THE PROBLEMATISATION OF AUTONOMOUS WEAPON SYSTEMS – A CASE STUDY OF THE US DEPARTMENT OF DEFENSE

MIKOLAJ FIRLEJ

ST CROSS COLLEGE

Thesis submitted to the Faculty of Law at the University of Oxford

for the degree of

Doctor of Philosophy in Socio-Legal Studies

Michaelmas Term 2022/23

ABSTRACT

THE PROBLEMATISATION OF AUTONOMOUS WEAPON SYSTEMS – A CASE STUDY OF THE US DEPARTMENT OF DEFENSE

Robotics systems play an increasingly important role in armed conflicts and there are already weapons in service that replace a human being at the point of engagement. The United States (US) is the first country to have adopted a policy on autonomous weapon systems (AWS) in the Directive 3000.09. The US policy on AWS is however poorly understood in the academic and policy circles. This thesis addresses the question of how the US Department of Defense (DoD) problematises the concept of AWS.

By applying a Bacchi's poststructuralist approach to policy analysis, the thesis asks how the US DoD constructs the 'problem' of AWS, what assumptions underlie this representation of the 'problem', how has it come about, what effects it produces, what is left out of problem representation, and how could it be questioned.

The US DoD problematisation of AWS does not only clarifies the Department's approach, but also it explores the role of human involvement over the use of AWS. The US policy states that AWS shall be used by 'appropriate levels of human judgment'. This term is, however, open to different interpretations, and some argue that it prohibits a lethal use of AWS, while others disagree.

The thesis focuses not only on content of the US concept of human judgment, but primarily on how this concept relates to the wider US military understanding of 'control.' In that, it unpacks the concept of human judgment and distinguishes it from the concept of human control. I argue that both concepts are important in the debate on AWS as they represent alternative policy approaches to the use of such weapons. By making these

concepts more explicit, my thesis contributes to the specific and emerging academic debate about the role of human involvement over the use of AWS.

Word count: 99,479

ACKNOWLEDGEMENTS

I would like to thank my supervisors from the Faculty of Law, Bettina Lange and Jessie Blackbourn, for their constant support, invaluable insights, and friendly encouragement, which have led me to complete this thesis, a work which at times made somewhat slow progress. Bettina, thank you for setting me the best example of how to be a good scholar.

This research took me to many interesting places where I met incredible people. I am thankful to my colleagues in the defense community, policymakers from the UN CCW meetings, AI researchers and legal scholars. I would also like to thank my friends in Oxford, especially from the Polish Society, for creating a friendly and inspirational atmosphere in which to conduct academic research. My gratitude is particularly extended to Mikołaj Barczentewicz, Wodzik Kiciński, Miłosz Palej, Mikołaj Kasprzak, Krzyś Bar, Paweł Borowski, Mati Firlej, Monk Jarek Kurek, Mateusz Kusio, Monica Kaminski, Lucas Kello, Tsvetelina van Benthem, Stergios Aidinlis, Ngaire Woods, Kalypso Nicolaïdis, and many others for your friendship and inspiring conversations about work and life. In my forever memory I will hold dear Zbig Pelczynski and Andrzej Krzeczunowicz.

I have also benefited from being a member of two inspiring research environments: the Centre for Technology and Global Affairs and the Centre for Socio-Legal Studies, both at the University of Oxford. I would also like to single out the Congregation of Ursuline sisters who provided me excellent conditions in ‘Matulinek’ in the final year of writing this thesis. Their humble service is a testimony to the work larger than life.

I am indebted to my wonderful tutors and mentors: my parents Barbara and Krzysztof Firlej, Krzysztof Ślezinski, Wiesław Staśkiewicz, George Swirski, Jan Krzysztof Bielecki, and Artur Kluz. I would also like to thank Helene Zaleski, Edward Lucas, The Grabowski Family Trust, Ewa Rozwadowska, and many others who supported my journey at Oxford. In undertaking a DPhil at Oxford and writing this thesis, I have been significantly supported by an ESRC grant.

Last but not least:

My special and most heartfelt thanks are due to my companion and wife, Ola, for her wisdom, patience, and loving support.

I dedicate this thesis to three wonderful ladies in my life: Ola, Zosia, and Hania.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SHORT TERMS

AARS	The Advanced Airborne Reconnaissance System
AFSOC	Air Force Special Operations Command
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AOC	Air Operations Center
APA	Administrative Procedure Act
AP I	Additional Protocol I to the Geneva Conventions
ATO	Air Tasking Order
AWS	Autonomous Weapon Systems
C2	Command and Control
Campaign	Campaign to Stop Killer Robots
CAPTOR	Encapsulated Torpedo
CAS	Close-Air Support
CDAO	Chief Digital and Artificial Intelligence Officer
CID	Combat Identification
CNAS	Center for a New American Security
CFR	Code of Federal Regulations
Col	Colonel
CRS	Congressional Research Service
CUP	Cambridge University Press
DARPA	Defense Advanced Research Projects Agency
DCGS	Distributed Common Ground System
DIB	Defense Innovation Board

DIU	Defense Innovation Unit
DSB	Defense Science Board
EBP	Evidence-Based Policy
F2T2EA	Find – Fix – Track – Target – Engage – Assess
FY	Financial Year
GAO	Government Accountability Office
GCS	Ground Control Station
Gen	General
GPS	Global Positioning System
HCI	Human–Computer Interface
HMT	Human–Machine Teaming
HRW	Human Rights Watch
HSC	Human Supervisory Control
IAI	Israel Aerospace Industries
ICRAC	International Committee for Robot Arms Control
ICRC	International Committee of the Red Cross
IEEE	Institute of Electrical and Electronics Engineers
IHL	International Humanitarian Law
IHRL	International Human Rights Law
ISR	Intelligence, Surveillance, and Reconnaissance
JADC2	Joint All Domain Command and Control
JAIC	Joint Artificial Intelligence Center
JCIDS	Joint Capabilities Integration and Development System
JFACC	Joint Forces Air Component Commander

JFC	Joint Forces Commander
JTAC	Joint Terminal Attack Controller
JTF	Joint Task Force
JUON	Joint Urgent Operational Needs
LAWS	Lethal Autonomous Weapon Systems
LOA	Level of Automation
LOAC	Laws of Armed Conflict
LRASM	Long-Range AGM-158C
LRE	Launch and Recovery Element
Lt	Lieutenant
Maj	Major
MAPRINT	Manpower and Personnel Integration
MBC	Management By Consent
MBE	Management By Exception
MCE	Mission Control Element
MHC	Meaningful Human Control
ML	Machine Learning
MOSA	Modular Open Systems Architecture
MUM-T	Manned-Unmanned Teaming
NASCAI	National Security Commission on Artificial Intelligence
NATO	North Atlantic Treaty Organization
NDAA	National Defense Authorization Act
NGO	Non-Governmental Organisation
NSTC	National Science and Technology Council

OA	Open Architecture
OODA	Orient, Observe, Decide, and Act
OUP	Oxford University Press
Phalanx CIWS	Phalanx Close-In Weapon System
PID	Positive Identification
PLC	Programmable Logic Controller
T&E	Test and Evaluation
TASM	Tomahawk Anti-Ship Missile
TDC	Target Development Cell
TST	Time-Sensitive Target
TTP	Tactics, Techniques, and Procedures
UAV	Unmanned Aerial Vehicle
UN CCW	United Nations Convention on Certain Conventional Weapons
UN GGE	United Nations Group of Governmental Experts
UNIDIR	United Nations Institute for Disarmament Research
USAF	United States Air Force
US DoD	United States Department of Defense
USDP	Under Secretary of Defense for Policy
USMC	United States Marine Corps
V&V	Verification and Validation
WPR	What's the Problem Represented to be?

TABLE OF CASES

TABLE OF PRIMARY LEGAL SOURCES

TABLE DIAGRAMS

INTRODUCTION

In the early twenty-first century, the defence sector in the US is on the cusp of significant transformation, particularly due to the growing advances in autonomy and artificial intelligence (AI).[1] There are already robotic weapon systems which are able to select and engage targets without any human intervention. In the academic literature and policy debate many authors argue that what is at stake is whether the determination about the release of force will be made by the machines. These authors postulate that the use of autonomous weapon systems (AWS) should be guided by a certain significant level of human control at the point of force engagement.

The US is the first country to have adopted a policy AWS. In November 2012, Ashton Carter, then Deputy Secretary of Defense for Policy of the US Department of Defense (US DoD) released a policy on autonomy in weapons systems called Directive 3000.09. The Directive states that such weapons 'shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force'.[2] The term 'human judgment over the use of force' is, however, open to different interpretations. Even within the US Department of Defense (US DoD) there are various, often contradictory, interpretations of what human judgment entails. Even though the Directive 3000.09 has already endured for 10 years, some senior US military representatives incorrectly claim that the requirement of human judgment prohibits the use of AWS. For

[1] The US DoD officials stressed the importance of autonomy and AI in enabling warfighters to achieve a battlefield advantage. See US DoD, 'National Defense Strategy' (2022) 8,19.
[2] Directive 3000.09 Autonomy in Weapon Systems 2012 4(a).

example, in February 2021, Col Marc Pelini, the Division Chief for Capabilities and Requirements in the US DoD's Joint Counter-Unmanned Aircraft Systems Office, said:

> Right now we don't have the authority to have a human out of the loop. Based on the existing US DoD policy, you have to have a human within the decision cycle at some point to authorize the engagement.[3]

In April 2021, Gen Mike Murray, the four-star commander of Army Futures Command, said, 'Where I draw the line – and this is, I think well within our current policies – [is], if you're talking about a lethal effect against another human, you have to have a human in that decision-making process.'[4] Both statements are false. Directive 3000.09, which is the only US policy on AWS, does not prohibit an autonomous machine from realising a force against human targets. Such a weapon system should generally go through a detailed senior review process where the Chairman of the Joint Chiefs of Staff, the Under Secretary of Defense for Policy (USDP); and the Under Secretary of Defense for Acquisition, Technology, and Logistics should grant consent.[5] The requirement of 'appropriate levels of human judgment over the use of force' can be satisfied by *effectuating the intentions of commanders and operators* in the machine's programming and sensors, rather than by having direct human input at the point of target engagement.[6] Depending on various conditions, the requirement of human judgment may include exercising no direct human input at the level of engagement at all. In this respect, Directive's 3000.09 consideration of

[3] C. Todd Lopez, 'Defense Official Discusses Unmanned Aircraft Systems, Human Decision-Making, AI' (*Department of Defense News*, 3 February 2021) <https://www.defense.gov/News/News-Stories/Article/Article/2491512/defense-official-discusses-unmanned-aircraft-systems-human-decision-making-ai/>. accessed 26 December 2022.
[4] Sydney J. Freedberg Jr, 'Artificial Intelligence, Lawyers And Laws Of War' *Breaking Defense* (23 April 2021).
[5] Directive 3000.09 Autonomy in Weapon Systems Enclosure 3 (1).
[6] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (US Government 2018) CCW/GGE.2/2018/WP.4.

the human role over the use of AWS differs significantly from the principle of 'meaningful human control' (MHC) articulated by other institutions and individuals.[7]

Statements from US DoD representatives illustrate that Directive 3000.09 is poorly understood even among high-ranking members of the US Armed Forces. Academics also tend to make statements that are too far-reaching. For example, some authors credit the US DoD policy on AWS as the first legal exemplification of the requirement of human control by emphasising that there is no meaningful difference between the concept of human control and human judgment.[8] In later parts of this thesis, I consistently make a point that one should not conflate the concept of human control with the US DoD policy of human judgment over the use of AWS.

The general confusion over the legal limitations of AWS makes informed public debate difficult. Autonomous weapons are often portrayed as 'killer robots', and various policy groups have advocated for an immediate international treaty prohibiting their development and use. This is hardly surprising. As reflected in popular media from Isaac Asimov novels to the *Terminator* movies, people have long been worried about the possibility of autonomous machines ultimately turning on their human creators.

While the lethal use of AWS has been incidental so far,[9] autonomy in weapon systems is becoming more prevalent, and governments risk lagging behind in strategic adaptation to this 'new' technology. The US drone strikes in Afghanistan between January

[7] Dan Saxon, 'A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of "Appropriate Leveles of Human Judgment over the Use of Force"', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 201.
[8] Heather Roff, 'Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons' (2016) 4.
[9] Theorists and practitioners disagree whether lethal AWS [LAWS] already exist and to what extent they have been already used in the military operations. See Bonnie Docherty, 'Losing Humanity. The Case Against Killer Robots' (Human Rights Watch 2012) 978-1-6231–32408.

2012 and February 2013, although not yet autonomous, vividly illustrate the problem. Ben Emerson, special investigator for the United Nations Human Rights Council, expressed concerns that the US lethal drone actions may have violated international humanitarian law (IHL) because the engagement was indiscriminate.[10] *The Intercept* reported that, of 200 people killed, only 35 were the intended targets.[11] The use of autonomous unmanned systems and the removal of manual human control only exacerbates this risk. Further concerns arise regarding international stability. The use of self-learning algorithms that derive their choices at least partially from the environment increases the unpredictability of their behaviour, and potential misjudgements by a robotic weapon system may lead to an escalation of conflict.[12] Other concerns relate to a dangerous precedent that the operations of lethal autonomous weapons could set. An example of lethal drones again demonstrates such a case. After the US military drone operations, other state and non-state countries adopted the practice of targeted killings, including the Islamic State in Iraq and Syria.[13]

The question of what rules or regulations are required to govern, restrict, or even prohibit the use of AWS is one of the most pressing issues in the military policy of the early twenty-first century and the answers may greatly inform the legal, ethical, and political ramifications of future policy development and statecraft considerations. This DPhil thesis aims to provide more clarity about the first and, so far, the only structured attempt to regulate AWS. It addresses the question of how US DoD problematises the concept of AWS. In order to answer this question, my thesis applies a poststructuralist

[10] Ben Emmerson, 'Report of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism' (2014) A/HRC/25/59.

[11] Jeremy Scahill, 'The Assassination Complex' (*The Intercept*, 2015). <https://theintercept.com/drone-papers/the-assassination-complex/> accessed 26 December 2022.

[12] Jürgen Altmann and Frank Sauer, 'Autonomous Weapon Systems and Strategic Stability' (2017) 59 Survival 117–142.

[13] WJ Hennigan, 'Islamic State's Deadly Drone Operation Is Faltering, but U.S. Commanders See Broader Danger Ahead' (*LA Times*, 2017).<https://www.latimes.com/world/la-fg-isis-drones-20170928-story.html> accessed 26 December 2022.

governmentality approach to policy analysis derived primarily from Carole Bacchi's work. Bacchi developed a framework to study policy problematisations, focusing on questions such as: what is a problem represented to be in a specific policy? What presuppositions or assumptions underlie this representation of the 'problem'? How has this representation of the 'problem' come about? What effects are produced by this representation of the 'problem'? What is left out of the problem's representation, and how could it be questioned and disrupted. This DPhil thesis not only addresses all these questions in the context of the US DoD problem representation of AWS, but it also contributes to the refinement of Bacchi's model. The thesis's answer to the final question – how the US DoD conceptualisation of AWS can be questioned and challenged – invites not only to contests the existing problem representation, but also presents alternative ways to regulate AWS.

My unit of analysis, which is the major entity being analysed, is US DoD's policy on AWS between 2009 and 2022. This specific timeframe has been selected on the basis that, in 2009, US DoD has started to work internally on the formal document regulating the development and use of AWS, which culminated in Directive 3000.09 in 2012. I focus particular attention on the statements regarding the role of human involvement over the use of AWS. A unit of observation, or the entity at the level at which I collected most data, are US DoD laws, rules, and policy documents applicable to AWS, as well as communications of US DoD representatives. The document analysis is supplemented by an analysis of original primary qualitative empirical data based on 12 elite interviews with both drafters of the US DoD policy, weapon's operators, pilots, military lawyers, and US DoD contractors. In describing the genealogy of the US DoD problem representation of AWS and some of its specific effects, of such problem representation I have decided to focus on the United States Air Force (USAF) as a nested case study as the aerial branch of the US Armed Forces is arguably the most advanced adopter of AWS, and no other branch of the

military has been deployed in such a consistent, essential fashion across the range of potential conflicts.

I focus on the US DoD case study because the US is the first country that has formulated its own regulation and policy on AWS. In describing specific military administration practices, I focus on USAF because it allows me to examine in more detail how the concept of human involvement over the use of AWS is exercised in the decision-making process in the context of a specific type of military missions: air-to-ground missions. Airpower is on the brink of a major technological transformation due to advances in autonomy; a focus on USAF is thus a critical nested case study. Moreover, no other branch of the military has deployed autonomy in such a consistent and essential fashion across the range of potential conflicts.[14] USAF is thus a nested case study of the use of US Armed Forces military power, but in the concluding remarks I will discuss the generalisability of the findings from the USAF case study.

My approach does not simply describe the US DoD policy on AWS, but it also explores how US DoD 'constructs the problem' associated with the development and use of such weapons, and indeed how the government 'addresses' this particular problem construction. In other words, I explore the specific risks of such weapons that are of concern to US DoD and how these risks have influenced the 'remedy', or a certain course of policy action. The framing of this question assumes a poststructuralist approach to the analysis. 'The risk' or 'the problem' as such do not exist as 'given facts' waiting to be discovered. Rather, what constitutes a 'risk' or 'problem' is the result of contingent outcomes of a struggle between competing discourses which transform 'what is out there' (e.g. weapon

[14] Douglas Birkey, Lt Gen David Deptula, USAF (Ret.) and Maj Gen Lawrence Stutzriem, USAF (Ret.), 'Manned-Unmanned Aircraft Teaming: Taking Combat Airpower to the Next Level' (2018) 15 Mitchell Institute Policy Papers 2.

systems able to select and engage targets with direct human action) into a socially, policy, and politically relevant issue, e.g. legal permissibility of the use of AWS.[15]

Thus, an overall analytical socio-legal issue this thesis addresses is the relationship between certain types of social phenomena that are scrutinised, defined, and constructed as 'problems' by the government, and the role of legal norms and regulations designed as 'answers' to these problems. This thesis argues that the study of policy problematisation may not only result in a more detailed understanding of a specific policy; importantly, it can also open up a perspective that makes politics, understood as the complex strategic relations that shape lives, visible. In other words, the study of policy problematisation sheds light on the variety of institutions and actors that play a role in shaping and re-shaping a specific policy. It also sheds light on their assumptions regarding various matters, such as the role of geopolitical considerations or perceived views on the maturity of technology, which ultimately result in a particular problem construction. In this respect, the thesis illustrates another socio-legal issue: that is the role of 'non-normative aspects', such as military considerations, geopolitical ramifications, or technical feasibility, in the norm creation process. The US DoD policy on AWS is an example of such a normative act in which many of various non-normative aspects have played an important role. By spotlighting these factors that led to the establishment of Directive 3000.09 one can open the potential critical discourse of contesting the specific assumptions that led to the US DoD problem construction of AWS. This could in turn lead to alternative problem constructions and alternative policy measures to address these problems.

[15] Herbert Gottweis, 'Theoretical Strategies of Poststructuralist Policy Analysis: Towards an Analytics of Government', *Deliberative Policy Analysis: Understanding Governance in the Network Society* (Cambridge University Press 2003) 249.

The thesis also can be considered as a critical account of IHL. The dominant view is that IHL, also known as the laws of armed conflict (LOAC), regulates the conduct of war, and it has a significant impact on the parties' use of methods and means of warfare.[16] Yet my argument is that IHL in fact has limited reach. What matters more than IHL are established military practices, particularly belonging to a dominant military power like the US. While IHL is in part based on customary law, understood as a uniform and consistent practice in the relationship between nations that serves as evidence of a generally accepted law,[17] my thesis refers to a very different set of practices – not between countries, but *within* the military organisation of a single country. Established practices of an organisation – in this case, US DoD –could modify or even precede the general rules expressed in the provisions of the international law treaties. A good example is US DoD's longstanding policy requiring the legal review of the intended acquisition or procurement of weapon systems.[18] The policy is considered as a cornerstone of the weapons acquisition process, and the US Armed Forces generally adhere to the document by implementing it in various regulations.[19] IHL, in Article 36 of Additional Protocol I to the Geneva Conventions (AP I), also requires state parties to conduct a weapons review of new weapons and new means and methods of warfare. However, the US is not formally a party to AP I. Further, Art. 36 AP I, contrary to US DoD regulations, does not contain a subjective standard that needs to be fulfilled for its implementation.[20] As a result, very few states have acknowledged that

[16] Jakob Kellenberger, 'The Relevance of International Humanitarian Law in Contemporary Armed Conflicts' (Committee of legal advisers on public international law, 28th meeting Lausanne, 13-14 September 2004, 14 September 2004); ICRC, 'What Is International Humanitarian Law?' (2004).
[17] The Statute of the International Court of Justice 1946 Article 38.
[18] Directive 5000.01 The Defense Acquisition System 2003.
[19] Department of the Army Regulation 27-53, Review of Legality of Weapons Under International Law 1979; Department of the Navy, 'Secretary of the Navy Instruction 5000.2E, Department of the Navy Implementation and Operation of the Defense Acquisition System and the Joint Capabilities Integration and Development System'; USAF, 'Department of the Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities'.
[20] Anne Dienelt, 'The Shadowy Existence of the Weapons Review and Its Impact on Disarmament' *S+F Sicherheit und Frieden / Security and Peace* 128.

they have put in place a domestic weapons review procedure rendering international law not particularly impactful.[21] Established US DoD military practices can also function as de facto rules in the absence of any formally recognised international rules. An example is a policy pertaining to AWS, as states were unable to agree on any treaty regulating AWS and there are limited examples of practices of using AWS by states to constitute the basis of customary law. Thus, Directive 3000.09 provides only applicable rules in the absence of any IHL AWS-specific considerations.

The thesis's main contribution to the academic literature on AWS is twofold. First, by evaluating a US DoD policy on AWS through the lens of a poststructuralist approach, I present that their problem representation is not grounded in any 'objective social problems', but rather contingent on certain assumptions regarding the military conflict and the development of technology more broadly. I critically reflect on the alternative problem representations which could challenge the US DoD conceptualisation of AWS, and which could result in different policy responses relative to Directive 3000.09.

Second, my major contribution is an in-depth evaluation of the concept of human judgment, particularly in the relation to the concept of human control, and more generally, to the notion of 'control' in the US military. In public policy and academic discussions, the concepts of human control and human judgments are often conflated and considered as synonymous. By making these concepts more explicit, I clarify how I and other authors can challenge them by considering the US DoD conceptualisation.

[21] ICRC, 'A Guide to the Legal Review of New Weapons, Means and Methods of Warfare' (2006); Vincent Boulanin and Maaike Verbruggen, 'SIPRI Compendium on Article 36 Reviews' (2017) 1.

PART I – THEORY AND CONCEPTS

Chapter 1: Mapping the Debate and Situating my Research Question

This chapter introduces the academic and public policy debate on AWS and justifies the thesis's main research question within this context. The chapter is divided into two sections and a summary. The first section offers a critical take on the academic and public policy debate on AWS. I illustrate how the debate on AWS came about, how it gained traction, and why it matters. I present key arguments for and against AWS. I argue that the focus on the role of human factors in the use of such weapons is critical in this debate, but that there are few in-depth explorations of what constitutes the concept of human involvement over the use of AWS beyond the general policy description. In particular, there are few studies about how governments have unpacked such a concept (or similar ones) and how it has been shaped by different institutions, actors, or narratives.

The second section justifies the main research question of this thesis in the context of this debate. I explain why the focus of this thesis is on the US DoD problematisation of AWS, rather than focusing on other countries. I justify the emphasis on USAF practices in a nested case study and present the wider implications of this thesis. I argue that, by specifically focusing on how the US military administration problematises AWS, particularly in relation to the role of human factors in the targeting and engagement process, I provide more clarity on the concept of human judgment, and the application of it. By making this concept more explicit, I believe that this thesis contributes to the specific and

emerging academic debate about the operationalisation of human factors over the use of AWS.

1. A Critical Analysis of Debates on Autonomous Weapons Systems (AWS) and Human Involvement over the Use of Such Weapons

This thesis focuses on an important nascent technology development that is likely to have a profound impact on future military affairs and in the dealing of states more generally – the development of machine autonomy. There is now an ongoing academic debate about the development and use of weapon systems with a high degree of autonomy. The debate is not restricted to academic circles. In recent years, a lively debate has developed about the potential restrictions applicable to AWS under the United Nations Convention on Certain Conventional Weapons (UN CCW). In 2014, there was a first, informal meeting of experts to discuss questions relating to emerging technologies specifically, in the area of the lethal use of AWS. For reference, I will refer to such weapons as 'LAWS', while I will refer to AWS which may not necessarily have a lethal effect as 'AWS'. At the 2016 Fifth CCW Review Conference, the High Contracting Parties decided to establish a United Nations Group of Governmental Experts (UN GGE) on LAWS to meet in subsequent years with a mandate to assess key issues applicable to LAWS. The UN GGE specifically works on exploring the possible recommendations for the governance of LAWS. The initial debate was largely dominated by a call from a coalition of non-governmental organisations (NGOs) seeking to pre-emptively ban LAWS.[22] A wide media campaign supported with

[22] Bonnie Docherty (n 9).

academic papers has led many countries to support an outright ban of LAWS. In total, since the first UN conference on LAWS, 30 States have endorsed such a ban.[23]

Since the establishment of UN GGE, the debate over whether LAWS should be banned has started to gain more academic traction. More theoretical contributions have been also invited by Christof Heyns, then UN Special Reporter on Extrajudicial, Arbitrary and Summary Executions, who has argued that LAWS does not require a prohibition so much as moratorium on their development.[24] Interestingly, one of the issues that requires further understanding and clarification is the very notion of AWS, in particular the extent to which this is a class of entirely new, future weapons or whether weapons that already exist can be included in the broad definition of the term.[25] In the next subsection, I present the current approaches towards the definition of AWS put forward by various organisations during the UN GGE format and beyond, and I situate the US Government approach in this context.

1.1. Various Definitions of AWS

This subsection presents three categories of AWS definition and situates the US Government's definition in the debate. The US Government approach to AWS differs from many others by focusing on the nature of human-machine interactions over the use of such weapons, as opposed to AWS capabilities or their compliance with international law.

[23] Campaign to Stop Killer Robots, 'Country Views on Killer Robots' (Campaign to Stop Killer Robots 2020). HRW, 'Stopping Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control' (2020).
[24] Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (2013) GE.13-12776.
[25] UN GGE, 'Report of the 2017 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (UN GGE 2017) CCW/GGE.1/2017/CRP.1. Chair's summary of the discussion.

One of the main points of contention in the academic scholarship and policy debate is whether AWS represent a unique, new type of weapons and, if so, what the nature of that novelty is. If there is anything novel in such weapons, this must be related to autonomy. However, the scope and characteristics of this autonomy are uncertain.[26] Despite these challenges, in the wider academic literature, there are at least three different groups of AWS definitions.[27]

The first category contains definitions that are based on capability parameters. For instance, the UK Government definition regards AWS as a system that is 'capable of understanding a higher-level of intent and direction […] It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may be present'.[28] Such a definition primarily aims to delineate AWS from unmanned aerial vehicles (UAVs), i.e. drones, which are often wrongly conceived as autonomous. Drones are often described as unmanned systems, but this can be misleading, as they are in fact controlled by a human via remote control. AWS, by contrast, are closer to truly unmanned weapons, as they are neither inhabited by a human, nor under the direct control of a human operator.[29] Where the problem arises with these definitions is at the level of specification of what exact parameters AWS should have to be classified as autonomous. Sometimes parameters are set so high that no existing weapons is capable of meeting them. 'Such systems are not yet in existence and are not likely to be for many years, if at all,'[30] said the UK Ministry of Defence. One may then argue that such

[26] See the discussion on two models of autonomous targeting: The Generating Model and the Execution Model in A Leveringhaus, *Ethics and Autonomous Weapons* (Palgrave Macmillan 2016) 53.
[27] Vincent Boulanin and Maaike Verbruggen, 'Mapping the Development of Autonomy in Weapon System' (2017) 8.
[28] UK Ministry of Defence, 'Joint Doctrine Publication 0-30.2. The UK Approach to Unmanned Aircraft Systems' (UK Ministry of Defence 2017).
[29] Leveringhaus (n 26) 49.
[30] UK Ministry of Defence (n 28).

a definition is not particularly helpful in providing more insights into the current challenges related to the autonomy in weapon systems.

The second type of definitions considers AWS in relation to law. An example here is the definition put forward by the Government of Switzerland, which refers to AWS as 'weapons systems that are capable of carrying out tasks governed by international humanitarian law in partial or full replacement of a human in the use of force, notably in the targeting cycle'.[31] These types of definition are more political declarations intended to constrain any new developments within the contours of LOAC, but they are not particularly helpful for delineating the boundaries between AWS and any other type of weapons.

Finally, the last group of definitions consists of definitions that are articulated on the basis of the nature of the human–machine relationship. An example of such a definition is the one put forward by US DoD, which states that AWS are 'weapon systems that, once activated, can select and engage targets without further intervention by a human operator'.[32] The US DoD definition shifts the conceptual problem of defining AWS onto the relationship between human and machine. It is a different approach from other presented definitions as it steers the direction away from delineating AWS from other weapons in order to focus on the general problem of human and machine interactions, which is undergoing important changes due to advances in robotics and computer technology.

In my thesis I refer to the US definition of AWS as I am exploring how the US – rather than other countries – problematises the development and use of AWS. Following a poststructuralist approach, I do not, however, consider the US DoD definition of AWS as a 'fact' or 'objective reality', but rather as a discursive concept that should be read together

[31] 'Towards a "Compliance-Based" Approach to LAWS' (Government of Switzerland 2016).
[32] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

with wider US DoD assumptions about the place of autonomous machines within the US military strategy and the US's views about the current stage of weapons development. The next chapters explore this topic in more detail. That said, it is important to emphasise that the US DoD definition of AWS opens up an important *context* for the investigation of AWS specifically in relation to human factors, rather than in relation to their potential legality or inherent capabilities. Moreover, this focus allows us to analyse how the human-machine interaction is changing and what potentially novel challenges this generates for the established military controls applicable to weapon systems.

Regarding the existence of AWS, the US definition specifies that it applies to already existing autonomous weapons, not to a future phenomenon. What is controversial is whether there are already existing autonomous weapons that can be used for *lethal purposes*[33] – in other words, whether *LAWS* have already been used. Despite contradictory statements from US DoD,[34] one can nevertheless argue that such weapons are already in limited use. This is why the US policy provides a specific review procedure of such systems. Further, some US DoD drafters of the US policy, such as Paul Scharre, have also confirmed that LAWS exist.[35] Potential examples of such weapons in the US are loitering munitions such as AeroVironment Switchblade[36] or anti-ship missiles, e.g. AGM-158C,[37] as both weapons, once activated, can select and engage targets without further intervention by a human operator.

[33] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment' [2014] Duke University 2.
[34] For example, Col Marc Pelini, the Division Chief for Capabilities and Requirements in the DOD's Joint Counter-Unmanned Aircraft Systems Office believes AWS do not exist within US DoD, while Paul Scharre, one of the drafters of the Directive 3000.09 believes such weapons are already in limited use. See Paul Scharre, *Army of None* (W W Norton & Company 2018); C. Todd Lopez (n 3).
[35] Paul Scharre, *Army of None* (W W Norton & Company 2018).
[36] AeroVironment, 'Switchblade' <https://www.avinc.com/uas/adc/ switchblade/> accessed 26 December 2022.
[37] Lockheed Martin, 'LRASM' <https://www.lockheedmartin.com/en-us/products/long-range-anti-ship-missile.html> accessed 26 December 2022.

1.2. The Human-Machine Interaction Problem

The relationship between AWS and human factors, known otherwise as the human-machine interaction problem has attracted a considerable interest from both academics and policy-makers, in particular participants of the UN GGE meetings.[38] The debate concentrates on the notion of 'control' – that is, how to use a weapon system according to human intentions in a safe and predictable manner. The objective of retaining human control over weapon systems has been critical for militaries since the beginning of human conflict.[39] Many weapons, ranging from simple bows and arrows to more modern weapons, require skill and experience to be used safely and effectively. With growing developments in the military technology, concerns have arisen about ensuring that highly advanced weapons are used in an appropriate fashion. Over the recent decades, various factors have contributed to the reduction of *direct* control exercised by humans over weapons. This process has primarily been driven by the military operational gains associated with more autonomous weapons, such as increases in safety and efficiency, and potentially decreases in personnel and administrative cost. These factors are often also cited as the main arguments for the potential use of AWS. The next sub-section illustrates this point in more detail.

1.3.Operational and Ethical Arguments for the Autonomy of Weapon Systems

It is argued that delegating of some degree of control to machines is safer for the user of the weapons, who can achieve military objectives without risking his or her life directly. Arkin argues that it is unreasonable to expect humans to operate in a modern battlefield

[38] UN GGE, 'Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (UN GGE 2018) CCW/GGE.1/2018/3 5.
[39] Paul Scharre and Michael Horowitz, 'Meaningful Human Control in Weapon Systems: A Primer' (CNAS 2015) 5.

environment.[40] Historically, war fighters often had to be physically present to activate a simple weapon. The fighter sometimes had to wait many days in an inconvenient environment to detonate the explosive by way of pulling a trigger. This action was not only inconvenient, but also dangerous. Human control over many explosive devices was gradually automated by allowing detonation of a weapon when certain pre-set parameters were met, as illustrated by anti-personnel landmines.

