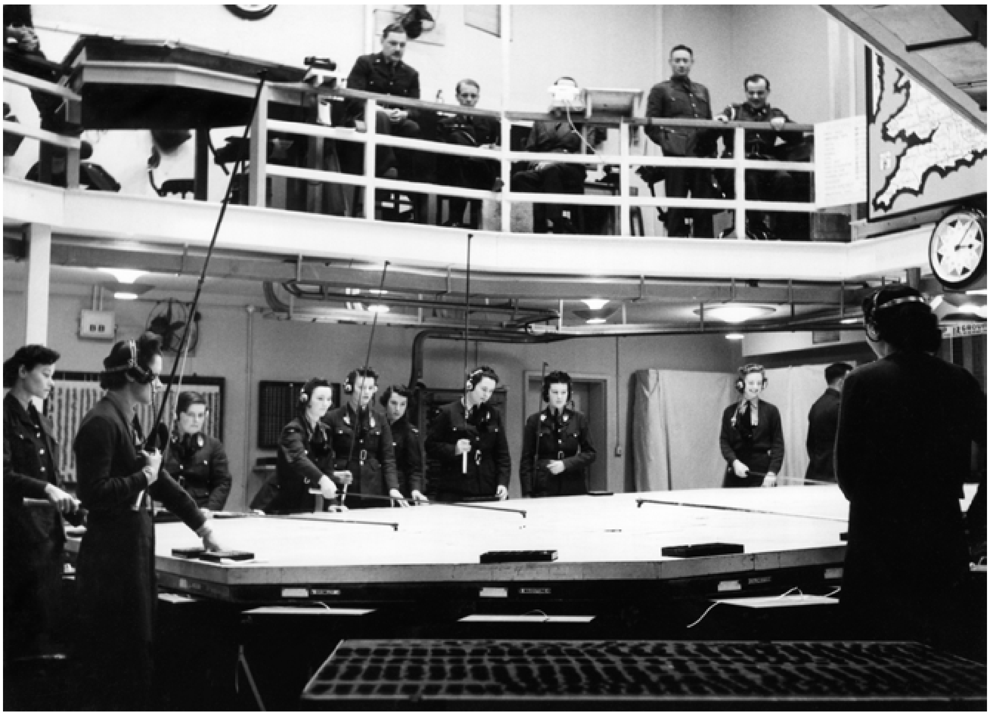


Photo of Women's Auxiliary Air Force plotters at the Operations Room at Fighter Command Head Quarters, during World War 2 **Credit** defenceimagery.mod.uk

Whilst the development of strategic air defence systems gained some momentum during the early years of the Cold War, by the 1970s it had stalled with the mutual agreement of the superpowers.[97] The development of tactical air defence systems continued, however. By this time, "it was recognized that reaction time, firepower, and operational availability in all environments did not match the threat".[98] Following the sinking of the Israeli destroyer *INS Eilat* by Soviet supplied Egyptian Komar-class attack boats armed with low flying anti-ship missiles during the Six Day War in 1967, these concerns were particularly acute in the naval sphere.[99] Developed thereafter, the American Phalanx CIWS was accepted into service at the end of the next decade, and has enjoyed considerable export success (figure 7). By 2017, over 850 Phalanx systems had been produced for over twenty states.[100]

---

[95] UK Ministry of Defence, "Defence. Outline of Future Policy," March 1957, 249, http://filestore.nationalarchives.gov.uk/pdfs/small/cab-129-86-c-57-84-34.pdf.

[96] UK Ministry of Defence, 249.

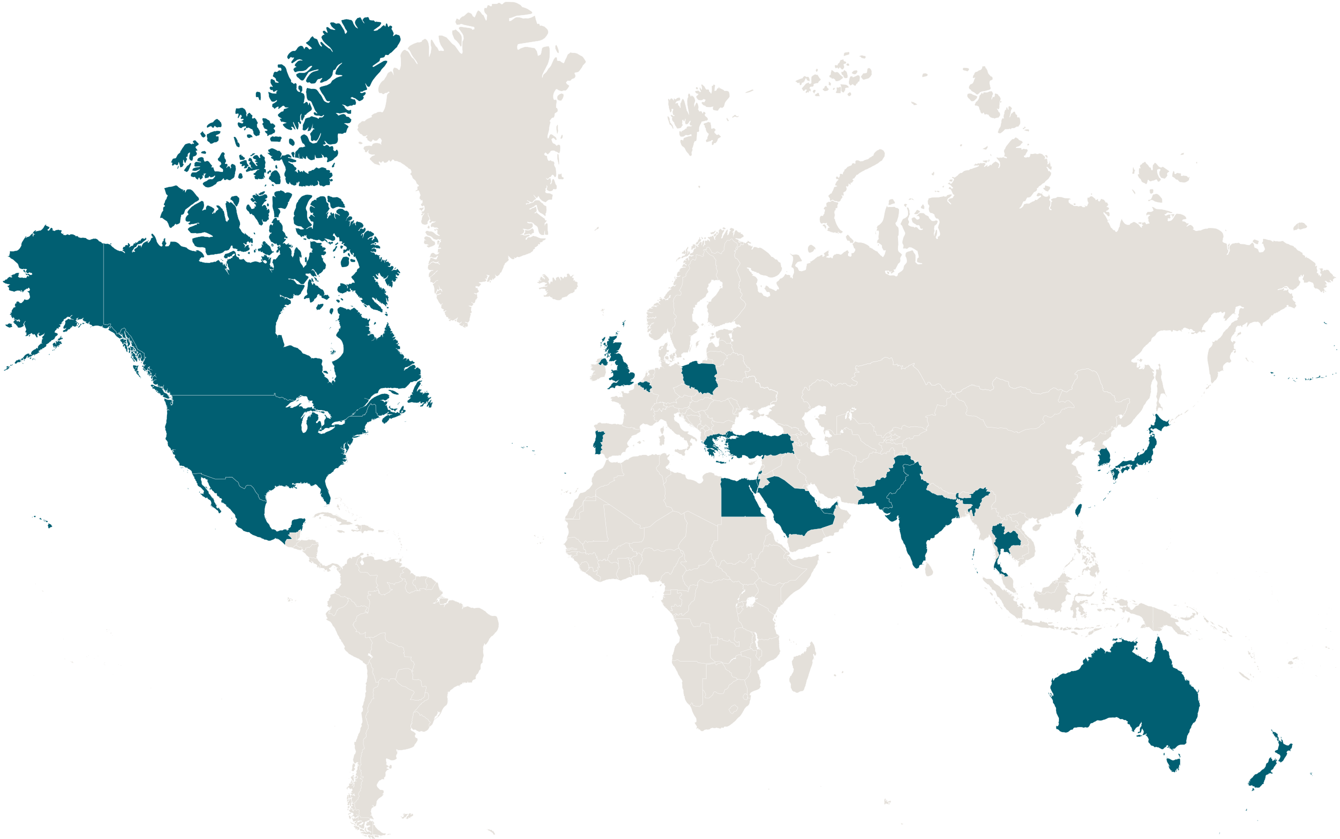
[97] The 1972 Anti-Ballistic Missile Treaty limited the US's and the Soviet Union's ability to develop sea-based, air-based, space-based, and mobile anti-ballistic missile systems and their critical subsystems. Their use was limited to the static defence of two locations: each state's capital city and an Intercontinental Ballistic Missile launch site. The George W. Bush administration withdrew the US from the Anti-Ballistic Missile Treaty in 2002 highlighting the perceived threat posed by 'rogue' states, such as Iran and North Korea. Steven A. Hildreth, "Ballistic Missile Defense: Historical Overview," CRS Report for Congress (Washington, DC: Congressional Research Service, July 9, 2007), 2, https://fas.org/sgp/crs/weapons/RS22120.pdf; Matt Korda and Hans M. Kristensen, "US Ballistic Missile Defenses, 2019," *Bulletin of the Atomic Scientists* 75, no. 6 (2019): 296.

[98] John Pike, "AEGIS Weapon System MK-7," FAS Military Analysis Network, December 31, 1998, https://fas.org/man/dod-101/sys/ship/weaps/aegis.htm.

[99] General Dynamics, "PHALANX Close-In Weapon System (CIWS)", General Dynamics, 2017, https://www.gd-ots.com/wp-content/uploads/2017/11/Phalanx.pdf

[100] Robert H. Stoner, "History and Technology: R2D2 with Attitude: The Story of the Phalanx Close-In Weapons", NavWeaps, 2009, http://www.navweaps.com/index_tech/tech-103.php

Figure 6 Past or Present users of the Phalanx Close in Defence systems[101]



Since the end of the Cold War, the mobility and survivability of air defence systems have further improved. The improved ability of Russian- and Chinese-made air defence systems to "hide, shoot and scoot", coupled with the global proliferation of these systems, has raised concern within Western governments.[102] Air defence systems are central to the anti-access/aerial denial strategies developed by China and Russia, and are perceived to reduce the distances at which Western states can project military power.[103] The proliferation of Russian air defence systems has also generated concern. In July 2019, NATO member Turkey was suspended from the American led F-35 Joint Strike Fighter project because of its intention to purchase the Russian manufactured S-400 air defence system.[104] As a White House press release explained: "the F-35 cannot coexist with a Russian intelligence collection platform that will be used to learn about its advanced capabilities".[105]

---

[101] Graphic created from the SIPRI dataset on autonomy in weapons systems and Scharre and Horowitz, "An Introduction to Autonomy in Weapon Systems," 21.

[102] Justin Bronk, "Modern Russian and Chinese Integrated Air Defence Systems. The Nature of the Threat, Growth Trajectory and Western Options," RUSI Occasional Paper (London: Royal United Services Institute, January 2020), 4. See also Carlo Kopp, "Proliferation of Advanced Air Defence Systems," Air Power Australia, March 2010, http://www.ausairpower.net/SP/DT-SAM-Proliferation-March-2010.pdf; Bronk, "Modern Russian and Chinese Integrated Air Defence Systems."

[103] David Ochmanek, "The Role of Maritime and Air Power in DoD's Third Offset Strategy. Testimony Presented before the House Armed Services Committee, Subcommittee on Seapower and Projection Forces on December 2, 2014." (Rand Corporation, December 2, 2014), 4, https://www.rand.org/pubs/testimonies/CT420.html; Maj. Peter W. Mattes, "What Is a Modern Integrated Air Defence Systems?," Airforce Magazine, October 1, 2019, https://www.airforcemag.com/article/what-is-a-modern-integrated-air-defense-system/; Bronk, "Modern Russian and Chinese Integrated Air Defence Systems."

[104] Aaron Mehta, "Turkey officially kicked out of F-35 program, costing US half a billion dollars", Defence News, 2019, https://www.defensenews.com/air/2019/07/17/turkey-officially-kicked-out-of-f-35-program/

[105] https://www.whitehouse.gov/briefings-statements/statement-press-secretary-64/

The proliferation of the S-400 system, pictured here, along with other recent Russian and Chinese air defence systems have engendered security concerns in Western governments **Source** Wikimedia Commons

The development of the next generation of air defence systems is currently being driven by multiple trends including the proliferation of unmanned aerial systems and advancements in hypersonic missile technologies.[106] In recent years, state and non-state actors across the Middle East have developed small, often slow-flying drones to strike their adversaries from a greater distance. Media reports suggest that some current air defence systems such as the Russian made Pantsir-S1 included in our catalogue have, when operated in Libya, struggled to intercept such threats and been destroyed.[107] The testing and development of hypersonic missiles which can travel at speeds in excess of Mach 5 also pose a technical challenge to current air defence systems.[108] To be clear however, the pressure to respond to increases in the speed and sophistication of aerial threats through greater automation and autonomy is not new. As the director of the Missile Defense Agency Vice Admiral Jon Hill noted in 2019, artificial intelligence in air defence is "an important part of the future and it's an important *part of now*". As he continued, "if you look back 15-20 years in a lot of our weapons systems you see artificial intelligence there, you know, where it was in that time in its evolution […] with the kind of speeds that we're dealing with today, the kind of reaction time that we have to have today, there's no other answer other than to leverage artificial intelligence".[109]

---

[106] According to reports, the US is looking to test and develop a new generation of anti-ballistic missiles that have a multi-warhead capable of detecting, tracking and attacking multiple enemy warheads from a single rocket. John Keller, "Three Defense Companies to Develop Ballistic Missile Defense Multi-Warhead Killer," Military & Aerospace Electronics, June 21, 2017, https://www.militaryaerospace.com/communications/article/16709759/three-defense-companies-to-develop-ballistic-missile-defense-multiwarhead-killer.

[107] https://nationalinterest.org/blog/buzz/does-russias-anti-drone-pantsir-s1-system-even-work-91251; https://www.realcleardefense.com/articles/2020/06/19/libya_a_catastrophe_for_russias_pantsir_s1_air_defense_system_115394.html

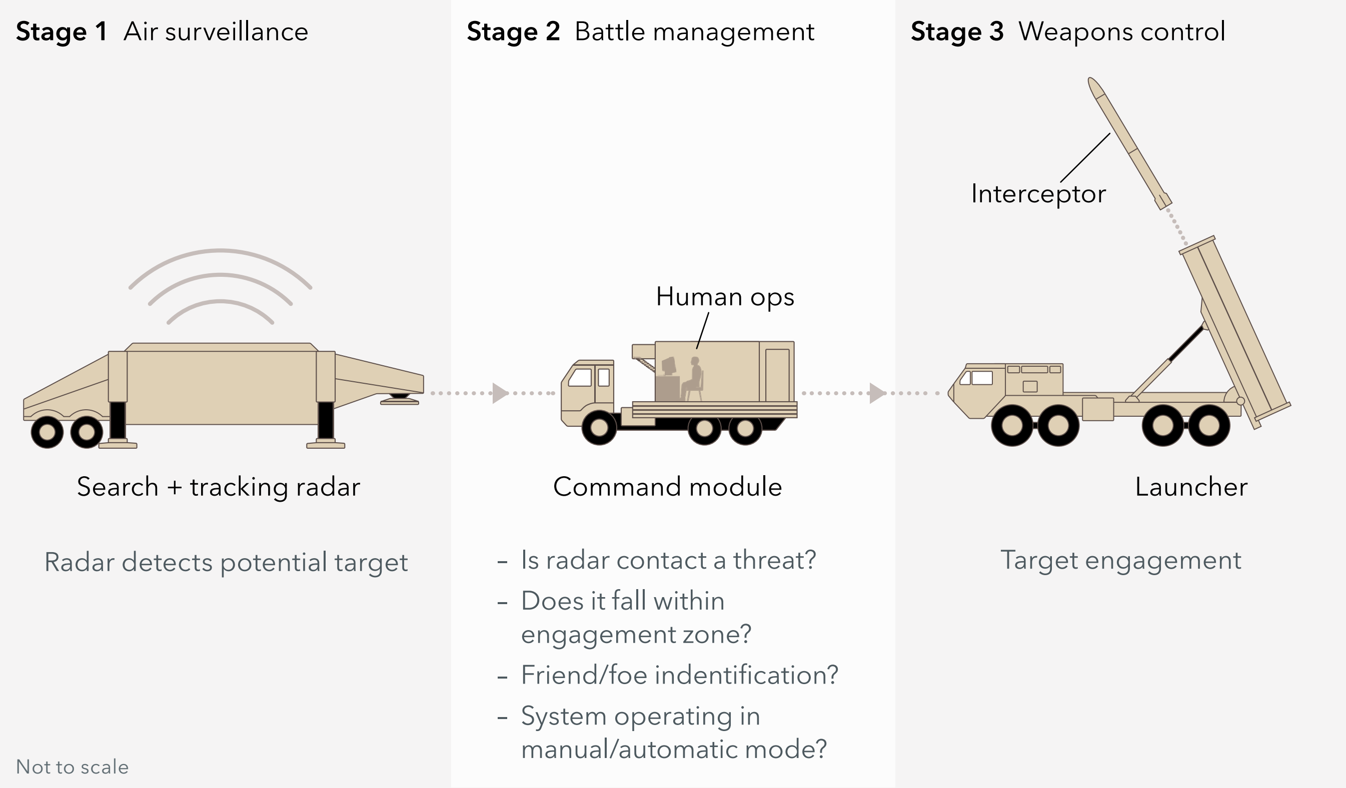
[108] Andrew W. Reddie, "Hypersonic missiles: Why the new "arms race" is going nowhere fast", Bulletin of the Atomic Scientists, 2020, https://thebulletin.org/2020/01/hypersonic-missiles-new-arms-race-going-nowhere-fast/

[109] CSIS, "A Vision for the Future of Missile Defense. A Conversation with Vice-Admiral Jon Hill," Center for Strategic and International Studies, October 7, 2019, quoted in, https://www.csis.org/analysis/vision-future-missile-defense.

3.1 How do air defence systems work?

Air defence systems are made up of four major sub-systems: (1) *search and tracking radars*, which are used to detect, identify, track, and locate incoming threats generally on the basis of their trajectory and velocity, and *fire control radars*, which are used to direct their interception if necessary; (2) *command modules*, which process this data, plot possible points of interception, and manage "the activities and engagement sequences of the various radars and missile launchers in each battery";[110] (3) *launchers*, which host the weapons system used to engage threats; and (4) *interceptors*, whether this be missiles or projectiles which are used to intercept threats.[111]

How do air defence systems works?