Thus, a weapon system with greater autonomy, particularly in terms of target selection and engagement, potentially offers increased capabilities in force protection, and it further removes the risk for the operators of the weapon system and their soldiers.[41] This factor has largely influenced the development of various military technologies that reduce the role that humans play on the battlefield in order to increase efficiency and reduce the risk of harm.[42] Robots are better equipped than humans for 'dull, dirty, or dangerous missions,' such as the mission that exposes humans to potentially harmful radiological material or the mission that requires intense mental concentration and a high degree of situational awareness.[43] Further, weapons with greater autonomy can act as a force multiplier, that is fewer human fighters are needed to complete a specific mission, and the efficacy of each fighter is greater.

Beyond operational gains, authors such as Ronald Arkin argue that weapons with a greater autonomy should be also ethically preferable to human fighters.[44] According to

[40] Ronald Arkin, 'Lethal Autonomous Systems and the Plight of the Non-Combatant' (2013) 137 AISB Quarterly.
[41] Thompson Chengeta, 'Defining the Emerging Notion of "meaningful Human Control"' (2016) 49 NYU Journal of International Law and Politics.
[42] ibid.
[43] Jason DeSon, 'Automating the Right Stuff? The Hidden Ramifications of Ensuring Autonomous Aerial Weapon Systems Comply with International Humanitarian Law' (2015) 72 Air Force Law Review 85; US DoD, 'Unmanned Systems Integrated Roadmap FY2007–2032' (2007)..
[44] Ronald Arkin (n 40) 5. Ronald Arkin, 'Warfighting Robots Could Reduce Civilian Casualties, So Calling for a Ban Now Is Premature' [2015] IEEE Spectrum.

Arkin, a human being is the weakest point in the killing chain because humans are fallible to a greater extent than machines. The modern battlefield is now increasingly outpacing a human fighter's ability to make sound rational decisions in the heat of combat. Robotic weapons can thus eliminate many atrocities in the conduct of war.[45] It is therefore morally desirable that humans should delegate their control to machines during combat operations. This moral argument, however, has been heavily criticised by many ethicists and military experts.[46] I discuss their arguments in the next subsection.

That said, the emergence of more autonomous weapons has led to a shift in traditional control checks and balances, as many of the tasks and functions that would ordinarily be performed by humans are being outsourced to machines. The most controversial aspect of machine autonomy is autonomy at the level of so-called 'critical functions,' that is the targeting and engagement of a weapon system.[47] Various authors have argued that the use of such weapons poses many ethical, strategic, and potentially legal challenges beyond considerations of military advantage. Let us explore four main arguments.

1.4. Arguments Against AWS

First, some oppose the use of AWS, especially their lethal use, on ethical grounds. However, the majority of authors argue that (L)AWS undermine human dignity when deployed to engage with human targets.[48] Dignity is a complex matter but usually related to the apparent special status attributed to humans from which certain rights and duties

[45] Ronald Arkin (n 44).
[46] Leveringhaus (n 26) 90–94. HRW and others, 'Killer Robots and the Concept of Meaningful Human Control' (2016).
[47] ICRC, 'Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons' (ICRC 2016).
[48] Christof Heyns, 'Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: An African Perspective', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016).

arise.[49] In the context of engagement, the question is whether something morally valuable is lost when a machine replaces a human in the use of force. The problem is whether there is a morally relevant difference between (1) ordering a human agent to kill another human and (2) programming an autonomous machine, an artificial agent, to kill another human. Alex Leveringhaus argues that the replacement of human agency with artificial agency at the point of force delivery is not morally desirable because human operators are not *fully* morally engaged. This is because to be a fully morally engaged human means more than just respecting someone else's rights. It is also to act for reasons that are not entirely rights-based, such as recognition of a common humanity, a concern for the vulnerable or pity and mercy. Thus, the replacement of human agency with artificial agency leads to at least partial moral disengagement.[50]

Second, Robert Sparrow argues that AWS can generate situations in which no one can be held responsible for what a machine does.[51] He differentiates between the systems with a low level of autonomy and full AWS. In the case of a system with low levels of autonomy, the creator of such a system is responsible for its actions. At full autonomy, only that fully autonomous agent is responsible. There is, however, a 'grey area' between the two, where there is some degree of autonomy. Here, the autonomous agent is acting, not with the competence sufficient for full autonomy, but with sufficient competence to absolve the creator of such a system. On this basis, Sparrow argues that one can still hold human operators or designers responsible, but only 'at the cost of allowing that they should sometimes be held entirely responsible for actions over which they had no control'.[52]

---

[49] Paolo Carozza, 'Human Dignity', *The Oxford Handbook of International Human Rights Law* (Oxford University Press 2015).
[50] Leveringhaus (n 26).
[51] Robert Sparrow, 'Robots and Respect: Assessing the Case Against Autonomous Weapon Systems' (2016) 30 Ethics and International Affairs 93.
[52] ibid.

Third, strategic challenges involve situations with potential unintended consequences of using AWS on the battlefield.[53] A failure that causes a weapon to engage an inappropriate target may result in mass fratricide.[54] A fully autonomous weapon, without having a human-in-the-loop, might not be able to prevent friendly forces from much greater destruction until it exhausts its ammunition. Such weapons can be especially dangerous when hacked by adversaries.[55] As the complexity of the weapon systems increases, it becomes progressively difficult to verify the system's behaviour under all possible conditions, particularly on the dynamic battlefield. In particular, the application of autonomous systems in the command, control, and communications of nuclear weapons raises the question of whether the decision to use nuclear weapons will be determined by humans or by an intelligent system. A fully automated nuclear command and control system may also increase the risk of accidental nuclear war.[56] In fact, such a danger has already existed.[57] In 1983, a false alarm by the Soviet early-warning satellite reported the launch of intercontinental ballistic missiles from bases in the US. Fortunately, the alarm was suspected to be false by Stanislav Petrov, an officer of the Soviet Air Defence Forces on duty at the command centre of the early-warning system. Petrov decided to wait for corroborating evidence, none of which arrived, rather than relaying only on early-warning system. His decision prevented a retaliatory nuclear attack against the US, which would likely have resulted in an escalation to a full-scale nuclear war.[58]

[53] Paul Scharre, 'Autonomous Weapons and Operational Risk' (CNAS 2016).
[54] Scharre (n 34) 38.
[55] Scharre (n 34).
[56] James Johnson, '"Catalytic Nuclear War" in the Age of Artificial Intelligence & Autonomy: Emerging Military Technology and Escalation Risk between Nuclear-Armed States' [2021] Journal of Strategic Studies.
[57] David Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy* (Anchor; 1st edition 2010).
[58] ibid.

Fourth, while the prevailing view is that existing AWS are not unlawful *per se* in light of the main principles of humanitarian law and human rights, it is very likely that such weapons might be used unlawfully.[59] The main argument is that today AWS are merely able to select and attack a specific target from a pre-selected group of potential targets in restricted circumstances, while sensitive, context-dependent, and standard-based assessments of proportionality and military necessity still require human judgment.[60] The delegation of military tasks to highly autonomous robots may therefore increase the number of crimes committed in military operations.

Fifth and finally, the development and use of increasingly autonomous weapon systems will likely lead to significant military advantages. Thus, many countries are interested in pursuing these capabilities, accelerating the global competition. This, in turn could undermine stability and security among nations. Robert Jervis has observed that many times the competition between nations leads to a 'security dilemma', a concept according to which 'many of the means by which a state tries to increase its security decrease the security of others'.[61] While it is not evident that an increase in one state's security must come at the expense of another's,[62] the problem often comes in the second- and third-order effects that could develop when another state reacts to having its security reduced relative to that of another.[63] Security competition could then leave both states worse off than before resulting in worsening global stability. For example, drawing on Cold War lessons and extrapolating insights from the current military use of remotely controlled

[59] David Akerson, 'The Illegality of Offensive Lethal Autonomy', *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff 2013) 85.
[60] Marcello Guarini and Paul Bello, 'Robotic. Warfare: Some Challenges in Moving from. Noncivilian to Civilian Theaters', *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2012) 386. Noel Sharkey, 'Killing Made Easy', *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2012).
[61] Robert Jervis, 'Cooperation Under the Security Dilemma' (1978) 30 World Politics 169.
[62] Charles Glaser, 'The Security Dilemma Revisited' (1997) 50 World Politics 174.
[63] Paul Scharre, 'Debunking the AI Arms Race Theory' (2021) 4 Texas National Security Review.

unmanned systems, authors such as Jürgen Altmann and Frank Sauer argue that AWS are prone to proliferation, resulting in increased crisis instability and the risk of escalation.[64]

1.5. The Origin of the Concept of Human Control

In response to the above arguments, many theorists have argued that the central area of concern regarding the use of AWS is that they may lack the necessary human control safeguards in terms of the critical functions of targeting and engagement. Article 36, the British NGO, coined the term 'meaningful human control', a figure of speech to express the core element that is challenged by the movement towards greater autonomy in weapons systems.[65] Article 36 defined MHC as an organising principle in the following terms:

> [...] those who plan or decide on an attack have sufficient information and control over a weapon to be able to predict how the weapon will operate and what effects it will produce in the context of an individual attack, and thus, to make the required legal judgements.[66]

This concept has quickly gained significant traction among state representatives, academics, and others engaged in the debate on the limitations of AWS.[67] It has been also included in an open letter signed by the world's leading robotics AI researchers who called

---

[64] Jürgen Altmann and Frank Sauer (n 12).

[65] Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (Article 36 2013) <http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf.>. accessed 27 December 2022; Article 36, 'Autonomous Weapons, Meaningful Human Control and the CCW' (Article 36 2014) <http://www.article36.org/weapons-review/autonomous-weapons-meaningful-human-control-and-the-ccw/>. accessed 27 December 2022; Article 36, 'Killing by Machine: Key Issues for Understanding Meaningful Human Control' (Article 36 2015) <http://www.article36.org/autonomous-weapons/killing-by-machine-key-issues-for-understanding-meaningful-human-control/>. accessed 27 December 2022 Article 36, 'Key Elements of Meaningful Human Control' (Article 36 2016) <https://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf>. accessed 27 December 2022

[66] Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (n 65).accessed 27 December 2022.

[67] HRW and others (n 46). UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward' (UNIDIR 2014). The research on 'meaningful human control' has also spread outside the military circles. See 'Meaningful Human Control over automated driving systems' project at Delft University of Technology. <https://www.tudelft.nl/en/technology-transfer/development-innovation/research-exhibition-projects/meaningful-human-control/> accessed 27 September 2022.

on the UN to ban the development and use of 'killer robots', as they referred to robots without any human oversight.[68]

At the most primary level, the requirement for MHC develops from two premises: (1) that a machine applying force and operating without any human control whatsoever is considered unacceptable; and (2) that a human simply responding to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense.[69] The second premise is particularly controversial because it is here that the word 'meaningful' comes into play. Critics of the MHC concept argue that the term 'meaningful' is undefined and vague. Various alternative terms have been suggested over recent years, such as 'appropriate', 'necessary', 'sufficient' or 'effective'.[70] None of these terms, however, has yet received a similar traction to 'MHC'. Article 36 defended the notion of 'meaningful' because it is general rather than context-specific (unlike 'appropriate') and derives from an overarching principle rather being outcome-driven (unlike 'effective' or 'sufficient'), and it implies human meaning rather than something administrative, technical, or bureaucratic.[71] This understanding of MHC as a kind of overarching principle resonates particularly among human rights theorists, but it is difficult to translate it into legal rules and procedures to ensure potential compliance. Some theorists have argued that such a broad formulation and acceptance of MHC necessarily comes with a legislative void, which will be difficult to overcome, if the principle cannot be clarified or operationalised.[72]

---

[68] Future of Life Institute, 'Open Letter on Autonomous Weapon' <https://futureoflife.org/open-letterautonomous-weapons/> accessed 27 September 2020.
[69] Article 36, 'Key Elements of Meaningful Human Control' (n 65).
[70] ibid.
[71] ibid.
[72] Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53.

One of the objectives of this thesis is to fill this intellectual and empirical data void by specifically studying the US DoD problematisation of AWS and their approach to problem of human-machine interaction. One of the important elements of this study is the relationship between the US DoD conceptualisation of human involvement over the use of AWS relative to the concept of MHC which will be discussed in the subsequent chapters.

2. The Thesis Research Question and its Contribution to the Debate on Human Involvement over the Use of AWS

This thesis addresses the question of how US DoD problematises the concept of AWS. As discussed, to answer this question, my thesis applies a poststructuralist governmentality approach to policy analysis, derived primarily from Carole Bacchi's work. In Chapter 3, I will present my detailed framework to study a policy problematisation of AWS. In this section, I focus on the thesis's contribution to the debate on human involvement over the use of AWS and I justify the focus and boundaries of my research question.

2.1. Problematisation of AWS Reveals the US DoD Approach to Human Involvement

I track back the US DoD problematisation of AWS from Directive 3000.09, the 'entry text' that constitutes what is the problem with AWS and sets out the practice to addresses this particular problematisation. The key practice set out by Directive 3000.09 to address the risks associated with AWS is the requirement that such weapons 'shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use

of force'.[73] I refer to this requirement as the policy of human judgment. As argued earlier, some authors do not see much difference between the concept of human judgment and human control.

The potential semantic differences on how to understand 'control' and 'judgment', however, go to the heart of the problem – that is, the current and future role of human involvement over the use of increasingly autonomous weapon systems. This lack of understanding of what actually constitutes such human involvement has profound practical implications. States are not able to reach any consensus regarding a basic legal framework for the use of (L)AWS while the pace of technological progress continues. Without further academic and policy progress, UN members risk either imposing unrealistic measures or failing to establish any legally binding international agreement on (L)AWS. As the report from one of the UN GGE meetings on LAWS states:

> […] it would be useful to continue discussions on reaching shared understandings on the extent and quality of the human-machine interaction in the various phases of the weapons system's life cycle as well as clarifying the accountability threads throughout these phases.[74]

This study aims to contribute to this understanding. Over the next chapters, I argue that US DoD's requirement of human judgment is different from the concept of human control, especially MHC as earlier defined. I further argue that both concepts represent two alternative policy proposals to address the risks associated with the use of AWS. Promoters of human judgment claim that the lethal use of AWS can be legal and that such weapon systems can be controlled through a variety of technical and legal safeguards, not

[73] Directive 3000.09 Autonomy in Weapon Systems 4(a).
[74] UN GGE, 'Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 38).

necessarily by direct and manual human control.[75] In contrast, those who favour the concept of human control argue that the lethal use of AWS is illegal, as all weapon systems should be manually controlled by human operators according to the international law and the requirements of morality.[76]

A study of the US DoD problematisation of AWS allows us not only to provide an in-depth account of the policy of human judgment, but also allows to shed a light on interdiscourse[77] - that is, the critical relationship that a discourse of human judgment has to a discourse of human control. In that, I argue that the study of policy problematisation can make the concept of human judgment more explicit. In other words, it can present how it relates to the US military practices of using increasingly autonomous weapon systems. MHC and similar terms referring to certain human involvement over the use of AWS have been consistent key terms in debates concerning AWS, but their usefulness as a policy or law-making tool is limited due to a lack of clarity on what these concepts encompass. During the 2018 April UN GGE meetings, the chair provided a summary of terms that are used to refer to the human element over the use of AWS (Table 1).[78]

[75] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[76] HRW, 'Stopping Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control' (n 23).
[77] Jean-Paul Metzger, *Discourse: A Concept for Information and Communication Sciences* (Wiley 2019) 61–91.
[78] Indian Ambassador Amandeep Singh Gill, 'Chart 2 Consideration of the Human Element in the Use of Lethal Force; Aspects of Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (UN GGE 2018) 2.

Table 1: Various terms referring to human involvement in the use of AWS

Some Terms			
Maintaining	Substantive		Participation
Ensuring	Meaningful		Involvement
Exerting	Appropriate		Responsibility
Preserving	Sufficient	Human	Supervision
	Minimum level of		Validation
	Minimum indispensable extent of		Control
			Judgment
			Decision

While the concept of human involvement over the use of AWS has been formulated in several different ways, many academics and policy-makers do not appreciate differences between them and use these terms interchangeably.[79] Those authors who recognise the importance of these differences focus predominantly on defining the key terms (e.g. by describing whether human involvement should be called 'human control' or 'judgment') and content of these terms (e.g. describing elements of human involvement).[80] There is much less focus on the context within which this human involvement is and ought to be exercised, more specifically on who exercise control, how, and over what. The study of

[79] An example of the confusion between human judgment and human control can be found in one of Article 36 reports. Article 36, 'Key Elements of Meaningful Human Control' (n 65) 2–3.
[80] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (PhD Thesis, Vrije Universiteit Amseterdam 2019) 133 <https://research.vu.nl/en/publications/the-distributed-conduct-of-war-reframing-debates-on-autonomous-we>.

AWS problematisation not only focuses on content of the concept of human judgment, but also on who should exercise control and over what elements specifically. In this respect, it unpacks the concept of human judgment and helps to draw critical distinctions with the concept of human control.

This policy analysis is different from a more dominant and conventional top-down or 'programmed' policy approach whereby policies are proclaimed by senior policymakers and then made operational by specific administration through their more detailed rulemaking activity.[81] The advocates of a top-down policy approach might first study the high-level content of a policy on human judgment and then analyse lower-level rules and procedures introduced by the US DoD administration with a view to assess the feasibility of implementing this policy. Usually, a top-down policy approach is based on a positivist analysis, according to which policies are considered as more or less self-evident responses to 'objective social problems.' According to this approach, the problem of AWS is well defined, but policymakers may have different measures of how to address it. The promoters of a top-down approach might then study how the content of human judgment, i.e. specific rules and procedures steaming from this policy, differs from the content of human control.

In contrast, according to policy problematisation research, it is not assumed that 'problems' are objective, but rather governmental practices constructs 'problems' as particular kinds of problems through their problematisations. For instance, the increased use of autonomy in weapon systems has produced the problem of the increased risks of unintended engagements associated with their operational use. The way in which these 'problems' are constituted in turn shapes the administrative practices related to the use of weapons, e.g. by introducing measures such as software validation and verification to

[81] Paul Sabatier, 'Top-down and Bottom-up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis' (1986) 6 Journal of Public Policy 21.

mitigate unintended consequences. The critical task, therefore, becomes the need to interrogate the particular problematisations within policies and the problematisations that arise from administrative practices.[82] By asking questions such as what specific problem is constructed, how it is produced, and with what effects, I have made the US DoD policy on AWS – and specifically the role of human involvement over the use of AWS – more explicit. I do not simply study the content of the requirement of human judgment, nor do I focus on specific rules or procedures designed to implement this requirement. Rather, I explore how the US DoD problematisation of AWS has influenced the creation of new norms and practices that guide specifically how such weapons should be controlled and by whom. For example, one of these new norms is the creation of a new decision-making process doctrine within USAF that distributes the control over weapon systems engagement across various entities, individuals, and technologies, which are placed in at least two different locations – in the US base and at the war theatre. The concept of human judgment, therefore, should be read in conjunction with this new USAF doctrine and other norms which are considered as effects of the US DoD problematisation but may not necessarily be present in Directive 3000.09. I argue that these other norms allow us to shed a light on the requirement of human judgment, making the concept more explicit and operational.

Further, for top-downers, policy operationalisation is primarily an administrative challenge of rulemaking activity. Researchers typically ignore or downplay broader public objectives and political aspects of the policy process.[83] As a result, researchers have arguably given too much weight to the importance of regulators, while ignoring the fact that policies are often a struggle between various competing actors and that even after the

[82] Carol Bacchi, 'Problematizations in Health Policy: Questioning How "Problems" Are Constituted in Policies' (2016) 6 SAGE Open.
[83] Richard Matland, 'Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation' (1995) 5 Journal of Public Administration Research and Theory 145, 147.

formal adoption of a document, the policy is being re-shaped, and its effects may be different than expected. In other words, researchers often ignore or downplay the importance of the institutional setting and other aspects of the context within which policy operationalisation occurs.[84] It is thus important to emphasise that the study of policy problematisation focuses on the variety of techniques and actors that play a role in shaping and re-shaping the US DoD policy on AWS. In that it reveals *political* considerations in the policy process.

The introduction of the US DoD policy on AWS was a result of internal tensions, notably between the US Army and USAF.[85] As USAF has arguably become the most vocal supporter of autonomy among the US DoD's military branches, this thesis's particular attention is directed towards USAF legal rules and practices of targeting and engagement with AWS. Specifically, I ask what novel norms, if any, have emerged to address problems associated with the use of AWS. This critical examination allows us to investigate how the US DoD problematisation of AWS has transformed the established USAF decision-making process. Thus, I do not study a concept of human judgment in isolation, but as a part of a set of measures put in place to transform US DoD practices over the use of increasingly autonomous weapon systems.

2.2.US DoD Case Study and USAF as a Nested Case Study

My analysis of the US DoD problematisation of AWS is a single-case study design, where the 'case' is the US DoD policy on AWS. My unit of analysis, or the major entity being analysed, is US DoD's policy on AWS over 10 years' time between 2009 and 2022.

[84] Mark Imperial, 'Implementation Structures: The Use of Top-Down and Bottom-Up Approaches to Policy Implementation', *The Oxford Encyclopedia of Public Administration : 2-Volume Set* (Oxford University Press 2022).
[85] Dan Saxon (n 7) 195.

Specifically, my unit of analysis focuses on the statements about the problems associated with the use of 'autonomy' in weapon systems, and I place special emphasis on the role of human involvement over the use of AWS. A unit of observation – that is, the entity at the level on which I will collect the majority of my data – are US DoD policy documents applicable to AWS, i.e. Directive 3000.09; more general directives applicable to all weapon systems, i.e. Directive 5000.01 *The Defense Acquisition System*; Title 10 of the US Code in the section regarding weapons development and procurement; *The Law of War Manual*, a US DoD-wide resource for US DoD personnel – including commanders, legal practitioners, and other military and civilian personnel – on the law of war; and communications of senior representatives of US DoD presenting US policy on AWS at UN CCW and GGE meetings, where states discuss what regulatory controls are applicable to such weapons.

The most relevant public document is Directive 3000.09, *Autonomy in Weapon Systems*. In the absence of congressional or executive action, Directive 3000.09 is considered, among others by the Congressional Research Service (CRS), as the de facto policy of the US on the role of autonomy in weapon systems.[86] Scharre, one of the architects of Directive 3000.09, has said openly that it is an official policy in this subject.[87] While Directive 3000.09 was introduced in 2012 and represents a single US policy on AWS, the topic of autonomy in weapon systems has been a recurring theme since at least 2009 in US DoD. Thus, this analysis is not limited to Directive 3000.09, but it also includes other key US DoD documents that ultimately stem from the White House National Security Strategies.[88] These policy documents are US DoD National Defense and Military

[86] Congressional Research Service, 'Lethal Autonomous Weapon Systems: Issues for Congress' (2016) 2.
[87] Scharre (n 34) 89.
[88] The White House, 'National Security Strategy' (2010); The White House, 'National Security Strategy' (2015); The White House, 'National Security Strategy' (2017); The White House, 'National Security Strategy' (2022).

Strategies,[89] The Quadrennial Defense Reviews[90] and Unmanned Systems Integrated Roadmaps.[91] These documents, read in conjunction, represent the US Government policy on AWS.

As discussed, I have deliberately decided to analyse the US DoD approach to AWS three years before the publication of Directive 3000.09, as the problematisation of AWS did not start with the adoption of that document, but can at least be dated to the beginning of the process of drafting it in 2009. This focus is also different from a top-down policy approach, which fails to consider the significance of actions taken earlier in the policy-making process before the formal adoption of a policy.[92] Robert Nakamura and Frank Smallwood argue that the policy formation process gives important insights into various interests at play, as well as the degree of consensus among those pushing for change. An analysis that takes policy as given and does not consider its past history might miss key assumptions and discursive connections behind policy formulation.[93]

The analysis of US policy on AWS focuses also on the statements of US DoD representatives presenting official communications on the topic of AWS at UN meetings on LAWS between 2014 and 2022.[94] Since the opening of discussions in 2014, 15 meetings have taken place under the format of UN CCW and GGE, all formal exchanges in-person at the UN in Geneva. The inclusion of data from these UN CCW and GGE meetings is important, as these are the only global for a, for discussing the potential legal and regulatory

[89] US DoD, 'National Defense Strategy' (n 1); US DoD, 'National Defense Strategy' (2018); US DoD, 'National Military Strategy' (2015); US DoD, 'National Military Strategy' (2011); US DoD, 'Defense Strategic Guidelines' (2012).
[90]US DoD, 'The Quadrennial Defense Review' (2014); US DoD, 'The Quadrennial Defense Review' (2010)..
[91] US DoD, 'Unmanned Systems Integrated Roadmap FY2007–2032' (n 43); US DoD, 'Unmanned Systems Integrated Roadmap FY2011-2036' (2011); US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (2014).
[92] Richard Matland (n 83) 147.
[93] Robert Nakamura and Frank Smallwood, *The Politics of Policy Implementation* (St Martin's Press 1980).
[94] For the reference, the UN meetings on LAWS have been initiated in 2014 and have not yet been concluded.

measurers applicable to AWS, while their mandate has been confirmed by the US and other countries.[95] During these meetings the specifics of the US policy were also rendered explicit through being contrasted with what some other countries presented as being their policy or views on AWS, most notably countries supporting the prohibition of (L)AWS. I expand the analysis of publicly available documents with data based on my critical observation of the debates, as I attended selected UN GGE meetings as a delegated expert from academia.[96] My analysis is also supplemented by academic literature which discusses the US position on AWS, specifically how US DoD approaches the issue of human involvement over the use of AWS in official policies and guidelines. I also enrich the conceptual research with an analysis of original primary qualitative empirical data based on 12 elite interviews with drafters of the US DoD policy, weapons operators, pilots, military lawyers, and US DoD contractors.

In describing the genealogy of the US DoD problem representation of AWS and some of the specific effects of such problem representation, I focus particularly on USAF military practices. I therefore decided to focus on a nested case study of the USAF, that is the aerial warfare service branch of the US Armed Forces. USAF is one of six military service branches organised within the Department of the Air Force and one of the three military departments of US DoD. The focus on USAF is particularly appropriate with reference to my research design. While it is theoretically possible to analyse all branches of the US military, assembling detailed information for various military departments exceeds the scope of this thesis. My focus on USAF allows me specifically to examine the

[95] UN GGE, 'Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems' (2014) CCW/MSP/2014/3.
[96] I attended the following sessions: 9-13 April 2018, 27-31 August 2018, 25-29 March 2019, 20-21 August 2019, 7–11 March 2022 and 25–29 July 2022.

effects of the US DoD problem representation of AWS on the specific set of practices applicable to the use of such weapons.

This exploration aims to examine how the US DoD problematisation of AWS influenced USAF regimes of practices in the area of growing autonomy of weapon systems. By studying the USAF decision-making rules and practices, I shed a light on the role of human factors over the use of weapon systems and make the policy of human judgment more explicit and operational. In other words, I show how this requirement has become possible as a measure to mitigate the risks associated with the use of AWS and how it can be read in conjunction with the wider USAF concept of 'control'. In the exploration of the USAF practices, I focus on immediate close-air support (CAS) missions that require dynamic targeting processes. Dynamic execution assumes a responsive use of air assets to exploit enemy vulnerability that are likely of limited duration.[97] I explain in more detail the reasons for selecting this specific mission in more detail in Chapter 7, but at this stage it is worth stating that I am interested in the more generic types of air missions involving dynamic targeting, rather than in any specific ones. Further, I have also had to consider the limitations in the available documentation and sources.

In a nested case study, a unit of analysis are USAF service members' views about the decision-making process involving a high degree of autonomy in weapon systems. Again, I place a special emphasis on the role of human factors in that decision-making process. A main unit of observation are USAF rules relating to the targeting and engagement process described in *Air Doctrine Publication 3-60,* a document that has recently been substantially updated. These rules describe the process of selecting targets and matching the appropriate response to them, taking account of command objectives,

[97] USAF, 'Air Doctrine Publication 3-60 Targeting' (USAF 2021) 23.

operational requirements, and capabilities.[98] Targeting helps translate more general strategies into actions against targets by linking ends, ways, means, and risks.[99] I also analyse *Joint Publication 3-30, Joint Air Operations*, which provides principles and guidance for the conduct of joint air operations – that is, the use of capabilities/forces from joint force components, e.g. in missions which require coordination between air and ground forces.[100] Strategies of senior USAF leaders also provide some relevant insights, particularly in the context of the air service main priorities.[101] Further, documents, such as the USAF *Strategic Master Plan* and *The Air Force Future Operating Concept*, translate general objectives into more specific goals,[102] while reports from the USAF Chief Scientist provide USAF with a framework and roadmap specifically for the use of autonomous systems.[103] Furthermore, I supplement my analysis with relevant academic literature[104] and I critically examine my findings through interviews with selected representatives from USAF, including, but not limited to, pilots, drone operators, and military lawyers.[105]

It is important to note that US military airpower capabilities are not limited to USAF, but include the US Navy, the US Army, the US Marine, and Allied or Coalition

[98] ibid 3.
[99] ibid.
[100] The Joint Chiefs of Staff, 'Joint Publication 3-30, Joint Air Operations' (2019).
[101] Gen Charles Brown Jr., USAF, 'Accelerated Change or Lose' (USAF 2020). Gen Mark Welsh III, USAF, 'America's Air Force: A Call to the Future' (USAF 2014). Gen Mark Welsh III, USAF, 'The World's Greatest Air Force—Powered by Airmen, Fueled by Innovation. A Vision for the United States Air Force' (2013). Gen Mark Welsh III, USAF, 'Global Vigilance, Global Reach, Global Power for America' (USAF 2013). USAF, 'Air Force Strategic Environment Assessment: 2014–2034' (2015).
[102] Gen Mark Welsh III, USAF and Deborah Lee James, USAF, 'USAF Strategic Master Plan' (USAF 2015). Gen Mark Welsh III, USAF, 'Air Force Future Operating Concept' (USAF 2015).
[103] Greg Zacharias, USAF, 'Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future, Volume I: Human-Autonomy Teaming' (USAF 2015) AF/ST TR 15-01; Greg Zacharias, USAF, 'Autonomous Horizons The Way Forward' (Air University 2019).
[104] See, among others, Scharre (n 34). Katherine Chandler, *Unmanning: How Humans, Machines and Media Perform Drone Warfare* (Rutgers University Press 2020). Ingvild Bode and Hendrik Huelss, *Autonomous Weapons Systems and International Law* (McGill-Queen's University Press 2022).
[105] My sampling strategy is that I conducted interviews with 12 participants aged between 35-80. The criteria for inclusion are significant expertise in policy and practical issues pertaining to the US policy on AWS as well as in the domain of targeting and engaging with the aid of AWS or weapons with significant autonomous capabilities. The interviewees were senior members in the organizational hierarchy of the USAF, such as Lieutenant Colonels who served as military pilots.

airpower capabilities. Each service has air-to-ground capabilities that are often critical to successful airpower operations. Although the military services offer different capabilities, such as equipment or personnel, under US DoD's joint operations doctrine, they fight as one force. A Joint Forces Air Component Commander (JFACC) oversees all airpower in a specific campaign without regard for the service that owns a particular capability.[106]

The collection of data is primarily focused on the period ranging from 2009 to 2022. In the genealogical parts of Chapter 6 of my analysis, I study data sources from even before 2009 in order to explore how the problem of AWS has come about.[107] In this analysis, I use predominantly secondary academic literature that can be traced back to the 1930s, when the first UAVs were constructed and US DoD introduced the concept of 'remote control', which predates the concept of 'autonomous operations'.[108]

2.3.Why Focus on the US Military Administration?

I have decided to focus on the US because it is the first and the only country that has published its own policy on AWS. It has also consistently communicated the content of this policy through numerous public announcements by government officials, particularly in UN meetings.[109] The US Government has presented rather coherent views on the role of human factors over the use of AWS by referring to the concept of 'appropriate levels of

[106] The Joint Chiefs of Staff, 'Joint Publication 3-30, Joint Air Operations' (n 100).
[107] By 'genealogical analysis' I refer to Foucauldian analysis whereby the goal is to present that a given system of thought, for instance penal practices or indeed targeting and engagement practices were the result of contingent turns of history, not the outcome of rationally inevitable trends. See: Stanford Encyclopedia of Philosophy, 'Michel Foucault' (*Stanford Encyclopedia of Philosophy*, April 2003) <https://plato.stanford.edu/entries/foucault/>; Michel Foucault, *Discipline & Punish: The Birth of the Prison* (Vintage Books 1995). In the genealogical parts of my analysis, I rely primarily on the secondary academic sources. See e.g. Katherine Chandler (n 104).
[108] Katherine Chandler (n 104); Greg Zacharias, USAF, 'Autonomous Horizons The Way Forward' (n 103) 9.
[109] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

human judgment' over the use of force.[110] Further, the US Government approach has also gained significant traction among both policymakers and academics.

2.3.Why Focus on the US Air Force (USAF)?

I have already argued why a nested case study on USAF is justified in terms of my research design and its academic significance. This section examines why the focus on USAF is justified from a military perspective.

The reason for focusing on the USAF rather than on other military branches of the US Armed Forces is that airpower is on the brink of a major technological transformation due to advancements in autonomy, processing power, and collaborative information.[111] These advancements have led many military practitioners to believe that the major shift in airpower will be the removal of military pilots from operations. Authors such as Timothy Schultz argue that, during the history of the pilot-aircraft relationship between 1903 and 2017, machines have increasingly assumed tasks previously performed by pilots and that the technology will soon make their work obsolete.[112] In their strategies, the USAF senior leaders have voiced support for growing autonomy in weapon systems through the gradual removal of direct human control.