Air defence systems do not generally operate independently of other systems, however. They form nodes in wider integrated air defence systems which are "an amalgamation of elements, organized to minimize threats in the air domain".[112] Integrated air defence systems are made up of three elements: (1) *air surveillance*, the detection, tracking, and identification of aircraft; (2) *battle management*, the assessment of whether an aircraft is a threat, the selection of a weapon to engage it, and the confirmation to attack it; and (3) *weapons control*, "where a particular weapon system performs the weapons pairing, acquiring, tracking, guiding, killing, and assessing functions".[113] Within an integrated air defence system, different types of air defence systems may be used to defend against different types of airborne threat. For example, shorter range CIWS may be paired with longer-range area defence systems to provide greater defensive coverage.[114] Targeting data may also be provided from a wider network of radars and sensors.[115]

---

[110] Bronk, 3.

[111] Bronk, 4; Arms Control Association, "Missile Defense Systems at a Glance," Arms Control Association, August 2019, https://www.armscontrol.org/factsheets/missiledefenseataglance.

[112] Mattes, "What Is a Modern Integrated Air Defence Systems?"

[113] Mattes.

[114] Mattes.

[115] Bronk, "Modern Russian and Chinese Integrated Air Defence Systems," executive summary.

Autonomy and automation are integrated into the critical functions of modern air defence systems. As others note, "autonomy in air defence systems has no other function than supporting targeting".[116] Autonomy is not needed for mobility, for example, because many air defence systems are transported in vehicles driven by human operators. Autonomous features are instead leveraged to "detect, track, prioritize, select and potentially engage incoming air threats" at speeds faster than human agents can manage.[117] The sequencing of an air defence systems use is presented, in a simplified form, in illustration 1 and figure 8.[118]

Figure 7 Sequencing of an air defence system operation

Stage	Task	Sub-system involved	Description
Stage 1	Target detection and assessment	Radars, command module, human operators	An air defence system's search and tracking radars detects a potential target. The trajectory and velocity of this target is then calculated. This is triangulated against the system's approved engagement zones – generated using data on the flight paths of the civilian aircraft and friendly military aircraft that it has been provided with – and, on some systems, an identification, friend or foe system designed to limit friendly fire. The human operators consult the Rules of Engagement under which they are operating. An assessment is then made on whether a potential target poses a threat.
Stage 2	Target prioritization	Command module	If multiple targets are detected, the command module will need to prioritise the order in which they are engaged. Target prioritisation is determined by the systems pre-programmed engagement parameters.
Stage 3	Target engagement	Launchers, interceptors, human operators	The system's launchers then release its interceptors, attempting to destroy the identified target. Human initiation or approval is needed depending on the system's mode of operation: when 'in the loop', the operator must approve weapons release; when 'on the loop', after having switched the air defence system on, the operator is limited to a supervisory role.

Air defence systems generally operate under one of two modes of human-machine interaction: (1) *manual mode*, where the operator authorises weapons launch and manages the engagement process; or (2) *automatic mode*, where, within its pre-programmed parameters, the system "can automatically sense and detect targets and fire upon them".[119] In automatic mode, the Phalanx CIWS, for example, has the capacity to "destroy incoming rockets in self-defence, using self-destruct rounds" within what are presented as "very limited parameters".[120] As examined in Section 4, whilst these "parameters" are set by human agents, in practice, they can still operate in unpredictable and problematic ways.

---

[116] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 37.

[117] Boulanin and Verbruggen, 37.

[118] Based on Boulanin and Verbruggen, 37–39.

[119] Roff, "Weapons Autonomy Is Rocketing."

[120] UK Ministry of Defence, "Joint Doctrine Publication 030.2," 42.

We can connect these two modes of operating air defence systems – manual and automatic – to different levels of human control (see figure 4, section 2). Human operators remain 'in the loop' in manual mode: they deliberate about specific targets before initiating an attack or choosing from a list of targets suggested by a program. In automatic mode, human operators are relegated to being 'on-the-loop' and their roles may range from approving pre-selected targets to being allocated a time-restricted veto. As Boulanin and Verbruggen argue, human supervisors only have a time-restricted veto in automatic 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".[121]

Irrespective of whether a human operator remains 'in the loop' or 'on the loop', the complexities of human-machine interaction intrinsic to the use of air defence systems should be accounted for. Ultimately, common descriptions of the human operator's role in air defence systems make two assumptions. First, that human operators can retain a sufficient level of situational awareness to make meaningful judgements. Second, that they have sufficient insight into the parameters under which the automated or autonomous parts of the command module select and prioritise threats to scrutinise target selection and, if necessary, abort the attack. Our research suggests that both of these criteria are not always met, circumscribing meaningful human control in specific targeting situations.

3.2 Air defence systems in the context of the AWS debate

As our discussion of air defence systems demonstrates, the *principle* that automated and autonomous features are used in warfare is not new, and nor are the uncertainties it raises about meaningful human control.[122] Some have labelled CIWS as "first-wave" autonomous weapons succeeded, in turn, by "second-wave" autonomous weapons such as the Harpy loitering munition.[123] Despite such observations, as the United Nations Institute for Disarmament Research notes, "[t]he international discussion on AWS has not been about these sorts of existing, already long deployed systems".[124] Rather, the focus of the discussion at the UN-CCW and elsewhere is directed toward *emerging* technologies. This, we argue, is problematic because it risks missing the important precedents for what counts as meaningful human control set by existing technologies.
As a Conservative Member of the House of Lords noted in November 2014:

> *The phrase meaningful human control is an emergent concept which the UK is mindful of and working to define with interested parties in step with technological and doctrinal developments. However, in practical terms, in UK operations every target is assessed by a human, and every release of weapons is authorised by a human; other than in a very small number of instances, all targets are also acquired by a human. The exception is in a small number of defensive anti-materiel systems e.g. Phalanx. However, in those instances a human is required to authorise weapons release.*[125]

---

[121] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 39, own emphasis.

[122] Hawley, "Patriot Wars," 2.

[123] Farrant and Ford, "Autonomous Weapons and Weapon Reviews: The UK Second International Weapon Review Forum," 396.

[124] UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches. A Primer.," UNIDIR Resources (Geneva: UNIDIR, 2017), 9.

[125] Lord Astor of Hever, "Autonomous Weapons. Question for Ministry of Defence. UIN HL2710, Tabled on 6 November 2014," UK Parliament Written questions, answers and statements, November 17, 2014, https://questions-statements.parliament.uk/written-questions/detail/2014-11-06/HL2710/ emphasis added.

Some early publications on AWS, such as Human Rights Watch's *Losing Humanity* report published in 2012,[126] included a brief description of automation in different air defence systems and scrutinised its problematic effects on human-machine interaction.[127] Broadly speaking however, this type of critical analysis has been marginalised within the vast and growing civil-society/policymaker dialogue on AWS. In other cases, such as in John Hawley's research, it has been narrowed to the study of a single air defence system such as the MIM-104 Patriot.[128] In some instances, air defence systems are not characterised as being problematic from a meaningful human control perspective. This can be seen in the Campaign to Stop Killer Robots' 2020 publication on *Frequently Asked Questions on Key Elements of a Treaty on Fully Autonomous Weapons*. Following a brief analysis of the technological and conditional dimensions of meaningful human control, the attention turns to discussing whether existing weapon technologies would be included in an international treaty on autonomous weapons systems. On the basis that air defence systems such as the Iron Dome and the Phalanx Close-In Weapon System "operate within tight parameters in relatively controlled environments", "target munitions rather than people" and offer "an opportunity for a human override", it was noted that they "*thus seem to function within the bounds of meaningful human control*". [129]

Iron Dome defence system operated by Israel Defense Forces **Source** Wikimedia Commons

The handful of studies which have examined air defence systems as part of the AWS debate provide useful technical information on the development histories and users of these technologies. More importantly for our purposes, they also help introduce their autonomous and automated features. As part of their influential report published with the Stockholm International Peace Research Institute (SIPRI), Vincent Boulanin and Maaike Verbruggen have generated a dataset which catalogues the commercial origins (developer and country of

---

[126] Human Rights Watch, "Losing Humanity: The Case against Killer Robots," 9–13. *Loosing Humanity* drew attention, among other air defence systems, to the US Navy's MK 15 Phalanx Close-In Weapons System and Israel's Iron Dome. The report noted that "[h]uman involvement, when it exists at all, is limited to accepting or overriding the computer's plan of action in a matter of seconds" and called for more research on such systems as important precursors to 'fully' autonomous weapons.

[127] ICRC, "Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems"; ICRC, "Expert Meeting: Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons." (Geneva: International Committee of the Red Cross, August 2016).

[128] Hawley, "Patriot Wars."

[129] Campaign to Stop Killer Robots, "Key Elements of a Treaty FAQ," 3 emphasis added.

manufacture), development status (whether the systems development was completed, and if so in what year), reported users (past and present), and "autonomous targeting"[130] capabilities of a total of 56 different air defence systems.[131] A further dataset compiled by Heather Roff coded the autonomous capacities of 284 weapons systems for the five top weapons exporters (the US, Russia, China, Germany, and France). This included a range of air defence systems and many missiles.[132] Working with a basic definition of autonomy as "the ability to undertake a particular task by itself",[133] Roff mapped 18 autonomous capabilities "rang[ing] from homing and navigation, to target identification, prioritization, fire control, auto-communication, and learning and adaptation"[134] and focussed coding on quantitative comparisons over time.[135] Writing for the *Centre for a New American Security*, Scharre and Horowitz have catalogued some commercial (manufacturer and manufacturer location) and historical (date of introduction) information on six air defence systems. Information is also provided on the operators of these systems.[136] Finally, in a 2017 publication, the Dutch NGO PAX produced a survey of 25 weapons systems featuring categories such as loitering munitions, unmanned combat aircraft, and unmanned ground systems. Some technical information is provided on "automation" in two air defence systems: the MIM-104 Patriot and the SeaRAM.[137] These entries are brief, but the report usefully notes how the evolution of the Patriot "highlight[s] the historical developments of a 'legacy' system through to 'futuristic' LAWS".[138]

3.3 Our catalogue of automated and autonomous features in air defence systems

Our catalogue of automated and autonomous features in air defence systems aims to provide greater *qualitative and analytical depth* to a debate largely characterised by its breadth. Rather than coding a quantitative dataset which reproduces the contributions made by SIPRI's and Roff's existing studies, our catalogue is organised around a more in-depth analysis of the integration of automated and autonomous features into the critical functions of air defence systems over time. In this way, whilst drawing from some of the technical and commercial information provided by SIPRI and Roff, our catalogue is closer to PAX's study in its design. It primarily aims to describe rather than code autonomy and automation.

Our catalogue nevertheless builds on existing studies in three ways: (1) it is focused solely on air defence systems rather than multiple types of AWS; [139]

---

[130] This is coded in four parts: "Autonomy in any critical function", "Autonomy in target engagement", "Autonomy in target selection" and "Autonomy in targeting assistance".

[131] The complete dataset includes 381 entries on different types of weapon systems including unmanned aerial systems, unmanned ground systems and loitering munitions. Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems", 36–40.

[132] While the dataset was originally available for download at the Arizona State University, it has since been removed. See also Heather M. Roff, "Weapons Autonomy Is Rocketing," Foreign Policy, September 28, 2016, https://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/.

[133] Roff.

[134] Roff.

[135] Heather M. Roff and Richard Moyes, "Lethal Autonomous Weapons, Artifical Intelligence and Meaningful Human Control," 2017, https://futureoflife.org/wp-content/uploads/2017/01/Heather-Roff.pdf?x64279.

[136] Scharre and Horowitz, 21–22.

[137] Frank Slijper, "Where to Draw the Line. Increasing Autonomy in Weapon Systems – Technology and Trends" (Utrecht: PAX, 2017).

[138] Slijper, 11.

[139] Although they include autonomous features, we chose not to include active protection systems in order to provide a clearer purpose to our report. We think that examining these systems more closely from the vantage point of human-machine interaction would represent a promising further research direction.

(2) we provide greater depth to the discussion of automation and autonomy in air defence systems, including two proxy indexes for meaningful human control (system response time and number of targets which can be simultaneously tracked/engaged); and (3) we examine system capability upgrades over time to assess whether trends in machine delegation have incrementally reduced the ‘meaningfulness’ of human control.

Russian military personnel operate the Command Centre of an air defence system
Source Wikimedia Commons

In total, we examine twenty-eight air defence systems which have been operated by at least sixty states. To help provide a clearer structure to our analysis, we distinguish between three different types of air defence systems: Close-in Weapons Systems, Area Defence Systems, and Ballistic Missile Defence Systems (see figures 8 and 9). These are defined in Figure 10 and mapped to the specific air defence systems included in our catalogue. There are limits to this categorisation. Some air defence systems are ‘dual hatted’ meaning that they could reasonably be placed in multiple categories. Moreover, as discussed earlier, air defence systems form constituent nodes within wider integrated air defence systems.[140] By distinguishing between Close-in Weapons Systems, Area Defence Systems, and Ballistic Missile Defence Systems however, we are able to (1) provide a clearer structure and order to our catalogue and (2) make more concise inferences about the integration of autonomy into the critical functions of air defence systems by capturing differences in engagement ranges, operational uses, and the types of target which they are principally designed to intercept.

---

[140] Mattes, “What Is a Modern Integrated Air Defence Systems?”; Bronk, “Modern Russian and Chinese Integrated Air Defence Systems.”

Figure 8 Air Defence Systems in catalogue by type

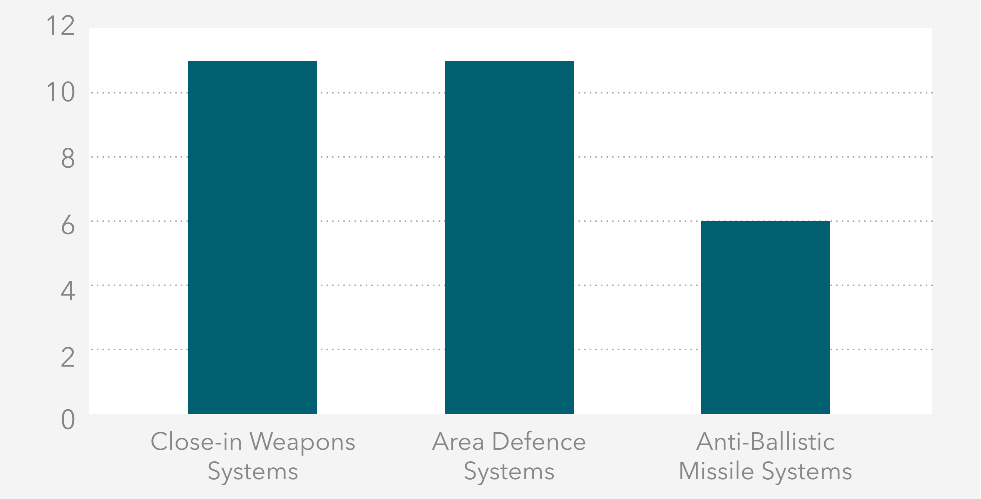


Figure 9 Overview of catalogue entries

Type of air defence system	Definition	Catalogue entries	
Close-in Weapons Systems (CIWS)	These systems provide short-range defence against incoming threats using either a gun and/or missile.[141] Their primary purpose is to provide warships or military bases a 'last line of defence' from aerial attack. CIWS detect, track, evaluate and, if necessary, engage incoming missiles/aircraft at a range of around 15kms.	AK-630M Goalkeeper NBS MANTIS Phalanx SeaRAM Type 730 (H/PJ12)	Crotale Next Generation Kashtan-M Pantsir-S1 Rapier FSC Sea Wolf
Area Defence Systems (ADS)	Operated at both land and sea, these systems are designed to use surface to air missiles to defend a geographical area, military formation or set of ships from aerial attack.[142] They have a greater range than CIWS.	BAMSE SRSAM HQ-9 KM-SAM (Cheongung) Sea Ceptor S-400 Triumf	Buk-M2 MIM-104 Patriot Iron Dome Sky Sabre Spada 2000 Tor-M1
Ballistic Missile Defence Systems (BMDS)	These systems are designed to intercept ballistic missiles at different stages of their trajectory. BMDS have the greatest range and operational ceiling of the air defence systems included in our catalogue. Their primary task is to detect, track, and engage immediate and medium range ballistic missiles.	Aegis Combat System BMD DRDO Ballistic Missile Defence System SAMP/T Air Defense System	Arrow Weapon System S-500 Prometey THAAD

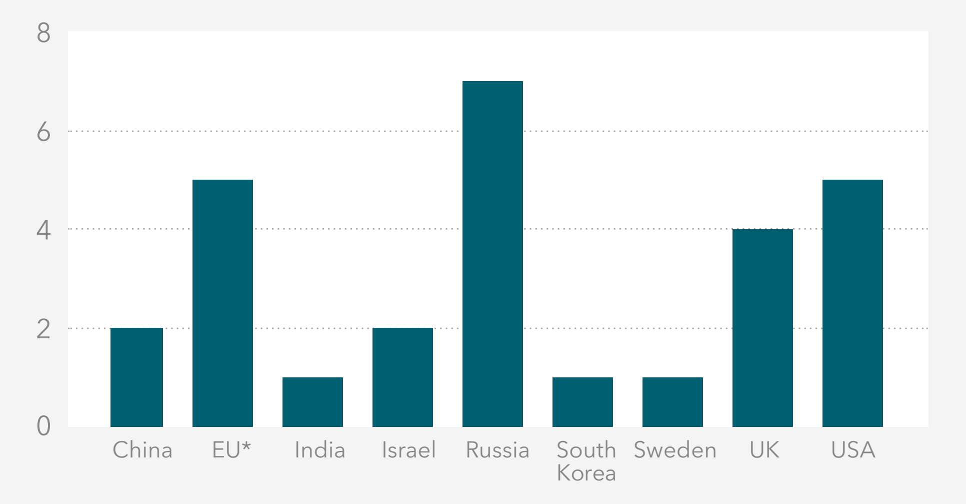
---

141 Some CIWS, such as the Phalanx, can also be deployed on land. This is usually to defend forward operating bases from enemy rocket, artillery and mortar attack. Raytheon, "Phalanx Weapon System," Raytheon Missiles and Defense, 2020, https://www.raytheonmissilesanddefense.com/capabilities/products/phalanx-close-in-weapon-system.

142 Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 37.

Our case selection was informed by the twin goals of including systems used by the states which are both major AWS *and* air defence systems developers. As illustrated in figure 10, our database includes air defence systems manufactured by all five of the recognised leaders in AWS development (the United States, China, Russia, South Korea and the European Union [France, Germany, the Netherlands and Italy]) in addition to other states such as India and Israel.[143] Consistent with both its controversial definition of autonomy and its funding into research supporting the development of armed autonomous drones,[144] we have also included four systems developed by the UK.

Figure 10 Country of origin of air defence systems included in our catalogue



Leadership in AWS development is broadly correlated to the development and export of air defence systems. The US and Russia, which make up twelve of our twenty-eight systems, are recognised as being "[c]ountries that have produced the largest variety of automatic air defence systems".[145] Two Chinese air defence systems – HQ-9 and the Type 730 (H/PJ12) – are included in our catalogue. This reflects the importance of air defence systems in China's anti-access area denial strategy, and the export of these systems across the world.[146]

In summary, our case selection was informed by three criteria: (1) to include cases of all three different types of air defence systems outlined above; (2) to capture the *global* development and usage of air defence systems including those manufactured by the current leaders in AWS technologies; and (3) to capture the modern development and testing of air defence systems from the 1970s through to the present day. Beyond this, our case selection was also informed by our decision to include air defence systems which have been suggested to be unproblematic from a meaningful human control perspective (the Phalanx and Iron Dome)[147] and that have been involved in high profile civilian airline disasters (the AEGIS, in the case of Iran Air IR655; the Buk, in the case of Malaysian Airlines MH17; and the Tor-M1, in the case of Ukrainian Airlines PS752) (see figure 11).

---

[143] Haner, J., & Garcia, D. (2019). The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development. *Global Policy*, *10*(3), 331-337.

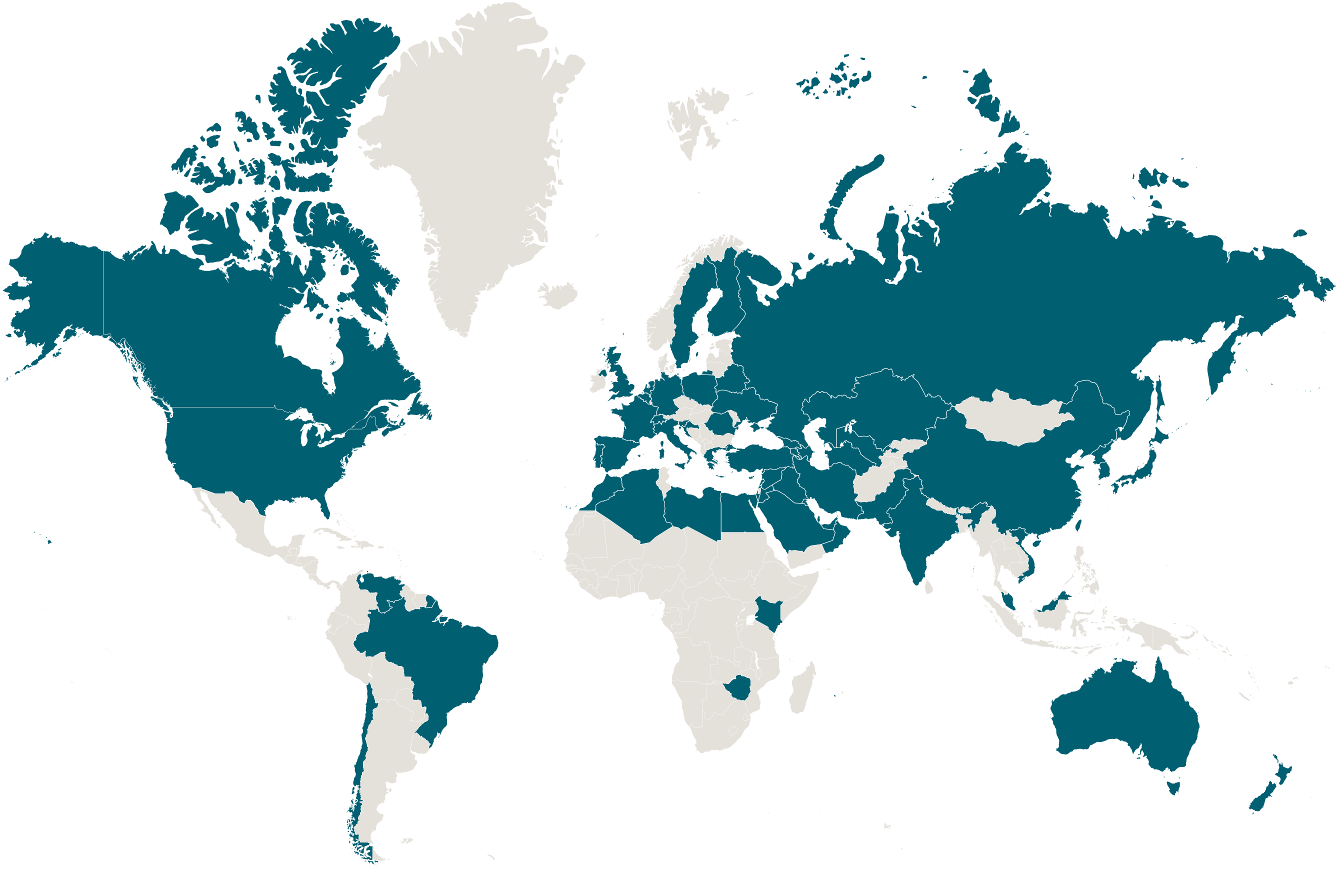
[144] Burt, "Off the Leash."

[145] Boulanin and Verbruggen, "Mapping the Development of Autonomy in Weapons Systems," 37. Whilst the US has provided significance financial funding to the Israeli government to support the development of the Arrow Dome and Iron Dome systems, we list these as Israeli systems.

[146] Bronk, "Modern Russian and Chinese Integrated Air Defence Systems"; Kopp, "Proliferation of Advanced Air Defence Systems."

[147] Campaign to Stop Killer Robots, "Key Elements of a Treaty FAQ," 3.

Figure 11 State users of air defence systems included in catalogue.



All catalogue entries are divided into nine sections, as summarised in figure 12. These have been designed to help the user navigate toward the information which they are most interested in. The full catalogue is available via https://doi.org/10.5281/zenodo.4485695. Appendix A to this report includes one sample system for the purpose of illustration.

Figure 13 Catalogue index

System manufacturer	Lists the system's principal developer and country of origin.[148]
System users	Lists states which have operated this air defence system.
System history	Briefly summarises the development history of the air defence system and any notable deployments.
Maximum system range	Lists in kilometres the maximum range at which the system can intercept aerial targets and, when data is available, the missile variant used to achieve this.[149]
Target type	Lists the different types of target which the system can engage e.g. aircraft, ballistic missiles, cruise missiles.
System updates and variants	Briefly summarises any major system upgrades or variants. This is important because many of the air defence systems included in our catalogue have been modernised over time, changing the character of human-machine interaction.
Automation and autonomy in critical functions	Presents information on the automated and autonomous features of the air defence system as reported by others including the system's manufacturer, analysts and the media.
Response time and simultaneous tracking capacity	Where data is available, we provide two proxy indexes for autonomy and automation: (1) system response time (e.g. the time the system can take between target detection to target engagement when operating in automatic mode); (2) and the maximum number of targets which the system can simultaneously track and engage. These are not the only areas of concern regarding human-machine interaction.[150] We have selected them for two reasons: (1) data availability; and (2) because a primary driver for integrating autonomy into air defence systems has been to enable a greater number of targets to be simultaneously tracked, and at greater speeds, than human operators can handle manually.[151]

---

[148] To minimise possible confusion for the reader, in the case of defence mergers between when a system was first developed and today, we use the current name of the defence company.

[149] As Justin Bronk noted in personal communication to the authors, the advertised maximum range of air defence systems is measured against "large, non-agile targets like tankers flying at medium-high altitudes". Against more manoeuvrable targets flying at lower altitudes, the range of air defence systems is likely to be significantly lower.

[150] There are also concerns about the length of time between when an air defence system is turned on and how long it operates with the human operator 'on the loop'. This is due to potential changes in the environment in which the air defence system is being used and the number of targets which an air defence system can engage before the before next human intervention. The authors are grateful to Maya Brehm for pointing this out.

[151] John K. Hawley, Anna L. Mares, and Cheryl A. Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control" (Army Research Laboratory, March 2005), 2.

3.4 Research approach and limitations

Our catalogue entries draw from a range of different types of open-source material. This includes:

1. Press releases and marketing material from weapons manufacturers such as Dassault, KBP Instrument Design Bureau, MBDA, Raytheon, and Thales Group. For Russian air defence systems, we also draw from the marketing material provided by Rosoboronexport: a state-owned subsidiary responsible for coordinating the export of Russian military equipment.
2. Press releases and factsheets published by defence ministries including the British Ministry of Defence, the American Department of Defence, and the French *Ministère des Armées*.
3. Technical and policy reports authored by researchers based at think tanks including the *Centre for a New American Security*, *Drone Wars UK*, the *Royal United Services Institute (RUSI)* and *SIPRI*.
4. Media reports from reputable international news and defence outlets such as *Army Technology*, *Defence News*, *Reuters* and *Jane's Defence Weekly*.
5. Air defence system databases including those published by *Army Technology*, the *Missile Defense Advocacy Alliance*, the *Center for Strategic and International Studies Missile Defence Project* and *SIPRI*.[152]

Building an open-source catalogue listing autonomous and automated features in air defence systems from these sources comes with at least *three* methodological challenges which we encourage the reader to keep in mind when reading our catalogue:

- Many of the technical capabilities of air defence systems are not publicly available. Without having physically observed the testing and development of these systems, or been involved in their operation, we cannot be sure of their exact capabilities.
- Given the political sensitivities concerning the definition of autonomy, some of the open-source information which is available on the autonomous and automated features of air defence systems is vague. Whilst not restricted to Chinese systems, these methodological barriers are often particularly acute for systems developed in this state. As noted in section 2, there can be significant differences in how autonomy, as well as automation, are defined by state and commercial actors. This complicates the creation of a catalogue of automated and autonomous features in air defence systems because what is listed as an 'autonomous' capability in one open-source reference could be different from what it is understood to be in another. Moreover, depending on the context in which it is used, the word "autonomous" can also refer to an air defence system's ability to operate independently of other nodes in an integrated air defence architecture.
- Whilst this catalogue has included information on the technical capabilities of air defence systems, this must be qualified by the uncertainty concerning the Rules of Engagement (ROE) under which human agents use these systems. The ROE are the "execute orders, deployment orders, operational plans or standing directives" that "provide authorisation for and/or limits on […] the

---

[152] Images: flaticon.com.

use of force, the positioning and posturing of forces, and the employment of certain specific capabilities".[153] ROE are, however, subject to constant change, depending on the conflict situation, as the case of the US *Vincennes* discussed in section 4.1 demonstrates. The opaqueness of ROE limits our understanding of meaningful human control within the specific case of air defence systems. Whilst they often have latent automated and autonomous capabilities, they may not necessarily be deployed in that capacity.

In building our catalogue of automated and autonomous features in air defence systems, we have been sensitive to the methodological barriers of determining the 'trustworthiness' of open-source material. Conceivably, the identity of the actor publishing the data – states, commercial actors, the media – and their intended audience – potential customers, policymakers or the public – may mean that automated and autonomous capabilities are over/under stated. Furthermore, given that defence ministries and weapons manufacturers are unlikely to publicly release information on failed tests, there may also be inconsistencies in how we have pieced together the historical trajectwories of specific systems. As with other existing databases, our entries should consequently be read as more indicative rather than as definite.

In the process of generating our catalogue, we have not included some entries because of data limitation issues.[154] To the greatest extent possible however, we have kept the entries into the database, even if there are gaps in the available open source data. This is with the belief that some – if limited – information about the integration of automation and autonomy into their critical functions is still of use to researchers and policymakers.

3.5 Summary of our findings

The development of modern CIWS – the most widely used type of air defence system – can be traced to the 1970s and the perception that the "reaction time, firepower, and operational availability" of existing air defence systems was inadequate to the threat posed by low flying anti-ship missiles.[155] Such concerns contributed toward a wave of CIWS development including the AK-630M (operational since 1979), the Crotale (operational since 1978), and the Phalanx (operational since 1980). Modernised variants of these systems – which can be operated in both manual and automatic modes – remain in service today and are widely recognised to have "fully automatic" tracking and targeting capabilities. In some cases, there is evidence to suggest that artificial intelligence has been leveraged to offset obsolescence issues. It has been reported that "partly due to new algorithms" a recent upgrade of the Thales Goalkeeper enabled the system to "detect targets faster, time shots better and change targets faster".[156] CIWS such as the Goalkeeper have widely proliferated. On some estimates, 300 Crotale systems, 63 Goalkeepers systems and 850 Phalanx systems have been

---

[153] Commander Alan Cole et al., "San Remo Handbook on Rules of Engagement" (International Institute of Humanitarian Law, November 2009), 1, http://iihl.org/wp-content/uploads/2017/11/ROE-HANDBOOK-ENGLISH.pdf.

[154] This included, for example, the Russian A-135 anti-ballistic missile system. Whilst information was available on this system's development history and sub-systems, in our minds reflecting the highly sensitive character of ballistic missile defence, the open-source information on the integration of automation and autonomy into its critical functions was negligible.

[155] John Pike, "AEGIS Weapon System MK-7," FAS Military Analysis Network, December 31, 1998, https://fas.org/man/dod-101/sys/ship/weaps/aegis.htm.

[156] Jaime Karremann, "Eerste Gemoderniseerde Goalkeeper Klaar Voor Tests," Marineschepen.nl, July 22, 2016, https://marineschepen.nl/nieuws/Gemoderniseerde-Goalkeeper-klaar-voor-tests-220716.html.

sold to date.[157] In these ways, the global operation of CIWS with automated and autonomous features can be argued to have contributed toward shaping an emerging, problematic international norm for what counts as 'appropriate' meaningful human control by diminishing the quality of human-machine interaction.

In the case of area defence systems, the delegation of decision-making tasks to machines has also generally been made in strides rather than in giant leaps. There are greater gaps in the available open-source data regarding the integration of automated and autonomous features into this type of air defence system. Consistent with our methodology however, the information which is available is still of use to researchers and policymakers. This can be seen, for example, in our discussion of the MIM-104 Patriot which is included in both our catalogue and as an appendix at the end of this report. This system has been described as being "nearly autonomous, with only the final launch decision requiring human interaction",[158] an example of a "human-supervised autonomous weapon system"[159] and a "system that is capable of applying lethal force with little or minimal direct human oversight".[160] The Patriot system can operate in automatic mode and is described in a 2015 NATO factsheet as having a "short response time [and] the ability to engage multiple targets simultaneously".[161] The S-400 Triumf, a system which entered service with the Russian military in 2007 and has been compared by some analysts to the Patriot, has similarly been described as operating with an "autonomous detection" capability.[162] This system has an estimated response time of between 9-10 seconds, and can reportedly engage up to 36 targets simultaneously.[163]

Although less studied in the existing debate on AWS and air defence systems, anti-ballistic missile systems are also significant from a meaningful human control perspective. As John Hawley notes in the context of the MIM-104 Patriot system, "[t]he nuts and bolts of the ballistic missile engagement process are too complex and time-limited for direct, in-the-loop human participation" and consequently require a greater degree of automation than systems operating in semi-automatic modes.[164] Six systems with different degrees of automation are included in our catalogue: the Aegis Combat System BMD, the Arrow Weapon System, the DRDO Ballistic Missile Defence System, the S-500 Prometey, the SAMP/T Air Defense System and the THAAD. Those involved with the DRDO Ballistic Missile Defence System's design have described it, for example, as being 'automated' to the degree that human intervention can only be needed to abort an interception.[165] The Israel Aerospace Industries Arrow System, like the Aegis

---

[157] Carlo Kopp, "Thomson-CSF (Thales) Crotale," Air Power Australia, January 27, 2014, http://www.ausairpower.net/APA-HQ-7-Crotale.html#mozTocId908954; Thales, "Thales Goalkeeper Scores Again and Again in Sea Acceptance Trials," THALES, March 20, 2018, https://www.thalesgroup.com/en/netherlands/press-release/thales-goalkeeper-scores-again-and-again-sea-acceptance-trials; General Dynamics, "Phalanx Close-In Weapon Systems (CIWS)," General Dynamics, May 2017, https://www.gd-ots.com/wp-content/uploads/2017/11/Phalanx.pdf.