> The absence of an onboard human may not only reduce size, cost, and complexity – it can increase range, endurance, and performance. [...] Future unmanned systems will be more autonomous and will place less demand on critical and vulnerable communications infrastructure.[113]

[110] Directive 3000.09 Autonomy in Weapon Systems 4(a).
[111] Douglas Birkey, Lt Gen David Deptula, USAF (Ret.) and Maj Gen Lawrence Stutzriem, USAF (Ret.) (n 14) 2.
[112] Timothy Schultz, *The Problem with Pilots: How Physicians, Engineers, and Airpower Enthusiasts Redefined Flight* (Johns Hopkins University Press 2018).
[113] Gen Mark Welsh III, USAF, 'America's Air Force: A Call to the Future' (n 101) 19.

Further, the Defense Science Board (DSB), in their report on the role of autonomy, states: 'With proper design of bounded autonomous capabilities, unmanned systems can also reduce the high cognitive load currently placed on operators/supervisors.'[114]

Moreover, airpower is a core component throughout the spectrum of military operations. In recent years, this has been even more evident through the operations against non-state actors in regions such as the Middle East or Africa. Airpower is an essential tool for a range of potential military operations, from peacetime to a hypothetical nuclear exchange. No other branch of the military is deployed in such a consistent, essential fashion across the range of potential conflicts.[115]

The focus on USAF does not automatically imply that other branches of the US military are moving in the same direction, although a deep dive into the experience of one military branch can at least illustrate the trends in the important part of the US military administration – and ultimately what may define their operations in the future.

2.4.Wider Implications of the Thesis

The thesis focuses on US DoD rules and administrative practices, but it has much wider implications. Building on the critical analysis of how US DoD policy constructs the problem of AWS, what the assumptions are behind this policy construction, and what the effects of this problematisation are on targeting and engagement rules and practices, I do not only flesh out the US policy in more detail, but also explore potential alternative problem representations and thus alternative policy formulations relative to US DoD

[114] Defense Science Board, 'Report of the Task Force on the Role of Autonomy in DoD Systems' (2012) 1.
[115] Douglas Birkey, Lt Gen David Deptula, USAF (Ret.) and Maj Gen Lawrence Stutzriem, USAF (Ret.) (n 14) 9.

concept of 'human judgment'. Thus, this thesis's contribution is applicable to the wider discussion about the current and future role of human factors over the use of AWS.

There are also relevant military reasons for transferring US findings to other countries. The US is the country that has demonstrated the most visible, articulated, and perhaps successful military research and development efforts on autonomy.[116] It also has a track record of using highly automated or autonomous weapons.[117] Therefore, this thesis has wider implications because other countries, such as China and most of the nine other largest arms-producing countries, are following the US's footsteps and conducting research and development projects focused on autonomy.[118] Some of them also deliberate on rules specifically applicable to AWS. Given the US is a key member of the North Atlantic Treaty Organization (NATO) and has the largest[119] and one of the most innovative militaries in the world,[120] the US problematisation of AWS and their policy will most likely influence the approach towards AWS in other members of NATO and beyond – as can already be observed.[121] For example, the US DoD's joint doctrine, which presents principles that guide the employment of US military forces in coordinated and integrated action towards a common objective, has been so influential that NATO modelled its own allied joint doctrine development system on it.[122] Thus, the significance of a study of the US DoD problematisation of AWS transcends a domestic context: it has a global significance.

[116] Vincent Boulanin and Maaike Verbruggen (n 27) 94–97.
[117] Scharre (n 34).
[118] Vincent Boulanin and Maaike Verbruggen (n 27) 104.
[119] '2022 United States Military Strength' (*Global Firepower 2022*, September 2022) <https://www.globalfirepower.com/country-military-strength-detail.php?country_id=united-states-of-america>. accessed 28 December 2022.
[120] Dan Steinbock, 'The Challenges for America's Defense Innovation' (2014).
[121] The US DoD definition of AWS has gained a widespread attention and is used widely by academics and researchers alike. See Thomas Bächle and Jascha Bareis, '"Autonomous Weapons" as a Geopolitical Signifer in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies' (2022) 10 European Journal of Futures Research 1, 5.
[122] George E. Katsos, 'U.S. Joint Doctrine Development and Influence on NATO' 101 Joint Force Quarterly.

Several countries are currently working on robotic weapons with autonomous capabilities, and many have already expressed a statement in support of additional restrictions over the use of AWS.[123] A better understanding of the US DoD problematisation of AWS can support or challenge subsequent analytical work in other countries.

3. A Summary of the Chapter

In Chapter 1, I have justified why the US DoD problematisation of AWS focuses on the role of human factors over the use of AWS and why such a focus is critical in this debate more generally beyond the domestic US policy context. I have argued that a focus on the role of human involvement over the use of AWS stems from the US Directive 3000.09 definition of AWS, which shifts the conceptual problem of defining AWS onto the relationship between human and machine over the use of weapon systems. It is a different approach from other presented definitions of AWS as it steers away from delineating AWS from other weapons in order to focus on the general problem of human and machines interactions in weapon systems. I further argue that, while the concept of human involvement over the use of AWS has been formulated in several different ways by both policymakers and academics, the various suggested terms – such as MHC or human judgment – are often conflated and lack in-depth operationalisation. Authors who recognise the importance of defining the role of human-machine interaction over the use of AWS focus predominantly on defining the key terms (e.g. by describing whether human involvement should be called 'human control' or 'judgment') and content of these terms (e.g. describing elements of human involvement), but there is little focus on the context within which this human involvement is and ought to be exercised and, more specifically,

[123]Robert Trager and Laura Luca, 'Killer Robots Are Here—and We Need to Regulate Them' *Foreign Policy* (11 May 2022)..

who should exercise control, how, and over what. By exploring the US DoD problematisation of AWS, this thesis not only focuses on the content of the concept of human judgment, but also on who should exercise control and over what elements specifically. In this respect, it unpacks the concept of human judgment and helps draw critical distinctions with the concept of human control. I argue that both of these concepts are important in the debate on AWS as they represent alternative policy approaches to the use of such weapon systems. By making these concepts more explicit, my thesis contributes to the specific and emerging academic debate about the operationalisation of human factors over the use of AWS.

In Chapter 1, I have also argued that my focus on the US problematisation of AWS is justified in the context of AWS debate. I have decided to focus on the US because it is the first and the only country that has published its own policy on AWS. In the study of administrative practices, I put a special emphasis on USAF practices, as the air branch of the US military is arguably the biggest supporter of growing autonomy in weapon systems and has very recently updated their targeting and engagement doctrine. Airpower is also a core component throughout the spectrum of military operations in today's world. I have further argued that the implications of this thesis go beyond the US domestic context, as I not only flesh out US policy in more detail but also explore potential alternative problem representations, and thus alternative policy formulations relative to US DoD's concept of 'human judgment'. Thus, this thesis's contribution is applicable to the wider discussion about the current and future role of human factors over the use of AWS. Further, as the US has the largest arsenal of weapons and one of the most innovative weapon systems in the world, the US problematisation of AWS and their policy will most likely influence the approach towards AWS in other members of NATO and beyond.

Chapter 2: A Critical Exploration of the Academic Debate on AWS

This chapter further expands the considerations from Chapter 1 by focusing specifically on the critical exploration of the academic debate on AWS. It is divided into two substantive sections and a summary. The first section situates the question of what is the US DoD problematisation of AWS is in the context of academic scholarship and existing academic gaps. I specifically focus on gaps in policy, legal, and socio-legal studies. In the second section, I point to the costs of the current scholarly neglect of the issue of human involvement over the use of AWS.

In Chapter 1, I briefly presented the scholarship gap in the debate on AWS. I argued that there are few in-depth explorations of what constitutes the concept of human involvement over the use of AWS beyond the general policy description. Specifically, there is a scarcity of studies regarding how governments have unpacked such a concept (or similar ones) and how it is shaped by different institutions, actors, or narratives. In Chapter 2, I analyse why the literature gap persists and I present my contribution to the existing scholarship.

1. The Scholarship Gap: The Scarcity of In-depth Studies Concerning the Practices of Human Involvement over the Use of Autonomous Weapons

In this section, I situate the question of what the US DoD problematisation of AWS is in the context of existing academic gaps. I specifically focus on gaps in policy, legal, and socio-legal studies.

The academic debate on AWS has recently been the subject of growing interest and of numerous studies.[124] Armin Krishnan is one of the first authors to write a book solely dedicated to AWS.[125] He puts the development of a robotic military in historical context and discusses the legality of AWS according to international law and the question of their compatibility with generally accepted principles and customs of war. He also explores the ethical considerations underpinning the potential use of AWS, such as the problem of the 'moral disengagement' of humans, as robotic weapons can allow human soldiers to stand even further back from the action.[126] This argument has been further explored and refined by Alex Leveringhaus in his comprehensive account about ethics and AWS.[127] Leveringhaus provides a map of conceptual problems associated with LAWS and discusses the notion of responsibility gaps in depth. His main contribution, however, is a novel reason for rejecting the lethal use of AWS based on the Argument from Human Agency:

> There needs to be space in armed conflict where individuals can exercise agency and choice […] as the replacement of human agency with artificial agency at the point of force delivery is not morally desirable.[128]

In addition to Leveringhaus's more specific contribution to the debate, U. C. Jha has written an account similar to Krishnan's book in which he explores the main moral and legal

---

[124] See e.g. Ingvild Bode and Hendrik Huelss (n 104); Nehal Bhuta and Stavros-Evdokimos Pantazopoulos, 'Autonomy and Uncertainty: Increasingly Autonomous Weapons Systems and the International Regulation of Risk', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016). Stuart Russel, *Human Compatible* (Viking 2019). Louis Del Monte, *Genius Weapons: Artificial Intelligence, Autonomous Weaponary, and the Future of Warfare* (Prometheus Books 2018); Nehal Bhuta and others, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016). Austin Wyatt, *The Disruptive Impact of Lethal Autonomous Weapons Systems Diffusion* (Routledge 2022). Jai Galliott, Duncan MacIntosh, and Jens Ohlin, *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021).
[125] Armin Krishnan, *Legality and Ethicality of Autonomous Weapons* (Routledge 2009).
[126] ibid 127–128.
[127] Leveringhaus (n 26).
[128] ibid 89–117.

challenges of LAWS.[129] A helpful guide to the general state of autonomy has been written by David Mindell, who explores existing machine developments operating on land, in the air, under sea, in space, and in the military.[130] He specifically argues that the current debate should move beyond the 'the myth of full autonomy' – that is, 'the utopian idea that robots, today or in the future, can operate entirely on their own'.[131] In contrast, authors such as Nick Bostrom or Max Tegmark explore longer-term developments of growing automation more generally, including the potential emergence of 'superintelligence', a system that 'greatly exceeds the cognitive performance of humans in virtually all domains of interest'.[132] The topic of superintelligence has received some traction among military authors, and in particular the recent book by Louis Del Monte presents various scenarios where robotic weapons are wirelessly controlled by superintelligence or where robotic weapons with superintelligence are embedded as a part of weapons.[133] The most widely discussed book on AWS, *Army of None*, has been authored by Scharre, a defence expert who led the US DoD working group between 2009 and 2012 that produced the US DoD Directive 3000.09. Scharre sheds some light on current developments in the US military and discusses various military, strategic, and ethical considerations involved in the use of AWS.[134] Finally, and more recently, Ingvild Bode and Hendrik Huelss have written a comprehensive account of the international rules applicable to the development and use of AWS from a legal doctrinal perspective.[135] The analysis of these contributions reinforces

[129] U C Jha, *Killer Robots Lethal Autonomous Weapon Systems Legal, Ethical and Moral Challenges* (Vij Books India 2016).
[130] David Mindell, *Our Robots, Ourselves* (Penguin 2015).
[131] ibid 23–24.
[132] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014); Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Knopf 2017).
[133] Louis Del Monte (n 124).
[134] Scharre (n 34).
[135] Ingvild Bode and Hendrik Huelss (n 104).

the view that legal and ethical challenges have become the main themes in the debate on (L)AWS.

2. Situating my Thesis in the Literature Gap

Despite these recent publications, there are some significant gaps in existing literature concerning legal, socio-legal, and policy aspects of AWS. A useful way to illustrate these gaps is by alluding to a few distinctions. The first distinction refers to the division between the positivist and poststructuralist approach to policy analysis.[136] It is primarily grounded in the contested epistemic status of the 'real' – that is, whether we can speak about the 'facts' and 'objective structures' or rather only discursively 'constructed objects'. The second distinction is based on the various legal regimes that are applicable to the use of AWS. Either theorists study the concept of human involvement over the use of force by AWS through the lenses of IHL, or they focus primarily on domestic rules that are usually produced by the military branch of the government administration. The third distinction pertains to the use of empirical material. The research is either conceptual – that is, based on already present information on a given topic – or at least to a certain degree empirical, wherein at least some of the data is derived from actual experience rather than entirely from theory.

Drawing from these distinctions, my thesis applies a poststructuralist approach to policy analysis to study the US domestic administrative rules and practices applicable to the use of AWS. It is in part based on interview data that allows me to critically examine the findings I have generated from my analysis of public policy and legal documents. Such

[136] See Cris Shore and Susan Wright, 'Technologies of Governance and the Politics of Visibility', *Policy worlds: Anthropology and the analysis of contemporary power* (Berghahn Books 2011). Carol Bacchi and Susan Goodwin, *Poststructural Policy Analysis – A Guide to Practice* (Palgrave Macmillan 2016) 7.

a focus helps shed light on the answers to different questions about the concept of human involvement over the use of force besides those present in the existing scholarship.

By problematising US DoD's approach to AWS, I focus on the military administration and their practices. I study the US military administration's targeting and engagement rules to understand how this specific problematisation of AWS has arisen. I also ask what assumptions and knowledge underpin this problematisation and how various actors are implicated in problem representations and produced as specific kinds of subjects. This specific set of questions helps unpack and contextualise the content and context of the US DoD policy on AWS, spotlighting implicit assumptions behind the policy and ultimately opening up the critical development of alternative policy formulations. Further, this approach has also allowed me to explore the problem from a specific socio-legal angle by conducting a discursive analysis[137] with qualitative empirical data to study the concept of human control over the use of LAWS.

The next subsections further clarify why I have decided to situate my research in this specific area. Let us begin with the various approaches to policy analysis.

1.1.Gaps in Policy Studies

A dominant approach to a policy analysis of AWS is a positivist policy analysis according to which policies are considered as more or less self-evident responses to 'objective social problems.' According to this approach, policies differ with respect to the remedies offered

[137] The term 'discourse' functions in contemporary policy analysis and social theory with diverse meanings, even among poststructuralists. Following Foucault, discourses are understood as socially produced forms of knowledge that set limits upon what it is possible to think, write or speak about a specific social object or practice. 'Knowledge' in this context is not truth, but it refers to what is accepted as truth—and is considered as a cultural product. The objective thus is to critically scrutinize the 'knowledge' that constitutes the 'problems', 'subjects', and 'objects' within specific policies. See Carol Bacchi and Susan Goodwin (n 136) 35–38.

to a specific social problem, but the problem, once recognised, is settled, and conceived as 'objective'. In the context of AWS, theorists largely agree that the problem is the delegation of authority over life and death decisions to machines,[138] but they usually differ with respect to the potential remedies to this problem. Authors such as Noel Sharkey argue that AWS should be prohibited because they are indiscriminate by their own nature,[139] while theorists such as Ronald Arkin and others argue that AWS should be regulated with a particular emphasis on defining guiding principles for human involvement in the use of force.[140] There are also contributions supporting the argument that existing IHL can be interpreted and applied in a way that accommodates challenges associated with the use of AWS, including their lethal use.[141]

In contrast to the positivist approach, poststructuralism has occupied a less prominent position in the field of policy analysis.[142] The potential reason for this is that a positivist approach has a much longer tradition in academic circles; particularly over the last two decades one has been able to see the resurgence of this approach under the broad theme of 'evidence-based policy' (EBP).[143] According to EBP, various public phenomena such as military defence, social welfare, or health systems are facts, 'given' structures from which policy analysts, trained in specific analytical techniques, can derive evidence and apply it to inform policy creation and policy assessment.[144] EBP, at its core, is nevertheless a return

[138] Such broad formulation of the problem is usually further specified by arguing that the use of AWS could result in excessive risk or generate ethical problems. See Leveringhaus (n 26).
[139] Noel Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016).
[140] Ronald Arkin and others, 'A Path Towards Reasonable Autonomous Weapons Regulation' (IEEE Spectrum 2019).
[141] Charles Trumbull IV, 'Autonomous Weapons: How Existing Law Can Regulate Future Weapons' (2020) 34 Emory International Law Review 533, 533–594.
[142] Carol Bacchi and Susan Goodwin (n 136) 6.
[143] Susan Goodwin, 'Women, Policy and Politics: Recasting Policy Studies', *Engaging with Carol Bacchi Strategic Interventions and Exchanges* (University of Adelaide Press 2012) 33.
[144] Carol Bacchi, 'Problematizations in Health Policy: Questioning How "Problems" Are Constituted in Policies' (n 82) 2.

to the traditional, positivist approach whereby policies are considered as rational and objective solutions to 'discovered' problems. Knowledge derived from 'facts' is seen as uncontested, capable of being translated into policy under the framework of 'best practices'. In the context of AWS, one can particularly see the prominence of the EBP approach in the debate about the risk of using machine autonomy. For example, both Scharre and Sharkey argue that AWS raise novel issues of risk, and this is the objective problem that poses the important question for policy-makers of how to retain an effective degree of human involvement over a machine's behaviour.[145] While their policy responses differ,[146] the problem formulation is perceived to be grounded on the evidence-based analysis of trained military authors who understand the risk level associated with the use of such weapons.

Poststructuralists, in contrast, argue that we must conceptualise policy phenomena such as 'technology risk' as the outcome of a contingent and complex process of representation rather than as a 'given fact'. The problem of 'technology risk' is not simply 'out there' waiting to be discovered; rather, what constitutes 'risk' is the result of complex and contingent outcomes of a struggle between competing discourses which transform 'what is out there' into a socially, policy, and politically relevant issue.[147] Instead of treating risk as an objective fact, authors such as Foucault would rather explore 'problematising moments' of risk by identifying times and places where he detects important shifts in practices of identifying and measuring risk. Authors such as Mitchell Dean or Carol Bacchi would focus more on problematising a policy response by spotlighting contingent

[145] Scharre (n 34).
[146] Noel Sharkey argues for the prohibition of AWS, while Paul Scharre argues for the regulation of their use. See: ibid; Noel Sharkey, 'Automating Warfare: Lessons Learned from the Drones' (2012) 21 Journal of Law, Information and Science 140; Noel Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 International Review of the Red Cross 787.
[147] Herbert Gottweis (n 15).

assumptions and knowledge used to develop a specific notion of risk in relation to a weapon's autonomy.[148]

While a poststructuralist perspective has influenced scholars to reflect on notions of policy problems that are taken for granted in various social fields,[149] including in the defence sector,[150] the problem of AWS has not yet attracted significant interest. Notable exceptions are selected publications which use the poststructural apparatus from Bruno Latour and Lucy Suchman to challenge the dichotomy between human and machine in the use of AWS.[151] Similar arguments have been advanced through a feminist poststructuralist approach represented by authors such as Emily Jones and Mary Manjikian. Specifically, Jones challenges dichotomies such as autonomy–automation and human–machine as fixed concepts and focuses rather on different ways in which the human and the machine are interconnected.[152] There have recently been some preliminary efforts by academics to apply Michel Foucault and Bacchi's approach to shed light on the problem representation of AWS in the EU context, but they lack more detailed analysis and they refer only to few general statements regarding ethical principles.[153] This being said, there is a clear scholarship gap

[148] Carol Bacchi and Susan Goodwin (n 136); Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (1st edition, Pearson 2009); Mitchell Dean, *Governmentality: Power and Rule in Modern Society* (2nd edition, SAGE Publications 2010).

[149] Stephen Ball, 'What Is Policy? 21 Years Later: Reflections on the Possibilities of Policy Research' (2015) 36 Discourse: Studies in the Cultural Politics of Education 306; Stephen Ball, *Politics and Policy Making in Education* (Routledge 1990); Peter Miller and Nikolas Rose, 'Governing Economic Life' (1990) 19 Economy and Society.

[150] See Emily Jones, 'A Posthuman-Xenofeminist Analysis of the Discourse on Autonomous Weapons Systems' (2018) 44 Australian Feminist Law Journal 93; Mary Manjikian, 'Becoming Unmanned: The Gendering Of Lethal Autonomous Warfare Technology' 16 International Feminist Journal of Politics 48.

[151] See Matthias Leese, 'Configuring Warfare. Automation, Control, Agency', *Technology and Agency in International Relations* (Routledge 2019).

[152] Emily Jones (n 150).

[153] See Nicole Beltran, 'Artificial Intelligence in Lethal Automated Weapon Systems - What's the Problem?: Analysing the Framing of LAWS in the EU Ethics Guidelines for Trustworthy AI, the European Parliament Resolution on Autonomous Weapon Systems and the CCW GGE Guiding Principles' (Uppsala Universitet 2020) <http://uu.diva-portal.org/smash/get/diva2:1436188/FULLTEXT01.pdf>.See also the application of Foucauldian bio-politics to the topic of AWS:Fred Martin Jr., 'Technologies of Sovereign Power? Private Military Corporations, Drones, and Lethal Autonomous Robots - A Critical Security Studies Perspective' (Ohio University 2015) <http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1428937559>..

regarding the application of a poststructuralist approach, and specifically Dean's and Bacchi's approach, to the problem of AWS and human control in the US context.

1.2.Gaps in Legal Studies

The legal scholarship dedicated to the problem of AWS focuses mainly on the potential international legal boundaries in relation to the development and use of such weapons.[154] In this respect, authors usually follow a legal doctrinal approach. According to this approach, the restraints of armed conflict should be derived in accordance with a formal test of pedigree that informs which principles/rules qualify as legal principles/rules and which do not. If a principle/rule meets this test, then it is applicable as a binding constraint of armed conflict. A doctrinal approach is thus concerned with the validity of the legal constraints on the use of force. Particular constraints are valid only if their justifications in the form of rules and principles are clearly expressed in formal sources of law.[155] Therefore, the majority of authors focus on the issue of the potential compliance of AWS with the well-settled principles of LOAC: principles of proportionality, distinction, military necessity, and humanity.[156] Academic papers usually draw the same conclusion: AWS are not illegal per se under existing LOAC, but it is likely that they can be used in an illegal manner, particularly in their lethal use against human targets.[157] Current assessments are

[154] See Ingvild Bode and Hendrik Huelss (n 104). Thompson Chengeta, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law' (2017) 45 Denver Journal of International Law and Policy.
[155] In the jurisprudence such approach is usually referred as 'legal positivism.' See Herbert Hart, *The Concept of Law* (2nd edition, Clarendon Press 1994).
[156] See Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law* (Cambridge University Press 2012).
[157] See Kenneth Anderson and Matthew Waxman, 'Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can' (2013) 11 American University Washington College of Law Research Paper; Michael Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' [2013] Harvard National Security Journal. For alternative views, see Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision Making' (2012) 94 International Review of the Red Cross 687; Bonnie Docherty, 'Making the Case. The Danger of Killer Robots and the Need for a Preemptive Ban' (2016).

rarely particularly complex, starting from the legal ramifications of abstract principles and then exploring the main difficulties related to the compliance of AWS with each of them. In part, the problem is an apparent lack of empirical accounts of AWS.[158] Many authors do not subscribe to the view that AWS represent weapons that already exist. They claim that AWS are potential future weapons and thus the assessment of their compliance with LOAC is based on hypothetical scenarios – although it is worth noting that authors such as Sharkey argue that even existing weapons that are autonomous to some degree do not comply with the main principles of LOAC.[159]

This being said, there is a prevailing lag in understanding how the principles of LOAC are operationalised in practice, either by the military administrative apparatus or by states, where the use of increasingly autonomous robotic weapons is concerned. Some useful insights, however, can be derived from authors who do not focus solely on AWS, but consider broad experiences of war fighters with reference to the specific principles of LOAC.[160] Such examples of analysis, informed by military practices, can be particularly helpful in the context of AWS. For example, by studying the military practices regarding the role of human involvement over the use of force with AWS, one can shed light on the content and context of the general requirement of human control or human judgment.

What many authors fail to advance in their academic work is that the operationalisation of general legal principles or policies applicable to AWS and other advanced weapons may reveal potential tensions between the perceived effect of general

[158] See James Walsh, 'Political Accountability and Autonomous Weapons' (2015) 2 Research and Politics.
[159] Sharkey wrote on drone attacks in Pakistan and Yemen and their 'questionable compliance with LOAC.' See Noel Sharkey, 'Automating Warfare: Lessons Learned from the Drones' (2011) 21 Journal of Law, Information and Science. See also Noel Sharkey, 'Weapons of Indiscriminate Lethality' [2009] FIfF-Kommunikation; Noel Sharkey, 'Grounds for Discrimination: Autonomous Robot Weapons' (2008) 11 RUSI Defence Systems.
[160] Maj John Merriam, US Army, 'Affirmative Target Identification – Operationalising the Principle of Distinction for U.S. Warfighters' (2016) 56 Virginia Journal of International Law.

principles or policies on regulated actors and actual accounts of governing practices, which may sometimes differ from an intended normative effect.[161] In this context, a rare example is a study by Monica Hakimi and Jakob Cogan on the relationship between two codes about the use of force: 'a formal code', understood as the output of formal decision-making processes that regulates the use of force, and 'an informal code', understood as the output developed through the practice of states.[162]

The requirement of human control or judgment over the use of AWS has sparked considerable debate within academic and military circles. In public debate, the notion of MHC is increasingly recognised as a general legal principle of LOAC.[163] Many state representatives have expressed explicit support for the establishment or recognition of this principle. Former German Foreign Minister Heiko Maas said in 2019: 'We want to codify the principle of human control over all deadly weapons systems internationally.'[164] Countries such as Austria, Brazil, and Chile have formally proposed the negotiation of a 'legally-binding instrument to ensure MHC over the critical function of weapon systems'.[165]

Despite these statements, the argument that human control is already a *legal principle* of LOAC is far-fetched. Many key military powers such as the US, the UK, China, Russia, or Israel have not recognised this principle. Some countries have postulated such a policy, but only the US has adopted a policy on AWS, and its formulation of the role of

[161] See Colin Gordon, 'Afterword', *Power/Knowledge: Selected Interviews & Other Writings 1972-1977 by Michel Foucault* (Pantheon Books 1980); Mitchell Dean (n 148).
[162] Monica Hakimi and Jacob Cogan, 'The Two Codes on the Use of Force' (2016) 27 European Journal of International Law 257.
[163] Thompson Chengeta (n 41).
[164] Deutsche Welle, 'Regulate Killer Robots, Says Heiko Maas' *Deutsche Welle* (15 March 2019). accessed 28 December 2022.
[165] Government of Austria, Brazil, and Chile, 'Proposal for a Mandate to Negotiate a Legally- Binding Instrument That Addresses the Legal, Humanitarian and Ethical Concerns Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (2018) CCW/GGE.2/2018/WP.7.

human involvement differs from the concept of human control. In the following part of this thesis, and consistently with US legal terminology, I refer to the US requirement of human judgment over the use of AWS as a policy, but not as a legal principle.

This being said, there is little academic research on how these requirements of human judgment or control are made operational or put into practice by the military administration. There are only few studies about the operationalisation of human involvement over the use of AWS either with reference to international law or domestic law. Two notable exceptions are works conducted by Mark Roorda and Merel Ekelhof who focus on understanding the role of human factors across various stages of NATO's targeting process.[166] Their focus is, however, slightly different. First, while their research goes beyond IHL and considers targeting and engagement practices, it is still based on a positivist approach to law. Both authors elaborate the rules applicable to the targeting process and explore the military practices of targeting. Yet these authors tend to assume that rules and practices exhibit correspondence, while in fact various military practices may not necessarily closely follow stated rules and procedures. The authors, for example, do not provide any insights about whether there are any discrepancies between stated policy objectives, administrative procedures, and military practices. Ekelhof argues that the conduct of war relies on distributed control at various phase of the targeting cycle, yet NATO doctrine states that control is 'centralized'.[167] Ekelhof's justification is that the doctrine of centralised control is 'not an either-or proposition; rather it is a question of

[166] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80); Mark Roorda, 'NATO's Targeting Process: Ensuring Human Control Over and Lawful Use of "Autonomous" Weapons' (2015) 13 Amsterdam Law School Research Paper.

[167] Allied Joint Doctrine for Joint Targeting, AJP-3.9 2021 1–5. By 'centralised control' I refer to the doctrine which states that a commander is responsible for a direction, coordination, and specific use of forces on the battlefield. See Lt Col Alan Docauer, 'Peeling the Onion Why Centralized Control / Decentralized Execution Works' [2014] Air & Space Power Journal.

balance'[168] – sometimes control is more centralised, sometimes less depending on the mission. Ekelhof does not consider that the existing NATO doctrine may not necessarily reflect the underlying military practices and that, instead of correspondence, the empirical material reveals a disjunction between stated rules and military practices. This disjunction is important in poststructuralist studies as it serves as the source of criticism of the established rules and provides the ground for a potential alternative policy or legal arrangements. Second, Ekelhof and Roorda's research focuses on NATO targeting rules, but NATO as an organisation, has not yet formulated the policy of human involvement over the use of AWS. Thus, the operationalisation of their rules is a useful exercise, but less directly related to the debate on human control and judgment, particularly in comparison to the study of US rules and practices.

To restate, legal studies about AWS focus primarily on the potential compliance of such weapons with LOAC. Various concepts of human involvement over the use of AWS, especially the concept of human control, have attracted considerable attention as a potential legal principle of LOAC, but these general concepts have rarely been addressed from a poststructuralist perspective that considers how they are related to military targeting and engagement. Therefore, what it lacks is a socio-legal perspective – that is, the analysis of laws directly connected to the exploration of the social situation of specific military practices in a specific context, e.g. US practices of targeting with the use of autonomous capabilities of weapon systems.

It is important to emphasise that socio-legal research does not merely use empirical data alongside conceptual analysis. What is required is a specific conception of how law

[168] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 165.

can be related to 'a social context'.[169] For instance, while the concept of human judgment is a policy guideline, its significance is of a quasi-legal nature, and it can be revealed when explored in the context of specific military practices of targeting and engagement applicable to the use of AWS. In the exploration of this policy, one nevertheless has to adopt a specific methodology that sheds a light on the relationship between the policy, legal rules, and administrative practices.[170] In my thesis, I refer to a poststructural policy analysis which studies how certain military practices have led to a discursive production of problems and how policy responses, in turn, have affected these practices. This analysis provides a framework for studying how policies, rules, and administrative practices establish 'regimes of practices' within the US military administration. By 'regimes of practices', I refer to the mechanisms through which the government's problematisation of AWS is realised in the daily conduct of military administration. For example, the problematisation of the risks associated with the use of AWS has led to the emergence of certain new rules and practices that mitigate these risks (e.g. the assessment of AI capabilities of weapon systems), new entities within US DoD (e.g. The Office of the Chief Digital and Artificial Intelligence Officer [CDAO]), etc. Taken together, these rules, policies, and practices constitute AWS regimes of practices within US DoD.

This approach ultimately allows us to extract insights regarding the role of human involvement over the use of AWS. It goes beyond the positivist definitions of law and considers broader socio-legal factors that affect the content of a concept such as human judgment. Such research is justifiable in socio-legal terms because the aims of a socio-legal

[169] David Schiff, 'Socio-Legal Theory: Social Structure and Law' (1976) 39 The Modern Law Review 288–289.
[170] ibid 289.

approach focus on exploring the significance of law and policy in the creation of social contexts.

Such a perspective is, however, absent in the current literature on AWS and the role of human involvement over their use. While there are authors that derive insights from interviews with military practitioners, they often do not integrate these findings through a coherent methodology characteristic for socio-legal research.[171]

1.3.**Reasons for the Scholarship Gap**

Despite considerable academic progress in recent years, gaps in policy and legal studies persist in the context of AWS. I have already mentioned the potential reason for why the poststructuralist analysis might be less prominent in the academic scholarship on AWS. There are, however, three more general reasons to explain why these gaps exist. First, the concept of MHC (or any similar terms referring to human involvement over the use of AWS), has been articulated relatively recently and the debate has not had time to mature yet.[172] Second, there is a scarcity of available data, particularly in relation to the empirical study of military practices.[173] Third, the prevailing perception is that humanity has not been exposed to autonomous weapons technology, so the academic debate refers at most to a future, uncertain phenomenon.[174] All of these claims are only justified to a certain extent.

The concepts of human judgment or MHC are indeed relatively new. The concept of human judgment was formulated in 2012 and MHC in 2013, when the debate about

[171] See e.g. Scharre (n 34).
[172] UN GGE, 'Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems' (2016).
[173] Nathan Leys, 'Autonomous Weapon Systems and International Crises' (2018) 12 Strategic Studies Quarterly 49.
[174] ICRC, 'Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects' (2014) 7; James Walsh (n 158) 2. See also UK Ministry of Defence (n 28).

AWS was at a nascent stage. Over the following years, both concepts and similar terms gained more attention in the policy and academic debate. Philosophers have considered these concepts when exploring the questions of responsibility and accountability over the use of AWS,[175] while legal theorists have contemplated whether MHC or any other similar term denoting human involvement on the use of weapons is addressed in existing LOAC or international human rights law (IHRL).[176] Regardless of the relatively significant media coverage of AWS, the academic debate on the role of the human in the context of increasing autonomy in weapons is far from flourishing. The academic community still has little exposure to existing military practice regarding highly automated or autonomous weapons, while the effects of current weapons operations – including, for example, recent swarming drone attacks[177] or autonomous cyber weapons[178] – are poorly understood. Thus, there is a critical need for a better understanding of how AWS are controlled and the role of human involvement over their use. Such studies also have a practical importance in the light of UN meetings on the potential regulation of AWS.