[158] CSIS Missile Defense Project, "Patriot," MissileThreat, 2018, https://missilethreat.csis.org/system/patriot/.

[159] Paul Scharre and Michael C Horowitz, "An Introduction to Autonomy in Weapon Systems" (Center for New American Security, February 2015), 12, https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf?mtime=20160906082257.

[160] John K. Hawley, "Patriot Wars. Automation and the Patriot Air and Missile Defense System," Voices from the Field (Center for New American Security, January 2017), 4.

[161] NATO, "Patriot Deployment. Fact Sheet," North Atlantic Treaty Organization, May 2015, https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2015_05/20150508_1505-Factsheet-PATRIOT_en.pdf.

[162] James Bosbitinis, "How Capable Is the S-400 Missile System?," Defence iQ, November 21, 2018, https://www.defenceiq.com/air-land-and-sea-defence-services/articles/how-capable-is-the-s-400.

[163] Army Technology, "S-400 Triumph Air Defence Missile System," Army Technology, 2020, https://www.army-technology.com/projects/s-400-triumph-air-defence-missile-system/.

[164] Hawley, "Patriot Wars," 6.

[165] PTI, "Missile Defence Shield Ready: DRDO Chief", The Hindu, May 6 2012, https://www.thehindu.com/news/national/missile-defence-shield-ready-drdo-chief/article3390404.ece.

Combat System, can in certain circumstances be operated in automatic mode.[166] System manufacturer MBDA have similarly advertised the SAMP/T as being "highly automated" and as having an "[e]xtremely quick response time".[167]

The step-by-step process of progressive software updates has over time led human operators to be asked to fulfil increasingly minimal but at the same time inherently complex roles. As discussed in Section 2, the concept of meaningful human control requires more than a human agent "simply pressing a 'fire' button in response to indications from a computer"[168] or, to paraphrase John Hawley, being a "warm body" passively monitoring a control station.[169] It needs "humans to deliberate about a target before initiating any and every attack".[170] Defined as the time a system can take between target detection to target engagement when operating in automatic mode, the response time of many CIWS included in our catalogue – the Crotale, the Kashtan-M, the Goalkeeper, the Rapier FSC, the Pantsir-S1 – is reported to be around six seconds. To what extent, it should be asked, can a human agent be reasonably expected to meaningfully deliberate and, if necessary, abort specific use of force decisions during this timeframe? Air defence systems must be actively switched onto operating in automatic mode and, even then, these systems can only operate "within specific parameters".[171] Nevertheless, human operators are incapable of operating at machine speeds. If they could, the pressure to integrate autonomy and automation into the critical functions of such systems would be lower because it has been the expressed 'need' for faster reaction times which has created, much in the way of a self-fulfilling prophecy, the impetus to "leverage artificial intelligence".[172]

Similarly, another key benchmark against which to measure whether human agents remain in meaningful control of weapons systems is *situational awareness*. This refers to "the perception of elements in the environment […], the comprehension of their meaning, and the projection of their status in the near future".[173] In order to retain situational awareness "operators must keep track of considerable information from a variety of sources over time and organize and interpret this information".[174] Again, however, it must be asked how can human operators retain situational awareness when the typical operation of modern air defence systems only reserves them minimal roles,[175] especially in automatic mode, and provides them little to no time to deliberate targeting decisions? These problems extend beyond the speeds at which the air defence

---

[166] IAI, "Arrow Weapon System: The world's most advanced missile defence system", 2020, https://www.iai.co.il/drupal/sites/default/files/2020-05/Arrow%20Brochure_0.pdf

Ozkan, B. et al quoted in Vincent Boulanin and Maaike Verbruggen, "Mapping the Development of Autonomy in Weapons Systems" (Stockholm: Stockholm International Peace Research Institute, 2017), 39-40, https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf.

[167] MBDA Missile Systems, "ASTER 30-SAMP/T", 2020, https://www.mbda-systems.com/product/aster-30-sampt/

[168] Heather M. Roff and Richard Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons" (Article 36, April 2016), 1, http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf.

[169] Hawley, "Patriot Wars," 9.

[170] Suchman quoted in Victoria Brownlee, "Retaining Meaningful Human Control of Weapons Systems," UN Office for Disarmament Affairs, October 18, 2018, https://www.un.org/disarmament/update/retaining-meaningful-human-control-of-weapons-systems/.

[171] Vincent Boulanin and Maaike Verbruggen, "Mapping the Development of Autonomy in Weapons Systems" (Stockholm: Stockholm International Peace Research Institute, 2017), 39, https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf.

[172] CSIS, "A Vision for the Future of Missile Defense. A Conversation with Vice-Admiral Jon Hill," Center for Strategic and International Studies, October 7, 2019, quoted in, https://www.csis.org/analysis/vision-future-missile-defense.

[173] Mica R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors* 37, no. 1 (1995): 36.

[174] John K. Hawley, Anna L. Mares, and Cheryl A. Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control" (Army Research Laboratory, March 2005), 5.

[175] For a more detailed discussion of minimal levels of meaningful human control see section 2.2.

systems included in our catalogue can make targeting decisions, also speaking to the complexities of human-machine interaction. There can be significant situational benefits for human operators remaining 'in the loop' via optical display, as the case of recent upgrades made to the Crotale CIWS included in our catalogue demonstates:

> *The camera gives you much more information than radar. Radar allows you to see something on a screen, but you don't know what it is. It can only tell you whether something is flying according to the flight plan. But with the infrared camera, you can see whether the object is, a dangerous aircraft, or just a toy drone. It can distinguish between a Finnish aircraft and a Russian, giving you 'friend or foe' capacity.*[176]

In summary, our catalogue demonstrates that automation and autonomy have increasingly been integrated into the critical functions of air defence systems thereby increasing the complexity of human-machine interaction. The cumulative effect of these processes of machine delegation appears to be a general trend toward the reduction in the range and substance of meaningful human control in some specific targeting decisions. When considered as part of the wider regulatory debate on AWS, this is important because the use of automated and autonomous features in existing air defence systems has arguably already shaped what is understood to be the human operator's 'appropriate' role in specific use of force decisions. To paraphrase Vice Admiral Jon Hill quoted earlier, the problems of complex human-machine interaction are "an important part of the future and an important part of now".[177]

---

[176] Thales, "Crotale: How the Leader in Air Defence Missile Systems Keeps Its Edge," THALES, n.d., https://www.thalesgroup.com/en/worldwide/defence/magazine/crotale-how-leader-air-defence-missile-systems-keeps-its-edge.

[177] CSIS, "A Vision for the Future of Missile Defense. A Conversation with Vice-Admiral Jon Hill," Center for Strategic and International Studies, October 7, 2019, https://www.csis.org/analysis/vision-future-missile-defense.

4

AIR DEFENCE SYSTEMS AND THE ROLE OF THE HUMAN OPERATOR IN TARGETING DECISIONS

Whilst air defence systems are technically capable of engaging manned fixed and rotary winged aircraft, according to some sources, they are "not used against human targets".[178] Incidents in which air defence systems have led to the loss of human life are the subject of considerable public and political scrutiny. By studying these incidents, we can deepen our analysis of the human operator's role in specific targeting decisions and highlight the significant challenges that human agents can face in exercising meaningful human control. Our focus is on the decision-making element of meaningful human control. As defined in Section 2 of this report, this concerns the quality of human supervision in human-machine interaction. It includes the requirement that human decision-makers *understand* how weapons systems function and, if necessary, can "deactivate" them.[179]

This section examines three civilian airline disasters: Iran Air IR655, shot down by an AEGIS system on USS *Vincennes* over the Persian Gulf in July 1988; Malaysian Airlines MH17, shot down by a Buk system over Eastern Ukraine in July 2014; and Ukrainian Airlines PS752, shot down over Tehran by a Tor-M1 in January 2020. These are arguably the most prominent incidents involving air defence systems, and are therefore those which have received the most significant media and analytical coverage. In our assessment, the dynamics of human-machine interaction revealed by the detailed study of these incidents are also representative of similar cases including the destruction of Siberia Airlines Flight 1812 (4 October 2001) and Korean Air Lines Flight 007 (1 September 1983). As such, their discussion provides a window into the larger challenges of maintaining meaningful human control in the context of air defence systems. We also study two friendly fire incidents involving the Patriot air defence system during the 2003 Iraq War. Examining the Patriot in particular detail is useful because the system has been central to the air defence strategy of the US since the 1980s and has also been used by four other states in combat operations.

Our analysis *explores the unintended outcomes that arise from how humans and machines operate together* and the *impossibly complex and time restricted roles human agents occupy within human-machine interaction*. Our analysis shows that whilst 'human error' appears as a prominent attribution of responsibility in the cases where air defence systems have destroyed civilian airplanes, the situation has been more complex. The focus on 'human error' is arguably a reflection of

---

[178] Chandrika Nath and Lorna Christie, "Automation in Military Operations," UK Parliament Post, October 22, 2015, https://post.parliament.uk/research-briefings/post-pn-0511/.

[179] Boulanin et al., "Limits of Autonomy in Weapon Systems," 27.

the perception that it is more appropriate to blame imperfect human agency for these incidents rather than to scrutinise the use of automated and autonomous features in air defence systems.

4.1 Iran Air Flight 655 (IR655) – 3 July 1988

On 3 July 1988, the guided missile cruiser USS *Vincennes* equipped with the AEGIS air defence system destroyed Iran Air Flight 655 (IR655) over the Strait of Hormuz, killing all 290 passengers and crew on board the Airbus A300.[180]

Described by Lockheed Martin as "the world's most advanced combat system",[181] the AEGIS is a complex air defence system which began development in the 1970s. The Aegis system is "capable of simultaneous operations"[182] across aerial and naval targets, with its current AN/SPY-1 radar reportedly capable of tracking over a hundred targets simultaneously.[183] The Aegis Combat System is described in an US Navy Fact file as a "centralized, automated, command-and-control (C2) and weapons control system that was designed as a total weapon system, from detection to kill"[184] – a description also used by others.[185] Since its original development, the system has been continuously upgraded. In 2005, an anti-ballistic missile capability was developed likely requiring a greater integration of automation into the system's core features. As Hawley, an expert on the Patriot air defence system notes "[t]he nuts and bolts of the ballistic missile engagement process are too complex and time-limited for direct, in-the-loop human participation".[186]

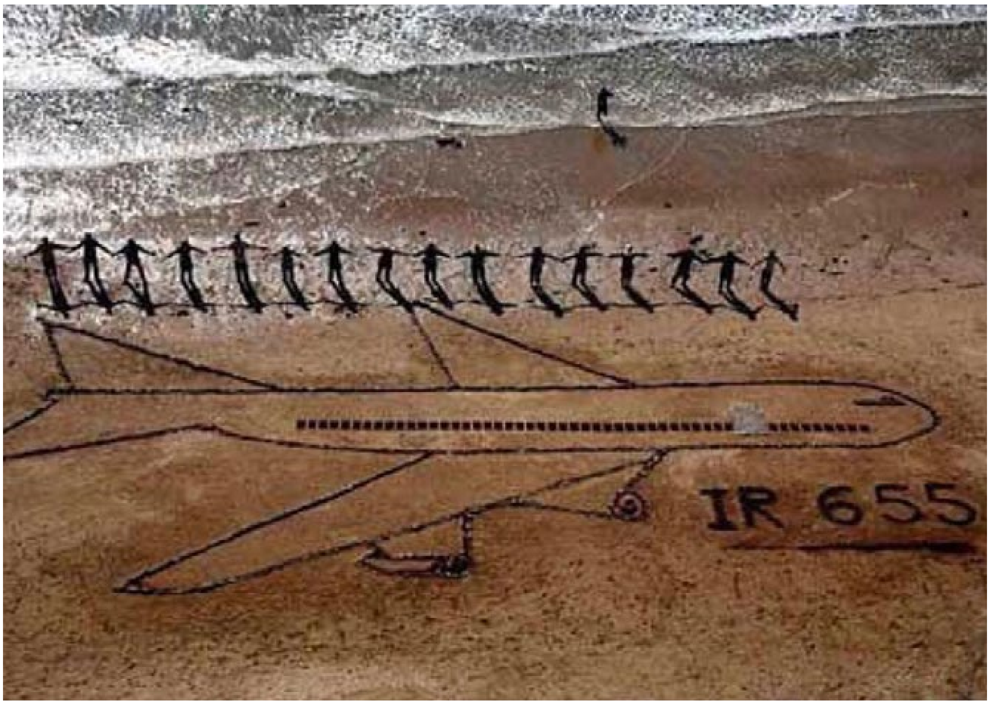


Image of River Art project commemorating the victims of Iran Airlines 655, January 2008
Source River Art **Credit** Raheleh Zomorodinia

---

[180] Justin Ling, "Canada's Path to Justice from Iran over Shot-Down Flight Will Be Hard," Foreign Policy, January 10, 2020, https://foreignpolicy.com/2020/01/10/ukraine-korea-plane-canada-path-justice-iran-shot-down-flight-hard/.

[181] Lockheed Martin, "Aegis. The Shield of the Fleet," Lockheed Martin, 2020, https://www.lockheedmartin.com/en-us/products/aegis-combat-system.html.

[182] US Navy Office of Information, "AEGIS Weapon System," America's Navy, January 10, 2019, https://www.navy.mil/Resources/Fact-Files/Display-FactFiles/Article/2166739/aegis-weapon-system/.

[183] Military.com, "Aegis Weapon System,"

[184] US Navy Office of Information, "AEGIS Weapon System."

[185] Military.com, "Aegis Weapon System."

[186] Hawley, "Patriot Wars," 6.

While the exact details of the incident are still contested, the destruction of IR655 took place in the context of the Iran-Iraq War (1980-1988). Washington was concerned about the potential disruption of oil exports from the Persian Gulf. Beginning in July 1987, the US, along with several other countries, tasked naval warships with guaranteeing the safe passage of oil tankers transiting through the Strait of Hormuz. In May 1987, the frigate USS *Stark* had been hit by two Exocet anti-ship missiles fired from an Iraqi Mirage F-1, killing 37 American sailors and wounding 21 others.[187] The *Stark* had observed the Iraqi F-1 for a hour, but its captain had decided not to engage the aircraft.[188] After this attack, the Rules of Engagement (RoE) for US warships operating in the region were loosened, providing commanders significantly greater latitude to use force.[189] This loosening of the ROE provides important context for the downing of IR655. As Admiral William J. Crowe Jr., then chairman of the Joint Chiefs of Staff, emphasised in a Congressional hearing: "each commanding officer's first responsibility was to the safety of his ship and his crew. […] Ship's captains are expected to make forehanded judgments, and if they genuinely believe to be under threat, to act *aggressively*".[190]

IR655 took off from Bandar Abbas International Airport – an airport used by both civilian and military aircraft[191] – en route to Dubai, United Arab Emirates. The USS *Vincennes* and the frigate USS *Montgomery* were in the middle of a skirmish with Iranian gunboats as IR655 departed.[192] As stated in the Pentagon's investigation report, the *Vincennes'* combat information centre (CIC) operators saw a "direct relationship to the ongoing surface engagement" and IR655.[193] Its radar signature was interpreted as an Iranian F-14 "head[ing] directly for Vincennes on a constant bearing at high speed, approximately 450 knots".[194] Thereafter, the *Vincennes* is claimed to have sent multiple warnings to IR655 on both military and civilian channels without response: the lack of response was taken as a further indicator of its hostile intent. Later reports instead suggest that the pilot of IR655 was likely not monitoring the channels these warnings were issued on.

Events unfolded differently from the perspective of the USS *Sides*, a guided missile frigate, which was also operating in the Strait of Hormuz. The *Sides* analysed the same radar data as the *Vincennes*,[195] but its captain did not authorise a missile strike against IR655 because it "simply had not behaved like a combat aircraft".[196] Unlike the *Vincennes*, the *Sides* was not equipped with the AEGIS. Tragically, "the electronic specialists in the *Sides*' combat information centre had correctly identified the aircraft's commercial transponder code at virtually the same instant that the *Vincennes* fired her missiles".[197]

---

[187] Global Security, "Operation Earnest Will," GlobalSecurity.org, n.d., https://www.globalsecurity.org/military/ops/earnest_will.htm.

[188] Dirk Maclean, "Shoot, Don't Shoot : Minimising Risk of Catastrophic Error through High Consequence Decision-Making" (Canberra: Air Power Development Centre, 2017), 23.