Scarcity of data is always a challenge in the field of highly innovative technologies, particularly in the context of their use by militaries. Yet some progress has been made in recent years by researchers who have actively been engaged in the study of US military practices. A comprehensive overview of various types of weapons in the US has recently been presented by Scharre, who complements earlier efforts on the topic by Heather Roff

[175] See Filippo Santoni de Sio and Jeroen van den Hoven, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account' [2018] Frontiers in Robotics and AI; Leveringhaus (n 26); Robert Sparrow (n 51).
[176] See e.g.Thompson Chengeta (n 154); Kenneth Anderson and Matthew Waxman (n 157); Michael N. Schmitt and Jeffrey S. Thurnher, '"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict.' (2013) 4 Harvard National Security Journal 231.
[177] Maziar Homayounnejad, 'Autonomous Weapon Systems, Drone Swarming and the Explosive Remnants of War' [2018] TLI Think!
[178] Lucas Kello, *The Virtual Weapon and International Order* (Yale University Press 2017).

and others.[179] Moreover, experiences with highly automated or autonomous technologies are not confined to ahistorical abstraction. Depending on the definition of AWS, there are examples of highly advanced weapons which have sparked controversy regarding the limited role of humans.[180] At most, the data gap reduces the degree of certainty of claims, but does not prevent reasoned debate about partly observable but important problems.[181]

As has been alluded to earlier, the academic debate has also been hindered by the perception that humanity has not been yet sufficiently exposed to AWS, and in effect academics mostly write about the AWS as a potential future phenomenon.[182] The current debate is largely framed by a fear that AWS could represent a new, dangerous category of weapons, fundamentally distinct from existing weapons, and that they could therefore potentially represent new challenges for international and domestic legal regimes. Such voices have been consistently raised by some of the world's leading robotics and AI researchers, who have called for a ban on the development and use of these weapons.[183] This focus on the future weapons, usually associated with the discussions of the potential emergence of 'general purpose' or 'humanlike weapon systems', is unhelpful for advancing current research on the role of human factors over the use of highly automated or autonomous weapons. First, it is unclear whether general purpose or humanlike weapon systems technology will ever be developed, and it is unlikely to be developed in the near

[179] Scharre (n 34); Paul Scharre and Michael Horowitz (n 39). Heather Roff, 'Survey of Autonomous Weapon Systems' <ttps://globalsecurity.asu.edu/robotics-autonomy>. accessed 28 December 2022.
[180] E.g. in the 1980s, the US Navy deployed a loitering anti-ship missile (The Tomahawk Anti-Ship Missile) that could hunt for, detect, and engage Soviet ships on its own. See Scharre (n 34).
[181] Lucas Kello (n 178).
[182] See Kenneth Anderson and Matthew Waxman (n 157).
[183] Future of Life Institute, 'Open Letter on Autonomous Weapon' <https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/>.

future.[184] Second, overemphasis on the potential of humanlike weapon technology hinders the effects of existing autonomous capabilities of weapon systems. [185]

Therefore, despite the above discussed limitations – the novelty of the subject, the scarcity of available data, and an apparent lack of direct exposure to AWS – further research is much needed as there is a clear gap in the academic literature related to the lack of in-depth studies about the role of human factors over the use of AWS. The next section assesses the consequences associated with the current state of research.

3. Costs of the Scholarship Gap: Dominance of Popular Knowledge and a Lack of Guidance for Statecraft

States and other actors will continue developing autonomous capabilities in the weaponry, and it is likely that their use will accelerate in years to come. Therefore, the lack of in-depth academic studies on the subject may undermine the relevance of existing theoretical concepts to understand ongoing technology transformations. The consequences of this neglect are significant. Let us proceed with two: intellectual costs and the consequences for statecraft.

I refer to the intellectual cost as a cost of scholarly neglect. The scholarly neglect may not necessarily be associated with a lack of studies or minimal interest in the subject.

[184] Vincent Müller and Nick Bostrom, 'Future Progress in Artificial Intelligence: A Survey of Expert Opinion', *Fundamental Issues of Artificial Intelligence* (Springer 2016).

[185] The concept of general-purpose or humanlike weapon system is usually associated with the debate on general versus applied AI. 'General AI' is usually defined as the ability of a machine to perform 'general intelligent action, similar to human-like cognitive abilities. 'Narrow' or 'applied' AI is the use of software to accomplish specific problem-solving tasks by using advanced computational methods such as machine learning and a class of techniques called deep learning. See e.g. Dustin Lewis, Gabriella Blum, and Naz Modirzadeh, 'War-Algorithm Accountability' (2016) 18.

It can also be related to the lack of theoretical originality, or the state of academic inertia. This specific scholarly neglect can be observed in the debate on AWS. The level of popularity of the concept of human control does not seem to correspond to equally strong intellectual foundations.[186] This intuitively appealing principle is so popular that various theorists and many state representatives have explicitly declared their support for it and started to question the lawfulness of weapons that operate without such control. A relative lack of opposition has led some to argue that it is either a newly developed customary norm or a pre-existing, recently exposed rule of customary international law, already binding on all states.[187] However, this broad support comes at a significant cost – the lack of an in-depth understanding of what this requirement of human control (or similar) actually contains and requires. Thus, more research is needed to clarify how various states problematise AWS and consider the role of human involvement over the use of such weapons.

The second cost involves statecraft. A scholarly neglect of the military practices related to the use of AWS reduces the capacity of security and military studies for policy guidance. While some scholars openly belittle the significance of theory in policy creation,[188] the area of military studies and ethics have very direct implications for statecraft. If there are any calls to disregard the theoretical contributions in such a domain, it is only because current academic concepts do no longer match the pressing needs of reality.[189] Therefore, it is even more desirable to challenge existing concepts grounded primarily in positivist studies and to provide a novel theoretical foundation which responds

[186] Evan Ackerman, 'Lethal Microdrones, Dystopian Futures, and the Autonomous Weapons Debate' *IEEE Spectrum* (2017).
[187] Rebecca Crootof (n 72) 53–54.
[188] Stephen Walt, 'The Relationship between Theory and Policy in International Relations' [2005] Annual Review of Political Science 41–42.
[189] Lucas Kello (n 178) 35.

to the most urgent military and security issues. In the debate about AWS, one can observe how policymakers struggle with basic terms applicable to the subject, including with the very definition of AWS and whether such weapons exist.[190]

This thesis aims to fill the existing literature gap by focusing on the US DoD problematisation of AWS by studying the US DoD rules, policies, and practices applicable to AWS, and integrating new empirical data from semi-structured interviews with military practitioners. As a result, the thesis presents the US DoD approach to the role of human involvement over the use of AWS in more detail and allows critically engagement with alternative policy proposals to regulate AWS within certain institutional and political settings that are characteristic for a US context.

4. A Summary of the Chapter

In Chapter 2, I argued that there are few in-depth explorations of what constitutes the concept of human involvement over the use of AWS beyond the general policy description. Specifically, there is a scarcity of studies about how governments have operationalised such a concept (or similar) and how it is shaped by different institutions, actors, or narratives. I have argued that the reasons why the literature gap persists is that the concept of human involvement over the use of AWS has been articulated relatively recently, and the debate has not yet had time to mature. Further, there is a relative scarcity of available data, particularly in relation to the empirical study of military practices associated with the use of AWS, or more generally, weapon systems with autonomous capabilities. Finally, some scholars argue that humanity has not yet been exposed to AWS,

[190] For example, the Campaign to Stop Killer Robots, an influential organisation aimed at prohibiting LAWS argue that such weapons do not yet exist. See Bonnie Docherty (n 9).

so the academic debate refers at most to a future, uncertain phenomenon and is thus less attractive.

I have further argued that my contribution to the academic scholarship is threefold. First, my thesis is the first comprehensive effort to apply a poststructural policy analysis to the problem of AWS, and thus it highlights the problem from a different angle. Rather than following a dominant, positivist approach to policy analysis, according to which policies are considered as more or less self-evident responses to 'objective social problems', I focus on unpacking and contextualising the US DoD approach to AWS, spotlighting and challenging assumptions behind the policy construction and exposing the heterogeneity of alternative policy formulations. Second, my contribution enriches the existing debate by presenting a more in-depth focus on the rules and practices of the US military administration, specifically USAF, targeting and engagement practices in contrast to a more prevalent approach focusing on general legal principles and the potential (non)compliance of AWS with IHL.[191] Third and last, my thesis builds on and expands the theoretical research by integrating empirical findings from 12 semi-structured elite interviews with military practitioners, and I subsequently integrate empirical material into a coherent socio-legal methodology. The next Chapter 3 explores my methodology in more detail.

[191] See e.g.Ingvild Bode and Hendrik Huelss (n 104); Thompson Chengeta (n 154); Charles Trumbull IV (n 141).

Chapter 3: Methodology Considerations

This chapter aims to provide more clarity about the question of 'how' one can study the problematisation of AWS. It situates problematisation studies within a wider poststructural perspective influenced by Foucault's work. Specifically, I use Dean's analytics of government as a general framework for my research, while Bacchi's 'What's the Problem Represented to be?' (WPR) approach provides several specific questions that allow me to open up a policy to a critical poststructural exploration. In doing so, I link Foucault's, Dean's, and Bacchi's approaches and contribute to the theoretical discussion about the application and limitations of their concepts.

I argue that Dean's analytics of government offers a novel approach to studying AWS by focusing on the discourses that constitute the problem and spotlighting the potential tensions or discrepancies within the US military administration's practices of governing these weapon systems. Specifically, Dean's approach elevates regimes of practice – that is, various mechanisms, such as rules, policies, and practices – through which the government's problematisation of AWS is realised in the daily conduct of military administration.[192] Dean, following Foucault, is particularly interested in how governing takes place. Based on a Foucauldian analysis of the governing of 'penal systems', 'madness', and 'sexuality', Dean argues that other policies can be subjected to the same kind of analysis. The goal is to access the 'thought' in governing practices. Thought here is conceived, not as what goes on in people's heads, but as regimes of practice, i.e.

[192] Mitchell Dean (n 148) 27–28.

mechanisms that constitute the regulated objects and subjects.[193] Drawing from these insights, I have decided to apply Foucault–Dean concepts as frameworks to investigate the US DoD problematisation of AWS.

Bacchi's approach broadens Foucault and Dean's agenda, and it is instrumental in my thesis. Bacchi argues that all policy proposals rely on problematisations which can be opened up and studied to gain access to the implicit systems of assumptions that underpin a specific policy.[194] Drawing from Foucault, the point is neither to declare a position 'pro' or 'contra' a specific stance, nor to identify the 'real problem', but to explore 'the system of limits and exclusions we practice without realizing it'.[195] The objective is to challenge concepts that are taken for granted to determine how they have come to be through studying the heterogeneous processes that have gone into their making – that is, regimes of practices. The analysis of regimes of practices begins with the policy documents which provide 'entry texts', setting out a practice that relies on a particular problematisation. When describing what to do with specific practices, policymakers indicate what they think needs to change and thus what constitutes a discursive problem.[196]

I, therefore, begin my analysis with Directive 3000.09 on AWS and 'work backwards' to deduce how it has produced a 'problem' of AWS. Drawing on a Foucauldian genealogy approach and in line with Bacchi, I explore the history of delegating the use of lethal force to autonomous systems and how the growing autonomy of weapon systems has affected the role of human factors in the use of weapons. When exploring what effects are produced by the US DoD problem construction of AWS, I focus specifically on USAF

[193] Mitchell Dean (n 148).
[194] Carol Bacchi, 'Why Study Problematizations? Making Politics Visible' (2012) 2 Open Journal of Political Science 1, 5.
[195] John Simon, 'A Conversation with Michael Foucault' (1971) 38 Partisan Review 198.
[196] Carol Bacchi, 'Why Study Problematizations? Making Politics Visible' (n 194) 4.

regimes of practices in order to extract more detailed insights on the use of AWS. I explore a variety of instruments that are characteristic of military regimes of practices: principles, tenets, and tactics, techniques, and procedures (TTP).[197] All of these instruments establish what is called a military doctrine.[198] In the US military discourse, principles are established ideas of the US military based on its past experience and provide general guidance on the application of military force. They provide a basis for incorporating new ideas, technologies, and organisational changes. For example, one of such principles in USAF is 'unity of command', which ensures concentration of effort for every objective under one responsible commander.[199] Tenets provide specific considerations for the employment of airpower. For example, 'mission command' is an approach that empowers subordinate decision-making in accomplishing a commander's intent.[200] Further, military doctrine consists of TTP, which translates principles into specific weapon's applications.[201] Tactics describe the employment of specific weapons individually or in concert with other assets, to accomplish military objectives. Techniques are methods used to perform missions, functions, or tasks. Procedures are detailed steps that prescribe how to perform specific tasks. In my analysis of regimes of practices, I also refer to strategic recommendations, guidelines, and opinions of US DoD and USAF researchers about AWS. Although such guidelines and opinions do not comprise an official doctrine, they nonetheless provide many useful details regarding US DoD and USAF problematisation of AWS, and particularly the role of human factors in the use of such weapons. Finally, I

[197] USAF, 'Air Force Doctrine Document (AFDP)' (2021).
[198] ibid; Department of the Army, 'ADP 1-01 Doctrine Primer' (2019); The Joint Chiefs of Staff, 'JP 1-02, DoD Dictionary of Military and Associated Terms' (2020).
[199] USAF, 'Air Force Doctrine Document 1 (AFDD)' (2021) 11.
[200] ibid 11–12.
[201] The Joint Chiefs of Staff, 'JP 1-02, DoD Dictionary of Military and Associated Terms' (n 198).

enrich my studies with empirical material from semi-structured interviews with experts from US DoD and USAF.

1. A Poststructural Approach to Policy Analysis

This thesis explores the US problematisation of AWS from a poststructural perspective, influenced by the writings of the late Foucault's. Specifically, I use Dean's analytics of government as a general framework of my research,[202] while Bacchi's WPR approach provides several specific questions that allow me to open up a policy to poststructural exploration.[203]

Both Dean and Bacchi are theorists drawing on a Foucault-influenced, poststructuralists perspective to policy analysis. Poststructuralism is a critical continuation of structuralism studies initiated primarily by Claude Lévi-Strauss and Roland Barthes.[204] In an essay titled *Structural Analysis*, Lévi-Strauss summarised the core components of structural linguistics:

> First, structural linguistics shifts from the study of conscious linguistic phenomena to study of their unconscious infrastructure; second, it does not treat terms as independent entities, taking instead as its basis of analysis the relations between terms; third, it introduces the concept of system. […] finally, structural linguistics aims at discovering general laws, either by induction or […] by logical deduction, which would give them an absolute character.[205]

[202] Mitchell Dean (n 148) 1–74.
[203] Carol Bacchi and Susan Goodwin (n 136).
[204] Claude Lévi-Strauss, *Structural Anthropology* (Allen Lane 1968); Roland Bathes, *Critical Essays* (Northwestern University Press 1972).
[205]Claude Lévi-Strauss (n 204) 33–34..

There are thus four key tenets of structuralism. First, the overall structure of linguistic relations often operates at the level of taken-for-granted, unconscious mechanisms that are beyond the control of the speaking agents. Second, the meaning in language derives from the relationships of difference and similarity between terms, and not from the terms themselves. A third component of structuralism is the idea that the relations of difference and similarity form a system. The fourth and the last tenet is that structural analysis can help discover general laws with universal character. This final assumption of structuralism was attacked by poststructural criticism.

Poststructuralists such as Foucault take three main tenets of structuralism but reject the idea that we could discover general laws. Rather than discovering a universal structure of meaning, they focus on the ambiguities in the system of meaning. As Harcourt puts it, 'The idea is not to find regularity, but instead to probe what the "discovered regularity" could possibly mean.'[206] Poststructuralists explore how our taken-for-granted assumptions and established regularities have emerged under specific historical contingencies and under what conditions they become dominant knowledge. They shed light on the process of shaping the dominant narrative or of determining what is taken for granted by asking what institutions and practices specifically lead us to believe that certain discourses are 'true' or 'accepted' at a particular time. By emphasising a heterogeneity of practices, it becomes possible to approach taken-for-granted narratives as contingent and open to challenge and change.[207]

Thus, the focus on heterogeneity and contingency offers a novel approach to policy analysis. Instead of taking 'things' for granted, a poststructural policy analysis explores

[206] Bernard Harcourt, 'An Answer to the Question: 'What Is Poststructuralism?' (2007) 156 University of Chicago Public Law & Legal Theory Working Paper.
[207] Mitchell Dean (n 148); Colin Gordon (n 161).

how particular 'things' have been constituted and brought into being. For instance, such an analysis may undermine the dominant international narrative about AWS, portraying them as autonomous 'killer robots' by challenging the assumption that such weapon systems are in fact 'autonomous.'

By applying a poststructuralist approach, I argue that the study of policy problematisation might not only result in a more detailed understanding of a US DoD policy on AWS or of the role of human involvement over the use of AWS, but also, importantly, that it can open up a perspective that makes politics, understood as the complex strategic relations that shape lives, visible. In other words, the study of policy problematisation sheds light on the variety of institutions and actors that play a role in shaping and re-shaping a specific policy. It also sheds light on assumptions regarding various matters, such as the role of geopolitical considerations, which have ultimately resulted in the US DoD problem construction. I argue that, by spotlighting these factors that led to the establishment of Directive 3000.09, one can open the potential critical discourse up to contest the specific assumptions that led to the dominant problem construction and a specific problem response. This, in turn, is where the greatest value of a poststructuralist approach lies – its *critical potential*. By making the US DoD approach more explicit, one can formulate alternative problem constructions and alternative measures to address problems associated with AWS within a certain framework of assumptions that are characteristic for a US DoD context. In this respect, one can formulate some practical policy solutions which may gain political traction and contribute to a policy change.

The next sub-sections explore two key methodological concepts behind my research: Dean's analytics of government and Bacchi's WPR approach, and what is my contribution to the theoretical framework in the application of these concepts.

1.1. Dean's Analytics of Government as a Research Framework

Dean outlined his concept of analytics of government to study public policies from a distinctively Foucauldian, poststructuralist perspective. For him, the study of policies is the study of 'programs', 'structured courses of government action'. Public policies refer to government programmes which aim to steer human conduct towards achieve specific outcomes. In order to situate his approach in the wider studies about government, he differentiates between 'government as a conduct' from 'government as a conduct of conduct'.[208] The former perspective concentrates on the question of 'what', by exploring what the role of government in various governance systems is. In this context, theorists discuss, among other things, what scope government should have, what authority it should possess, or what limitations should be imposed on it with respect to its role in a society. The analytics of government, however, approaches the studies of government activity from a different angle as it focuses on the government 'conduct of conduct'.[209] Here the focus is on the 'how' question of government, by exploring how the government attempts to shape, with some degree of deliberation, various aspects of human behaviour according to a certain vision and for a variety of ends.[210] 'Government' conceived as the 'conduct of conduct'[211] is defined in the following way:

> Government is any more or less calculated and rational activity, undertaken by a multiplicity of authorities and agencies, employing a variety of techniques and forms of

[208] Mitchell Dean (n 148) 17.
[209] Michel Foucault, 'Subject and Power' (1982) 8 Critical Inquiry 777, 789.
[210] Mitchell Dean (n 148) 18.
[211] Michel Foucault, 'Subject and Power' (n 209).

> knowledge, that seeks to shape conduct by working through desires, aspirations, interests and beliefs of various actors, for definite but shifting ends with a diverse set of relatively unpredictable consequences, effects and outcomes.[212]

There are three important features of this definition. First, the studies of government as the conduct of conduct are primarily concerned with the exploration of a government's attempt to deliberatively steer human conduct. This assumption underpins public policy studies. The government's activity is based on the premise that human conduct is something that can be governed and turned towards specific ends through a deliberate and rational calculation.[213] The term 'rational' in this context refers to the attempt to bring any form of rationality to the calculation of how to govern.[214] While terms such as 'rational' or 'rationality' have broad and ambiguous meanings in the social sciences, here the term refers simply to any form of thinking in a fairly systematic manner which strives to provide answers about how things are and how they ought to be – that is, how we think of governing.[215]

Second, the studies of analytics of government seek to engage with how both 'governed' and 'governors' regulate themselves.[216] It does not end with the perspective of those who establish policy or law but includes the perspective of those who are the object of specific government intervention. Assume a new governmental programme introduces UAVs in the military arsenal. Since the establishment of this programme, a human operator of a drone can be constituted as 'a remote supervisor' rather than as an active pilot. While detailed legal rules outline his role and scope of responsibilities, he or she also re-examines

[212] Mitchell Dean (n 148) 18.
[213] Mitchell Dean (n 148).
[214] ibid.
[215] ibid 18–19.
[216] ibid 19–21.

his/her role and develops new practices, routines, and habits. Government as the conduct of conduct, therefore, introduces the examination of self-government: it extends to cover the way in which an individual problematises his or her own conduct.[217]

Third, the analytics of government emphasises that a process of governing is a reflective enterprise. Those who govern and are being governed exercise capacity of thinking and are able to critically analyse various governmental programmes and call into question a kind of knowledge and expertise that have been drawn upon to legitimate these specific programmes. A human operator of a drone may also question his or her actions or express critical thoughts or feelings related to his/her responsibilities. Foucault noticed that these 'practices of the self' may sometimes take the form of so-called 'counter-conducts' – that is, actions that call for 'different form of conducts', for instance by 'other leaders' or in pursuit of 'other objectives' or with 'other procedures and methods'.[218] For example, drone supervisors do not physically fly the aircraft, but rather sit in front of a screen and operate the aircraft remotely from a safe base in the US. The Air Force may not recognise them as professional, career-track pilots, but drone supervisors may feel that their work is equally relevant as that of traditional air pilots in contemporary conflicts. Thus, they may challenge the established procedures.

The assessment of various modes of thinking underpinning how the government apparatus govern others and themselves has been termed as 'governmentality studies'[219] and has subsequently been popularised by the post-Foucauldian scholarship.[220]

[217] ibid 19.
[218] Michel Foucault, *Security, Territory, Population* (Palgrave Macmillan 2007) 194–202.
[219] Michel Foucault, 'Governmentality', *The Foucault effect: Studies in Governmentality* (University of Chicago Press 1991).
[220] Mitchell Dean (n 148); Michel Foucault, 'Governmentality' (n 219).

Governmentality is thus, the organised way of thinking about, calculating, and responding to a problem, which addresses a specific type of human conduct.

There are two important caveats here. First, according to Foucault, there is no pre-given notion of problem; what constitutes a 'problem' is the result of various discourses. These discourses can include internal deliberations within the government as well as external communications framed by private sector organisations, NGOs, or influential individuals. The discourses are based on the expertise, vocabulary, theories, and other forms of knowledge that are available and situated in a specific time and space. This point has been further advanced by Bacchi, who argues that 'problematisations' – that is, the way the government constructs problems – are critical to understand how the government aims to shape human conduct. By exploring different accounts of the problem associated with the use of AWS, one can derive more insights about the 'organised way of thinking about, calculating and responding to such problem'. This is precisely why I examine how the problem of AWS is construed by the US policy on AWS, what presuppositions and assumptions underlie this representation of the 'problem', how this representation of the 'problem' has come about, and what effects are produced by this representation of the 'problem', among other things. As a result, the thesis strives to present a US DoD 'governmentality of AWS'.

The second caveat is that, for Foucault, the word 'government' had a broad meaning concerned with the modes of thought underpinning any kind of governance, including the management of a prison or household.[221] In this thesis, however, I follow a Dean–Bacchi approach based on a narrower meaning of 'government' as state government – that is, the central bureaucratic administration. This focus does not necessarily go against a

[221] Michel Foucault, 'Subject and Power' (n 209) 789–790.

Foucauldian spirit. In fact, both Dean and Bacchi build on Foucauldian genealogical writings which trace back the development of power in Western European societies. Foucault observed that, in the early modern period, the art of government became separated from the theory and practice of sovereignty and emerged as an autonomous domain of power. Dean refers to this phenomenon as a *governmentalisation of the state*.[222] The specific form of rule over things exercised earlier by the will of the sovereign has been increasingly replaced by a rule produced by the bureaucratic and administrative apparatus of the state. Dean builds on this point and places special attention on the complexity of relations between institutions that constitute the modern state. He challenges the view that the state is a unified actor, both in the domain of international relations and within its internal system of authority.[223] The analytics of government assumes that government is accomplished through multiple actors and agencies rather than a centralised set of state apparatus, but Dean's focus remains on state government, particularly the activity of administrative apparatuses.[224]

1.2. The Place of Bacchi's Policy Problematisations in the Analytics of Government

The analytics of government does not aspire to provide a coherent method of analysing the US DoD problematisation of AWS. Rather, Foucault and Dean provide some guidance in the form of 'characteristic moves',[225] 'propositions', or 'game openings.'[226] Policy

[222] Mitchell Dean (n 148) 36.
[223] ibid 34.
[224] ibid 36–37.
[225] ibid 37.
[226] Michel Foucault, 'Questions of Method', *The Foucault effect: Studies in Governmentality* (University of Chicago Press 1991) 73–74.

problematisations play a prominent role in the analytics of government and wider governmentality studies. Dean considers problematisation – that is, the action of calling into question some aspects of the 'conduct of conduct' – as the starting point of an analytics of government. Bacchi goes further and argues that policy problematisations are *central* to understand how governments create and sustain regimes of practice.[227] As discussed earlier, governmentality studies explore how government conducts of conduct shape human action and constitute regulated entities as 'governable subjects' through promoting certain behaviours deemed to be desirable. Hence, authors such as Dean or Gordon focus on specific public programmes to explore how particular components of a policy shape regimes of practice. Bacchi accentuates more how the process of policy problematisation promotes, facilitates, and attributes various capacities and qualities to particular agents. She is particularly interested in 'modes of problem formation' which enable us to explore which of these problem representations indicate transformations or consolidations of specific regimes of practice.[228]

Bacchi's approach, similarly to Dean's analytics of government, is deeply rooted in Foucauldian writings on governmentality and his genealogical method. Foucault uses the term 'problematisation' to describe how and why certain phenomena become a problem.[229] For him, 'thinking problematically' is to examine how these phenomena are 'questioned, analysed, classified and regulated' at 'specific times and under specific circumstances'.[230] Studying how certain phenomena emerge in the process of problematisation brings their taken-for-granted status into question and allows us to trace the relations, 'connections,

[227] In the context of the state government, Dean uses the term 'regimes of government'. See Mitchell Dean (n 148) 28..
[228]Carol Bacchi and Susan Goodwin (n 136) 45. Nikolas Rose, Pat O'Malley and Mariana Valverde, 'Governmentality' (2006) 2 Annual Review of Law and Social Science 83, 88.
[229] Michel Foucault, *Language, Counter-Memory, Practice: Selected Essays and Interviews* (Cornell University Press 1977) 185–186..
[230] Roger Deacon, 'Theory as Practice: Foucault's Concept of Problematization' (2000) 118 Telos 127.

encounters, supports, blockages, plays of forces, strategies on so on' that result in their emergence as objects.[231] In Foucault's writings, concepts such as 'punishment', 'madness', or 'illness' do not exist as fixed essences; rather, they 'become', they 'emerge' as objects for thought in discursive practices. He describes 'the problematisation of madness and illness arising out of social and medical practices' and 'the problematisation of crime and criminal behaviour emerging from punitive practices.'[232] Similarly, the US DoD problematisation of AWS did arise from certain (mal)practices associated with the use of weapon systems with autonomous capabilities, specifically from the use of Patriot missile defence systems in an autonomous mode in Iraq in 2003. I explore this problem construction in detail in Chapter 4.

The 'problems' defined by public policies and associated with specific practices only come to be that way when they become part of a discourse. Post-Foucauldian authors focusing on the modern state challenge the assumption that before a policy intervention there is a discovery process that uncovers 'real' social problems.[233] Rather, specific practices are targeted and described by various actors and become an object of interest for policy-makers. Policy proposals, according to post-Foucauldian scholars, are prescriptive texts, setting out a practice that relies on a particular problematisation.[234] When describing what to do with specific practices, policy-makers indicate what they think needs to change and thus what constitutes a discursive problem. As discussed earlier, Bacchi argues that it is possible to take any policy proposal and 'work backwards' to deduce how it produces a 'problem'. Bacchi shares Dean's view that, to study problematisations, it is useful to open

[231] Michel Foucault, 'Questions of Method' (n 226) 73–86.
[232] See Michel Foucault, *Discipline & Punish: The Birth of the Prison* (n 107); Michel Foucault, *Madness and Civilization: A History of Insanity in the Age of Reason* (Vintage Books 1988).
[233] Rob Watts, 'Government and Modernity: An Essay in Thinking Governmentality' (1994) 2 Arena Journal 103, 116.
[234] Carol Bacchi, 'Why Study Problematizations? Making Politics Visible' (n 194).

them up for analysis by separating first the implied 'problem' – what is seen as being in need of 'fixing' – from a policy or policy proposal. A policy is designed as a response to a particular problem, and it characterises the problem in a specific, legal, or quasi-legal discourse.

By adopting and refining Dean's and Bacchi's approach, my thesis traces back the discussion of policy-makers in the US and during the UN GGE and explores why the use of AWS has become US DoD object of scrutiny. Directive 3000.09, the US DoD policy on the use of AWS, is the starting point of analysis. I focus specifically on exploring why Directive 3000.09 requires that autonomous and semi-autonomous weapon systems 'shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force'.[235] The approach follows Foucault's suggestion that 'practical' or 'prescriptive' texts provide entry points for identifying problematisations.[236] The analytics of government approach uses 'entry texts', such as Directive 3000.09, to open up reflections on the forms of governing, and associated effects, instituted through a particular way of constituting a 'problem'. This being said, it necessarily involves familiarity with other texts that cover the same or related topics or circumstances.[237] Foucault described such documents as 'texts written for the purpose of offering rules, opinions, and advice on how to behave as one should'.[238] Therefore, I analyse policy in the context of discourse – that is, within a wider meaning system, where specific concepts are expressed in language, but their meaning is created through particular language uses, framing, and accompanying values.

[235] Directive 3000.09 Autonomy in Weapon Systems 4(a).
[236] Carol Bacchi, 'Why Study Problematizations? Making Politics Visible' (n 194) 4.
[237] Carol Bacchi and Susan Goodwin (n 136) 18.
[238] Michel Foucault, *The Use of Pleasure. The History of Sexuality (Vol 2)* (Viking 1986) 12.

Policy is about meaning creation, but it is also often anchored in wider public debate, in which various texts influence each other. For instance, other documents of US DoD, USAF, and publications by NGOs or academic experts can heavily influence the government's policy proposal. From an intertextual perspective, meaning is not an inherent property of words or isolated texts; rather, it emerges from relationships with other texts and contexts.[239] Intertextuality refers to the phenomenon that other texts are drawn upon within a text, which is usually expressed through explicit surface textual features such as quotations marks.[240] Sometimes, however, the influence is more profound and involves the whole language system referred to in a text. Such a phenomenon, referred to as 'interdiscursivity', is concerned with the implicit relations between discursive formations rather than the explicit relations between texts. In describing interdiscursive relationships, one is concerned with specifying what discourse types are used in the domain in focus, but also what relationships there are between them – whether, for instance, they build and expand on themselves, or whether they undermine each other. The concept of interdiscursivity draws attention to the potential heterogeneity of discourses and spotlights the mechanisms through which a novel configuration of existing discourses emerges.

An example of such interdiscursivity is the debate between US DoD representatives and the Campaign to Stop Killer Robots (Campaign), a prominent coalition of think tanks and international organisations opposing AWS.[241] Their discourses deeply engage with each other, sometimes directly in the form of questions and answers at UN GGE meetings. The interdiscursivity is particularly evident regarding the role of human involvement over

[239] Norman Fairclough, 'Intertextuality in Critical Discourse Analysis' (1992) 4 Linguistics and Education 269, 272.
[240] ibid 273.
[241] Marta Kosmyna on behalf of the Campaign to Stop Killer Robots, 'Statement to the UN General Assembly First Committee on Disarmament and International Security' (2019).

the use of AWS, e.g. when US DoD representatives explained why the concept of 'human judgment over the use of force' is distinct from human control over the weapon.[242]

The exploration of problematisation is also associated with Foucauldian genealogy studies in the process of tracing back over time the specific developments that have contributed to the formation of identified problem representations.[243] The goal of genealogy is to upset any assumptions about the 'natural' evolution of the problem representation. By exploring specific points in time, one sees that the problem representation under scrutiny is contingent and hence susceptible to change. Genealogy thus has a destabilising effect on problem representations, which are often taken for granted.[244]

One of the criticisms of post-Foucauldian scholarship is that various aspects of Foucault's work are not easily transferable or applicable to other inquiries. Many authors tend to take Foucault's historically derived and therefore highly context-specific concepts as universal categories.[245] Colin Koopman and Tomas Matza argue that this is often the case with general Foucauldian concepts such as discipline, biopower, and security. These concepts are highly depended on the contexts in which they did originally develop, so one has to be mindful about importing them from Foucault's writings into one's own inquiries, which may involve very different contexts. For example, biopower in late Victorian England is surely different than biopower in the early twenty-first century.[246] On the other hand, some Foucauldian frameworks, or modes of analysis such as problematisations or genealogy studies are much more portable in their original form. It is worth stressing here

[242] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[243] Carol Bacchi and Susan Goodwin (n 136) 45–49.
[244] ibid 10–11.
[245] Colin Koopman and Tomas Matza, 'Putting Foucault to Work: Analytic and Concept in Foucaultian Inquiry' 39 Critical Inquiry 817, 819.
[246] ibid 824.

that, for Foucault, a problematisation is both an object of inquiry (that is, a study of what is problematic in today's world) and an act of inquiry (that is, an inquiry which renders seemingly neutral issues/phenomena more problematic).[247] This DPhil thesis adopts problematisation as an act of inquiry. This mode of analysis has been further clarified by Bacchi, who developed a useful framework to study problematisations in the form of six stages elaborated below:[248]

1) What is the problem represented to be in a specific policy?
2) What presuppositions or assumptions underlie this representation of the 'problem'?
3) How has this representation of the 'problem' come about?
4) What is left unproblematic in this problem representation? Where are the silences? Can the 'problem' be thought about differently?
5) What effects are produced by this representation of the 'problem'?
6) How/where has this representation of the 'problem' been produced, disseminated, and defended? How could it be questioned, disrupted, and replaced?