[189] Richard Halloran, "The Downing of Flight 655: US Downs Iran Airliner Mistaken for F-14; 290 Reported Dead; A Tragedy, Reagan Says; Action Is Defended," *The New York Times*, July 4, 1988, https://www.nytimes.com/1988/07/04/world/downing-flight-655-us-downs-iran-airliner-mistaken-for-f-14-290-reported-dead.html.

[190] Admiral Crowe quoted in US House of Representatives, "The July 3rd, 1988 Attack by the Vincennes on an Iranian Aircraft. Hearung before the Investigations Subcommitee and the Defense Policy Panel of the Committee on Armed Services.," July 21, 1992, www.loc.gov/item/93231140/ own emphasis.

[191] Halloran, "The Downing of Flight 655."

[192] US Department of Defense, "Investigation Report. Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988," 1988, 2.

[193] US Department of Defense, 2.

[194] Halloran, "The Downing of Flight 655."

[195] A data link "enabled the Sides and Vincennes computers to exchange tactical information in real time", allowing the Sides officers to have access to the same information that was displayed in the Vincennes. Evans, "USS Vincennes Case Study."

[196] Lieutenant Colonel David Evans, "USS Vincennes Case Study," *Proceedings* 119, no. 8/1086 (1993), https://www5.in.tum.de/persons/huckle/Naval_Science.htm.

[197] Evans.

Key here was the time which the commander had available to act: as the Pentagon's investigation noted, "the compression of time gave him an extremely short decision window".[198] IR655 first appeared on the radar screen at 10:47am – the *Vincennes* made its decision to fire four minutes later at 10:51am.

In the lead-up to this decision, Captain Rogers received faulty information from the CIC on the *Vincennes*. The most significant misreading identified IR655 as descending in altitude.[199] This appears to have been the result of how information was displayed to human operators, in particular a code-change for IR655's radar track. Initially, IR655 had been assigned track number (TN) 4474 but this was changed automatically by AEGIS to TN4131 along with its designation on the USS *Sides*.[200] Not only had Captain Rogers not been aware of this change, but TN4474 had also been "re-assigned to an US-Navy A6 making a carrier landing in the Arabian Gulf".[201] There were also many psychological factors at play: the misinterpretation that IR655 was descending was not double-checked by key personnel higher up in the decision-making hierarchy of the *Vincennes*, for example the anti-air warfare tactical action officer (AAW TAO). Had the AAW TAO checked, they would have seen that the system displayed IR655 as ascending rather than descending.[202]

AEGIS control centre on USS Vincennes, 1988 **Credit** Tim Masterson **Source** Wikimedia commons

Apart from the faulty readings highlighted above, the decisions made by human operators in the CIC of the *Vincennes* also displayed a significant lack of *situational awareness*: "the perception of elements in the environment […], the comprehension of their meaning, and the projection of their status in the

---

198 US Department of Defense, "Investigation Report. Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988," 6.

199 US Department of Defense, 5.

200 Evans, "USS Vincennes Case Study"; News reports published immediately after the downing of IR655 also include this piece of information, see Halloran, "The Downing of Flight 655."

201 Maclean, "Shoot, Don't Shoot : Minimising Risk of Catastrophic Error through High Consequence Decision-Making," 29; 63-65 Kristen Ann Dotterway, *Systematic Analysis of Complex Dynamic Systems: The Case of the USS Vincennes* (Monterey, CA: Naval Postgraduate School, 1992), see also, https://core.ac.uk/download/pdf/36723372.pdf.

202 US Department of Defense, "Investigation Report. Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988," 5.

near future".[203] This led the CIC operators to misinterpret information they received during the minutes that IR655 was in the air. Due to a lack of adequate foresight planning, CIC operators on the *Vincennes* did not have the background knowledge necessary to retain situational awareness and correctly assess incoming information. A key piece of such knowledge was, for example, that "failure of a track to respond to warnings would have an entirely ambiguous meaning – it could be a commercial flight or a hostile military aircraft. No store should be given to this when making a decision as to its identity or intent".[204]

The AEGIS system itself was reported to have "performed as designed" as it was not capable of determining aircraft types, a task which required "human judgment".[205] Notably, however, the AEGIS' longer-range radar was designed to give operators more time to make such judgments – a functionality that was negated in this case given that warships were "operating close-in to a land-based airfield".[206] As human operators misread or mistook some of the AEGIS' indications, the case of the *Vincennes* is sometimes referred to as an example of "under-trust" defined as "[…] the human operator ignor[ing] relevant information provided by the system or overrid[ing] its action without justification".[207] However, based on the brief summary of events provided above, we argue that the deciding factors for the incident come down to the complexities inherent to human-machine interaction.

The Pentagon's investigation identified "combat induced stress on personnel" as likely being a "significant" contributing factor to this failure. It consequently recommended more detailed investigations of the "[…] stress factors impacting on personnel *in modern warships with highly sophisticated command, control, communications and intelligence systems,* such as AEGIS."[208] Indeed, the deployment of the *USS Vincennes* marked the first combat operation of an AEGIS-equipped cruiser, meaning that its CIC operators had no prior experience of operating the highly complex AEGIS system under combat conditions. Reports indicate that several senior CIC personnel were unfamiliar and uncomfortable with the computerised exercise of their roles demanded by operating the AEGIS.[209] From a US Navy perspective, the mistaken identification of IR655 as an F-14 "was a professional disgrace".[210] It lead to a seven-year research project that included human factor analysis and resulted in "a series of design changes in the AEGIS user-interfaces to eliminate obvious sources of error" in human-machine interaction.[211] However, as we discuss in the following case studies, the challenges inherent to human-machine interaction are not easily solved. Human factor analysts continue to question "[…] whether even the best-trained crew could handle, under stress, the torrent of data that AEGIS would pour on them".[212]

---

[203] Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," 36.

[204] Maclean, "Shoot, Don't Shoot : Minimising Risk of Catastrophic Error through High Consequence Decision-Making," 40.

[205] US Department of Defense, "Investigation Report. Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988," 7.

[206] US Department of Defense, 7.

[207] Boulanin et al., "Limits of Autonomy in Weapon Systems," 19.

[208] US Department of Defense, "Investigation Report. Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988," 69 own emphasis.

[209] John Barry and Roger Charles, "Sea of Lies," *Newsweek*, December 7, 1992, https://www.newsweek.com/sea-lies-200118.

[210] Barry and Charles.

[211] Maclean, "Shoot, Don't Shoot : Minimising Risk of Catastrophic Error through High Consequence Decision-Making," 26.

[212] Barry and Charles, "Sea of Lies."

4.2 Malaysia Airlines Flight 17 (MH17) – 17 July 2014

On 17 July 2014, Malaysia Airlines Flight 17 (MH17) was destroyed over the contested Eastern Ukraine, killing all 298 people on board. A report authored by an international investigation team under the direction of the Dutch Safety Board established that a Buk air defence system (either Buk M1 or Buk M1-2) was responsible for MH17's destruction.[213] The investigation team also established that the Buk system entered and exited Ukraine from Russia and was fired from territory under the control of pro-Russian separatists.[214] Whilst Russia continues to deny any involvement, a criminal trial for three Russian nationals and an Ukrainian national accused of "co-operat[ing] to obtain and deploy" the Buk opened in the Netherlands in March 2020.[215] Coverage of the MH17 tragedy has been dominated by attributions of responsibility and the search for justice.

Wreckage from Malaysia Airlines Flight 17 in a field in eastern Ukraine, July 2014 **Source** AFP

The Buk (Russian designation 9K37, NATO designation SA-11) is an all-weather medium-range air defence system manufactured by Almaz-Antey. Its earliest variants were first operated by the Soviet military in 1979. Thereafter, its baseline capabilities have been improved through multiple system upgrades: Buk M1 1983, Buk M1-2 1988, Buk M2 2008 (included in our catalogue), Buk M3 2016.[216] Algeria, Azerbaijan, China, Egypt, Georgia, India, Iran, Syria, Ukraine, and Venezuela are among other states that operate export variants of this system.[217] Because of the warhead fragments found at the wreckage site, the Dutch Safety Board concluded that only the three older Buk variants (Buk, Buk M1, Buk M1-2) could have been used to down flight MH17.[218]

---

[213] Dutch Safety Board, "Buk Surface-to-Air Missile System Caused MH17 Crash," Crash MH17, 17 July 2014, October 13, 2015, https://www.onderzoeksraad.nl/en/page/3546/crash-mh17-17-july-2014.

[214] Luke Harding and Alex Luhn, "MH17: Buk Missile Finding Sets Russia and West at Loggerheads," The Guardian, September 28, 2016, https://www.theguardian.com/world/2016/sep/28/flight-mh17-shot-down-by-missile-brought-in-from-russia-ukraine-malaysia-airlines.

[215] Quoted in Anna Holligan, "Flight MH17: Trial Opens of Four Accused of Murdering 298 over Ukraine," BBC News, March 9, 2020, https://www.bbc.co.uk/news/world-europe-51725417.

[216] Dutch Safety Board, "Crash of Malaysia Airlines Flight MH17. Hrabove, Ukraine, 17 July 2014" (The Hague: Dutch Safety Board, October 2015), 134; ODIN, "9K317M Buk-M3 (9K37M3) Russian Medium-Range Air Defense Missile System," OE Data Integration Network (ODIN), Worldwide Equipment Guide V.2.9.2, 2020, https://odin.tradoc.army.mil/WEG/Asset/9K317M_Buk-M3_(9K37M3)_Russian_Medium-Range_Air_Defense_Missile_System.

[217] ODIN, "9K317M Buk-M3 (9K37M3) Russian Medium-Range Air Defense Missile System."

[218] Dutch Safety Board, "Crash of Malaysia Airlines Flight MH17. Hrabove, Ukraine, 17 July 2014," 132.

In a typical operation, “a Buk battery consists of three elements: an armored vehicle with a large radar device for target acquisition; the command vehicle, where there are monitors from which the battery is controlled; and finally, one or more mobile launching pads with four missiles each”.[219] According to documentation provided by the Dutch Safety Board, the missiles that destroyed MH17 were fired from a missile-launching vehicle that was operating independently in a field outside Snizhne in eastern Ukraine.[220] This suggests that “someone simply started firing from a missile-launching vehicle”[221] without requisite command, control, and radar support.

Speaking to a BBC journalist, Lieutenant Colonel Sergey Leshchuk of the Ukrainian Air Force presents the Buk air defence system as being capable of engaging six different targets within 90 seconds.[222] In the same segment, whilst the Buk system was reported as possessing an automatic friend/foe identification system, it was explained that it came down to the “expertise and experience” of the operator to determine whether unidentified planes were civilian or military in nature.[223] These assessments are confirmed by other experts on Russian military technology, such as Steve Zaloga: “When those guys are looking at a target they don’t have the same sort of information that the air traffic controllers have. […] All they know is a target is travelling at 33,000 ft. That’s it”.[224]

As the Buk’s missile-launching vehicle was operating independently, it is probable that its operators “[…] didn’t know what they were shooting at because they would not have been connected to the civilian air traffic system that helps identify what is civilian and what is a military target”.[225] Older Buk variants are suspected to have track classification problems related to civilian aircraft because, given their envisaged usage in a ‘hot’ war against NATO forces, they were not originally designed with this capability.[226] Some experts indicate that the Buk system’s operators probably misidentified MH17 as an Ukrainian military transport aircraft.[227]

In the four weeks preceding the downing of MH17, more than ten Ukrainian military aircraft were shot down over the region. It is therefore hard to understand why the Ukrainian authorities kept the airspace open for civilian aviation.[228] As the airspace remained open, aircraft operators, with a single exception, did not deviate their routes.[229]

The precise level of human control under which the Buk system was operated remains uncertain. Reports point to different capabilities and the inclusion of automated and autonomous features: “On the screen there would be a target identified using a symbol and the Buk would do the rest. […] This happens at such speeds that a human couldn’t control it. It’s all automatic *after* the launch starts.”[230]

---

[219] SPIEGEL staff, “The Tragedy of MH17. Attack Could Mark Turning Point in Ukraine Conflict,” SPIEGEL International, July 21, 2014, https://www.spiegel.de/international/world/a-deadly-error-with-global-consequences-shooting-down-flight-mf17-a-982114.html; Christopher Harress, “Blowing MH17 Out The Sky Was No Easy Task,” *International Business Times News*, July 18, 2014, http://www.lexisnexis.com/hottopics/lnacademic; Ajey Lele, “MH17 and Its Aftermath,” *MINT*, July 28, 2014, Web edition, http://www.lexisnexis.com/hottopics/lnacademic.

[220] BBC News, “MH 17 Plane Crash: What We Know,” BBC News, February 26, 2020, https://www.bbc.com/news/world-europe-28357880.

[221] SPIEGEL staff, “The Tragedy of MH17. Attack Could Mark Turning Point in Ukraine Conflict.”

[222] BBC News, “How Does a BUK Missile System Work?,” September 28, 2016, https://www.youtube.com/watch?v=PlcmziopqZA.

[223] BBC News.

[224] Quoted in Harress, “Blowing MH17 Out The Sky Was No Easy Task.”

[225] Harress.

[226] Harress.

[227] Lele, “MH17 and Its Aftermath.”

[228] Dutch Safety Board, “Crash of Malaysia Airlines Flight MH17. Hrabove, Ukraine, 17 July 2014,” 244.

[229] Dutch Safety Board, 245.

[230] Zaloga quoted in Harress, “Blowing MH17 Out The Sky Was No Easy Task” own emphasis.

The question of whether the human operators of the Buk received adequate training looms particularly large in this case. From the available information, it is clear that they must have received at least some rudimentary training in order to operate the system. Three operators with at least a month's training are needed to properly operate the Buk system.[231] According to US Secretary of State John Kerry, "[…] the separatists have a proficiency that they've gained from training from Russians as to how to use these sophisticated SA-11 systems".[232] It is also possible that the Buk system was operating in automatic or semi-automatic mode, compounding the operators' lack of training and operational experience.[233]

4.3 Ukraine International Airlines Flight 752 (PS752) – 8 January 2020

On 8 January 2020, two missiles fired from an Iranian Tor-M1 (NATO designation SA-15 Gauntlet) air defence system brought down Ukraine International Airlines Flight 752 (PS752), killing all 176 of the passengers and crew on board the Boeing 737-800.[234]

Iran imported 29 Tor-M1 systems from Russia in 2005 as part of a US$700 million contract. The Tor system has been described as "an all-weather low to medium altitude, short-range surface-to-air missile system designed for engaging airplanes, helicopters, cruise missiles, precision guided munitions, unmanned aerial vehicles, and short-range ballistic threats".[235] Having achieved initial operational capability with the Soviet Union in 1986, its baseline capabilities have been improved to counter cruise missiles and other forms of precision munitions.[236] These upgrades have in all likelihood been enabled by a greater integration of automation and autonomy into its critical features.[237] Like most air defence systems, Tor-M1 can be operated in manual and automatic mode. When operating in the latter, "the system constantly scans the operational airspace and automatically targets all objects not recognized as friendly via a 'friend or foe' radar-based identification system".[238] It can reportedly take as little as eight seconds from target identification to missile launch.[239] This reaction window sets serious limitations to the exercise of 'meaningful' human control.

A closer examination of the circumstances in which PS752 was shot down provides us with useful additional information on the Tor-M1 and insight into wider limits to meaningful human control via human-machine interaction. The incident happened hours after Iranian attacks on two military bases housing US-American troops in Iraq, an action precipitated by the Trump administration's assassination of the Iranian General Soleimani. During this time, Iran's air-defence systems were on high alert. According to the commander of the Revolutionary

---

[231] SPIEGEL staff, "The Tragedy of MH17. Attack Could Mark Turning Point in Ukraine Conflict."

[232] SPIEGEL staff.

[233] Richard quoted in SPIEGEL staff.

[234] BBC News, "Iran Plane Crash: Why Were so Many Canadians on Board," BBC News, January 11, 2020, https://www.bbc.co.uk/news/world-us-canada-51053220.

[235] ODIN, "SAM-15 (SA-15 Gauntlet) Iranian Short-Range Surface-to-Air Missile (SAM) System," OE Data Integration Network (ODIN), n.d., https://odin.tradoc.army.mil/mediawiki/index.php/SAM-15_(SA-15_Gauntlet)_Iranian_Short-Range_Surface-to-Air_Missile_(SAM)_System.

[236] Global Security, "9K331 Tor SA-15 GAUNTLET SA-N-9 HQ-17," GlobalSecurity.org, n.d., https://www.globalsecurity.org/military/world/russia/sa-15.htm.

[237] ODIN, "SAM-15 (SA-15 Gauntlet) Iranian Short-Range Surface-to-Air Missile (SAM) System."

[238] Illia Ponomarenko, "Explainer: How Could an Iranian Tor-M1 Missile System down Flight PS752?," Kyiv Post, January 10, 2020, https://www.kyivpost.com/ukraine-politics/explainer-how-could-an-iranian-tor-m1-missile-system-down-flight-ps752.html?cn-reloaded=1.

[239] Military Today, "Tor. Short-Range Air Defense System," Military Today, 2006, http://www.military-today.com/missiles/tor.htm.