While the above framework is useful, I have refined it in the context of my research question. First, I have changed the order of Bacchi's questions. I have decided to focus first on the effects of the problem representation, and then to consider what has been left out of the problem representation, as the latter question uncovers a critical part of my analysis while initial four questions in the revised order are more descriptive in nature (see Table 2 below). Second, I have decided to combine a set of questions from (4) with a set of questions from (6) as I argue that the answer to the question of what is left unproblematic in a specific problem representation immediately invites a critical potential which should be further developed and expanded. Similarly, the answers to the question of whether the

[247] ibid 827–828.
[248] Carol Bacchi and Susan Goodwin (n 136) 2.

'problem' can be thought about differently naturally invites a reflection on how one can disrupt the dominant problem representation. Further, I do not also think it is particularly helpful for my research question to answer how and where the representation of the 'problem' has been produced, disseminated, and defended. This sub-question, listed in a point (6) by Bacchi, does not directly refer to the content of the problem representation, but rather to the ways in which this content has been communicated. Thus, a further sub-question from a point (6) inviting a critical analysis can be read as a follow-up to the question of how a problem has been communicated. Yet I do not think it is helpful to 'question' or 'disrupt' the problem representation predominantly in the context of how such a problem has been produced, disseminated, and defended. Rather, I believe that what has been left out of a problem representation as 'unproblematic' or 'silenced' in the problem construction as well as how the assumptions behind the representation have been framed, provides a more fertile ground for a critical contestation of the specific problem representation. Taking these considerations together, I have slightly revised Bacchi's framework in the context of my thesis, as illustrated in Table 2.

Table 2: The thesis's main research question and key sub-questions

The Thesis's Main Research Questions and Key Sub-Questions	
Main Question	What is the US DoD problematisation of AWS, specifically in the context of the role of human involvement in the use of such weapons?
A Descriptive Part	
Sub-Question 1	What is the problem represented to be in the US DoD policy on AWS?
Sub-Question 2	What presuppositions or assumptions underlie this representation of the 'problem'?
Sub-Question 3	How has this representation of the 'problem' come about?

Sub-Question 4	What effects are produced by this representation of the 'problem'?
A Critical Part	
Sub-Question 5	What is left unproblematic in this problem representation, and how could these omissions be questioned?

Further, in the next subsection, I discuss the role of empirical data in my application of Dean and Bacchi's concepts. After this, I present the key stages of my analysis, guided by Bacchi's questions. The application of Dean's analytics of government and Bacchi's WPR approach sheds light on the specificity of the US DoD's problematisation of AWS and specifically on role of human involvement in the use of such weapon systems.

1.3. The Significance of Empirical Data in my Research

It is important to clarify the role of empirical data in my application of Dean and Bacchi's concepts. The interview data collected from the current and former US DoD and USAF officials allows me to critically examine the findings I have generated from the analysis of the publicly available documents. Based on Bacchi's approach, however, I do not consider statements from the members of the military administration to be an explanation of the 'true' meaning of discussed issues. This assumption is at the heart of the difference between interpretivists and poststructuralists. The work of interpretivists stems from a hermeneutic tradition that considers people's self-interpretations and their interpretations of social phenomena as central to understanding a specific phenomenon, whereas Foucault-influenced poststructuralists support a post-humanist analysis that questions the existence

of an independent subject who can access 'true' meaning.[249] Thus, my collection and analysis of interview data was not driven by the motivation to understand the problem or policy construction by specific actors, but rather to interrogate governmental problematisations and how specific actors are constituted within them.

This being said, empirical data can play an important role in poststructuralist analysis. One of the limitations of my research is the lack of an ethnographic approach which could contribute information about the daily practices of military administration, specifically informal norms. Such data can perhaps help present 'the established ways of doing things' more accurately. Only by paying close attention to the day-to-day operations of US DoD and USAF individuals, particularly their interactions with weapon systems, can one identify their regimes of practice. The lack of an ethnographic analysis does not, however, hinder my exploration, as the purpose of the interviews I have undertaken is precisely to get a better understanding of the daily challenges of human–machine interactions related to the use of weapon systems with significant autonomous capabilities. My thesis also benefits from other empirical studies in the field of AWS, particularly in the context of USAF practices.[250] Further, in my research, I refer to ethnographic fragments of knowledge, such as my critical observation of the US delegation at the UN GGE meetings or data from interviews that extends the analysis of publicly available rules and guidelines.

2. How Do I Apply Bacchi's and Dean's Concepts?

[249] Carol Bacchi, 'Meanings of Problematization' (18 February 2018) <https://carolbacchi.com/2018/02/18/meanings-of-problematization/>. Accessed 31 December 2022.
[250] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80); Katherine Chandler (n 104); Madeleine Clare Elish, 'Remote Split: A History of US Drone Operations and the Distributed Labor of War' (2017) Vol. 42(6) Science, Technology, & Human Values 1100.

I apply Dean's analytics of government and Bacchi's WPR approach over the next five chapters to answer my overall research question: that is how does US DoD problematises AWS and the role of human involvement in the use of such weapons?

Chapter 4 tackles the thesis's first research sub-question. It begins the analysis by focusing on the exploration of what specific problem has been identified by US DoD policy on AWS and why. To identify a problem representation, I work backwards from Directive 3000.09 and other US DoD documents to explore what kind of practices related to the use of weapon systems with autonomous capabilities are scrutinised and I investigate why these practices have become an object of interest for the US military administration.

The the most important data source for Chapter 4 is US DoD Directive 3000.09, which I read together with other US DoD documents, including US DoD National Defense and Military Strategies,[251] Quadrennial Defense Reviews,[252] Unmanned Systems Integrated Roadmaps,[253] and the communications of US DoD representatives at the UN GGE meetings on LAWS in Geneva.[254] In addition, I explore various other documents, produced mainly by the Defense Science Board (DSB) and Congressional Research Service (CSR) related to the use of autonomy in weapon systems.[255] Interviews with the drafters of Directive

[251] US DoD, 'National Defense Strategy' (n 1); US DoD, 'National Defense Strategy' (n 89); US DoD, 'National Military Strategy' (n 89); US DoD, 'National Military Strategy' (n 89); US DoD, 'Defense Strategic Guidelines' (n 89).
[252]US DoD, 'The Quadrennial Defense Review' (n 90); US DoD, 'The Quadrennial Defense Review' (n 90)..
[253] US DoD, 'Unmanned Systems Integrated Roadmap FY2007–2032' (n 43); US DoD, 'Unmanned Systems Integrated Roadmap FY2011-2036' (n 91); US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (n 91).
[254] United Nations Office for Disarmament Affairs, 'Background on LAWS in the CCW' <https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/>; Reaching Critical Will, 'CCW Group of Governmental Experts on Lethal Autonomous Weapon Systems' <https://reachingcriticalwill.org/disarmament-fora/ccw>.
[255] Directive 3000.09 Autonomy in Weapon Systems; Defense Science Board, 'Summer Study on Autonomy' (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics 2016) 20301–3140; Defense Science Board, 'The Role of Autonomy in DoD Systems' (2012); Congressional Research Service, 'Lethal Autonomous Weapon Systems: Issues for Congress' (n 86); Congressional Research Service, 'Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems' (2020).

3000.09 and current US DoD representatives responsible for implementing this document were instrumental in reconstructing the US DoD problematisation of AWS.[256]

The Chapter 5 deals with the second thesis's research sub-question and it considers what presuppositions underlie US DoD representation of the 'problem'. The term 'presuppositions' refers to background 'knowledge', usually taken-for-granted and not questioned assumptions.[257] It is important to note that this analysis does not attempt to elicit the assumptions or beliefs held by policy-makers or identify their biases. Rather, the goal is to uncover deep-seated presuppositions that 'lodge within problem representations'.[258] In order to do so, one sometimes has to go deeper than the level of public discourse and investigate the opinions of key architects of Directive 3000.09 and other senior US DoD leaders. The examination explores possible patterns in problematisation that might indicate the operation of a particular 'governmental rationality'. By governmental rationality, I refer to the rationales produced to justify a particular problem and policy construction, to make 'some form of that activity thinkable and practicable both to its practitioners and to those upon whom it is practiced.'[259]

Similarly to the Chapter 4, important sources in Chapter 5 are publicly available US DoD documents regarding the use of autonomy in weapon systems and interviews with current and former US DoD representatives. In Chapter 5, I also use additional legal sources to place the US DoD's policy on AWS within the wider context of the US administrative law. In that, I explore the US Administrative Procedure Act (APA), the Attorney General's Manual on the APA, and selected judgments of the US Supreme Court.

[256] Interview with Paul Scharre, 'Interview' (5 February 2021); Interview with Shawn Steene, 'Interview' (12 January 2021).
[257] Carol Bacchi and Susan Goodwin (n 136) 4–5.
[258] ibid 5.
[259] C Gordon, 'Governmental rationality: an introduction' in G Burchell, C Gordon and P Miller (n 202) 3.

Chapter 6 deals with the thesis's third research sub-question. It examines how the application of a lethal force in an autonomous way does has come to be a policy problem for the US Government. The analysis draws on the Foucauldian genealogy approach and shows that the application of lethal force in an autonomous way, and the subsequent legitimisation of such a practice, has been the result of contingent turns of history, not the outcome of rationally inevitable trends. My examination focuses primarily on the evolution of practices within USAF because the air service branch of the US military has played the most significant role in this genealogy and air operations continue to be at the forefront of delegating lethal authority to autonomous machines.

As part of the genealogical examination, I study three types of source: US DoD and USAF documents, academic literature, and data collected from interviews with senior DoD and USAF officials. In terms of the US Government documents, I found many useful insights about US DoD thinking in the *Patriot System Performance* report, published by the DSB, and the commentaries on it by academics and policy theorists. In the part related to the evolution of doctrine within USAF, I studied Air Force doctrine documents, manuals, unmanned aircraft systems roadmaps, and commentaries published primarily by Air University. In the examination of the origins of delegating lethal authority to autonomous machines, I rely on several excellent academic publications, including Katherine Chandler's work on the genealogical evolution of drone warfare and Madeleine Elish's work on the history of US drone operations.[260]

The Chapter 7 tackles the thesis's fourth research sub-question. It explores what specific effects the problematisation of AWS has had on the US DoD and USAF regimes of practices. The question naturally invites a broad set of considerations. I have decided to

[260] Katherine Chandler (n 104); Madeleine Clare Elish (n 250).

limit the scope of investigation in two ways. First, as discussed earlier, I have decided to focus exclusively on USAF to present an in-depth study of at least one of the six US DoD military branches. Second, I have narrowed the analysis to the specific set of effects that the problem representation has had on *the emergence of norms* associated with the use of AWS. The sole focus on norm emergence is consistent with the objective of the thesis. My interest lies in a deeper understanding of the governmentality of AWS in US DoD, and particularly of how requiring a human element in the use of force is operationalised in the area of increasingly autonomous weapons.

In Chapter 7, I explore both historical and current USAF targeting doctrine in more detail.[261] In the study of USAF regimes of practices, I also relay on some academic literature that has generated relevant primary data on certain types of air operations.[262] Further, my own interviews with drafters of US DoD's policy, weapons operators, pilots, military lawyers, and selected US DoD contractors has provided me critical data to better understand the effects of the US DoD problematisation of AWS on the potential emergence of new norms applicable to the use of such weapon systems.

Finally, Chapter 8 tackles the thesis's two remaining research sub-questions. I explore what has been left unproblematic in the US DoD problem representation about AWS – in other words, the issues which have often been raised in academic or public discourse about AWS but which were not addressed in Directive 3000.09. For each issue, I reflect on how this specific omission could be questioned and ultimately on how it could disrupt the dominant US DoD problem representation. This critical part of the analysis presents my own 'self-problematisation' of AWS in the context of US DoD discourse. [263]

[261] USAF, 'Air Force Doctrine Document (AFDP)' (n 197).
[262] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80).
[263] Carol Bacchi and Susan Goodwin (n 136) 19, 38–41.

The rationale for this commitment to self-problematisation is grounded in a Foucauldian concept of 'reflexivity', which assumes that the author should also subject his or her own thinking to critical scrutiny within the historically and culturally entangled context of knowledge.[264] My self-problematisation of AWS and my normative declarations associated with the use of such weapons stem from the historically and culturally entangled context of knowledge that I have generated about the US DoD problematisation of AWS. The major sources in the Chapter 8, beyond the US DoD publications, are academic contributions.[265]

3. A Summary of the Chapter

In Chapter 3, I have explored the thesis's methodology in more detail. I have situated problematisation studies within a wider poststructural perspective influenced by Foucault's work. Specifically, I justify the use of Dean's analytics of government as a general framework for my research, and argue that Bacchi's WPR approach provides several specific questions that allow me to expose a policy to a critical poststructural exploration.

I have argued that Dean's analytics of government offers a novel approach for studying AWS by focusing on the discourses that constitute the problem and spotlighting the potential tensions or discrepancies within the US military administration's practices of governing these weapon systems. Specifically, I have made a point that Dean's approach illuminates regimes of practice – that is, various mechanisms, such as rules, policies, and practices – through which the US Government problematisation of AWS is realised in daily conduct of military administration. I have justified why Bacchi's approach broadens Foucault and Dean's agenda and serves as an instrumental framework in my thesis. Bacchi argues that it is possible to take any policy proposal and 'work backwards' to deduce how

[264] Carol Bacchi, 'The Issue of Intentionality in Frame Theory: The Need for Reflexive Framing', *The Discursive Politics of Gender Equality: Stretching, bending and policymaking* (Routledge 2009).
[265] Scharre (n 34); Leveringhaus (n 26); Ingvild Bode and Hendrik Huelss (n 104); Lucas Kello (n 178).

it produces a ‘problem’. She has developed a useful framework to study problematisations in the form of specific research questions which ask, among other things, what the problem is represented to be in a specific policy, what presuppositions or assumptions underlie this representation of the ‘problem’, how this representation of the ‘problem’ has come about, what effects are produced by this representation of the ‘problem’, and so on.

In applying Bacchi’s framework, I have modified her model in the context of my main research question. Specifically, I have separated a descriptive analysis of the US DoD problematisation of AWS from a critical part of my analysis. The first four research sub-questions aim to present the US DoD governmentality of AWS, while the final fifth question provides arguments regarding how the US DoD problematisation of AWS could be questioned, disrupted, and replaced.

In Chapter 3 I have also discussed the role of empirical material in my thesis. I have made a point about the importance of primary, empirical data in the context of the Dean-Bacchi methodology. In my thesis, I refer to ethnographic fragments of knowledge, such as my critical observation of the US delegation at the UN GGE meetings, as well as data from interviews with the current and former US DoD and USAF officials. Specifically, the interview data allows me to critically examine the findings I have generated from the analysis of publicly available documents.

PART II – US DOD GOVERNMENTALITY OF LAWS

Chapter 4: How is the Problem of AWS constructed in the US DoD Policy

This chapter tackles the thesis' first sub-question and explores what is the 'problem' is represented to be in the US policy on AWS. It is based on the assumption that, since all policies are problematising endeavours, they contain implicit problem representation.[266] As discussed earlier, the US policy on AWS is predominantly outlined in Directive 3000.09, which should be read in conjunction with other US DoD documents, including US DoD National Defense and Military Strategies,[267] Quadrennial Defense Reviews,[268] Unmanned Systems Integrated Roadmaps[269] and communications of US DoD representatives at the UN GGE meetings on LAWS in Geneva.[270] In addition, I explore various other documents, produced mainly by the DSB (a US DoD independent commission) and the CSR (a public policy research institute of the US Congress), related to the use of autonomy in weapon systems.[271] Interviews with the drafters of the Directive 3000.09 and the current US DoD

[266] Carol Bacchi and Susan Goodwin (n 136) 19.
[267] US DoD, 'National Defense Strategy' (n 1); US DoD, 'National Defense Strategy' (n 89); US DoD, 'National Military Strategy' (n 89); US DoD, 'National Military Strategy' (n 89); US DoD, 'Defense Strategic Guidelines' (n 89).
[268] US DoD, 'The Quadrennial Defense Review' (n 90); US DoD, 'The Quadrennial Defense Review' (n 90).
[269] US DoD, 'Unmanned Systems Integrated Roadmap FY2007–2032' (n 43); US DoD, 'Unmanned Systems Integrated Roadmap FY2011-2036' (n 91); US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (n 91).
[270] United Nations Office for Disarmament Affairs (n 254); Reaching Critical Will (n 254).
[271] Directive 3000.09 Autonomy in Weapon Systems; Defense Science Board, 'Summer Study on Autonomy' (n 255); Defense Science Board, 'The Role of Autonomy in DoD Systems' (n 255); Congressional Research Service, 'Lethal Autonomous Weapon Systems: Issues for Congress' (n 86); Congressional Research Service, 'Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems' (n 255). Defense Science Board Task Force, 'Fulfillment of Urgent Operational Needs' (2009) 20301–3140.

representatives responsible for implementing this document are instrumental in reconstructing the US DoD problematisation of AWS.[272]

The chapter is divided into three sections. The first section explores how the US Government constructs the problem of AWS. I argue that the key problem associated with the use of weapon systems with autonomous capabilities is their potential 'lethal use' – that is, particularly the use of such weapon systems to kill other humans. Specifically, the potential application of a lethal force in an autonomous way increases the risks of 'unintended engagements.' This increased risk relates to the growing sophistication of weapons, particularly the introduction of autonomous supervised weapon systems, known otherwise as 'semi-autonomous.' The use of such weapons has become a complex socio-technical system that requires trust and deep integration of human and automation factors. I argue that the notion of trust has a particular significance in addressing the risks of unintended engagements.

The second section explores in more depth various types of weapons which have not been qualified as 'lethal autonomous weapon systems' according to the US Directive 3000.09. Specifically, the section discusses potentially controversial features of semi-autonomous weapons such as the Phalanx Close-In Weapon System (Phalanx CIWS),[273] self-guiding anti-ship missiles such as the Long-Range AGM-158C (LRASM),[274] and loitering munitions. I argue that Directive's 3000.09 'legalisation' of their use is not without controversy in the wider academic literature.

[272] Interview with Paul Scharre, 'Interview' (5 February 2021); Interview with Shawn Steene, 'Interview' (12 January 2021).
[273] Raytheon, 'Phalanx Weapon System' <https://www.raytheonmissilesanddefense.com/what-we-do/naval-warfare/ship-self-defense-weapons/phalanx-close-in-weapon-system>.
[274] Lockheed Martin, 'Long Range Anti-Ship Missile (LRASM)' <https://www.lockheedmartin.com/en-us/products/long-range-anti-ship-missile.html>.

The third section explores two alternative policy responses towards the problem of AWS –US DoD and the Campaign, the prominent opponent of AWS. US DoD policy does not prohibit AWS, including their lethal use, but Directive 3000.09 introduces the requirement of 'appropriate levels of human judgment' in the use of such weapons. The Campaign, in contrast, proposes to ban AWS – irrespective of their lethal or non-lethal use – because they are beyond 'human control'.[275] In this section, I explicate how these concepts are different.

1. The Problem of 'Unintended Engagements' of AWS

Directive 3000.09 has become an important piece of regulation for two reasons. First, it has codified existing US military practices related to the use of autonomous systems since at least 1980s. Second, it was the first policy globally to address the issue of AWS, including the use of such weapons for offensive and lethal purposes.[276]

Regarding the existing US military practices, US DoD's policy on AWS delineates three types of robotic system that received the 'green light' for approval:[277] (1) semiautonomous weapons, such as homing munitions; (2) defensive supervised autonomous weapons, such as the ship-based Aegis weapon system or land-based air and missile defence systems such as Patriots; and (3) non-lethal, non-kinetic autonomous weapons, such as electronic warfare to jam enemy radars.[278] These three classes of autonomous system are in wide use today, and Directive 3000.09 was introduced to confirm the 'legality' of their use.[279] These weapons are subject to the usual US DoD acquisition

[275] Bonnie Docherty (n 9); Bonnie Docherty (n 157).
[276] Interview with Paul Scharre (n 256); Interview with Shawn Steene (n 256).
[277] Scharre (n 34) 89.
[278] Directive 3000.09 Autonomy in Weapon Systems 4.c.(1)-(3). Scharre (n 34) 89.
[279] Scharre (n 34). Currently, the US military fields defensive AWS, such as the Aegis at sea and the Patriot on land, both designed to defend against missile attacks.

rules and do not require any additional approval.[280] However, any future weapons that might use autonomy in a *novel way* outside those three types get a 'yellow light' and *must* go through an additional procedure. Those systems are subject to a lengthy, senior-level review process. This senior-level review requires the USDP, the Chairman of the Joint Chiefs of Staff, and either the Under Secretary of Defense for Acquisition and Sustainment or the Under Secretary of Defense for Research and Engineering to approve the system before formal development.[281] A potential novel way of using autonomy that falls outside the specified instances refers to any kind of weapon systems able to *apply lethal force in an autonomous way, in particular for offensive purposes*.[282] In other words, potential use of existing AWS against human targets as part of an offensive action could theoretically require additional approval.

It is worth stressing, however, that the difference between offensive and defensive weapon systems in academic literature and military practice is often blurry and that most weapons can be used for offence and defence.[283] Thus, one can argue that it is political strategies rather than weapon systems that can be either offensive or defensive.[284] Further, in US DoD, there is no official document outlining this distinction, and the particular classification of weapon systems depends on the label an administration chooses to give it according to the policies that guide its deployment.[285] Thus, one can argue that the distinction between offensive and defensive weapons is not particularly important for

[280] Directive 5000.01 The Defense Acquisition System.
[281] Directive 3000.09 Autonomy in Weapon Systems 4d.
[282] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment', *Lethal Autonomous Weapons* (Oxford University Press 2021) 274.
[283] John Mearsheimer, *Liddell Hart and the Weight of History* (Cornell University Press 1988) 36. John Mearsheimer, *Conventional Deterrence* (University of Chicago Press 1987) 25–26.
[284] Sean Lynn-Jones, 'Offense-Defense Theory and Its Critics' [1995] Security Studies 674.
[285] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment' (n 282) 276.

understanding the US DoD problematisation of AWS. Rather, the question is whether AWS use for lethal or non-lethal purposes triggers a potential senior review process.

While lethal AWS are not prohibited by Directive 3000.09, the policy introduced additional restrictions because of the higher degree of risk associated with the potential development and use of such weapons. The reason these restrictions are in place is the stated purpose of the policy: to 'minimise the probability and consequences of failures in autonomous and semiautonomous weapon systems that could lead to *unintended engagements*'.[286] 'Unintended engagements' are defined as 'the use of force resulting in damage to persons or objects that human operators did not intend to be the targets of US military operations.'[287] The most significant type of damage, according to the US DoD, is 'unacceptable levels of collateral damage beyond those consistent with the law of war, rules of engagement, and commander's intent'.[288] The US delegation to the UN GGE on LAWS provides an example of accidental attacks that kill civilians or friendly forces, which would be considered to be 'unintended engagements' under US DoD Directive 3000.09.[289] Failures are defined as 'an actual or perceived degradation or loss of intended functionality or inability of the system to perform as intended or designed'.[290] Directive 3000.09 states that failures can result from various causes, including human error or human–machine interaction failures, as well as from cyber-attacks.[291]

1.1.The Problem of Unintended Engagements is the Problem of Trust

[286] Directive 3000.09 Autonomy in Weapon Systems Enclosure 3.
[287] ibid Glossary Part II Definitions.
[288] ibid.
[289] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[290] Directive 3000.09 Autonomy in Weapon Systems Glossary Part II Definitions.
[291] ibid Glossary Part II Definitions.

While Directive 3000.09 does not provide more context for why AWS attracted the interest of US DoD policy-makers, other US DoD documents are more eloquent. The DSB report on autonomy is the key document outlining in detail how machine autonomy may offer the US Government a competitive advantage relative to other countries. The report states that the major challenge for broader adoption of autonomous technologies is the issue of trust.[292]

While the notion of trust is complex and multidimensional in philosophy,[293] the ambition of this thesis is not to argue for what trust is or can be, but rather to follow a genealogical approach and to explore how US DoD conceives of the role of trust. In the DSB report, we read about the 'need to build trust in autonomous systems while also improving the trustworthiness of autonomous capabilities'[294]. This sentence reveals that trust has a double significance for US DoD. First, in the socio-technical system of using autonomous weapons, there must be trust between humans and machines. Even if the weapon systems can act in an autonomous way, the use of such machine still occupies a certain place in the wider chain of military command, with human commanders who exercise general oversight known in military language as 'command and control' (C2).[295] Further, the deployment of AWS in any mission is executed in accordance with an order that specifies the mission objectives and methods, while human commanders are responsible for implementing this order. Thus, there is no completely autonomous AWS behaviour; rather, the use of such weapon systems can assist commanders and their staff to develop situational awareness and plan missions, and sometimes with the application of force. Autonomous weapons must then be designed in such a manner that humans can trust the system with respect to the missions for which they were designed. US DoD refers to a

[292] Defense Science Board, 'Summer Study on Autonomy' (n 255) 14–24.
[293] Paul Faulkner and Thomas Simpson, *The Philosophy of Trust* (Oxford University Press 2017).
[294] Defense Science Board, 'Summer Study on Autonomy' (n 255) ii.
[295] William Barker, 'Guideline for Identifying an Information System as a National Security System' (National Institute of Standards and Technology 2003) NIST Special Publication 800–59 8.

deep integration between autonomous systems and humans as 'human–machine teaming and collaboration'.[296] In the USAF report on the role of human–machine teaming (HMT) in airpower, we read the following: 'airmen [will] need to develop informed trust – an accurate assessment of when and how much autonomy should be employed, and when to intervene'.[297] While US DoD sees the military potential of HMT for war fighting missions, they argue that the implementation of autonomous capabilities will require significant changes in the current US DoD doctrine of decision-making.[298] I will elaborate on this issue in Chapter 7, when I discuss the effects of the US DoD problem construction of AWS on the emergence of norms.

The second layer of trust is that there must also be trust in a machine's autonomous capabilities to produce predictable outcomes, 'within its envelope of competence'[299]. In other words, the system should have trustworthy autonomous capabilities such as a high level of reliability, transparency, and traceability in its situational awareness and decision-making. Again, I discuss the effects of such a problematisation on the emergence of norms in this area in Chapter 7.

Both dimensions of trust assessment are prone to miscalibration – whether over-trust or under-trust – during design, development, or use of AWS.[300] Under-trust can occur in situations such as defending against large numbers of incoming ballistic missiles, when a high level of automation (LOA) or autonomous operations would be desirable, but human operators could be overwhelmed by the situation and might not permit the system to control

[296] Defense Science Board, 'Summer Study on Autonomy' (n 255) 9.
[297] Greg Zacharias, USAF, 'Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future, Volume I: Human-Autonomy Teaming' (n 103) 8.
[298] Defense Science Board, 'Summer Study on Autonomy' (n 255) 16.
[299] ibid 1.
[300] ibid 23.

the engagement.[301] This is the reverse of over-trust, or unwarranted reliance on a system's automatic or autonomous operating mode. Overreliance on automation has in the past been the contributing factor to fratricidal engagements by US forces.[302] In fact, US DoD's past experience with automation over-trust involving autonomous supervised weapons was the main reason for the administration to start the process of drafting US DoD Directive 3000.09, as I discuss in more detail in Chapter 6.

1.2.The US DoD Directive 3000.09 Leaves the Door Open to LAWS Development

Despite the problem with trust Directive 3000.09 does not prohibit the use of AWS, even for lethal purposes. On the contrary, it says that if LAWS meet all the necessary criteria, such as 'reliability, effectiveness, and suitability under realistic conditions', then in principle they could be authorised.[303] Directive 3000.09 does not explicitly pose any limits regarding the development or use of autonomy in weapon systems. It is thus an example of a problem construction that minimises the problem significantly – in other words, while the DSB report recognises the increased uncertainty in the operation of AWS and their current inability to exhibit *operational trustworthiness*, Directive 3000.09 assumes that the AWS can, in principle, be developed and deployed if they meet 'necessary criteria'.[304] Specifically, Directive 3000.09 belittles one aspect of operational trustworthiness: the concept of human–machine interaction during combat operation. Another DSB report says that the urgent deployment of unmanned systems in military operations leaves 'little time to refine *concepts of operation* which, when coupled with the lack of assets and time to support pre-deployment exercises, created operational challenges'.[305] Directive 3000.09

[301] John K. Hawley, 'Patriot Wars: Automation and the Patriot Air and Missile Defense System' (CNAS 2017).
[302] Defense Science Board, 'Patriot System Performance' (2005) D.C. 20301-3140.
[303] Directive 3000.09 Autonomy in Weapon Systems; Scharre (n 34).
[304] Defense Science Board, 'Summer Study on Autonomy' (n 255) 1,14.
[305] Defense Science Board, 'The Role of Autonomy in DoD Systems' (n 255) 2.

only states that relevant 'training, doctrine, and TTP will be established',[306] but it has been over a decade since Directive 3000.09 was introduced and TTP have not yet been officially established.[307] As a result, US forces may not have enough guidelines to deal with the use of systems not anticipated by developers or currently unknown limitations in system capabilities.

This being said, based on the most recent data points from the DoD, there has not yet been an instance of a weapon that has had to go through the additional review applicable to the offensive and lethal use of AWS.[308] This view has also been publicly stated by Frank Kendall, former Under Secretary of Defense and one of the architects of the US policy on AWS: 'We have not had anything that was even remotely close to autonomously lethal.'[309] Two early conclusions follow. First, it means that US DoD does not consider any existing lethal weapons with a high degree of autonomy as 'autonomous'. Second, more importantly, Directive 3000.09 has essentially legalised the use of *all existing lethal weapons with a high degree of autonomy* that have been within the DoD since at least the 1980s.

These statements illustrate how the US policy constructs the problem of AWS. Directive 3000.09 excludes a plethora of potentially controversial weapons from the increased regulatory oversight by differentiating semi-autonomous weapons from fully autonomous weapons, thus legitimising the use of many *near*-autonomous weapons of today. Full LAWS exist as a concept that may never materialise not because US DoD is not advancing the use of autonomy in weapon systems, but because the department has established such a high bar for qualifying any lethal weapon as 'autonomous.' Another

[306] Directive 3000.09 Autonomy in Weapon Systems 4a (1).
[307] Interview with Shawn Steene (n 256).
[308] ibid.
[309] Interview with Frank Kendall, 'Interview' (7 November 2016) in Scharre (n 34) 91.

explicit example of a high bar of autonomy is the UK Government's definition of LAWS as a weapon system that 'will, in effect, be self-aware'.[310] In another, similarly broad definition, the UK Government has defined AWS as weapons 'capable of understanding higher level intent and direction'.[311] According to this definition, weapon systems would need to achieve a general human level of intelligence in order to be classified as 'autonomous'. That, however, seems unlikely, as many authors doubt whether humanlike AI technology will soon be developed.[312] US DoD's definition of 'autonomy' is not quite the same that put forward by the UK Government, but one can argue that it performs the same function: it establishes a bar too high for any weapon to reach for the considerable future.

2. A Problem Construction that Legitimises All Existing Weapon Systems

This high bar of *full autonomy* in turn allows US DoD to construct the problem in such a way that it legitimises the current use of what is portrayed as 'limited autonomy' in weapon systems. To illustrate the arbitrary and controversy of such problem construction, let us critically discuss various types of existing weapon systems which are by some authors credited as AWS.[313]

2.1. Semi-autonomous Weapon Systems and their Supervisory Control

The US Aegis combat system and Phalanx CIWS are called autonomous anti-ship missiles, but the name 'autonomous' should be taken with a degree of reservation, as most of these missiles are in fact supervised systems. These systems are usually differentiated as

[310] UK Ministry of Defence, 'Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems' (2011).
[311] UK Ministry of Defence (n 28) 72.
[312] Vincent Müller and Nick Bostrom (n 184).
[313] Scharre (n 34).

having a human either in-the-loop or on-the-loop. Supervised systems with a human in the loop refer to weapon systems in which humans set the parameters of the weapon and the system engages with targets only after direct authorisation from a human. In US DoD, such systems are often described as 'supervised autonomous weapon systems'. Supervised systems with a human-on-the-loop refer to weapon systems that are autonomously able to target incoming threats with a human operator present and ready to intervene at any time to stop the engagement. In the US DoD such systems are often described as 'semi-autonomous weapon systems'.[314] Examples of such semi-autonomous weapons are some homing munitions where the human operator does not directly control the trajectory of the munition but does control the weapon's aimpoint in real time. This function allows the human to redirect the munition in-flight or abort the attack.[315]

Both types of supervised weapon have introduced a control architecture known as human supervisory control (HSC).[316] I trace the emergence of this type of control in Chapter 6, as it is a key 'innovation' that opened the door to the application of force in an autonomous way. At this stage, I am concerned with the rudimentary description of the concept. HSC is the process by which a human operator interacts with a computer by receiving feedback from, and providing commands, to a system with different degrees of embedded automation.[317] According to Mary Cummings, HSC systems vary in terms of the sophistication of embedded autonomy, and different systems have different allocations of human element relative to automation across various system functions.[318] US DoD has

[314] Directive 3000.09 Autonomy in Weapon Systems.
[315] Scharre (n 34) 40.
[316] P.J. Mitchell, Mary Cummings, and Thomas Sheridan, 'Human Supervisory Control Issues in Network Centric Warfare' (Massachusetts Institute of Technology 2004) HAL2004-01.
[317] Thomas Sheridan, *Telerobotics, Automation and Human Supervisory Control* (MIT Press 1992); P.J. Mitchell, Mary Cummings, and Thomas Sheridan (n 316) 1.
[318] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment' (n 282) 275.

various supervisory autonomous weapons in its current arsenal, and weapons such as Phalanx CIWS or Patriot are widely used.

Many scholars and policy activists, particularly those arguing for restrictions on the development and use of LAWS, claim that such supervisory autonomous weapons already cross the line and do not meet the principles of international law; they should therefore be prohibited.[319] They argue that a supervisory control architecture introduces a number of challenges in military operations.

First, it brings automation bias – the tendency to over-rely on automated decision support systems.[320] While humans are rather effective in naturalistic decision-making scenarios in which they leverage experience to solve real world problems under stress, they are nonetheless prone to fallible heuristics and various decision biases that are also heavily influenced by their experience, their framing of things, and the presentation of information.[321] Automated decision support systems can also serve as a new source of bias, when a human decision-maker disregards, or does not search for, contradictory information considering a computer-generated solution. Automation bias is particularly problematic in the execution of missions where there is a significant time-pressure, such as emergency path planning and resource allocation.[322] One of the consequences of automation bias is skills degradation.[323]

[319] Noel Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons' (n 139).
[320] ibid 32–35.
[321] Don Harris and Wen-Chin Li, *Decision Making in Aviation* (Routledge 2017) 290–291.
[322] Mary Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (Routledge 2015).
[323] Andreas Haslbeck and Hans-Juergen Hoermann, 'Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills Among Airline Pilots' (2016) 58 Human Factors 533.

Second, supervisory systems pose moral questions.[324] Supervision introduces a distance between the human and the target of engagement which, according to some authors, allows the separation of moral reactions from inhumane conduct and disables the mechanism of self-condemnation.[325] Further, Shannon Vallor argues that the supervisory system, particularly with a limited human-on-the-loop role will not only lead to practical skills degradation, but also to the 'moral deskilling' of the military.[326] She argues that the mere ability of a human to intervene in a weapon's operations is in some cases largely meaningless, as the human is only given a fraction of a second to stop a weapon's engagement, as is the case with several systems already in operation.[327]

The arguments about the potential moral disengagement or deskilling were also discussed within US DoD prior the introduction of Directive 3000.09, according to one of its drafters.[328] This being said, all existing supervisory weapons in use have been legalised by the introduction of the Directive 3000.09.

2.2. Self-guiding Long-Range Anti-Ship Missiles

Up to this point, we discussed human-supervised weapons (semi-autonomous) that are not deemed to be fully autonomous. The difference between various semi-autonomous weapons and fully autonomous, according to US DoD, is when the weapon is 'activated'. In human-supervised weapons, autonomy may be used to search for and detect targets and carry out the engagement, but the human chooses the target or specific target group and then activates the weapon. In autonomous weapons, the human activates the weapon and

[324] Leveringhaus (n 26).
[325] Armin Krishnan (n 125); Leveringhaus (n 26).
[326] Shannon Vallor, 'Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character' (2015) 28 Philosophy and Technology 107.
[327] ibid.
[328] Interview with Paul Scharre (n 256).

the weapon system itself selects and engages its target.[329] Thus, in AWS, the entire engagement loop – searching, detecting, deciding to engage, and engaging it – is delegated to the autonomous function. This being said, advanced anti-ship cruise missiles such as LRASMs often blurs the lines between supervised and fully autonomous weapon system.[330]

What is novel in the LRASM is that it can select and engage a specific target on its own.[331] The predefined target criteria are set by a human who launches the weapon against a highly constrained target enemy identified by satellites. This being said, the weapon, once activated, actively 'selects and engage targets without further intervention by a human operator'. Yet the LRASM, according to US DoD, does not qualify as LAWS because in the broader decision cycle of the weapon system, it is a human who decides about the targets that will be engaged.[332]

What, however, is meant by 'the broader decision cycle of the weapon system' and whether it is a human or a machine who 'decides' in operations with weapons like LRASM requires further evaluation, which will be carried out in Chapter 7, where I focus on the effects of the US DoD problem representation of AWS on the military targeting and engagement practices. For now, it is worth emphasising that US DoD does not consider and LRASM to be a LAWS. Thus, Directive 3000.09 essentially legitimises the novel use of autonomous features that the LRASM exhibits at the level of targeting and engagement.

The use of self-guiding algorithms to pinpoint specific targets on its own are not without controversy in the academic scholarship and public debate. Publicly available data

[329] Robert Work, 'Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities' (CNAS 2021) 7.
[330] Lockheed Martin (n 274).
[331] ibid; *Long Range Anti-Ship Missile (LRASM)* (Directed by Lockheed Martin, 2016) <https://www.youtube.com/watch?v=h449oIjg2kY>.
[332] Scharre (n 34) 68.

on LRASM functionality is ambiguous with respect to their degree of autonomy in selecting and engaging targets. DARPA and Lockheed Martin declined to comment on how the weapon engages with targets, saying the information is classified.[333] Thus, some commentators such as Roff argue that the uncertainty of the process by which the LRASM track and automatically follow specific target leaves too much room for error.[334] This is because the use of self-guidance targeting requires advanced image recognition technology, which at this stage of development is still nascent and prone to errors.

2.3. Loitering Munitions and the Autonomous Target Decision Function

There are, however, limited examples of weapons that satisfy the US DoD criteria for 'autonomous weapons'. These weapons can be categorised as either static search weapons or bounded search weapons.[335] An example of a static search weapon is the Mark 60 encapsulated torpedo (CAPTOR), a deep water mine fielded in 1979 during the Cold War and which remained in service until 2001.[336] The weapon system was to be placed in deep water, anchored in the ocean, and could last from weeks to months underwater. CAPTOR had its own upward-looking sonar system that identified and tracked the difference between hostile submarine signatures, surface vessels, and friendly submarines. When detecting a hostile submarine, the torpedo was activated and the weapon system could detect, classify, and attack its own target without any further human oversight or intervention.[337]

Bounded search weapons are weapons that can monitor a defined search area called a 'kill box' to hunt down and attack imprecisely located targets or classes of targets. These

[333] John Markoff, 'Fearing Bombs That Can Pick Whom to Kill' *The New York Times* (11 November 2014) <https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html>.
[334] Heather Roff, 'Meaningful or Meaningless Control' (*The Duck of Minerva*, 25 November 2014).
[335] Robert Work (n 329).
[336] Weapon Systems, 'Mark 60 CAPTOR' <https://weaponsystems.net/system/449-Mark+60+CAPTOR>.
[337] Robert Work (n 329) 6.

are called ‘loitering weapons’, colloquially known as ‘kamikaze drones’, and they represent the most striking example of autonomy in weapon systems. They are weapon systems in which the munition loiters around the kill box for some time, searches for targets, and attacks once a target is located. Loitering munitions, unlike homing munitions, do not require precise intelligence on enemy targets before launch.[338] A human can launch a loitering munition into a clearly defined geographic area to search for enemy targets without the knowledge of any specific targets beforehand. Loitering munitions can circle overhead for extended periods of time, searching for potential targets over a specified area, and engage beyond line-of-sight ground targets with an explosive warhead. Some loitering munitions are monitored by a human operator via a radio connection, and ultimately it is a human who approves each target. However, weapons such as the Israeli Harpy operate fully autonomously, as no human approves the specific target before engagement.[339] The Harpy is found in the arsenal of various countries today, including China, India, South Korea, Chile, and Turkey. It is also reported that the Chinese have reverse engineered their own variant.[340] Harpy loitering munitions were used several times in 2018 and 2019 by the Israel Defense Forces to destroy Syrian Pantsir-S1 SAM batteries.[341]

US DoD currently owns a miniature loitering munition called the AeroVironment Switchblade.[342] .[343] The Switchblade was used by the US Army in Syria and Afghanistan to target ‘high value targets’, such as insurgent leaders, mortar teams, or insurgents

[338] Scharre (n 34).
[339] IAI, ‘Harpy Autonomous Weapon for All Weather’ <https://www.iai.co.il/p/harpy>.
[340] Scharre (n 34) 47.
[341] Charlie Gao, ‘Loitering Munitions (A.K.A. Suicide Drones) Are Getting Deadlier Every Year’ (26 November 2020) <https://nationalinterest.org/blog/reboot/loitering-munitions-aka-suicide-drones-are-getting-deadlier-every-year-173454>.
[342] AeroVironment, ‘Switchblade 600 Loitering Missile’ <https://www.avinc.com/tms/switchblade-600>.
[343] ibid.

travelling in vehicles.[344] Switchblade still keeps humans in the loop via a functioning radio link to approve targets before engagement, making them semiautonomous weapons, but it can be potentially deployed without direct human intervention. According to the latest data points from US DoD, the US military have not yet used Switchblade with an autonomous target decision function.[345] The former US DoD official I approached who is knowledgeable in the area did not, however, exclude the possibility that either Switchblade or Harpy might already have been used in that manner to target specific objects.[346] Further, loitering munitions are becoming increasingly popular, with a number of countries either developing such weapons or reported to have purchased them. Beyond the improved precision compared to equivalent weapons, loitering munitions may also be cheaper than some guided missiles.[347]

What is novel in terms of advanced loitering munitions such as Harpy or Switchblade is their autonomous target decision function.[348] While in the LRASM example a human operator is responsible for the initial selection of targets, loitering munitions with autonomous targeting capabilities operate without any human supervision. A human is not even on-the-loop, as the missile, once activated, can select and engage targets without any

[344] Stavros Atlamazoglou, 'This New Technology May Be the Future of Close Air Support' (*Sofrep*, 2 February 2019) <https://sofrep.com/news/this-new-technology-may-be-the-future-of-close-air-support/>.
[345] Interview with fomer Senior US DoD official, 'Mikolaj Firlej Interview' (7 February 2021).
[346] ibid.
[347] The Switchblade price depends on various weapon's functions and other factors. The Switchblade is generally estimated to cost around $70,000 a piece, roughly two-thirds the cost of the AGM-114 Hellfire. However, some articles report the price as low as $6,000 per unit. See Ken Dilanian, Dan De Luce, and Courtney Kube, 'Biden Admin Will Provide Ukraine with Killer Drones Called Switchblades' (*NBC News*, 16 March 2022) <https://www.nbcnews.com/politics/national-security/ukraine-asks-biden-admin-armed-drones-jamming-gear-surface-air-missile-rcna20197>; Center for the Study of the Drone, 'Loitering Munitions' (2017); Rich Smith, 'AeroVironment Will Upgrade the Switchblade' (*The Motley Fool*, 11 May 2016) <https://www.fool.com/investing/general/2016/05/11/aerovironment-will-upgrade-the-switchblade.aspx>.
[348] Scharre (n 34).

intervention by a human operator. Thus, the weapon is an example of a human-out-of-the-loop system.

According to one of the drafters of US DoD Directive 3000.09, loitering munitions with autonomous targeting capabilities are examples of LAWS.[349] Interestingly, there is no evidence that Switchblade has gone through the additional review procedure applicable for LAWS, despite the possibility of operating with an autonomous targeting function. This being said, a former US DoD employee said that, if the Israeli Harpy had been developed in the US, it would have gone through the additional review, as it qualifies as LAWS.[350] This does not mean, however, that the US does not have LAWS in its arsenal. The US Navy deployed the loitering missile know as a Tomahawk Anti-Ship Missile (TASM) that could search for, select, and attack Soviet ships on its own. Although the TASM was never used in combat and taken out of Navy service in the early 1990s, it was likely the first bounded search lethal autonomous weapon.[351]

The TASM example nevertheless illustrates an important point which is often missed in the debate about LAWS: it is not 'intelligence' that makes a weapon 'autonomous', but rather it is freedom that matters, according to US DoD's definition. A LAWS is a weapon system that, once activated, is intended to search for, select, and engage targets where a human has not decided which specific targets are to be engaged. Therefore, LAWS can be relatively simple weapons, as the TASM was and the Harpy is today. However, where intelligence comes into play is in expanding their usefulness. As was the case for TASM in the past, Harpy munitions still have limited computer vision technology, which restricts their operation in a more complex environment. This was also the reason

[349] ibid.
[350] Interview with fomer Senior US DoD official (n 345).
[351] Robert Work (n 329) 6.

why more modern version of TASM in the US military arsenal – the Low-Cost Autonomous Attack System (LOCAAS) – was never fielded.[352] Currently, US DoD is more enthusiastic about the prospect of fielding autonomous loitering munitions. Besides the use of Switchblade, the US Government has also used Altius multi-purpose mini-drones that can easily be turned into loitering munitions.[353] The manufacturer of Altius loitering munitions, Anduril Technologies, has not yet provided details about the targeting system used in its loitering munitions, but a post on the company's blog says that 'Altius has demonstrated autonomous coordinated strike, target recognition and collaborative teaming.'[354] Palmer Luckey, a founder and Chief Technology Officer of Anduril, said in an interview that AWS exist today and one 'can't have a person literally be responsible for pulling the trigger in every instance'. [355]

After reviewing these examples of various weapon systems, it is clear that US DoD Directive 3000.09 was a major document as it legitimised and, at least according to Scharre, 'legalised' some controversial uses of weapons that have largely operated over recent decades in a legal vacuum. As argued, US DoD Directive 3000.09 has legitimised the use of autonomous supervisory weapons (semi-autonomous weapons), as well as the autonomous targeting capabilities of weapon systems. Further, as there is no clear rule or policy guidance regarding the distinction between defensive and offensive weapon systems

[352] 'Low Cost Autonomous Attack System (LOCAAS) Miniature Munition Capability' <https://man.fas.org/dod-101/sys/smart/locaas.htm>; Robert Haddick, 'Stopping Mobile Missiles: Top Picks For Offset Strategy': *Breaking Defense* (23 January 2015). Robert Work (n 329) 6.
[353] Andrew Eversden, 'Meet Anduril's New Loitering Munitions, the Firm's First (but Not Last) Weapons Program' *Breaking Defense* (6 October 2022); Anduril Industries, 'Altius' <https://www.anduril.com/hardware/altius/>; Garrett Reim, 'Anduril Introduces Loitering Munition Warheads For Altius Drones' *Aviation Week Network* (7 October 2022) <https://aviationweek.com/shows-events/ausa/anduril-introduces-loitering-munition-warheads-altius-drones>.
[354] Anduril Industries, 'Anduril Announces Best In Class Loitering Munition' (6 October 2022) <https://blog.anduril.com/anduril-announces-best-in-class-loitering-munition-8b00a72aba2a>.
[355] Steven Levy, 'Palmer Luckey Says Working With Weapons Isn't as Fun as VR' *Wired* (14 March 2022) <https://www.wired.com/story/palmer-luckey-drones-autonomous-weapons-ukraine/>.

in US DoD, some 'defensive' weapon systems can exhibit a significant level of autonomy that falls under the broad category of 'semi-autonomous weapons.'

A useful illustration of the spectrum of autonomy within US DoD is the LOA in Table 3, authored by Thomas Sheridan and William Verplank. According to the terminology in Directive 3000.09, there are no offensive weapons within US DoD operating above LOA 5, the level of supervisory control. However, US DoD does have various defensive weapons that operate at LOAs 6 and above.[356] As the distinction between offensive and defensive weapons is blurry, Directive 3000.09 has given a lot of discretionary power to US officials to develop and use weapon systems with a high degree of autonomy.

Table 3: LOAs[357]

Classification	Automation Level	Automation Description
No autonomy	1	The computer offers no assistance: human must make all decisions and actions.
Partial autonomy	2	The computer offers a complete set of decisions / action alternatives, or:
	3	Narrows the selection down to a few, or:
	4	Suggests one alternative, and:
Supervisory Control	5	Executes that suggestion if the human approves, or:

[356] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment' (n 282) 276.
[357] ibid; Thomas Sheridan and William Verplank, *Human and Computer Control of Undersea Teleoperators* (MIT Press 1978); Thomas Sheridan (n 317).

	6	Allows the human a restricted time to veto before automatic execution, or:
Full Autonomy	7	Executes automatically, then necessarily informs humans, and:
	8	Informs the human only if asked, or:
	9	Informs the human only if it, the computer, decides to
	10	The computer decides everything and acts autonomously, ignoring the human

The sharp distinction between the review of 'existing practices' regarding the use of autonomy in weapon systems and the review LAWS that is the weapons that falls outside 'normal review' illustrates the mechanics of the problem construction of AWS by US DoD's policy. The US policy-makers have put all existing instances of using autonomy in weapon systems in the one bucket and differentiated it from the *potential unspecified instances of using novel weapon systems*. They have not specified, for instance, that some elements of the current use of autonomy may pose some novel and important challenges. On the contrary, they described that existing weapon systems do not differ much from each other, as they all simply go through a 'normal review process'. Thus, the 'legalisation' of existing practices is explained purely as a *confirmation* of existing practices. Further, when asked whether any controversy has been associated with any specific use of autonomy in weapon systems or specific autonomous capabilities prior the introduction of Directive 3000.09, one of the drafters said that no US DoD branch wanted any of their current

weapons to be subject to additional review because it would render the process *too bureaucratical*.[358]

Directive 3000.09 is recognised in the literature as the first policy on AWS, with some authors going even further by wrongly interpreting that the Directive essentially prohibits such weapons.[359] Directive 3000.09 is, however, first and foremost a way to legitimise all existing weapons with a high degree of autonomy by delineating all existing weapon systems from the potential future weapons. This appears to suggest that the US DoD problem construction here is one where the incremental autonomisation of weapon systems is not considered to be a negative feature, but that this process can easily be absorbed into the projected account of what the problem is – in other words, that only novel uses of autonomy, particularly the lethal use of AWS, constitute a problem. In that sense, the autonomisation of weapon systems of the last few decades has been framed as a desirable development that should be continued or even accelerated. While US DoD recognises the increased problem of potential unintended engagements of such weapons, the department believes that the focus should be on establishing trust measures between humans and machines, as well as the trustworthiness of autonomous capabilities.

3. Different Problem Constructions Lead to Alternative Policy Responses – Human Judgment and Meaningful Human Control

The distinctiveness of the US DoD problem construction of AWS is evident when compared with the problem construction of the Campaign. As discussed earlier, the

[358] Interview with Paul Scharre (n 256).
[359] C. Todd Lopez (n 3).

Campaign is a coalition of NGOs that seeks to pre-emptively ban AWS.[360] The Campaign's problem construction of AWS is that such weapons by 'their own nature' do not comply with LOAC.[361] Specifically, it argues that AWS are unable to distinguish between combatants and civilians and thus violate LOAC's principle of distinction prohibiting weapons if they cannot be directed at a specific military object or if their effects cannot be limited as required by international law.[362]

This is a different problem construction than the one outlined by US DoD. While the US Government focuses on obstacles to establishing trust in AWS, specifically in the context of potential unintended engagements, they do not claim that the mere potential of such engagements, even in the lethal use of these weapons, should make AWS illegal under LOAC. In the next chapter, I will discuss the different assumptions underlining these competing views in more detail, but at this stage I would like to pinpoint the policy implications of the two alternatively framed problematisations.

The Campaign's policy response is that all weapons which do not allow for MHC should be prohibited. This is because only the exercise of MHC allows weapons to comply with the principle of distinction and other principles of LOAC. LAWS are thus all weapon systems that 'by their nature select and engage targets without meaningful human control'.[363] Which weapons fall under such a definition is unclear. Article 36, a think tank now part of the Campaign, argued that MHC has three requirements:

[360] Bonnie Docherty (n 9); Bonnie Docherty (n 157).
[361] Bonnie Docherty (n 157).
[362] The principle of distinction arises from international customary law. It is also codified in art. 48 of the Protocol I, with supplementary rules in art. 51 and art. 52. See Shane Darcy, *Judges, Law and War: The Judicial Development of International Humanitarian Law* (Cambridge University Press 2014). For the Campiagn's argument see Bonnie Docherty (n 157).
[363] Campaign to Stop Killer Robots, 'Key Elements of a Treaty on Fully Autonomous Weapons' (2019).

> Information – a human operator, and others responsible for attack planning, need to have adequate contextual information on the target area of an attack (...)
>
> Action – initiating the attack should require a positive action by a human operator.
>
> Accountability – those responsible for assessing the information and executing the attack need to be accountable for the outcomes of the attack.[364]

US military experts, however, have pointed out that many existing weapons, including supervised weapon systems with a human on the loop (semi-autonomous weapons), would fail to meet these requirements. For instance, fire-and-forget missiles such as CIWS are specifically designed for situations where the time of engagements is too short for humans to adequately respond. After launching the weapons, humans do not need to push a button to fire at each target and the weapons can hit targets that are not necessarily in the line of sight of the human operator. In such situations, human control is exercised only by determining the system's rules of engagement, in the initial decision to activate the system, and in real-time human supervision of its operation with the option to deactivate the system if such engagement is no longer appropriate for use.[365] Thus, US DoD refused to refer to the concept of MHS as a way of distinguishing AWS, particularly LAWS, from other types of weapon.

However, despite reservations regarding the requirements of MHC, US DoD states in Directive 3000.09 that 'it is DoD policy that: autonomous and semiautonomous weapon systems shall be designed to allow commanders and operators to exercise *appropriate levels of human judgment* over the use of force'.[366] It is also noticeable that the origins of

[364] Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (n 65) 4.
[365] Paul Scharre and Michael Horowitz (n 39) 10–13.
[366] Directive 3000.09 Autonomy in Weapon Systems 4a.

both concepts – MHC and appropriate human judgment – are closely intertwined and can be traced back to 2012. The phrase 'human control' was formulated in the widely discussed report by the Campaign, published just *two days before* the introduction of US Directive 3000.09.[367] The report says:

> Humans should therefore retain control over the choice to use deadly force. Eliminating human intervention in the choice to use deadly force could increase civilian casualties in armed conflict.[368]

During the UN GGE discussions, many authors pointed out that the concept of 'human control' is elusive and open to different interpretations. In the report from 2018, the UN GGE expressed fear that 'there may be no single touch point or notion that can fully describe the role of humans throughout the life cycle of a weapons system'.[369] This view has been confirmed by the US delegation to the UN, who argued that US DoD Directive's 3000.09 notion of 'appropriate levels of human judgment over the use of force' has been purposefully constructed as a flexible term to reflect the fact that there is no fixed, one-size-fits-all level of human judgment that should be applied to every context. The content of what is 'appropriate' can differ across weapon systems, domains of warfare, types of warfare, and operational contexts, as well as across different functions in a weapon system. This is because some functions might be better performed by a computer, while other functions should be performed by humans.[370]

[367] Bonnie Docherty (n 9).The concept of human control with the adjective 'meaningful' has been formulated in more structured way later in 2013 by Article 36 and then adopted by HRW.
[368] ibid 37.
[369] UN GGE, 'Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 38).
[370] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

This reasoning suggests that US DoD's contextual understanding of the role of human factors in the use of weapon systems is a way of constructing the problem of AWS that is instrumental – in other words, it is about delineating what is possible in specific operations and within the realm of 'effective management. There might even be situations in which a machine might perform specific targeting and engagement functions more effectively than a human, leaving the door open to exercising no direct human control at all.[371]

In contrast, the Campaign's problem construction is conceptual: they construct the problem of AWS as one where human control needs to be exercised in order to render weapon systems acceptable according to specific normative criteria (LOAC). Therefore, the discussions about the Campaign's concept of MHC usually focus on content, or specific elements of human control, while the discussions about US DoD's concept of appropriate human judgment focus on context: how human judgment ought to be exercised, and specifically *who* should exercise human judgment over *what* so a weapon system can complete its mission effectively.[372]

This being said, the concepts of human control and human judgment appear to be very similar, and some authors have even equated these terms.[373] While it might be tempting to unify these concepts, particularly for policy-makers, they nevertheless differ in important ways. This view has been explicitly expressed by the US delegation to the UN GGE. Specifically, US DoD opposes the use of the word 'control.'[374]

---

[371] Dan Saxon (n 7).
[372] Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control' (*ICRC Blog*, 15 August 2018).
[373] Article 36, 'Key Elements of Meaningful Human Control' (n 65) 36.
[374] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

The US delegation has distanced itself from the notion of human control because they feel that framing the debate on the use of AWS on 'control' is too restrictive and may imply a so-called *direct* human control. Humans have historically exercised a 'direct control' over weapons because weapons have been seen merely as tools in the hands of fighters. In a sense, humans have been 'masters' of their weapons.[375] This is echoed in the definition of AWS put forward by the CRS, which states that AWS are 'a special class of weapon systems that use sensor suites and computer algorithms to independently identify a target and employ an onboard weapon system to engage and destroy the target without *manual human control* of the system'.[376]

The idea of 'direct control' is also reflected by the International Committee of the Red Cross (ICRC), which argues that the decisions regarding when to employ a weapon, against whom, and at what level of severity remain the domain of fighters.[377] According to international law theorists such as Thompson Chengeta, the concept of 'direct control' of weapons by humans was also embedded in the 1949 Geneva Conventions and in their 1977 Additional Protocols, whose provisions invoke the idea that, without human control or use, a weapon is nothing but a mere tool.[378] As an example, in armed conflict, participating in hostilities is shown by the 'bearing of arms'. Thus, persons 'who have laid down their arms' are considered to be 'taking no active part in the hostilities'.[379] Such an interpretation of LOAC suggests that all types of weapon should be guided by 'direct control' to be fully compliant with the law.[380] This interpretation has also been the

[375] Thompson Chengeta (n 41) 839.
[376] Congressional Research Service, 'Artificial Intelligence and National Security' (2019) 15.
[377]HRW, 'Shaking the Foundations. The Human Rights Implications of Killer Robots' (2014)..
[378] Thompson Chengeta (n 41) 840.
[379] Geneva Convention Relative to the Treatment of Prisoners of War (Third Geneva Convention) 1949 Article 3.
[380] Thompson Chengeta (n 41) 840.

backbone of the Campaign's narrative to prohibit LAWS, which will be explored in more detail in Chapter 5.

The US delegation to the UN GGE on LAWS does not agree with this notion; they cite various examples to support a broader understanding of human–machine interaction on the battlefield. One of the examples is the Automatic Ground Collision Avoidance System developed by USAF, which has helped prevent so-called 'controlled flight into terrain' accidents. The system assumes control of an aircraft when an imminent collision with the ground is detected, and returns control back to the pilot when the collision is averted.[381] Based on such examples, US DoD prefers to place emphasis on the design requirements of the weapons and the machine's communication with a human commander or operator rather than on the notion of 'control', which is often limited and interpreted too rigidly.[382]

Interestingly, however, US DoD Directive 3000.09 expands the notion of human judgment beyond the design stage in situations where a weapon is unable to complete engagements consistently with the operator's intentions. In such cases, we read that a weapon should seek additional human operator input before continuing the engagement or terminating the engagement. This is a classical reference to the concept of direct control.

3.1.Control-By-Design Supersedes Direct Control

One may therefore argue that there are two human factor dimensions in the US DoD problematisation of AWS. The first type is 'a control-by-design'; the second type of

[381] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[382] ibid.

control is related to the ability of a human to directly exercise control by terminating a weapon's engagement. This 'finger on the button' control represents a different dimension of control, even though prior design may determine it. It is different because a designed mechanics of control is determined ex-ante engagement, while a 'finger on the button' is the ability to exercise control ex-post engagement – that is, in a realm of warfare during a real-time operation. The exact phrasing of these two notions of control in Directive 3000.09 is as follows.

Human control-by-design is formulated as: 'The system design incorporates the necessary capabilities to allow commanders and operators to exercise appropriate levels of human judgment in the use of force.'[383]

Direct human control ('finger on the button') is formulated as: 'The system is designed to complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, to terminate engagements or seek additional human operator input before continuing the engagement.'[384]

US DoD argues that this understanding provides a more 'flexible' approach than that of the Campaign to the problem of controlling LAWS because it does not restrict the notion of control to direct human control. However, the notion of direct human control in US DoD operations is in fact not a rigid requirement. Section 2 of Enclosure 3 of Directive 3000.09 gives an opportunity to waive the requirement of direct human control in cases of 'urgent military operational need'.[385] Directive 3000.09 does not define 'urgent military operational need', which makes the phrase vulnerable to broad interpretation,

[383] Directive 3000.09 Autonomy in Weapon Systems Enclosure 3, Section 1a(1).
[384] ibid Enclosure 3, Section 1a(2).
[385] ibid Enclosure 3, Section 2.

whereas the DSB found that current approaches to implementing rapid responses to urgent needs have not been sustainable.[386]

Let us highlight this distinction between the Campaign's concept of human control and US DoD's human judgment in the most radical context: in the context of LAWS being used in response to an urgent military operation need. According to the Campaign's requirement of human control, humans must always retain direct control over life and death decisions, which means that delegating lethal authority for a machine to make its own decision should not be allowed, even in such special situations. This contrasts with the requirement of human judgment, whereby the development and use of such weapons can be allowed in such situations in principle. As Robert Work, former Deputy Secretary of Defense between 2014–2017, explicitly said:

> In fact, DoDD 3000.09 does not mandate human-in-the-loop or on-the-loop control schemes. Instead, it establishes broad policies and an internal bureaucratic process for senior leaders to approve or reject novel uses of autonomy in weapons, including fully autonomous weapons.[387]

It is worth emphasising how this subtle semantic difference plays a transformative role. By using the word 'judgment', Directive 3000.09 steers the focus away from direct control at the level of engagement and targeting to the design requirement of weapons that allow human to make informed decision about their potential deployment. The appreciation of design requirements generates two subsequent positive obligations: (1) that humans deploying a system must understand how the weapons operate in realistic environments so that humans can make informed decisions regarding their use, and (2) to

[386] Defense Science Board Task Force (n 271).
[387] Robert Work (n 329) 9.25/07/2023 13:15:00

satisfy this obligation, AWS require adequate levels of operational testing, verification, validation, and evaluation. As long as these two positive obligations are satisfied, the policy of human judgment does not in principle prohibit the potential deployment and use of autonomous weapon for *lethal purposes*.[388] On the contrary, the requirement of human control, at least in the form specified by the Campaign and ICRAC, explicitly states that 'humans should retain control over the choice to use deadly force'.[389]

To restate, both the concept of human control and the concept of human judgment – stem from the recognition of the risks that are posed by AWS, yet they imply different propositions. Both concepts have been introduced as a response to the fact that the incremental development of robotic weapons has arrived at the point where weapon systems are able to produce a lethal effect in an autonomous way, which can lead to unintended engagements. Yet the concept of human control is used to support the prohibition of AWS used for lethal purposes (and of some existing semi-autonomous weapons as well), while the concept of human judgment is used to leave the door open for the potential development and use of such weapons.

4. A Summary of the Chapter

In Chapter 4, I have argued that the US DoD problem construction of AWS relates to their potential 'lethal use', particularly the use of such weapon systems to kill other humans. Specifically, the potential application of lethal force in an autonomous way increases the risk of 'unintended engagements.''. This increased risk relates to the growing sophistication of weapons, particularly the introduction of so-called autonomous supervised weapon systems, known otherwise as 'semi-autonomous'. I argue that the notion of trust has a

[388] Mary Cummings, 'The Human Role in Autonomous Weapon Design and Deployment' (n 282).
[389] Bonnie Docherty (n 9).

particular significance in addressing the risk of unintended engagements. US DoD identifies two kinds of measure designed to build trust in AWS. First, in the socio-technical system of using AWS, there must be trust and deep integration between humans and machines. Second, there must also be trust in a machine's autonomous capabilities to produce predictable outcomes – in other words, that the system has trustworthy autonomous capabilities. Thus, I have argued that the potential increased likelihood of unintended consequences associated with the use of autonomous weapons is in fact rooted in a deeper, underlying problem of how trust can be established in the decision-making process of complex socio-technical systems.

US DoD addresses the potential increased risks associated with the lethal use of AWS through an additional senior review mechanism. Directive 3000.09, however, does not in principle restrict the development and use of such weapons, contrary to the Campaign's discourse, that the lethal use of AWS should be prohibited. I have argued that the US military has established such a high bar for qualifying any lethal weapon as 'autonomous' that the concept of autonomy is in fact used in an indeterminate fashion. This problem construction allows, in turn, for the exclusion of a plethora of already existing weapon systems with advanced autonomous capabilities from increased regulatory oversight, in other words the additional senior review mechanism established by Directive 3000.09. I then explored various types of weapons which have not been qualified as 'lethal autonomous weapon systems' according to US Directive 3000.09: Phalanx CIWS, LRASM, and loitering munitions such as Switchblade. I have argued that the US DoD Directive 3000.09 'legalisation' of their use is not without controversy in the wider academic literature. Semiautonomous weapons are based on supervisory systems that are prone to automation bias, that may cause military skills degradation, and that generate moral dilemmas associated with the practice of killing from a distance. Weapon systems

such as LRASM are based on advanced AI and image recognition techniques, which are prone to a trade-off between performance and interpretability (or explainability). Thus, at least to a certain extent, developers and users of such weapon systems either have to sacrifice certain performance accuracy or the ability to explain how the AI models behind such weapon systems arrive at a decision. Further, weapons such as loitering munitions can operate without any human supervision even at the level of targeting and engagement, which only exacerbates the challenges mentioned above.