Guards Air Defence Network, his forces "were totally prepared for a full-fledged war".[240] Given this, it is likely that the Iranian air defence forces operated under looser rules of engagement, perhaps similar to those that we discussed in the case of IR655.[241]

Initially, the head of the Iranian Civil Aviation Organization commented: "scientifically, it is impossible that a missile hit the Ukrainian plane".[242] The US challenged such claims maintaining that it had evidence that the Tor-M1 had locked onto PS752's radar signature prior to its destruction[243] and had identified "infrared signals from two suspected missiles" via US satellites.[244] Both European and North American leaders were quick to point out that PS752 could have been shot down by mistake (i.e. human error).[245] Capitalising upon the diplomatic wiggle-room such statements created, Iranian officials conceded their culpability for the destruction of PS752. Commander of the Islamic Revolutionary Guards Corps' airspace unit, General Amir Ali Hajizadeh, admitted that "[t]he plane was flying in its normal direction without any error and everybody was doing their job correctly".[246] He added that "if there was a mistake, it was made by one of our members".[247] Other Iranian leaders also attributed the failure to human error: on January 11th, Iranian Foreign Minister Mohammad Javad Zarif took to Twitter to attribute the incident to "human error at time of crisis caused by U.S. adventurism led to disaster".[248]

But does that mean that the system was operated under *meaningful* human control? As discussed in the previous case studies, there is a range of human-machine interaction challenges associated with operating air-defence systems in high-pressure environments including, most importantly, target identification. In the words of a former European air defence officer: "Shooting down a hostile aircraft is easy. It's identifying the aircraft and not shooting down friendlies that are the challenges".[249] The Tor-M1 targeting software relies upon a combination of radar, visual identification, and signals from the plane's tracking transponder[250] – "a radar beacon that transmits flight data and an aircraft's identity back to ground controllers".[251] While the system is performing the latter

---

[240] IFP Editorial Staff, "IRGC Releases Details of Accidental Downing of Ukrainian Plane," Iran Front Page – IFP News, January 11, 2020, https://ifpnews.com/irgc-releases-details-of-accidental-downing-of-ukrainian-plane.

[241] Jeremy Bogaisky, "If Iranian Troops Really Thought Ukraine Flight 752 Was A Cruise Missile, They Made A 'Hail Mary' Shot," Forbes, January 15, 2020, https://www.forbes.com/sites/jeremybogaisky/2020/01/15/if-iranian-troops-really-thought-ukraine-flight-752-was-a-cruise-missile-they-made-a-hail-mary-shot/#1b0f1ab02270.

[242] Quoted in Dan Sabbagh and Michael Safi, "Iran Crash: Plane Shot down by Accident, Western Officials Believe," The Guardian, January 9, 2020, https://www.theguardian.com/world/2020/jan/09/tehran-crash-plane-downed-by-iranian-missile-western-officials-believe.

[243] Jim Sciutto et al., "Video Appears to Show Missile Strike as Canada and UK Say They Have Intel Iran Shot down Ukrainian Plane," CNN, January 10, 2020, https://edition.cnn.com/2020/01/09/politics/is-iran-ukraine-plane/index.html.

[244] Sabbagh and Safi, "Iran Crash: Plane Shot down by Accident, Western Officials Believe."

[245] Sky News, "Iran Plane Crash: Canada Says Evidence Shows Jet Was Shot down by Iranian Missile – but Iran Denies It," Sky News, January 10, 2020, https://news.sky.com/story/iran-plane-crash-downing-street-looking-into-reports-ukrainian-jet-was-shot-down-by-missile-11904698; Sabbagh and Safi, "Iran Crash: Plane Shot down by Accident, Western Officials Believe."

[246] The New York Times, "Plane Shot Down Because of Human Error, Iran Says," The New York Times, January 11, 2020, https://www.nytimes.com/2020/01/11/world/middleeast/plane-crash.html.

[247] The New York Times.

[248] Javad Zarif, "Tweet: A Sad Day," Twitter, January 11, 2020, https://twitter.com/JZarif/status/1215847283381755914.

[249] Gerry Doyle, "Explainer: Missile System Suspected of Bringing down Airliner – Short Range, Fast and Deadly," Reuters, January 10, 2020, https://uk.reuters.com/article/uk-iran-crash-missiles-explainer/explainer-missile-system-suspected-of-bringing-down-airliner-short-range-fast-and-deadly-idUKKBN1Z90A0.

[250] The New York Times, "Plane Shot Down Because of Human Error, Iran Says."

[251] Nolan Peterson, "Iran Admits to Shooting Down Ukrainian Airliner," The Daily Signal, January 10, 2020, https://www.dailysignal.com/2020/01/10/why-it-looks-like-iran-shot-down-ukrainian-airliner/?utm_source=rssutm_medium=rssutm_campaign=why-it-looks-like-iran-shot-down-ukrainian-airliner.

task automatically, without human triangulation, "[…] everything becomes an enemy to the missile – unless you can identify it by sight and turn the missile off".[252] Even without transponder signals, PS752's "flight speed, altitude, and the fact that it was in a civilian corridor" should have prompted the system's operator to identify it as a civilian plane.[253]

Candlelight vigil to commemorate the victims of Ukraine Airlines 752, January 2020 **Credit** AP/TASS

The Tor-M1's operators appeared to be "operating without a solid picture of the known traffic in Iranian airspace as whole".[254] This is typical for this type of short range air defence system characterised as a "stand-alone system, meaning it is mounted on the back of a vehicle and not typically plugged into a country's broader air defence radar network".[255] Whilst the system operator had sought authorisation for the attack from higher up in the chain of command, they were unable to communicate this request due to either jamming or the high level of traffic across the system.[256] As importantly, the short-range radar of the Tor-M1 gave "a very short reaction time" of about ten seconds "to interpret the data".[257]

Interestingly, on January 11th, Commander Ali Hajizadeh claimed that an Iranian air defence operator had misidentified flight PS752 as a cruise missile:

> *At several stages, the Alert Level 3, which is the highest level, is communicated and emphasized to the entire network. So all air defence systems were at highest alert level. For several times, these systems including the one involved in the incident were notified by the integrated network that cruise missiles have been fired at the country. For a couple of times, they received reports that 'the cruise missiles are coming, be prepared'. […] So you see the systems were at the highest alert level, where you should just press a button. They had been told cruise missiles were coming, and the air defence unit engaged in this incident and fired a missile.*[258]

---

[252] Glen Grant quoted in Ponomarenko, "Explainer: How Could an Iranian Tor-M1 Missile System down Flight PS752?"

[253] Melnyk quoted in Peterson, "Iran Admits to Shooting Down Ukrainian Airliner."

[254] Justin Bronk, "What Happened to Flight PS752?," The Telegraph, Jan 11, 2020, https://www.telegraph.co.uk/news/2020/01/10/badly-trained-iranian-defence-team-could-have-made-mistakes/.

[255] Peterson, "Iran Admits to Shooting Down Ukrainian Airliner."

[256] IFP Editorial Staff, "IRGC Releases Details of Accidental Downing of Ukrainian Plane."

[257] IFP Editorial Staff; Harmer quoted in Peterson, "Iran Admits to Shooting Down Ukrainian Airliner."

[258] Quoted in IFP Editorial Staff, "IRGC Releases Details of Accidental Downing of Ukrainian Plane" own emphasis.

Given these observations, some commentators have concluded that "a badly-trained or inexperienced crew […], scared of being hit as part of a retaliatory US strike following the ballistic missile attacks on bases in Iraq, made a series of tragic and incorrect assumptions when PS752 appeared on their radar screen."[259] This observation suggests that the Tor-M1 crew lacked the necessary *combat* experience to properly operate the system: "taking the time to cross-reference or confirm the status of a radar contact under those circumstances takes a level of discipline uncommon to operators with no combat experience, and that no longer exists in the Iranian military".[260] But it should also be acknowledged that the kind of 'snap decision' that led to the downing of PS752 is a typical part of how human agents operate air defence systems. The focus on human 'error' or human 'mistakes' distracts from how the automated and autonomous technology *structures* the use of force.[261] Even well-trained crews are subject to the limited situational awareness and increased complexity that operating an air defence system with automated and autonomous features brings with it (see also section 4.4). This means that, in some situations, the individual human operators at the bottom of the chain of command often can bear the responsibility for structural failures in how air defence systems are designed and operated.

It is also unclear whether the Tor-M1 was operating in manual or in automatic mode. If operating in automatic mode, this "could have led to an accidental launch by an inexperienced ground crew".[262] In this case, inexperience does not refer to target identification specifically but rather to a broader understanding of the system's operation: "any system that can work automatically is always a danger of the crew does not fully understand the merits and limitations of that ability".[263] The official claim that PS752 was misidentified as a cruise missile,[264] however, suggests that the Tor-M1 operated in manual mode: the flight behaviours of airliners and cruise missiles differ so significantly that PS752 would not have been captured within the system's algorithmic parameter classification for cruise missiles.[265]

4.4 The Patriot and fratricides

In the early stages of the 2003 invasion of Iraq, the MIM-104 Patriot was involved in two fratricidal engagements which destroyed a RAF Tornado fighter jet (24 March 2003) and a US Navy F-18 fighter jet (2 April 2003), killing three crewmembers in total.[266] The RAF Tornado was wrongly identified as an Iraqi anti-radiation missile: "The track [of the intended target] was interrogated for IFF [Identification Friend or Foe] but there was no response. Having met all classification criteria, the Patriot crew launched the missile".[267] Two Patriots downed the US Navy F-18 about a week later.[268] Compounding matters, there

---

[259] Bronk, "What Happened to Flight PS752?"

[260] Venable quoted in Peterson, "Iran Admits to Shooting Down Ukrainian Airliner."

[261] The authors would like to thank Maya Brehm for this point.

[262] Peterson, "Iran Admits to Shooting Down Ukrainian Airliner."

[263] Grant quoted in Ponomarenko, "Explainer: How Could an Iranian Tor-M1 Missile System down Flight PS752?"

[264] The New York Times, "Plane Shot Down Because of Human Error, Iran Says."

[265] The authors thank Justin Bronk for drawing our attention to this point.

[266] US Department of Defense, "Report of the Defense Science Board Task Force on Patriot System Performance. Report Summary." (Washington, DC: Office of the Under-Secretary of Defense for Acquisition, Technology, and Logistics, January 2005), 2, https://dsb.cto.mil/reports/2000s/ADA435837.pdf.

[267] UK Ministry of Defence, "Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710," Military Aircraft Accident Summary (London: Directorate of Air Staff, May 2004), 2, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/82817/maas03_02_tornado_zg710_22mar03.pdf.

[268] These friendly fire incidents led to the creation of new Rules of Engagement for the Patriot system, which now had (and continues to have) to receive specific authorization by the Air Force controlling authority before engagement – a "decision [that] took Patriot out of the fight" due to the short engagement timelines for tactical ballistic missiles. Patriot crew members now also receive slightly longer, but not substantially different forms of, training before deployment. Hawley, "Patriot Wars," 6, 8.

was a further close-call friendly fire incident on 25 March 2003 when a Patriot battery locked onto an US Air Force F-16. In this case, the pilot "was alerted to the fact that he had been targeted by radar" and launched a counter-attack that destroyed the Patriot battery.[269] The Patriot's two fratricidal engagements amounted to 18 percent of the system's 11 engagements during the operation: an "unacceptable" rate of fratricides according to the US Army.[270]

Patriot missile battery at a Turkish army base, 2013 **Source** defense.gov

The MIM-104 Patriot "is a long-range, all-altitude, all-weather air defence system to counter tactical ballistic missiles, cruise missiles, and advanced aircraft".[271] First fielded by the United States in the mid-1980s, the system has since been upgraded multiple times.[272] Export variants of the system are operated by the Dutch, Egyptian, German, Israeli, Japanese, Jordanian, Kuwaiti, Saudi, South Korean, and the United Arab Emirate armed forces. The Patriot's command module comprises two operators[273] and is the only sub-system of a Patriot battery that involves direct human control. As the central sub-system, the command module can communicate and coordinate actions with launching stations, other Patriot systems, and command headquarters. As Army Technology note on the Patriot's operation sequence:

> *[t]arget engagement can be carried out in manual, semi-automatic or automatic mode. When the decision has been made to engage the target, the engagement control station selects the launch station or stations and pre-launch data is transmitted to the selected missile. After launch, the Patriot missile is acquired by the radar.*[274]

---

[269] Lester Haines, "Patriot Missile: Friend or Foe?," The Register, May 20, 2004, https://www.theregister.co.uk/2004/05/20/patriot_missile/.

[270] Paul D. Scharre, "Autonomous Weapons and Operational Risk," Ethical Autonomy Project (Washington, DC: Center for New American Security, February 2016), 32.

[271] Army Technology, "Patriot Missile Long-Range Air-Defence System US Army," *Army Technology* (blog), 2019, https://www.army-technology.com/projects/patriot/.

[272] John K. Hawley and Anna L. Mares, "Human Performance Challenges for the Future Force: Lessons from Patriot after the Second Gulf War," in *Designing Soldier Systems: Current Issues in Human Factors*, ed. Pamela Savage-Knepshield et al. (Burlington, VT: Ashgate, 2012), 3; Charles Piller, "Vaunted Patriot Missile Has a 'Friendly Fire' Failing," Los Angeles Times, April 21, 2003, https://www.latimes.com/archives/la-xpm-2003-apr-21-war-patriot21-story.html.

[273] Piller, "Vaunted Patriot Missile Has a 'Friendly Fire' Failing."

[274] Army Technology, "Patriot Missile Long-Range Air-Defence System US Army."

In simple terms, the Patriot's radar tracks objects in air and its engagement algorithm "identifies those objects, and then displays them as symbols on a screen".[275] What then happens depends on whether the Patriot is operating in semi-automatic or automatic mode. In semi-automatic mode, the Patriot functions as a 'human-in-the-loop' system. While the human operators receive "more computer-based engagement support",[276] they make all critical decisions regarding the use of force and play an essential part in the control loop. That said, even for this human-in-the-loop setting, problems with validating the accuracy of the system's recommendations and performance remain.

In automatic mode, however, Patriot becomes a 'human-on-the-loop' system, being "[…] nearly autonomous, with only the final launch decision requiring human interaction".[277] Here, once the human operator has put the system into 'ready' status in response to a reported incoming threat, it can open fire without further human action.[278] This sets serious limits to meaningful human control in specific targeting decisions. According to Hawley, an engineering psychologist at the US Army Research Laboratory with a long history working with the Patriot system, "there are few 'decision leverage points' that allow the operators to influence the system's engagement logic and exercise real-time supervisory control over a mostly automated engagement process".[279] Whilst human agents monitor the command module, the Patriot system is "capable of applying lethal force with little or minimal direct human oversight".[280] This effectively reduces the human agent's role to a veto power in engagement decisions.[281] As with other air defence systems, Patriot operators only have a few seconds to exercise a veto.[282]

The Patriot was employed by the US in automatic mode in both Gulf Wars. This was intended to defend against tactical ballistic missiles (TBMs), a capability that the manufacturer Raytheon introduced to the Patriot shortly before the First Gulf War (1990-1991).[283] Official statistics claimed that during this conflict the system had destroyed 79 percent of the Scud missiles launched at Saudi Arabia and 40 percent of the Scuds launched at Israel.[284] The veracity of these numbers has, however, been questioned, including by an independent Congressional report commissioned by the House Committee on Government Operations that found the Patriot had downed less than 9 percent of Scuds launched.[285] Despite anecdotal reports of a series of "close-call" fratricide incidents during the First Gulf War,[286] the publicised success of the Patriot system whilst operating in automatic mode paved the way for its subsequent use during the 2003 invasion of Iraq.

---

[275] Rebecca Leung, "The Patriot Flawed?," CBS News, February 19, 2004, https://www.cbsnews.com/news/the-patriot-flawed-19-02-2004/.

[276] Hawley, "Patriot Wars," 4.

[277] Missile Defense Project, "Patriot," Missile Threat, 2018, https://missilethreat.csis.org/system/patriot/.

[278] Scharre, "Autonomous Weapons and Operational Risk," 31.

[279] Hawley, "Patriot Wars," 4.

[280] Hawley, 4.

[281] Peter W. Singer, *Wired for War. The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2010), 125.

[282] Leung, "The Patriot Flawed?"

[283] John K. Hawley, "Looking Back at 20 Years on MANPRINT on Patriot: Observations and Lessons" (Army Research Laboratory, September 2007), 1.

[284] Fred Kaplan, "Patriot Games," Slate, March 24, 2003, https://slate.com/news-and-politics/2003/03/how-good-are-those-patriot-missiles.html.

[285] US Congress, House Committee on Government Operations, "Activities of the House Committee on Government Operations. 102nd Congress. First and Second Sessions, 1991 – 1992 Report 102-1086. Performance of the Patriot Missile in the Gulf War.," 1992, http://www.turnerhome.org/jct/patriot.html. This was partially because of the way intercepts were officially classified as when the Patriot missile "got within lethal range of the Scud and its fuse exploded" rather than when it actually destroyed the missile. Kaplan, "Patriot Games."

[286] Hawley, "Patriot Wars," 6.