I also explored two alternative policy responses towards the problem of AWS represented by US DoD and the Campaign. US DoD's policy does not prohibit the lethal use of AWS, but Directive 3000.09 introduces the requirement of 'appropriate levels of human judgment' in the use of such weapons. The Campaign, in contrast, proposes a ban on LAWS because they are beyond 'human control'. I have explicated that the concept of 'human judgment' is different from the requirement of 'human control', as the latter relates to so-called 'direct control' in the form of a human manually exercising control by terminating a weapon's engagement. US DoD argues for a broader understanding of 'control' that includes both manual control and control-by-design, or the *ex ante* determination of a weapon's capabilities.

Thus, the US DoD problematisation of the role of human factors in the use of AWS is instrumental – in other words, it is about delineating what is possible in specific operations. I have argued that there might even be situations in which a machine might perform specific targeting and engagement functions more effectively than a human, leaving a door open to exercise no direct, manual human control at all. This is the case, for example, in the situation of an urgent military operational need, where I have demonstrated that US DoD's policy allows for direct control to be superseded by a

control-by-design. In contrast, the Campaign's problem construction is conceptual – in other words, they construct the problem of AWS as one where human control needs to be exercised in order to render such a weapon system acceptable according to specific normative criteria (LOAC). Therefore, the discussions about the Campaign's concept of MHC usually focus on content, or specific elements of human control, while the discussions about the US DoD concept of appropriate human judgment focus on context: how human judgment is and ought to be exercised and, specifically, who should exercise human judgment over what so the weapon system can complete its mission effectively. In the next chapter, Chapter 5, I reconstruct in more detail the assumptions that underpin this representation of AWS and the role of human factors in the use of such weapons.

Chapter 5: What Presuppositions Underlie the US DoD Approach to LAWS?

This chapter deals with the thesis's second research sub-question. It considers what presuppositions underlie the US DoD representation of the 'problem'. The major data sources in Chapter 5 are publicly available US DoD documents regarding the use of autonomy in weapon systems and interviews with current and former US DoD representatives. In this chapter, I use additional legal sources to place the US DoD policy on AWS within the wider context of US administrative law. I explore the US APA, the Attorney General's Manual on the APA, and selected judgments of the US Supreme Court.

The chapter is divided into three sections. In the first section, I argue that the introduction of US DoD Directive 3000.09 is an attempt to strike a balance between two competing assumptions underlying the US approach to AWS.

In the second section, I argue that, although US DoD recognises the increased risks associated with the adoption of autonomy in weapon systems, the department does not consider that the application of a lethal force via AWS is necessarily illegal according to LOAC and US domestic law. I explore interdiscursive connections between US DoD and Campaign's narrative in more detail.

In the third section, I argue that Directive 3000.09 concentrates only on the risks associated with autonomy conceived as independence from human operator, leaving considerations regarding AI-augmented weapons unaddressed.

1. A Policy of Human Judgment as a Balancing Act

In the first section, I argue that the key assumption behind the US DoD problematisation of AWS, particularly the role of human judgment over their use, is the effort to establish a balancing act between two major competing interests: the strength of military deterrence and the safety requirements of the military innovations.

1.1. The Geopolitical Ramifications of the US DoD Policy on AWS

The main reason behind the introduction of US DoD Directive 3000.09, according to its drafters, was to fill the regulatory void regarding the development of autonomy in weapon systems, an area increasingly dominated by the international competition between countries.[390] In the CRS Brief, we read that 'the United States may be compelled to develop LAWS in the future if potential US adversaries choose to do so.'[391] Robotics systems with greater autonomy have been highlighted as a key component of the future strength of the US military by all current major documents outlining US strategic decisions about defence and security.[392] LAWS were specifically considered as a vital element of US DoD's 'Third Offset Strategy'.[393] In US national defence circles, an 'offset' refers to the necessary action that needs to be taken by the US Armed Forces to compensate for enemy superiority. The objective of the Third Offset Strategy is to ensure a continued asymmetric combat advantage for the US, with a particular focus on the development and deployment of

[390] Interview with Paul Scharre (n 256).
[391] Congressional Research Service, 'Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems' (n 255).
[392] US DoD, 'National Defense Strategy' (n 1) 8; US DoD, 'National Defense Strategy' (n 89) 3,7; US DoD, 'National Military Strategy' (n 89); The White House, 'National Security Strategy' (n 88); The White House, 'National Security Strategy' (n 88).
[393] US DoD, 'The Defense Innovation Initiative (Memorandum)' <https://defenseinnovationmarketplace.dtic.mil/wp-content/uploads/2018/04/DefenseInnovationInitiative.pdf>.

autonomous and semiautonomous weapons which produce superior effectiveness and efficiency.[394]

In the *AI Strategy 2018*, US DoD argued that the US 'must adopt AI to maintain its strategic position, prevail on future battlefields, and safeguard this order', particularly given that China and Russia 'are making significant investments in AI for military purpose […] [that] threaten to erode US technological advantages and destabilize the international order'.[395] US DoD's argument has been echoed by Vladimir Putin, who has explicitly said that whoever reaches a breakthrough in developing AI will come to dominate the world.[396] Russia has introduced its *National Strategy for AI*, which aims to accelerate the development of AI through significant investments.[397]

Over the last few years, US DoD has started to prioritise China as an even greater threat than Russia to US supremacy. Since Barack Obama's 2011 'Pivot to East Asia', the US Government has officially recognised the rise of China as an emerging superpower,[398] while from 2017 onwards China has been treated as a 'long-term strategic competitor' in official US Government strategy documents.[399] More recently, US President Joe Biden signed the National Defense Authorization Act (NDAA) for the fiscal year 2023. The act allocates the highest budget in history of $11.5 billion for the Pacific Deterrence Initiative, one of the key initiatives to counterbalance China and which can be traced back to Obama's

[394] Cheryl Pellerin, 'Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence' (*DoD News*, 31 October 2016) <https://www.defense.gov/News/Article/Article/ 991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence>. Robert Work, 'Remarks by Deputy Secretary Work on Third Offset Strategy' (28 April 2016) <https://www.defense.gov/News/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/>.
[395] US DoD, 'AI Strategy: Harnessing AI to Advance Our Security and Prosperity' (2018) 5.
[396] Associated Press, 'Putin: Leader in Artificial Intelligence Will Rule the World' (1 September 2017).
[397] Government of the Russian Federation, 'National Strategy for Artificial Intelligence Development' (2019).
[398] Brent Reininger and others, 'Assessing the Obama Administration's Pivot to Asia' (2016).
[399] The White House, 'National Security Strategy' (n 88); The White House, 'National Security Strategy' (n 88).

Pivot.[400] These measures have been put in place as the US Government, since at least early 2012, has regarded China as a revisionist power whose long-term aim is global supremacy.[401] The US considers Chinese large infrastructure projects as 'geopolitical projects' which will allow China to undermine a US-led world order, diminish US partnerships, and weaken the US influence.[402]

The US Government is particularly worried about China's advances in AI. China's *AI Development Plan* states that 'AI is a strategic technology that will lead the future military revolution' and calls for China to be the world leader in AI by 2030.[403] Some former US DoD officials, such as Nicolas Chaillan, the USAF first Chief Software Officer, and Robert Spalding, a retired Air Force brigadier general who served as defence attaché in Beijing, have voiced concerns that the US Government has already lost AI supermacy to China.[404]

US DoD experts argue that China is also developing advanced autonomous weapons that could threaten US military superiority.[405] The most recent report of the National Security Commission on AI (NASCAI) states:

> China is not only actively pursuing increased autonomous functionality across a range of military systems, but it is also currently exporting armed drones with autonomous

[400] National Defense Authorization Act for Fiscal Year 2023 2022; United States Senate Committee on Armed Forces, 'Summary of the Fiscal Year 2023 National Defense Authorization Act' (2022).
[401] Brent Reininger and others (n 398). The White House, 'National Security Strategy' (n 88); The White House, 'National Security Strategy' (n 88).
[402] Lt Col Daniel Lindley, USAF, 'Assessing China's Motives: How the Belt and Road Initiative Threatens US Interests' [2022] Journal of Indo-Pacific Affairs.
[403] Graham Webster and others, 'Full Translation: China's "New Generation Artificial Intelligence Development Plan"' (Stanford Cyber Policy Center 2017).
[404] *US Government Must "Wake up Now" to AI Threat from China, Says Former Air Force Software Chief* (Directed by Government Matters on 7 News, 2021) <https://www.youtube.com/watch?v=7tqS9_AN9y0>; Katrina Manson, 'US Has Already Lost AI Fight to China, Says Ex-Pentagon Software Chief' *Financial Times* (10 October 2021) <https://www.ft.com/content/f939db9a-40af-4bd1-b67d-10492535f8e0>.
[405] Eric Schmidt, Robert Work, and others, 'Final Report of the National Security Commission on Artificial Intelligence' (2021) 96.

> functionalities to other nations. This includes systems (...) capable of conducting autonomous, lethal, targeted strikes.[406]

Further, US DoD experts are sceptical about the potential compliance of China' and Russia's use of AWS with LOAC. The NASCAI report on AI states:

> There is little evidence that U.S. competitors have equivalent rigorous procedures to ensure their AI-enabled and autonomous weapon systems will be responsibly designed and lawfully used.[407]

In contrast to China and Russia, US DoD experts argue that US DoD's weapon review process is consistent with IHL,[408] and cite the ICRC appreciation of 'the strength and transparency' of this review by listing the US as 'one of eight countries that have national mechanisms to review the legality of weapons and that have made the instruments setting up these mechanisms available to the ICRC'.[409]

The US Government considers autonomous and AI weapons as strategic technologies that can enable China to achieve supremacy, and thus it is essential for the US to have more advanced military capabilities that will maintain the competitive advantage over adversaries.

1.2.Addressing Safety Concerns by the Soft Law

As discussed earlier, the main reason behind the introduction of US DoD Directive 3000.09 was to legitimise the development of autonomy in weapon systems given the increased importance of AWS in future military conflicts. This being said, the underlying assumption

[406] ibid.
[407] ibid 95.
[408] ibid 94.
[409] ICRC, 'A Guide to the Legal Review of New Weapons, Means and Methods of Warfare' (n 21) 5.

behind the policy action was the need to address specific safety concerns that have been evident in the past malfunctions of highly advanced robotic systems on the battlefield.[410] A US DoD working group that was tasked to draft Directive 3000.09 carried out an exhaustive review of various military practices as well as of different types of weapon systems, paying particular attention to the failures and missteps that have occurred in testing, training, and deployment.[411] The next chapter, Chapter 6, traces the genealogy of US DoD Directive 3000.09 and discusses some of these instances in detail, but if there was a single military event that particularly affected the drafting of Directive 3000.09, it was Operation Iraqi Freedom in 2003.[412] During the invasion of Iraq, the US electronic Patriot missile defence system, created to shoot down incoming ballistic missiles, was a contributing factor in three incidents of fratricide. Patriot missiles twice engaged friendly coalition aircraft, resulting in the death of three crew members, and in a third case it fired on a Patriot battery believed to be an Iraqi surface-to-air missile.[413]

These experiences, among others, have spotlighted the need to regulate the use of AWS, in particular in the area of human–machine interaction. Thus, the objective of Directive 3000.09 is to establish US DoD policy for the development and use of AWS, while the cornerstone of this policy is the *guidance* that AWS 'shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force'.[414] The word 'guidance' may incline readers to think that Directive 3000.09 does not establish a new legal obligation for the US military administration. Indeed, according to the APA, policy statements are considered as 'non-legislative rules', which means that they fall within the definition of 'rules' but are not required to be promulgated through the

[410] Interview with Paul Scharre (n 256).
[411] ibid.
[412] Dan Saxon (n 7) 190.
[413] John K. Hawley (n 301).
[414] Directive 3000.09 Autonomy in Weapon Systems 4a.

use of legislative rulemaking procedures.[415] Thus, policy statements do not have the force of law of interpretative rules. What differentiates interpretative rules from policy statements is that the former are rules issued to clarify or explain existing laws, while the latter are issued to 'advise the public of the manner in which the agency proposes to exercise a discretionary power'.[416] Legislative rules, contrary to interpretative rules and policy statements, have the 'force and effect of law' and may be promulgated only if they have gone through a public notice and comment procedure – a process by which the public is given an opportunity to comment on a proposed version of the rule and the agency responds to those comments.[417]

However, rules that involve 'military functions' are exempt by the APA[418] from the notice and comment procedure, which sometimes makes it difficult to determine whether a particular rule constitutes a new law, clarifies an existing law, or merely represents advice about the exercise of an agency's discretionary power. Courts in the US focus on the particular language used in documents when making this determination. For instance, mandatory language delineating an agency's obligation in the policy statement can serve as strong evidence of an intention to bind the agency itself.[419] Agency statements 'couched in terms of command' may be read to eliminate agency discretion when applying a policy, transforming the statement into a legislative rule.[420] Following this approach, if a 'so-called policy statement is in purpose or likely effect one that narrowly limits administrative discretion, it will be taken for what it is – a binding rule of substantive law'.[421]

[415] Administrative Procedure Act 1946 5 U.S.C. § 553(b)(A).
[416] Tom Clark, 'Attorney General's Manual on the Administrative Procedure Act' (1947) 30.
[417] Administrative Procedure Act 5 U.S.C. §553(b), (c).
[418] ibid 5 U.S.C. §553(a) (1).
[419] Congressional Research Service, 'General Policy Statements: Legal Overview' (2016) 9.
[420] 'American Bus Association v. United States' 627 F.2d 525.
[421] 'Guardian Federal S & L Association v. Federal Savings and Loan Insurance Corporation' 589 F.2d 658, 666.

Interestingly, the US DoD Directive on AWS does contain some mandatory language. For instance, it requires AWS to go through a detailed review process before development and fielding.[422] However, it is unclear how to interpret the main issue at stake –the concept of human judgment over the use of AWS. Directive 3000.09 states that AWS '*shall* be designed […] to exercise appropriate levels of human judgment over the use of force',[423] but the word 'shall' is confusing because it can mean 'may, will or must'. Although various parts of the Code of Federal Regulations that govern federal departments use the word 'shall' to establish mandatory requirements, the US Supreme Court has held that 'shall' could also mean 'may'.[424] Thus, the wording of the Directive 3000.09 does not necessarily suggest whether the requirement of human judgment over the use of AWS is a legislative rule or a soft policy intent that leaves the US military departments with a wide degree of discretion. Directive 3000.09 has also been purposefully left undefined. As argued by Dan Saxon:

> By choosing not to define its standard of 'appropriate levels of human judgment over the use of force', the United States keeps alive all possible options for the exercise of a commander or operator's judgment as long as they fall within the bounds of IHL.[425]

I will revisit this point in Chapter 6 when exploring the effects of the US DoD problem construction of AWS and its AWS policy on USAF regimes of practice related to weapon targeting. For now, an important point is that Directive 3000.09 is at minimum a declaration of intent presenting how US DoD aims to exercise a discretionary power over the potential use of AWS. Whether Directive 3000.09 establishes any new law is, however, uncertain at

[422] Directive 3000.09 Autonomy in Weapon Systems.
[423] ibid 4a.
[424] See 'Gutierrez de Martinez v. Lamagno' 515 U.S. 417, 434 n.9. 'Though "shall" generally means "must," legal writers sometimes use, or misuse, "shall" to mean "should," "will," or even "may."'
[425] Dan Saxon (n 7).

best. 'The Directive [3000.09] is as legal as any other US directives', said the Senior Force Developer for Emerging Technologies at the Office of USDP, who later stressed that the Directive should be taken into account before a weapon system will be approved.[426] He continued:

> […] we are trying to *balance* [my emphasis] these operational challenges that we're trying to cope with, in which autonomy might offer a capability that helps us address those challenges. But there is also a recognition that […] we want these weapons to do what, what we want them to do, [that is] we want them to strike the targets, but only those targets and we do not want them to do, you know, other things.[427]

One of the drafters of Directive 3000.09 emphasised the fact that the Directive requirements are deliberately ambiguous:

> The Directive is vague, but it is vague on purpose. Many principles of international law are similarly vague, [as] these are general concepts that informs how weapons operators and developers should think about AWS.[428]

While Directive 3000.09 must be read in conjunction with other applicable laws, in particular LOAC and US DoD's general weapons review procedures,[429] the Directive's legal influence largely depends on the discretionary interpretation of US DoD officials, who have been left with a text that is often general and ambiguous.

---

[426] Interview with Senior Force Developer for Emerging Technologies at the Office of the Under Secretary of Defense for Policy, 'Mikolaj Firlej Interview' (5 February 2021).
[427] ibid.
[428] Interview with Paul Scharre (n 256).
[429] See Directive 5500.15 Review of Legality of Weapons under International Law, US Department of Defense 1974. The Directive 5500.12 has since been superseded by updated directives and individual service instructions. See Department of the Army Regulation 27-53, Review of Legality of Weapons Under International Law; USAF, 'Department of the Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities' (n 19); Department of the Navy (n 19).

Directive 3000.09 on AWS (2012) was announced after Obama's Pivot to Asia (2011) and before the Third Offset (2014). While I do not argue that there is a direct relationship between these texts, it is clear that the US DoD problematisation of AWS is grounded in the assumption that AWS are considered as one of strategic weapon systems in the global competition with China. Yet, as I have argued, Directive 3000.09 also responds to the safety concerns related to the past failures involving the use of weapons with autonomous capabilities by introducing new requirements in the use of such weapons, in particular the requirement of human judgment. A closer examination of the choice of words in Directive 3000.09, reaffirmed by the findings from interviews with senior US DoD officials, suggests that US DoD did not want to limit themselves in developing and using AWS, which are increasingly important in the modern theatre of war. Thus, the requirements introduced in Directive 3000.09 pertaining to the role of human involvement are considered as soft policy guidelines rather than as a new law.

This being said, US DoD acknowledges that the use of AWS should not only comply with Directive 3000.09 and other US domestic rules, but also with 'fundamental principles of LOAC'.[430] In the next section of this chapter, I explore how US DoD justifies the compliance of AWS, including their lethal use, with existing LOAC.

2. Moving Beyond a Legal Discourse: LAWS Compliance with LOAC Subjugated to Technical Analysis

[430] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

In this second section, I explore how the US Government justifies the compliance of AWS, including their lethal use, with existing LOAC. Specifically, US DoD takes a qualitatively different interpretation of LOAC's principle of distinction relative to the Campaign's legal discourse. Similarly, US DoD differs from the Campaign regarding the problem of responsibility for wrongdoings committed using LAWS.

2.1. Is the Lethal Use of AWS Non-Compliant with the LOAC?

Directive 3000.09, which legitimises the development and use of AWS, is not the only legal document in the US regulating these weapon systems. As argued earlier, the US Government recognises that the weapons review of AWS 'should be guided by the fundamental principles of the law of war'.[431] The LOAC obligation to conduct a weapon's review process is codified in Article 36 of AP I and is considered by the UN and other agencies as the legal safeguard for preventing the development of unlawful AWS. Although the US is not a party to AP I, US DoD's longstanding policy requires a legal review of the intended acquisition of a weapon system to ensure its development and use is consistent with IHL. This policy dates back to 1974 and predates the adoption of AP I by other states.[432] The responsibility for conducting reviews resides within US DoD and all individual military branches. According to US DoD's legal review of weapons procedure, prior to fielding or deploying any weapon systems, the weapons are reviewed in accordance with all international legal obligations of the US, as well as with US DoD's domestic weapons review procedure, in order to ensure compliance with LOAC and other applicable US domestic laws and policies, such as Directive 3000.09. The US Government argues that

[431] ibid.
[432] Directive 5500.15 Review of Legality of Weapons under International Law, US Department of Defense. DoD, 'Law of War Manual' (Office of General Counsel DoD 2015) 6.2.3.

such precautions allow for checking the legality of each new weapon system classified as AWS or LAWS on a case by case basis, rather than considering them all under one single umbrella.[433]

The view of US DoD representatives is informed by the domestic process of the legal review of weapons, which consists of three steps to determine whether the acquisition or procurement of a weapon is prohibited. The process starts with the question of whether there is a specific rule of law, either as a treaty obligation or as customary international law, prohibiting or restricting the use of the weapon.[434] In answering this question, the US Government points out that there is no specified international law treaty that focuses exclusively on AWS. This leads to the conclusion that AWS as such are not illegal, but that a specific application of AWS may not comply with specific LOAC principles. As there is no specific prohibition or restriction, the answer to the second question should determine whether the specific example of the AWS's intended use might cause superfluous injury. Finally, the third question is whether this specific weapon is inherently indiscriminate.

In contrast to US DoD, the Campaign argues that AWS already do not comply with the main principles of LOAC. The Campaign's argument is that *AWS as such* are unable to sufficiently distinguish between combatants and civilians, irrespective of their potential use.[435] The difference between the Campaign's argumentation and the US DoD position is that the former argues that one should consider AWS as a single type of weapon system, not as individual weapons. AWS, according to the Campaign, are a special class of weapons

[433] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[434] DoD (n 432) 6.2.
[435] Bonnie Docherty (n 9).

which are fully autonomous from a human operator and thus 'inherently indiscriminate' because a lack of MHC makes indiscriminate engagements more likely.[436]

The Campaign further argues that AWS should be regulated in the form of a new international treaty that would introduce the requirement of MHC over the use of such weapons.[437] As discussed earlier, US DoD opposes the Campaign's definition of AWS on the ground that it focuses on the narrow concept of direct control. Yet the fact that states in the UN GGE have not agreed with the Campaign's definition of AWS allows US DoD to nullify the Campaign's arguments that the mere possibility of using AWS for lethal purposes makes indiscriminate engagements more likely.

2.2. The DoD's Discourse that Autonomous Weapons Could Comply with the Principle of Distinction

US DoD has developed their own alternative discourse regarding the potential compliance of AWS with the LOAC's principle of distinction.

While US DoD agrees that all weapons should comply with the basic principles of LOAC, including the principle of distinction, they argue that existing AWS are not 'inherently indiscriminate' and in fact often *enhance* the compliance with law, rather than violate it.[438] US DoD representatives point out that LOAC does not require that a weapon should determine whether its target is a military objective, but rather that the weapon should be capable of being employed consistent with the principle of distinction. They argue that this logic also applies to the use of weapons that 'may be characterized as capable of taking

[436] Bonnie Docherty (n 157) 9–10.
[437] Bonnie Docherty (n 157).
[438] US DoD, 'Autonomy in Weapon Systems' (UN GGE 2017) CCW/GGE.1/2017/WP.6.

some form of action or decision in a given moment in the absence of direction by a human being such as whether to fire the weapon or to select and engage a target'.[439] For US DoD, persons must comply with LOAC by employing weapons in a discriminate and proportionate manner. For instance, even if the weapon autonomously selects and engages targets, 'its use would be precluded when expected to result in incidental harm to civilians or civilian objects that is excessive in relation to the concrete and direct military advantage expected to be gained'.[440] One should not confuse the prohibition on weapons that are indiscriminate - because they cannot be aimed at a lawful target - with the prohibition on the use of discriminate weapons in an indiscriminate fashion.

Michael Schmitt, a Professor at West Point, who served 20 years in USAF as a judge advocate, gives an example that illustrates the difference by referring to SCUD missiles launched by Iraq during the 1990-1991 Gulf War. The missiles were not unlawful per se, because special conditions existed in which they could be launched discriminately. The missiles were capable of use against troops in open areas such as the desert, and they actually struck very large military installations without seriously harming the civilian population. However, when targeted in the direction of cities, their use was found to be unlawful.[441] This was because the missiles were insufficiently accurate to reliably strike any legitimate military objects. Similarly, AWS can be aimed at a legitimate target depending on specific operations.

In fact, existing AWS are designed to be discriminating by their very nature. They can only attack specifically designated targets that meet set criteria determinable by an

[439] *Ibid*.
[440] US DoD, 'Autonomy in Weapon Systems' (n 438).
[441] Michael Schmitt, 'The Principle of Discrimination in 21st Century Warfare' (1999) 2 Yale Human Rights and Development Law 148.

algorithm among previously predefined targets.[442] Schmitt further argues that advancements in technologies, doctrine, and tactics continue to heighten the quality of the targeting process, and the result is a growing resort to precision attack.[443] Schmitt emphasises that the growing prevalence of precision operations have profound implications for the application of LOAC principles. The more precise the strike, the more likely it is that the right target will be hit, and increased accuracy allows the use of a smaller charge to achieve the desired probability of damage, thereby risking less collateral damage and incidental injury.[444]

Further, both Schmitt and Thurnher, from the Office of the US Army Judge Advocate General, argue that there are already instances of AWS being used without violating LOAC principles. There are situations in which AWS could satisfy this rule even with a considerably low level of ability to distinguish between civilian and military targets.[445] First, in well-defined circumstances 'without placing civilians at excessive risk'. An example of such a situation is a battlefield where combatants are strictly separated from civilians and occupy only a specific territory. Secondly, existing AWS might only target enemy weapons, as opposed to the individuals operating them, until that individual poses a potential threat. Third and last, such weapons might also operate where no civilians are present.[446]

US DoD further argues that while emerging LAWS may today pose certain challenges related to the technical ability to discriminate civilians, further improvements of

[442] Michael Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (n 157); Michael Schmitt, 'Precision Attack and International Humanitarian Law' (2005) 87 IRRC.
[443] Michael Schmitt, 'Precision Attack and International Humanitarian Law' (n 442) 453.
[444] Michael Schmitt, 'Precision Attack and International Humanitarian Law' (n 442).
[445] Jeffrey S. Thurnher, 'The Law That Applies to Autonomous Weapon Systems' (2013) 17 ASIL Insights; Michael N. Schmitt and Jeffrey S. Thurnher (n 176).
[446] Michael N. Schmitt and Jeffrey S. Thurnher (n 176).

technology will 'fix' these challenges. This idea of 'technology fixing' is echoed by the statement of US DoD representatives during a UN GGE meeting:

> Emerging technologies are difficult to regulate because technologies continue to change as scientists and engineers develop advancements. A best practice today might not be a best practice in the near future. Similarly, a weapon system that, if built today, would risk creating indiscriminate effects, might, if built with future technologies, prove more discriminating than existing alternatives by reducing the risk of civilian casualties.[447]

In order, however, to move beyond the debate about the technical feasibility of designing a weapon to comply with LOAC, the Campaign has turned to another important problem - the issue of responsibility for AWS's wrongdoings. The next sub-section explores this.

2.3. The Legal Problem of a Responsibility Gap Again Shifts Attention Towards Risk Analysis

According to Campaign representatives, if the killing were to be done by a fully autonomous weapon, the problem would become whom to hold responsible. Authors associated with Campaign argue that there might potentially be situations where no one is held responsible for a machine's wrongdoings. This situation is called the 'responsibility gap'.[448] While the notion of responsibility gaps in the use of highly advanced machines appeared long time ago,[449] the most widely discussed proposition in the context of AWS has been formulated by Sparrow, another prominent supporter of the Campaign.[450] Sparrow argues that the more autonomous a weapon system becomes, the less it will be possible to

[447] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[448] Bonnie Docherty (n 9) 42.
[449] See Daniel Dennett, 'When HAL Kills, Who's to Blame? Computer Ethics', *Hal's Legacy: 2001's Computer as Dream and Reality* (MIT Press 1997); Andreas Matthias, 'The Responsibility Gap - Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175.
[450] Robert Sparrow, 'Killer Robots' (2007) 24 Journal of Applied Philosophy 69–70..

hold those who designed it or ordered its use properly responsible for their actions. Yet the impossibility of punishing the artificial agent means that we cannot hold a machine responsible.[451] The responsibility gap arises when, in the execution of a targeting decision, LAWS does something that the operator did not directly programme it to do, and thus the operator cannot be blamed for misapplications of force.[452] On the one hand, this element of unpredictability often guarantees a machine's flexible adjustment to a dynamic environment. Programmers deliberately design these systems to allow them to respond to changes in real time, rather than to anticipate every possible eventuality that may arise. On the other hand, this element of unpredictability appears to be particularly problematic in the application of force to a target.[453] The problem is that LAWS deliberately distance operators from the enforcement of targeting decisions. Operators appear in the first stages of the causal chain leading to the application of force to a target, but not in the final stage.

The Campaign's discourse frames the responsibility gap as a legal problem.[454] It argues that existing mechanisms for legal accountability are ill-suited to address the unlawful harms AWS might cause', and as a result humans involved in the development or use of AWS would 'escape liability for the suffering caused by such weapons.'[455]

US DoD responded to this legal argument by narrowing it down to purely technical considerations regarding risk analysis. US DoD strategy has been to downplay the

[451] *Ibid*.
[452] A Leveringhaus (n 16) 79.
[453] *Ibid*.
[454] In the debate about AWS the words 'responsibility' and 'accountability' are often used interchangeably. Both the UN GGE and the Campaign refer to the term 'accountability' as an umbrella term to describe various forms of legal responsibility, i.e. state responsibility, administrative proceedings undertaken in response to violations of IHL, civil liability, and individual criminal responsibility. In this thesis, I refer to the broad notion of 'responsibility gap', but I discuss both state and individual responsibility. See: HRW, 'Mind the Gap: The Lack of Accountability for Killer Robots' (2015); UN GGE, 'Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 38).
[455] HRW, 'Mind the Gap: The Lack of Accountability for Killer Robots' (n 454).

relevance of the 'responsibility gap'. The department argues that the 'responsibility gap' does not occur because, in principle, LOAC deals primarily with states, not individuals.[456] Unlike individual criminal responsibility, state responsibility is not based on the concept of personal culpability, but on the attribution to the state of the mis(conduct) of its organs or agents.[457] Individual criminal responsibility can only be attributed under LOAC to the most serious breaches of the law such as war crimes, crimes against humanity, aggression, and genocide.[458]

The Campaign argues, however, that the use of AWS has the potential to commit such crimes, and thus the responsibility gap appears.[459] According to the concept of individual criminal responsibility, criminal offences are either caused intentionally or by negligence. When an artificial agent is intentionally directed to harm or harm is caused by negligence, the human operator is criminally liable. When the lack of intent (*mens rea*) of the operator cannot be established, then – the Campaign argues – the responsibility gap arises.[460]

US DoD representatives disagree and claim that, in the entire chain of command, a human element can always be found, either at the programming level or at the operating level.[461] Thus, the programmer may be held responsible if he or she programmed LAWS in such a way that they intentionally breach LOAC, or the operator may be held responsible if he or she decides to operate LAWS in an unlawful manner.[462] In situations where neither programmer nor commander can be initially identified, the impossibility of ascribing

[456] US DoD, 'Autonomy in Weapon Systems' (n 438).
[457] Roberta Arnold, 'Legal Challenges Posed by LAWS: Criminal Liability for Breaches of IHL by (the Use of) LAWS', *Lethal Autonomous Weapon Systems* (German Federal Foreign Office 2016) 10.
[458] ibid.
[459] HRW, 'Mind the Gap: The Lack of Accountability for Killer Robots' (n 454).
[460] ibid.
[461] US Government (n 378).
[462] R Arnold (n 396) 10.

criminal responsibility to a person is not caused by the fact that the harmful conduct was committed by LAWS; rather, the problem is the impossibility of *collecting evidence* allowing for the proper identification of a relevant human element responsible for the machine's wrongdoings.[463]

The defenders of the Campaign could argue that the US DoD approach ignores the fact that certain AWS are so advanced that they are able to execute a wide discretionary decision-making power at a speed surpassing human ability to intervene. For instance, if a weapon can identify and engage with a target faster than a human can determine whether it is a legitimate object, the operator's ability to intervene is rendered meaningless. Sparrow then argues that one can still hold human operators responsible but only 'at the cost of allowing that they should sometimes be held entirely responsible for actions over which they had no control'.[464] One could respond to this argument that a degree of responsibility gap may indeed occur, but a higher standard of care for both weapon operators and designers should minimise such accidents. Legal authors refer in this context to the doctrine of command responsibility that establishes the negligence standard of care for the legal responsibility. Negligence appears only when military commanders fail to know what they should have known. It covers situations where risks were not properly recognised, and as a result, have been falsely disregarded.

However, the introduction of the negligence condition in practice subjugates the discussion to technical analysis, as negligence can be avoided or minimised with a sufficient standard of care and a risk assessment. Again, the US DoD view:

[463] *Ibid*.
[464] Robert Sparrow (n 51).

> In this regard, training on the weapon system and rigorous testing of the weapon system can help commanders be advised of the likely effects of employing the weapon system. These measures, found in DoD policy, can help promote good decision-making and accountability.[465]

According to such analysis, responsibility gaps are considered only if a weapon's operator did not intend to use a weapon to engage with an illegitimate target, but the weapon nevertheless violated its mission specification. 'Rigorous testing' should, according to US DoD, minimises such chances to zero.