In the twelve years between the two Gulf Wars, the US Army had become so confident of the Patriot system's automatic mode that it de-skilled its operators, "reduc[ing] the experience level of their operating crews [and] the amount of training provided to individual operators and crews".[287] The experience level of the Patriot crew involved in the Tornado fratricide underlines this: "the person who made the call […] was a twenty-two-year old second lieutenant fresh out of training".[288] This underlines the level of confidence invested in the Patriot system's capabilities. These actions are consistent with a typical myth of autonomous systems: "the erroneous idea that once achieved, full autonomy obviates the need for human-machine collaboration".[289] In reality however, *operating the Patriot in automatic mode has increased the complexity of its management, demanding more "human expertise and adaptive capacity"*.[290]

A series of interrelated factors were found to have contributed toward the fratricide incidents. The British Board of Inquiry's report on the downing of the Tornado lists the following six factors among others: (1) the Patriot system's anti-radiation missile classification criteria; (2) the Patriot system's firing doctrine and crew training; (3) autonomous Patriot battery operation; (4) Patriot IFF procedures; (5) the Tornado's IFF serviceability; as well as (6) orders and instructions.[291] Because of our focus on the decision-making element of meaningful human control, we focus our analysis on the nature of human-machine interaction.

Track classification problems in the Patriot system were a major factor leading to both incidents. The Patriot system classifies "tracks" and targets as aircraft, different kinds of missiles (ballistic, cruise, anti-radiation), or other categories based on "flight profiles and other track characteristics such as point of origin and compliance with Airspace Control Orders".[292] When there is a misclassification, "[…] the system-generated category designation does not match the track's actual status".[293] It is important to note that the target profiles of air defence systems like the Patriot are not programmed to defend against a *specific* set of target profiles but around an envelope of possible target profiles.[294] If the target's parameters are defined too precisely, the risk of false-negatives increases.[295] These are situations when the system fails to recognise a target object because one of its prerequisite target conditions are not met, e.g. a missile may be flying slightly slower or from a different angle than the defined profile.[296] The Patriot system's engagement algorithm suffered from both general and specific track classification problems.

(A) In general terms, the system's track classification was not completely reliable – and this had been known prior to the incidents.[297] In fact, there are limits to the reliability of track classification due to the "brittleness" associated with

---

[287] Hawley, 8.

[288] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York; London: W. W. Norton, 2018), 166.

[289] Jeffrey M. Bradshaw et al., "The Seven Deadly Myths of 'Autonomous Systems,'" *IEEE Intelligent Systems* 28, no. 3 (2013): 58.

[290] Matthew Johnson, John K. Hawley, and Jeffrey M. Bradshaw, "Myths of Automation Part 2: Some Very Human Consequences," *IEEE Intelligent Systems* 29, no. 2 (2014): 84.

[291] UK Ministry of Defence, "Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710," 2–3.

[292] Hawley and Mares, "Human Performance Challenges for the Future Force," 6–7.

[293] Hawley and Mares, 7.

[294] The authors want to thank Peter Burt for drawing our attention to this point. See Moyes, "Target Profiles."

[295] Personal communication with Richard Moyes.

[296] Moyes, "Target Profiles," 5.

[297] Hawley and Mares, "Human Performance Challenges for the Future Force," 7.

algorithms and AI, which are characterised by their inability to contextualise[298] and have "little capacity to handle gray or ambiguous situations".[299] Rather than directly addressing these deficiencies by communicating them to the system's operators, the US Army framed these as a software problem and repeatedly claimed "[…] that a 'technical fix' […] was just around the corner".[300] Amongst other factors, this omission contributed toward an *unwarranted over-trust in the system* (see also later discussion of automation bias).

(B) The algorithm governing the Patriot system's targeting selection was trained on a data set that lacked the specificity to prevent false identifications. They "were based on the many different Anti-Radiation Missiles available worldwide," rather than "on the known threat from Iraq".[301] Informed by his personal experience with developing the system, Hawley concludes that the Patriot's engagement algorithms were consequently not specific enough to "handle unusual or ambiguous tactical situations reliably" that invariably present themselves in the context of countering conventional air threats.[302] Due to these inaccuracies, the oncoming data points received by the system confused the friendly jet for an imminent missile attack. In the words of a reporter who was embedded with Patriot batteries in the Second Gulf War:

> *This was like a bad science fiction movie in which the computer starts creating false targets. And you have the operators of the system wondering is this is a figment of a computer's imagination or is this real. They were seeing what were called spurious targets that were identified as incoming tactical ballistic missiles. Sometimes, they didn't exist at all in time and space. Other times, they were identifying friendly U.S. aircraft as incoming TBMs [tactical ballistic missiles].*[303]

Compounding these track classification problems, it also appears as if the identification friend or foe (IFF) system did not perform as expected. This was a known problem, as indicated by earlier near-miss fratricidal engagements involving Patriot batteries in the First Gulf War and in training.[304] As the Defence Science Board Task Force's report stated: "This is not exactly a surprise; this poor performance has been seen in many training exercises. The Task Force remains puzzled as to why this deficiency never garners enough resolve and support to result in a robust fix."[305]

As a consequence, the Patriot crew lacked the time, and, crucially, it also lacked the necessary information, understanding, and expertise to overrule the targeting decisions made by the system. This included known limitations of the Patriot system and the potential conditions under which it might fail, such as the IFF system and track classification problem. In the British Ministry of Defence's assessment, "Patriot crews are trained to react quickly, engage early and to *trust* the Patriot system. […] The crew had about one minute to decide whether to engage".[306] These aspects of training and trust are characteristic of a wider problem with existing patterns of human-machine interaction across a range of air defence systems.

---

[298] Jason Pontin, "Greedy, Brittle, Opaque, and Shallow: The Downsides to Deep Learning," Wired, February 2, 2018, https://www.wired.com/story/greedy-brittle-opaque-and-shallow-the-downsides-to-deep-learning/.

[299] Hawley, "Patriot Wars," 4.

[300] Hawley and Mares, "Human Performance Challenges for the Future Force," 7.

[301] UK Ministry of Defence, "Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710," 3.

[302] Hawley, "Patriot Wars," 4.

[303] Riggs quoted in Leung, "The Patriot Flawed?"

[304] Scharre, "Autonomous Weapons and Operational Risk," 30.

[305] US Department of Defense, "Report of the Defense Science Board Task Force on Patriot System Performance. Report Summary.," 2.

[306] UK Ministry of Defence, "Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710," 3.

The concept of over-trust,[307] also known as automation bias, or "automation complacency",[308] refers to human operators being overly confident in the reliability of automated and autonomous systems and the accuracy of their outputs. This manifests in a "psychological state characterized by a low level of suspicion".[309] Trusting the system was deeply ingrained into the Patriot's way of functioning, as the Defence Science Board's review of the Patriot noted: "The operating protocol was largely automatic, and the operators *were trained to trust* the system's software; a design that would be needed for heavy missile attacks".[310] These circumstances of over-trust contrast with the under-trust discussed earlier in the context of the USS Vincennes: the system's instrumentation was correctly displaying that IR655 was descending, but this information was disregarded by the operators.

These observations point to the increasing challenges faced by the human operators tasked with remaining 'on the loop' for the Patriot system. Hawley refers to this as the "humans' residual role in system control, and how difficult that role can be to prepare and perform".[311] The human operator has to step in where the system fails, requiring both "sustained operator vigilance, […] broad-based situation awareness",[312] and adequate expertise/experience in hands-on battle management so that they can scrutinise the system's decisions. In other words, human operators must know *when to trust* the system and *when to question* its outputs.[313] This is a case of human judgement which operators have to get exactly right: both too much or too little trust in the system can create problems. This requires human operators to understand how the system works and what its weaknesses are. But these requirements for meaningful human control are typically lacking, if not impossible to meet, in specific targeting situations like the ones discussed.[314]

The friendly fire incidents discussed demonstrate the extent to which the inclusion of automated functions renders human-machine interaction incredibly complex – and in the process contributes to setting *emerging,* circumscribed standards for a key component of exercising meaningful human control. To illustrate the system's technical and tactical complexity, the "Patriot currently employs more than 3.5 million lines of software code in air battle management operations".[315] Moreover, systems such as the Patriot do not operate in isolation but rather as part of an integrated air defence system. This means that the Patriot works in close association with other, equally complex air defence systems that also include autonomous features, such as the AEGIS or the THAAD.[316] For human operators, the Patriot is therefore "knowledge-intensive in terms of the amount of information required to characterise and comprehend the system".[317] The fratricides illustrate how "the complexity of the system contributed to human operators' misperceiving or misunderstanding the system's behaviour".[318]

---

[307] Boulanin et al., "Limits of Autonomy in Weapon Systems," 19.

[308] Raja Parasuraman and Dietrich H. Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52, no. 3 (June 2010): 381–410, https://doi.org/10.1177/0018720810376055.

[309] E. L. Wiener, "Complacency: Is the Term Useful for Air Safety," in *Proceedings of the 26th Corporate Aviation Safety Seminar* (Denver, CO: Flight Safety Foundation, Inc., 1981), 117.

[310] US Department of Defense, "Report of the Defense Science Board Task Force on Patriot System Performance. Report Summary.," 2.

[311] Hawley, "Patriot Wars," 2.

[312] Hawley, 8.

[313] Hawley and Mares, "Human Performance Challenges for the Future Force," 7.

[314] Hawley, "Patriot Wars," 4.

[315] Hawley and Mares, "Human Performance Challenges for the Future Force," 4.

[316] Hawley and Mares, 4.

[317] Hawley, "Looking Back at 20 Years on MANPRINT on Patriot," 1.

[318] Scharre, "Autonomous Weapons and Operational Risk," 32.

This has significant repercussions for operators' *situational awareness* – a key issue in the diminished capacity for exercising meaningful human control through human-machine interaction. Following an established definition in human factor analysis, situational awareness refers to "the perception of elements in the environment […], the comprehension of their meaning, and the projection of their status in the near future".[319] In order to retain situational awareness and "behave appropriately […] operators must keep track of considerable information from a variety of sources over time and organize and interpret this information".[320] But, human operators of air defence systems with autonomous features, such as the Patriot, are ill-equipped to retain situational awareness because they have been transformed from being the active controllers of a weapons system (in-the-loop) to system monitors (on-the-loop).[321] This observation is fully in line with what human factor researchers have long argued: automation (and autonomy) "change the nature of the work that humans do, often in ways unintended and unanticipated".[322]

The human agent's modified role from active controller to system monitor implies the delegation of cognitive skills, not just motor and sensory tasks, to machines.[323] This produces two distinct but interrelated problems regarding the retention of situational awareness: first, in their role as monitors, human agents are either overloaded or underloaded with tasks vis-à-vis those delegated to the system.[324] This means that humans are either incapable of competently performing the tasks allocated to them (overload) or that these tasks are so menial that retaining appropriate vigilance increases in difficulty over the required period (underload).[325] The underload problem, manifest in a lack of vigilance, was identified as one of the key factors contributing to the Patriot fratricide incidents.[326]

Second, and more fundamentally, human operators may not have a workable model of how the machine makes decisions as well as the control process that informs this operation. In this way, they lack an understanding of the logic underpinning the tasks they are expected to perform.[327] With many of their decision-making tasks now being performed by the system, humans are relegated to monitoring. This may mean that they are potentially second-guessing a decision-making process that they are no longer familiar with. As the operator does not have "something reasonable to do when the system is operating normally, it is unlikely that he or she can function effectively [both] in manual backup mode"[328] or when something out of the ordinary occurs. Given these restrictions to the situational awareness of human operators, "calling for reliable supervisory control over a complex automated system is an unreasonable performance expectation".[329]

---

[319] Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," 36.

[320] Hawley, Mares, and Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control," 5.

[321] Hawley, "Patriot Wars," 10.

[322] Raja Parasuraman and Victor Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no. 2 (1997): 231.

[323] Hawley, Mares, and Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control," 3.

[324] B. H. Kantowitz and R. D. Sorkin, "Allocation of Functions," in *Handbook of Human Factors*, ed. G. Salvendy (New York: Wiley, 1987), 355–69.

[325] Hawley, Mares, and Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control," 8.

[326] Hawley, "Looking Back at 20 Years on MANPRINT on Patriot."

[327] Hawley, Mares, and Giammanco, "The Human Side of Automation: Lessons for Air Defense Command and Control," 8.

[328] Hawley, Mares, and Giammanco, 8.

[329] Hawley, "Patriot Wars," 8.

These aspects were confirmed in an internal review of the Patriot system conducted in the aftermath of the fratricide incidents at the US Army Research Laboratory. Two conclusions were reached: first, functions had been automated in the Patriot system both at the design and the implementation stages "[…] without due regard for the consequences for human performance";[330] and second, that the Patriot system was fielded and operated with a significant automation bias, evident in a "blind faith in technology".[331]

What is more, in addition to hindering human comprehension, the complexity of systems with automated and autonomous features, such as the Patriot, makes them susceptible to failure. This is because it is not possible to ascertain and test how the system and it sub-systems will behave across all possible conditions and situations since "the number of potential interactions within the system and with its environment is simply too large".[332] Consequentially, operating weapons systems with automated and autonomous features comes with significant risks of failure – and that "risk can be reduced but never entirely eliminated".[333]

The case of the Patriot demonstrates how the operation of air defence systems has contributed toward establishing emerging standards of appropriateness regarding the use of force, setting norms for what human-machine interaction looks like and what its acceptable quality is. In most cases, these emerging norms go unnoticed. They only become subject to scrutiny in the case of failures.[334] Since 1987, US law has required that ballistic missile systems, such as the Patriot, cannot use "[…] lethal fire except by *affirmative human decision* at an appropriate level of authority".[335] Yet, the exact meaning of "affirmative human decision" has never been defined. According to Hawley, this law has only minimally impacted how air defence systems are developed and operated.[336] As our analysis suggests, "affirmative human action" has been interpreted in a *minimal* way: "the requirement for positive human control is met even if that means not much more than having a warm body at the system's control station".[337] While human operators have to remain 'on-the-loop', they lack "substantive situational understanding" and often have only a few seconds to make decisions.[338] This is tantamount to a loss of meaningful human control in specific targeting decisions, making the Patriot system the *de facto* "ultimate decision-maker in engagement decisions".[339]

The Patriot system's complexity and its significant challenges to exercising meaningful human control via human-machine interaction are representative of a wider range of air defence systems as well as other weapons systems with automated and autonomous features that the US Army may field in the future.[340] Moreover, the Patriot case-study highlights the incremental way in which meaningful human decision-making via human-machine interaction has

---

[330] Hawley and Mares, "Human Performance Challenges for the Future Force," 6.

[331] Hawley and Mares, 6.

[332] Scharre, "Autonomous Weapons and Operational Risk," 5.

[333] Scharre, 25.

[334] The Patriot fratricides led to a significant review exercise conducted by experts at the Army Research Laboratory. Based on conducting a series of training tests, these experts were tasked, inter alia, with securing human control for the Patriot system. However, the training tests were not conducted as planned and still featured many of the issues that had been previously identified as problematic. As a consequence, the test never lead to significant changes in Patriot system training overall. See Hawley, " Looking Back at 20 Years on MANPRINT on Patriot," 7.

[335] United States of America, *United States Code. 2012 Edition. Volume 5, Title 10 Armed Forces §§1431-7921* (Washington, DC: Government Printing Office, 2013), 618.

[336] Hawley, "Patriot Wars," 9.

[337] Hawley, 9.

[338] IPRAW, "Focus on the Human-Machine Relation in LAWS," Focus On Report (Berlin: International Panel on the Regulation of Autonomous Weapons, March 2018), 17.

[339] Hawley and Mares, "Human Performance Challenges for the Future Force," 10.

[340] Hawley and Mares, 4–5.

diminished over time.[341] System developers and users edged into situations where friendly-fire incidents occurred by degrees. Rather than delegating decision-making tasks to machines in one sweep, this was an incremental process of progressive software updates that has culminated in human operators being asked to fulfil minimal, but at the same time impossibly complex, roles.

4.5 Summary: Challenges inherent to human-machine interaction

The detailed analysis of severe incidents involving air defence systems reveals a long list of overlapping challenges to human-machine interaction which are characteristic of operating complex systems with automated or autonomous features:

- **Automation bias/Over-trust**. This leads human operators to uncritically trust system outputs without subjecting them to deliberative or critical reasoning. This makes them more likely, for example, to not question algorithmic targeting parameters, despite the potential existence of track classification problems.
- **Lack of system understanding**. Human operators do not understand the precise functioning of automated and autonomous features in air defence systems, including their target profiles and how they calculate target assessments. This is partly due to the system's complexity creating a barrier to understanding. But the incidents have also shown that operators were not aware of *known* system weaknesses, e.g. IFF performance in the case of the Patriot.
- **Lack of situational understanding**. In the move from being active controllers (being 'in the loop'), to supervisory controllers ('on the loop'), human operators lose situational understanding. This makes it more difficult to question system outputs and make reasoned deliberations about selecting and engaging specific targets.
- **Lack of time for deliberation**. The current engagement window of air defence systems provides human operators only a few seconds to make decisions. This places impossible demands on any potential critical deliberation.
- **Lack of expertise**. Operating complex systems such as air defence systems competently requires extensive training and experience. Examples from the Patriot system suggest that human operators lack this expertise due, in part, to a misguided faith in the capabilities of weapons technologies and management policies.
- **Inadequate training.** Training of human operators for air defence systems, as evidenced by the Patriot example, appears to focus on inappropriate mechanistic approach as opposed to simulating scenarios key for retaining meaningful human control in targeting decision-making.
- **Operation under high-pressure combat situations exacerbates challenges inherent to human-machine interaction**. All of the discussed incidents happened in the context of a period of international tension when militaries were on high alert. The pressure on the individual human operators involved has been a factor in the incidents and exacerbated the challenges inherent to human-machine interaction. It is important to remember that such high-pressure combat situations are the *default* situation in times of war, which is when weapons systems can be expected to be used.[342]

---

[341] Hawley, "Looking Back at 20 Years on MANPRINT on Patriot," 5.

[342] The authors want to thank Peter Burt for drawing our attention to this point.

Inside a Patriot's control station, October 2007 **Source** defensie.nl

These challenges are well-known to scholars of human factor analysis. The fact that they severely compromise the exercise of meaningful human control in specific targeting situations involving air defence systems *is therefore an expectable outcome*.

However, despite this knowledge there has been no fundamental reflection on the appropriateness and the unintended consequences of the continued integration of automation and autonomy in air defence systems. In fact, continued upgrades of automated and autonomous features in air defence systems have made the human-machine interaction challenges more acute. States using air defence systems have therefore implicitly accepted the compromised role of human agents in modern air defence systems. On our reading, these challenges to human-machine interaction have made human control *meaningless*. This is particularly the case for human-on-the-loop situations, which are often presented as a form of meaningful human control and acceptable during the supervision of weapons systems with automated and autonomous features.[343] In some ways, human operators appear to serve as a "form of ethical cover for what is in reality, an almost wholly mechanical, robotic act".[344] These circumscriptions of meaningful human control have generally not been openly acknowledged or discussed. Nevertheless, they are characteristic of how modern air defence systems are operated. This means that meaning*less* human control has become accepted by a group of states as an emerging, 'appropriate' understanding of how force can be used in specific targeting situations.

---

[343] The authors want to thank Peter Burt for drawing our attention to this point.

[344] Mike Ryder, "Killer Robots Already Exist, and They've Been Here a Very Long Time," The Conversation, March 27, 2019, https://theconversation.com/killer-robots-already-exist-and-theyve-been-here-a-very-long-time-113941.

5

CONCLUSION: KEY FINDINGS AND RECOMMENDATIONS

Automated and autonomous features have been integrated into the critical functions of air defence systems for decades, long preceding the current international dialogue on AWS at the UN-CCW in Geneva. As our report shows, this has come with far-reaching repercussions for the human operator's role as a meaningful decision-maker in specific targeting situations. Air defence systems have been cast as 'unproblematic' from the perspective of meaningful human control. Applying the recent understanding of the concept developed jointly by the ICRC and SIPRI as well as separately by the Campaign to Stop Killer Robots, they are operated under three sets of measures that could conceivably increase the prospects of meaningful human control: through the weapons system's operating parameters, control can be exercised over *what* is targeted (munitions or hostile aircraft), *where* force is used (the environment), and *when* force is used ("there is an opportunity for a human to override").[345]

But our detailed study of human-machine interaction in air defence systems – a central component of operationalising meaningful human control – reveals deeply problematic practices. Increased machine delegation has opened the door for the incremental and largely overlooked reduction of meaningful human control. While human operators often formally retain the final decision about specific targeting decisions, as we have shown, this 'decision' is in practice often *meaningless*. The complexities of human-machine interaction does not allow for situational awareness, it does not provide human operators with the knowledge/expertise necessary to understand and question the system's logic, and it does not give them the time to engage in *meaningful* deliberation. In current air defence systems, humans can in some situations be unintentionally effectively set up to fail as meaningful operators.

As the integration of automation and autonomy into the critical functions of air defence systems has widened and accelerated, an emerging standard of appropriateness – or norm – attributing humans a diminished role in specific targeting decisions has emerged. This process can and should be paused to allow for critical reflection, debate, and legal regulation.

Key findings

- The analysis of air defence systems shows that there is already an operational (and problematic) understanding of the human-machine interaction component of meaningful human control.

---

[345] Campaign to Stop Killer Robots, "Key Elements of a Treaty FAQ," 3.

- The decades long operation, development, and testing of automated and increasingly autonomous air defence has diminished the capacity of human operators to exercise meaningful human control over specific targeting decisions.
- The implicit operational understanding of meaningful human control is deeply problematic because human operators are required to fulfil a minimal but at the same impossibly complex role.
- Systems with both automated and autonomous features obstruct the exercise of meaningful human control because they increase the complexity of the systems operation. Due to the character of human-machine interaction in these systems, the human operator can lack a sufficient understanding of the decision-making process, leading to over-trust in the system and a lack of sufficient situational understanding.
- This problematic operational quality of human-machine interaction is yet to be fully publicly acknowledged, scrutinised, or deliberated.
- If this continues, there is a risk that deliberative efforts to codify meaningful human control as a vital component of new international law on AWS will be undermined.

On this basis, we make the following recommendations for stakeholders involved in the international debate on LAWS at the UN-CCW. These begin from the premise that positively codifying an operationalised version of meaningful human control is the most promising avenue for creating a regulatory framework on LAWS' development and usage.

- **Current practices of how states operate weapons systems with automated and autonomous features in specific use of force situations should be brought into the open and scrutinised. As we argue, such operational practices shape what constitutes 'meaningful' human control, especially the quality and type of human-machine interaction.** Our examination of control through human-machine interaction in air defence systems has highlighted significant challenges to its meaningful exercise. Further analysis of this and other existing weapons systems can help make the regulatory dialogue on LAWS less abstract and push the debate toward a greater awareness of the problematic consequences of meaningless human control. Similarly, we encourage civil society groups to call policymakers to interrogate their socialisation into, and internalisation of, assumptions regarding the appropriateness of integrating automation and autonomy into the critical functions of weapons systems of all types.
- **More in-depth studies of the emerging standards set for meaningful human control produced by the use of other existing weapons systems with automated and autonomous features beyond air defence systems is required. Such studies can provide practical insights into the existing and future challenges to human-machine interaction created by autonomy and automation that, if not explicitly addressed, may shape silent understandings of appropriateness.** We support calls by key stakeholders, such as the ICRC and SIPRI, for the detailed study of existing AWS. All "weapons systems that select targets on the basis of sensor input"[346] should be assessed for whether they allow for meaningful human control. Such systems include, but are not restricted to active protection systems, counter-drone systems,[347] and 'fire-and-forget' missiles. These and other case studies

---

[346] Campaign to Stop Killer Robots, 4.

[347] Arthur Holland Michel, "Counter-Drone Systems. 2nd Edition" (Centre for the Study of the Drone, Bard College, December 2019), https://dronecenter.bard.edu/files/2019/12/CSD-CUAS-2nd-Edition-Web.pdf.

would help bridge the gap between deliberation and operational practices that otherwise risks undermining efforts to establish a robust regulatory framework.

- **Our study of air defence systems highlights that while all three components of meaningful human control (technological, conditional, and human-machine interaction) are important, control through human-machine interaction is the decisive element in ensuring that control remains meaningful. This is not least because human-machine interaction highlights meaningful human control at the specific point of using a weapons system, rather than the exercise of human control at earlier stages, such as during research and development.** Air defence systems are often deemed unproblematic from a meaningful human control perspective because states can theoretically limit where, how and when they are deployed by setting a weapons system's parameters of use and controls on the environment. But our close study of control through human-machine interaction demonstrates how, under certain conditions, this can render the human operator's role in specific targeting decisions essentially meaningless and that restrictions on the use of air defence systems intended to achieve human control can fail.
- **Control through human-machine interaction should be integral to any codification of meaningful human control in disarmament debates. We identify three prerequisite conditions needed for human agents to exercise meaningful human control: (1) a functional understanding of how the targeting system operates and makes targeting decisions including its known weaknesses (e.g. track classification issues); (2) sufficient situational understanding; and (3) the capacity to scrutinise machine targeting decision-making rather than over-trusting the system. Of course, human operators should also have the possibility to abort the use of force.** In short: human operators must be able to build a mental model of the system's decision-making process and the logic informing this. This includes, for example, access to additional (intelligence) sources beyond the system's output, allowing them to triangulate the system's targeting recommendations.
- **These three prerequisite conditions (functional understanding, situational understanding and the capacity to scrutinise machine targeting decision-making) of ensuring meaningful human control in specific targeting situations set hard boundaries for AWS development that should be codified in international law. In our assessment, they represent a technological Rubicon which should not be crossed as going beyond these limits makes human control meaningless. Adhering to these conditions does not only ensure that human control remains meaningful, it also has the potential of easing the pressure put on human operators of air defence systems who are currently, unintentionally, set up to fail.** At the moment, individual human operators at the bottom of the chain of command are often held accountable for structural failures in how automation and autonomy in air defence systems is designed and operated. Retaining meaningful human control in the true sense of the word therefore also has the potential to make their tasks *doable*. Regulating meaningful human control can make a constructive contribution for states operating air defence systems with automated and autonomous features. Normative constraints in using weapons systems do not have to stand against state interests. Militaries want to enhance control and the failures described in relation to current ways of operating air defence systems are something that states want to avoid. In this way, positively codified standards of how to retain meaningful human control in specific use of force situations can be helpful for states.

- **The complexity inherent to human-machine interaction means that there will be limits to exercising meaningful human control in specific targeting decisions. Ensuring the stringent training of human operators is a necessary precondition for maintaining meaningful human control but is not a panacea. This inconvenient truth should be made clear to all relevant stakeholders.**

APPENDIX

Appendix A **MIM-104 Patriot catalogue entry**

MIM-104 Patriot	
System manufacturer	Raytheon (US)
System users	US, Germany, Greece, Israel, Japan, Kuwait, the Netherlands, Saudi Arabia, South Korea, Poland, Sweden, Qatar, Romania, Spain, Taiwan and the UAE.[348]
System history	The MIM-104 "Phased Array Tracking Radar to Intercept on Target" (Patriot) system is a surface-to-air air defence system. It is described by the Missile Defence Agency as "[p]rovid[ing] simultaneous air and missile defense capabilities as the Lower Tier element in defense of U.S. deployed forces and allies".[349] Development on the SAM-D programme, as it was then called, began in 1976 and was designed to replace the Nike Hercules and the MIM-23 Hawk air defence systems. The first Patriot system was declared fully operational by the US Army, its principal user, in 1985.[350] Shortly thereafter, the system rose to international prominence during Operation Desert Storm when it was deployed to Israel and Saudi Arabia to protect against Iraqi SCUD missile attacks.[351] An upgraded variant of the Patriot system – PAC-3 – was deployed to Kuwait during the 2003 invasion of Iraq and intercepted several Iraqi surface-to-surface missiles.[352] During this conflict, a MIM-104 Patriot system was involved in two fratricidal engagements which destroyed a RAF Tornado fighter jet (24 March 2003) and a US Navy F-18 fighter jet (2 April 2003), killing three crewmembers in total.[353] There was a further close-call friendly fire incident on 25 March 2003 when a Patriot battery locked onto an US Air Force F-16. In this case, the pilot "was alerted to the fact that he had been targeted by radar" and launched a counter-attack that destroyed the Patriot battery.[354] The US has operated Patriot systems in South Korea since 1994 to protect against possible North Korean missile attacks.[355] Between January 2013 and the end of 2015, five NATO Patriot batteries were deployed to Turkey to help defend against possible ballistic missile attack from Syria. The United Arab Emirates and Saudi Arabia have also deployed variants of the Patriot system to defend against aerial attacks from Houthi rebels fighting in Yemen.[356]

---

348 Army Technology, "Patriot Missile Long-Range Air-Defence System," 2020, https://www.army-technology.com/projects/patriot/.

349 Missile Defense Agency, "Patriot Advanced Capability-3 Fact Sheet," 2020, https://www.mda.mil/global/documents/pdf/pac3.pdf.

350 Missile Threat, "Patriot," *CSIS Missile Defence Project*, 2020, https://missilethreat.csis.org/system/patriot/.

351 Ibid.

352 Army Technology, "Patriot Missile Long-Range Air-Defence System."

353 US Department of Defense, "Report of the Defense Science Board Task Force on Patriot System Performance. Report Summary." (Washington, DC: Office of the Under-Secretary of Defense for Acquisition, Technology, and Logistics, January 2005), 2, https://dsb.cto.mil/reports/2000s/ADA435837.pdf.

354 Lester Haines, "Patriot Missile: Friend or Foe?," *The Register*, May 2004.

355 Missile Defence Advocacy Alliance, "Patriot Missile Defense System," 2019, https://missiledefenseadvocacy.org/defense-systems/patriot-missile-defense-system/.

356 Ibid.

System range	20km interception range against ballistic missiles. The Patriot's radar has a reported range of 150km.[357]
Target type	Aircraft, short range ballistic missiles, cruise missiles, UAV.[358]
Major system updates	According to the systems manufacturer Raytheon, "[e]ach and every time Patriot is tested or live fired, engineers uncover new ways to further improve or enhance the system".[359] To this end, Raytheon claim to have "continually embraced new technologies to stay ahead of evolving threats".[360] For example, upgrades to the Patriot system's software made during the late 1980s enabled it to track and intercept short-range ballistic missiles (Patriot PAC-1 upgrades).[361] This anti-ballistic missile capability is reported to have required a greater level of automation because "[t]he nuts and bolts of the ballistic missile engagement process are too complex and time-limited for direct, in-the-loop human participation".[362] A series of different Patriot missile variants have also been developed, with the most recent being the PAC-3: "a high velocity interceptor that defeats incoming targets by direct, body-to-body impact".[363]
Automation and autonomy in critical functions	Raytheon press releases from 2005 described the Patriot as having "automated operations – including man-in-the-loop (human) override" as a key feature.[364] The CSIS Missile Defence Project describe the Patriot system as being "nearly autonomous, with only the final launch decision requiring human interaction". The AN/MSQ-104 Engagement Control Station – described as "essentially the brain of the Patriot system" – is the only crewed part of the Patriot system and is operated by three people.[365] According to the BBC, the Patriot system "can be automated, although an operator is able to override it".[366] Others have labelled the Patriot as an example of an "human-supervised autonomous weapon systems".[367] John Hawley, an expert who has both researched this system and been involved with its development, describes the Patriot as being "highly automated".[368] According to Hawley, the Patriot was one of the first systems operated by the US to be capable of "lethal autonomy" defined as a "system that is capable of applying lethal force with little or minimal direct human oversight".[369] When operating in automatic mode, Hawley describes the Patriot system as being capable of "the conduct of near-autonomous operations".[370]

---

[357] NATO, "PATRIOT Deployment," 2015, https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2015_05/20150508_1505-Factsheet-PATRIOT_en.pdf.

[358] Raytheon Missiles & Defence, "Global Patriot Solutions," 2020, https://www.raytheonmissilesanddefense.com/capabilities/products/global-patriot-solutions.

[359] Ibid.

[360] Ibid.

[361] Missile Defence Advocacy Alliance, "Patriot Missile Defense System."

[362] J. Hawley, "PATRIOT WARS: Automation and the Patriot Air and Missile Defense System," *Center for a New American Security*, 2017, 6, https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf.

[363] Army Recognition, "Patriot MIM-104 Surface-to-Air Defense Missile System," 2019, https://www.armyrecognition.com/united_states_american_missile_system_vehicle_uk/patriot_mim-104_surface-to-air_defense_missile_data_sheet_specifications_information_description.html.

[364] Raytheon, "PATRIOT MISSILE SYSTEM," 2005, https://web.archive.org/web/20050826224850/http://www.raytheon.com/products/patriot/.

[365] Missile Threat, "Patriot."

[366] BBC, "Fact File: Patriot Missile Defence," 2012, https://www.bbc.co.uk/news/world-20594466.

[367] Paul Scharre and Michael C. Horowitz, "An Introduction to AUTONOMY in WEAPON SYSTEMS," *Center for a New American Security*, 2015, 12, https://www.files.ethz.ch/isn/188865/Ethical Autonomy Working Paper_021015_v02.pdf.

[368] Hawley, "PATRIOT WARS: Automation and the Patriot Air and Missile Defense System," 1.

[369] Ibid., 4.

[370] Ibid., 6.

Response time and simultaneous tracking capacity	**Response time**	Whilst the authors were unable to find the exact response time for this system, according to a 2015 NATO factsheet, the Patriot system has a "short response time [and] the ability to engage multiple targets simultaneously".[371]
	Number of targets which can be simultaneously tracked/engaged	This system can reportedly "track up to 100 targets and can provide missile guidance data for up to nine missiles".[372]

Soldiers from the 31st Air Defense Artillery Brigade load a Patriot missile onto a transfer vehicle at McGregor Test Range, N.M., during Exercise Roving Sands 97 on April 18, 1997 **Source** Senior Airman Jerry Morrison, U.S. Air Force

---

[371] NATO, "PATRIOT Deployment."

[372] Army Technology, "Patriot Missile Long-Range Air-Defence System."

DRONE WARS

Shining a spotlight
on military drones

CENTER FOR
WAR
STUDIES

SDU

University of
Southern Denmark