Thus, in the debate on the responsibility gap the question is whether delegating control to a machine is 'too risky,' or whether the risk associated with deploying LAWS is 'reasonable.' US DoD argues that the risk of harm to civilians and other persons or objects can be mitigated in various ways. Proper monitoring could stop the operation of a weapon in the event that it malfunctions, or that circumstances change.[466] Sometimes, however, it might be more appropriate to consider 'whether it is possible to programme or build mechanisms into the weapon that would reduce the risk of civilian casualties while in no way decreasing the military advantages offered by the weapon'.[467]

Controversy then arises over the standards by which one assesses whether there was a consciously disregarded substantial risk. Various people in the military chain may have different views regarding what factors constitute risks that are too excessive.[468] US DoD experts refer to the argument that strategies of risk minimisation have a comparative nature.[469] Specifically, the question is whether the use of AWS is more-or-less risky than

[465] US DoD, 'Autonomy in Weapon Systems' (n 438).to
[466] ibid.
[467] ibid.
[468] Vincent Müller and Thomas Simpson, 'Autonomous Killer Robots Are Probably Good News' (2014) 273 Frontiers in Artificial Intelligence and Applications 297.
[469] Larry Lewis, 'Killer Robots Reconsidered: Could AI Weapons Actually Cut Collateral Damage?' [2020] Bulletin of Atomic Scientists.

the use of alternative technologies or soldiers. Authors such as Larry Lewis, a member of the US delegation to the UN GGE, argues that AWS may actually save more civilian casualties than operations where a human is present.[470] Lewis explores various examples where civilian casualties were the result of collateral damage from the engagement of a valid military target by a human and argues that humans often make errors under these circumstances, misidentifying civilians as valid targets.[471]

The discussion about the responsibility gap follows the same pattern as the debate about the compliance of AWS with the principle of distinction. This started as a legal discourse, but then developed primarily into a debate about the reasonable risk threshold and risk standards. The risk analysis, in turn, is largely dominated by the technical knowledge about the current and potential future performance of autonomous systems, their testing, and their training specifications. Here again, US DoD and its experts shift the focus from legal and moral arguments towards mere technical considerations. 'The standard of care or regard that is due in conducting military operations with regard to the protection of civilians is a complex question to which the law of war *does not* provide a simple answer,'[472] states the US DoD in UN GGE. The US DoD further states that this standard must be assessed based on the general practice of states and common standards of the military profession in conducting operations, particularly regarding 'training on the weapon system and rigorous testing.'[473]

Let us then explore the place of training and testing of new weapons in the context of the US DoD problematisation of AWS.

[470] ibid.
[471] ibid.
[472] US DoD, 'Autonomy in Weapon Systems' (n 438).
[473] ibid.

2.4. An Urgent Operational Military Need and a Weapons Review

The requirement for training and testing constitutes an integral part of the broader US DoD process of reviewing any new weapon before its potential use in combat.

Specifically, each new weapon systems which shall apply force in an autonomous way will need to go through a full weapons review process, which consists of the following checklist:

(1) The system design capabilities to allow commanders and operators to exercise appropriate levels of human judgment in the use of force.
(2) The system design capabilities to complete engagements in a required timeframe and consistent with operators' intentions.
(3) The system design capabilities that allow to terminate weapon's engagements or seek additional human operator input before continuing the engagement, if the weapon is unable to do so.
(4) The system design specific safeties, anti-tamper mechanisms, and information assurance that helps to minimize the probability or consequences of failures.
(5) V&V and T&E to establish system reliability, effectiveness, and suitability under realistic conditions, to a *sufficient standard* [my emphasis] consistent with the potential consequences of an unintended engagement or loss of control of the system.
(6) A preliminary legal review of the weapon system.[474]

Points (1)–(4) outline the consideration regarding the role of human factors in the use of such weapons. Point (1) establishes the requirement of a direct human control, while Points (2)–(4) focus on human control-by-design. As argued earlier, the notion of control-by-design supersedes direct human control because Directive 3000.09 provides an opportunity

[474] Directive 3000.09 Autonomy in Weapon Systems. Enclosure 3.

to waive the requirement of direct human control in cases of 'urgent military operational need'. Point (5) establishes the duty of training and testing LAWS before their development and use. Directive 3000.09 further specifies that weapon systems will need to go through hardware and software V&V and realistic system developmental and operational test and evaluation (T&E).[475] This process includes a regression test of the software to validate whether critical safety features have not been degraded. The purpose of regression testing is not simply to detect software bugs, but rather to identify any unwanted changes in functionality caused by changes to the software or its environment. Each change will need to undergo T&E to characterise the system behaviour in that new operating state.[476] Directive 3000.09 also stipulates that training, doctrine, and TTP regarding the use of LAWS will be established.[477] A closer examination of Directive 3000.09 and US DoD manuals, however, reveals challenges with the US DoD approach that have not quite been exposed by the Campaign's discourse.

First, Enclosure 3 to Directive 3000.09 states that the USDP and the Under Secretary of Acquisition, Technology, and Logistics may request a Deputy Secretary of Defense waiver for weapons review requirements (with the exception of the requirement for a legal review) in cases of urgent military operational need.[478] This waiver relates to all Points (1)–(5) enumerated earlier, including the V&V and T&E procedures. Recall that Directive 3000.09 also allows to weave the requirement of direct human control in the same cases, as discussed earlier. This means there might be instances when only the *legal* weapon review if effectively binding for the use of LAWS. Yet, as has been discussed, US DoD has consistently subjugated legal analysis to technical assessments and, in the case of urgent military operational need, the lack of details on V&V and T&E outcomes will not help

[475] ibid. 4 (1) (a)-(c)
[476] ibid.
[477] ibid.
[478] ibid. Enclosure 3, Section 2.

determine whether the use of a particular weapon can lead to indiscriminate effects or a responsibility gap. Rather, one can argue that US DoD will take a more permissive approach in such cases.

Second, leaving aside edge cases of urgent military operational need, one can argue that it is difficult to subordinate legal analysis to technical standards alone, particularly given the fact that US DoD itself does not have strong confidence in the current T&E and V&V capabilities applicable to LAWS. Current T&E and V&V capabilities are specified by the Joint Capabilities Integration and Development System (JCIDS) – that is, the formal US DoD process which defines acquisition requirements and evaluation criteria for future defence programmes. The T&E and V&V processes are based on the 'safety assurance concept'. According to this concept, the primary objective of JCIDS is to ensure that the capabilities required by the military are identified, along with their associated operational performance criteria, in order to successfully execute the assigned missions. This means that, before any new capability can enter the development process related to reviewing and validating its requirements, the originating sponsor organisation (a weapon manufacturer) is first obliged to identify the AWS system capability requirements related to its functions, roles, mission integration, and operations. However, over the recent decade, US DoD has produced a number of reports which stress that weapon systems able to deploy a force in an autonomous way raise new issues that challenge current T&E and V&V practices.[479] According to the established approach, each new software and hardware acquisition was procured as a vertically integrated, vendor-proprietary solution consisting of the vehicle system, control station, communications channels, and encryption technologies. These single-system variants were usually 'closed' systems utilising proprietary interfaces. While

[479] William Wagner, *Lightning Bugs and Other Reconnaissance Drones: The Can-Do Story of Ryan's Unmanned Spy Plane* (Aero Publishers 1982); Gen Mark Welsh III, USAF, 'America's Air Force: A Call to the Future' (n 101); USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (2017).

T&E and V&V practices that focus on achieving a specific mission or capability may be optimal for a single system, they are increasingly less applicable to the current state of weapon systems. In the Unmanned Systems Integrated Roadmap published by US DoD in 2011, we read:

> Today's V&V processes will be severely stressed due to the growth in the amount and complexity of software to be evaluated. They utilize existing industry standards for software certification that are in place for manned systems (e.g., DO-178B). Without new V&V processes, such as the use of trust audit trails for autonomy, the result will be either extreme cost growth or limitations on fielded capabilities.[480]

Yet despite the importance of problem, the existing processes are still far from satisfactory. In the latest available Unmanned Systems Integrated Roadmap from 2017, we read:

> For the most demanding adaptive and non-deterministic systems, a new approach to traditional TEVV will be needed. For these types of highly complex autonomous systems, an alternate method leveraging a run-time architecture that can constrain the system to a set of allowable, predictable, and recoverable behaviors should be integrated early into the development process. Emergent behaviors from large-scale deployment of interacting autonomous systems poses a difficult challenge.[481]

It is further argued in the report that additional measures, beyond V&V, will be required to ensure safe operation of AWS. No V&V process can guarantee 100% error-free operation of complex systems. As software complexity increases, predicting the precise behaviour of AWS in dynamic and unstructured environments will become increasingly difficult.

[480] William Wagner (n 479) 50.
[481] USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479) 10.

Moreover, today's weapon systems consist of many interconnected platforms – some of them manned, some unmanned – but the progress in network communications allows elements of these systems to affect one another. Failures often occur at the interfaces between various elements thought to be separate. Thus, such systems require a more holistic approach to testing and evaluation in order to reveal common components rather than simply focusing on a single capability.

US DoD has recognised this problem, and the current major focus is on interoperability and manned-unmanned (MUM) teaming to ensure that the systems will be capable of operating well with each other. The problem with longstanding and siloed T&E procedures was identified even before the introduction of US DoD Directive 3000.09. In the Unmanned Systems Integrated Roadmap published by US DoD in 2011, we read:

> […] silence about the lack of interoperability and standards failed to foster dialog on how to overcome them. […] As the unmanned systems industry matures, however, the acquisition process must evolve in parallel. Addressing and enabling interoperability within unmanned systems will help accomplish this goal.

Despite noticeable progress regarding the new acquisition practices and interoperability policies, the problem with interoperability and MUM teaming remains critical and has not yet been fully addressed.[482] In the DSB report, the authors argue that current conventional testing capabilities are still 'inadequate for testing software that learns and adapts'.[483] The longstanding approach focusing on specific system capabilities often led to siloed solutions, with significant functional overlap and no method for understanding the common components of each system. The potential remedy to this challenge is open architecture

[482] DoD, 'Unmanned Systems Integrated Roadmap 2017-2042,' 5.
[483] {Citation}

(OA), which facilitates interoperability between systems by leveraging common capability descriptions, components, and common and open data models, standards, interfaces, and architectures in the system design. The current focus on interoperability is, however, limited to the messaging layer.[484] In order to establish more effective and holistic interoperability, the so-called *plug-and-play interface*, one has to migrate current systems to OA and focus on establishing standards-based interoperability based on a common data repository.[485]

Another challenge with existing T&E procedures in US DoD relates to US DoD culture regarding the use of data and acquisition practices. Lt Gen Shanahan, the former head of the Joint AI Center (JAIC), warned of the problem of 'stove-piping' data within US DoD. US DoD's longstanding historical practice is for military branches to keep data for their own use without sharing it in a common platform.[486] Stove-piping is particularly problematic with AI because it may lead to the proliferation of 'edge cases'; in other words, the weapon will encounter scenarios in which the systems do not perform as required or as expected due to a lack of required data. Further, the culture of acquisition has also been historically dominated by short-term needs.[487] Many of the existing weapon systems have been rapidly acquired and immediately fielded for war fighter use through the so-called Joint Urgent Operational Needs process. The weapon systems acquired through this process have not undergone a rigorous requirements review and joint coordination through the standard JCIDS process, which includes systems interdependencies and interoperability considerations.[488]

[484] USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479). 8.
[485] ibid.
[486] 'Starting Project Maven with Lt. Gen. Jack Shanahan' <https://aneyeonai.libsyn.com/episode-45-jack-shanahan>.
[487] Gen Mark Welsh III, USAF, 'America's Air Force: A Call to the Future' (n 101) 20.
[488] ibid.

The Campaign has generally been silent about these challenges, except in the most recent article from Laura Nolan criticising 'Project Convergence,' the US Army's experimentation event aimed at testing the integration of the Army's weapon systems and C2 systems with those of the rest of the US military. The major innovation tested during the event was a prototype plug-and-play interface – an early version of the future Modular Open Systems Architecture (MOSA) – that allows sharing data between various systems on everything from target coordinates to engine diagnostics.[489] US Army representatives hope that MOSA will be essential in modern warfare, as it will connect various aircraft and drones by transmitting data directly from one machine to the next without the 'intermediary of human voices over the radio or human hands on a keyboard'.[490] MOSA will also be essential for maintenance purposes, as currently the process for replacing various weapon systems components is considered to be expensive and time-consuming. Rather than ask contractors to develop their own, often incompatible proprietary solutions, the US Army can now leverage OA and dictate common standards and interfaces, which are made available to all – hence 'open'. US DoD hopes this should allow the replacing of a piece of code from one vendor with better code from another without having to rewrite the rest of the system – this is why architecture is also called 'modular'. According to the US Army, some elements of the new open network technology are already in use and designed to improve existing communications. MOSA has not yet been used to transmit real-time targeting data from spy satellites, AI command hubs, or drone swarms.[491] However, Maj Gen Peter Gallagher is enthusiastic about the potential of this approach. 'We're pushing

[489] Sydney Freedberg Jr., 'Kill Chain In The Sky With Data: Army's Project Convergence' *Breaking Defense* (14 September 2020) <https://breakingdefense.com/2020/09/kill-chain-in-the-sky-with-data-armys-project-convergence/>.
[490] Sydney Freedberg Jr., 'MOSA: The Invisible, Digital Backbone Of FVL' *Breaking Defense* (13 March 2020) <https://breakingdefense.com/2020/03/mosa-fvls-invisible-digital-backbone/>.
[491] Sydney Freedberg Jr., '"Improvised Mode": The Army Network Evolves In Project Convergence' *Breaking Defense* (21 September 2020) <https://breakingdefense.com/2020/09/improvised-mode-the-army-network-evolves-in-project-convergence/>.

them [network technologies communication] to limits that we never envisioned,'[492] Gallagher said. 'It's a mesh network solution with some advanced networking waveforms that significantly improves the war fighting capability of our manoeuvre brigades, but it was not fielded to do the things we're doing.'[493]

In her article for the Campaign, Nolan points out that Project Convergence does not mention anything about human factors in the design of software interfaces, what training users get about the targeting systems related to OA, or how US DoD aims to ensure operators have sufficient context and time to make decisions. She also stresses that this is not surprising, as the lack of these considerations is consistent with the Defense Innovation Board (DIB)'s *AI Principles*, which also do not mention human factors, computer interfaces, or how to deal with the likelihood of automation bias.[494]

Nolan herself represents a wider group of academics who support the Campaign and argue that the T&E and V&V challenges of a specific weapon only illustrate the wider *normative* problem of weapons with the ability to deploy lethal force in an autonomous way – that is, the transformations of established decision-making processes undermining direct human control. Nolan is a computer programmer who is a member of ICRAC and a founding member of the Campaign. Her profile spotlights the role of two important groups of actors – academics on the one hand and private companies on the other – in the production of two different 'rationalities' to justify either the functionalist focus on a single weapon or the normative focus on the class of weapons called LAWS.

Nolan, before joining ICRAC, resigned from Google over 'Project Maven'. The project, formally known as the Algorithmic Warfare Cross-Functional Team, is US DoD's

---

[492] ibid.
[493] ibid.
[494] Laura Nolan, 'AI Enabled Kill Webs and the Slippery Slope towards Autonomous Weapons Systems' *Campaign to Stop Killer Robots* (12 October 2020) <https://stopkillerrobots.medium.com/ai-enabled-kill-webs-and-the-slippery-slope-towards-autonomous-weapons-systems-68440dbf8423>.

initiative to apply computer vision algorithms to tag objects identified in images or videos captured by surveillance aircraft or reconnaissance satellites and thus reducing the manual collection of data. As an example, such a data processing system can tag data feeds from full motion videos of a Chinese Fighter Jet, and it can 'learn' to identify it in a fraction of the time it would take a human. The programme received increased attention after Google, one of several technology companies participating in the programme, publicly withdrew amid negative reaction from employees about the 'weaponisation' of AI.[495] Google has been under pressure from over 1000 academics who signed ICRAC's *Open Letter in Support of Google Employees and Tech Workers* and urged Google's executives to join other academics in calling for an international treaty to prohibit AWS.[496] Another example of the significant engagement of academics against the development of LAWS came after over 8,000 academics signed Open Letter, produced by Future of Life Institute.[497] The Campaign's reliance on the academic support not only helps to extend their discourse and justify the specific course of action, i.e. the prohibition of AWS and the regulation of MHC. It also emphasises the narrative shift from conventional lobbying advocacy to more 'expertly-grounded' arguments.

The US DoD policy on AWS is further justified outside the traditional contours of government by the prominent role of professionals and academics who have had prior engagement with US DoD. A good example is Lt Gen Jack Shanahan, now Senior Fellow at the Centre for National American Security (CNAS), who oversaw Project Maven and later became the inaugural Director of JAIC, the US DoD centre focused developing AI capabilities by partnering across US DoD branches, academia, and commercial AI industry. Most of the US DoD 'expertise' comes from CNAS, a relatively small but influential

[495] Nick Statt, 'Google Reportedly Leaving Project Maven Military AI Program after 2019' *The Verge* (1 June 2018).
[496] ICRAC, 'Open Letter in Support of Google Employees and Tech Workers'.
[497] Future of Life Institute (n 183).

Washington DC-based foreign policy think tank.[498] Scharre, who previously worked in the Office of the Secretary of Defense and led the US DoD working group that drafted US DoD Directive 3000.09, is now the Vice President and Director of Studies at the CNAS. Robert Work, who is the Distinguished Senior Fellow for Defense and National Security at CNAS, previously served as the Deputy Secretary of Defense, where he was responsible for overseeing the Pentagon's work, including the drafting process of Directive 3000.09. He is also credited as an author of the 'Third Offset Strategy'. Prior to working as a Deputy Secretary, Work was a CEO of CNAS. Among CNAS's main organisational donors are US weapon manufacturing companies such as Northrop Grumman Aerospace Systems, Lockheed Martin Corporation, Raytheon Company, and the US Government. The close and fluid interactions between US DoD and CNAS shed light on the wider US governmentality of AWS. The US conceptualisation of AWS is produced and re-produced by the network of interrelated agencies, professionals, and 'experts' who may one day draft a policy, and the next justify it while wearing the hat of an 'expert'.

3. The US DoD Approach to the Weaponised AI

In this third section, I focus on US DoD assumptions regarding the concept of 'autonomy' and 'AI.' I argue that US DoD assumes that 'autonomy' is not necessarily interchangeable with AI, but that Directive 3000.09 only concentrates on the risks associated with autonomy, leaving considerations regarding AI-augmented weapons unaddressed. Further, the US DoD problem representation uses an unbounded notion of autonomy which does not prescribe any limitations to the application of autonomy to machines. While US DoD representatives are aware of the longer-term risks associated with the application of

[498] See Center for a New American Security, 'People' <https://www.cnas.org/people?group=full-time-staff>.

autonomy and AI to weapon systems, they argue that, as current AWS are technically controllable, future autonomous weapons will be too.

3.1. The Assumption that Autonomy is a Risk, Not AI

US DoD representatives repeatedly asserted that 'AI and autonomy are not interchangeable. While some autonomous weapon systems use AI, this is not always the case.'[499] The US DoD view assumes that 'autonomy' refers to a machine's degree of independence from a human, rather than to the self-learning abilities of a weapon system. This is well illustrated by the Levels of Autonomy framework in Table 3 where 'full autonomy' is considered to be when a machine 'executes the mission *automatically*' and then either informs or does not inform a human. It is worth acknowledging that Sheridan and Verplank's 10 levels of autonomy have been widely influential in US DoD. One of the applications of LOA framework by US DoD is the measurement of a machine's level of dependence on humans while executing the Orient, Observe, Decide and Act (OODA) Loop.[500] According to OODA, when comparing two competing forces, the one which moves faster between these phases will control the initiative of the conflict, forcing the opponent to react rather than initiate. While the 'loop' is not a clean linear process as it includes constant feedback and integration among the different stages, the OODA Loop concept allows for relatively straightforward comparisons of systems based upon their technological capabilities.[501] The greater a machine's ability to observe, orient, decide, and act on its own, the greater its autonomy. Three factors influence the degree to which a machine is considered to be

[499] 'The Policy and Law of Lethal Autonomy with Michael Meier and Shawn Steene' <https://madsciblog.tradoc.army.mil/305-the-convergence-the-policy-and-law-of-lethal-autonomy-with-michael-meier-and-shawn-steene/>; Interview with Shawn Steene (n 256).
[500] W Marra and S McNeil, 'Understanding "the Loop" Regulating the Next Generation of War Machines' (2013) 36 Harvard Journal of Law and Public Policy 1151.
[501] ibid 1146.

automatic, automated or autonomous:[502] (1) the frequency of operator interaction that the machine requires to function; (2) the machine's ability to execute tasks despite environmental uncertainty; and (3) the machine's level of responsiveness regarding various operational decisions that allow the machine to complete its mission.[503]

Yet the US DoD conceptualisation of autonomy is not a universally accepted approach in academic literature and policy discourse. Some argue that the application of advanced AI, such as ML techniques, is what makes *real autonomy* possible, not the independence from human as such. In this context, authors generally distinguish autonomous systems from both 'automatic' and 'automated' weapons.[504] The term 'automatic' is usually associated with simple, mechanical responses to environmental input. The term 'automated' refers to more complex, rule-based systems, while the term 'autonomous' is reserved for machines that execute self-learning ability.[505] In this context, AWS are are based on advanced AI such as ML, which enables systems to solve various problems by learning and improving from experience without being explicitly programmed to do so.[506] I deliberately refer to ML as an advanced AI, because authors often adopt a very broad definition of AI which also includes rule-based systems.

The difference between these two types of systems is that a rule-based solution will analyse the inputs to predict whether a given output can be achieved or not based on a set of if-then style rules, while an ML system will learn from user inputs and data from the environment to predict the best possible outcome for a given scenario.[507] Both solutions can be classified as types of AI, as they mimic human intelligence through different means.

---

[502] See Paul Scharre and Michael Horowitz, 'An Introduction to Autonomy in Weapon Systems' (CNAS 2015).
[503] Marra and McNeil (n 500) 1151.
[504] Paul Scharre and Michael Horowitz (n 502).
[505] ibid.
[506] Tom Mitchell, *Machine Learning* (McGraw-Hill Education 1997).
[507] Harry Surden, 'Artificial Intelligence and Law: An Overview' (2019) 35 Georgia State University Law Review 1305.

The advantage of a rule-based system is that it captures the way people tend to reason given a set of known facts and their knowledge about the particular problem domain. The advantage of an ML system is that it models the associations between inputs and outputs where people are less certain about their specific connections. The application of ML allows a machine to exhibit self-learning capabilities, in other words to adjust to various scenarios and improve its performance. Some authors, therefore, even within US DoD, argue that autonomy is inherently related to ML.[508] On the other hand, according to the mainline US DoD conceptualisation of autonomy, there might be weapon systems which are autonomous but based on simple rule-based systems – only because they are independent from a human operator.[509] In this respect, Directive 3000.09 does not seem to distinguish the concept of 'autonomy' from that of 'automation' in weapon systems.

The US DoD mainline conceptualisation of autonomy has two implications. First, the levels of autonomy framework assumes that increases in automation must come at the cost of lowering direct human control. On the other hand, the conceptualisation of autonomy as advanced AI does not necessarily mean less direct human control. It could simply mean that the system is able to exhibit self-learning capabilities, but that a human can still supervise this process. In fact, human supervision can take various forms. It can relate specifically to the human decision to limit the use of such a system to target recognition, but not to target engagement. Humans also can supervise the way a model makes decisions. In fact, one of the dominant approaches in ML is called supervised learning, where the system learns an association between human labelled input data samples and corresponding outputs after performing multiple training data instances.[510] The

[508] David Gunning and others, 'DARPA's Explainable AI (XAI) Program: A Retrospective' (2021) 2 Applied AI Letters 1.
[509] Interview with Shawn Steene (n 256).
[510] Solveg Badillo and others, 'An Introduction to Machine Learning' (2020) 107 Clinical Pharmacology & Therapeutics 871. There are also solutions that allows for the automation of data labelling process.

learning process is then accelerated by human experts who supervise the model in real time. For example, if the ML model recognises a piece of data it is uncertain about, a human can be asked to assess it and give feedback. The model then learns from this input and uses it to make a more accurate prediction next time. Therefore, shifting the narrative from autonomy to AI considerations of weapon systems could liberate thinking by amplifying, augmenting, enhancing, and empowering the role of human operators, rather than replacing them.

Second, the US DoD mainline conceptualisation of autonomy generates a certain degree of confusion over which weapon systems should be classified as AWS. A good example are mines, which can be considered as 'automatic weapons.' Mines can 'sense' and 'act' on their own, but they have very limited methods for 'deciding' whether to fire or not. Mines are of course not self-learning systems, yet they can be considered as LAWS as they are independent of human operators and, unless specifically designed to self-deactivate, they can detonate long after an explosion is expected. Should mines then be considered as LAWS and go through a senior review according to Directive 3000.09? This is doubtful, even though drafters of the directive consider mines as autonomous.[511] Scharre therefore adds another condition to differentiate mines from other types of AWS: 'the ability to complete the engagement cycle – searching for, deciding to engage, and engaging targets on their own.'[512] While mines are generally not able to actively search for targets, there is an exception: the encapsulated torpedo mine, a special type of anti-submarine mine, which once activated, homes in on a target, similar to loitering munition. As discussed, the US Government in the past developed such mines, called US Mk 60 CAPTOR. Would

[511] Scharre (n 34) 50–52; ICRC, 'What You Need to Know about Autonomous Weapons' (26 July 2022) <https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons#:~:text=Mines%20can%20be%20considered%20rudimentary,anti%2Dpersonnel%20mines%20in%201997.>.
[512] Scharre (n 34) 52.

CAPTOR mines be classified as weapons that should go through a senior review process? Again, this is doubtful, as all existing weapon systems with various degree of autonomy have been legitimised by Directive 3000.09.

It is uncertain whether CAPTOR mines using a simple deep water sensor network[513] would fall under the senior review process, but it is certain that existing AI-augmented UAVs such as Reaper drones[514] do not go through this procedure, as they are considered as semi-autonomous weapons with humans acting as passive supervisors.[515] Yet one could argue that the use of such drones with self-learning capabilities generates greater challenges for militaries around the globe. For example, Reaper drones can be used for urban missions where civilians can be present. Their AI-augmented image recognition technology will collect data from the environment and recognise which targets should be engaged. Assume that the AI model that Reaper is using has suffered an adversarial ML attack, such as data poisoning, whereby malicious users inject false data with the aim of corrupting the learned model. Data plays a critical role in the security of an ML system. This is because an ML system learns directly from data. If an attacker can intentionally manipulate the data being used by an ML system in a coordinated fashion, the entire system can be compromised. As a result, Reaper drones may attack civilians or engage friendly forces. The risk of misjudging a target can be significant, irrespective of the fact that there is a human acting as a supervisor.

To be clear, I am not arguing that the US DoD problem representation of AWS does not consider the risk associated with the unintended consequences of a weapon's capabilities as such. As discussed in the previous chapter, a second layer of trust is that

[513] US Mk 60 CAPTOR torpedo mines used acoustic duct for sound propagation called the Reliable Acoustic Path.
[514] General Atomics Aeronautical, 'MQ-9A "Reaper"' <https://www.ga-asi.com/remotely-piloted-aircraft/mq-9a>.
[515] Interview with Senior Force Developer for Emerging Technologies at the Office of the Under Secretary of Defense for Policy (n 426).

there must also be trust in a machine's autonomous capabilities to produce predictable outcomes. However, US DoD Directive 3000.09 does not *specifically* address the challenges posed by AI capabilities of weapon systems. It is expected that the system's autonomous capabilities will exhibit a high level of 'reliability, effectiveness, and suitability under realistic conditions,'[516] but there is no requirement, for instance, to specifically address the risk of adversarial ML activity before the deployment of a weapon system. As it may appear that there is a substantial difference between, say, landmines and advanced AI-augmented drones, Directive 3000.09 grants 'a green light' to all existing uses of autonomy, including those which depend on advanced AI capabilities. The lack of focus on AI capabilities in Directive 3000.09 is problematic, as the USAF senior leaders have argued in their strategies for the development of AI-equipped AWS:

> The accelerated development of AI (...) will revolutionize the concept of autonomy. Whereas we view autonomous systems as those able to execute a set of pre-programmed functions, future systems will be better able to react to their environment and perform more situational-dependent tasks (...) with other autonomous systems.[517]

3.2.The Use of an Unbounded Notion of Autonomy

US DoD's assumption behind the problem construction of AWS is that, even though using AWS could result in unintended engagements, there should be no limits regarding the development of such weapon systems. As discussed in the previous section, the assumption is that current weapon systems are technically controllable, irrespective of their level of autonomy. As argued earlier, the requirement of human judgment can be satisfied even in the context of AWS used for *lethal purposes*. I have also discussed that, according to the

[516] Directive 3000.09 Autonomy in Weapon Systems Enclosure 3, 1b (2).
[517] Gen Mark Welsh III, USAF, 'America's Air Force: A Call to the Future' (n 101) 19.

US DoD's definition, there are already limited examples of weapon systems that can be categorised as LAWS, e.g. loitering munitions with autonomous targeting capabilities and weapons such as LRASMs. While US DoD representatives recognises the increased challenges of AWS, Directive's 3000.09 uses an unbounded notion of autonomy according to which there are not hard limitations to the application of autonomy in weapon systems.[518] This was illustrated, for example, by the US delegation during the UN GGE meeting:

> DoD Directive 3000.09's requirements that weapons be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force reflect a deliberate decision to permit weapons that are programmed to make "decisions" that relate to targeting.[519]

US DoD authorities set the potential development limits of AWS – if there are any limits at all – very low, particularly relative to the Campaign's discourse, which argues against the use of any weapons with autonomous targeting and engagement capabilities. One can interpret the US DoD discourse as deliberately leaving the door open to building and deploying even more sophisticated weapons than, say, autonomous loitering munitions, particularly as Directive 3000.09 does not stipulate any specific 'red-lines' regarding the use of autonomy in weapon systems.

Specifically, the US DoD problem representation of AWS does not clarify whether the use of autonomy will be only to achieve specific goals, such as the autonomous target engagement, or rather US DoD is interested in applying autonomy to all different stages of targeting process. Directive 3000.09 applies to the 'critical functions' of weapon systems –

[518] Defense Science Board, 'Report of the Task Force on the Role of Autonomy in DoD Systems' (n 114) 23.
[519] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

that is, to the process of selecting and engaging with a target. US DoD is nevertheless pursuing many other applications of autonomy across the wider targeting process, for example, for identifying targets or navigating the movement of military assets. Thus, some fear that the unbounded application of autonomy across a broad spectrum of tasks may lead one day to the emergence of a general-purpose weaponised machine which will take over the role of human soldiers. I explore this topic critically in Chapter 8, where I discuss how one can think about alternative problem representations of AWS.

4. A Summary of the Chapter

In this chapter I have analysed the assumptions behind the US DoD problematisation of AWS. I have discussed how the policy on AWS, institutionalised by Directive 3000.09, aims to strike a balance between safety considerations associated with the use of autonomy in weapon systems and the need to maintain the asymmetric combat advantage. I have argued that the requirement of human judgment from Directive 3000.09 fits well with this narrative, as it leaves the US military departments with a wide degree of discretion in implementing this requirement.

I have argued further that, although US DoD recognises the increased risks associated with the adoption of autonomy in weapon systems, the department does not consider that the delegation of lethal authority to AWS is necessarily illegal according to LOAC and the US domestic law. In this respect, US DoD discourse adopts a qualitatively different take on the legal problems of discrimination and responsibility relative to the Campaign's narrative. The US administration shifts the legal ramifications of these problems into a technical issue by arguing that existing AWS can be controlled primarily through technical measures such as T&E and software V&V. The potential novel legal and ethical challenges for human–machine controls arising from AWS are predominately, if not

exclusively, subjugated to technical expertise and military 'know-how'. Therefore, existing LAWS are not considered to be weapons that inherently render too excessive risks.

I have argued, also, that US DoD's understanding of 'autonomy' refers to the system's degree of independence from a human operator. An alternative approach to autonomy focuses on the machine's ability to exhibit self-learning capabilities based on advanced AI. The US DoD Directive 3000.09 concentrates only on the risks associated with autonomy when conceived as an independence from human operator, leaving considerations regarding AI-augmented weapons unaddressed. I have argued that the levels of autonomy framework adopted by US DoD assumes that increases in automation must come at the cost of lowering a direct human control. On the other hand, the conceptualisation of autonomy as advanced AI does not necessarily mean less direct human control, as it focuses on the self-learning capabilities of weapon systems. This recognition opens the way for an alternative problem representation of AWS, which is discussed in Chapter 8. I have argued, also, that the US DoD conceptualisation of autonomy generates a certain degree of confusion over which weapon systems should be classified as AWS, and I have discussed the contested status of mines. Finally, I have explored how the US DoD problem representation uses an unbounded notion of autonomy which does not prescribe any limitations in the application of autonomy to machines. While US DoD representatives are aware of the longer-term risks associated with the application of autonomy and AI to weapon systems, they argue that, as current AWS are technically controllable, future autonomous weapons will be too.

Chapter 6: Unmanning Human Control – A Genealogy of Lethal Autonomy

Chapter 6 deals with the third research sub-question. It examines how the application of lethal force in an autonomous way has come about as a policy problem for the US Government. The chapter examines the origins of the problem of the increased risk of unintended engagements of AWS.

The analysis draws on the Foucauldian genealogy approach and shows that the application of lethal force in an autonomous way, and the subsequent legitimisation of such practices by Directive 3000.09, has been the result of contingent turns of history, not the outcome of rationally inevitable trends. My examination focuses primarily on the evolution of practices within USAF, because the air service branch of the US military has played the most significant role in this genealogy and air operations continue to be at the forefront of delegating lethal authority to autonomous machines.

In my genealogical examination, I study three types of source: US DoD and USAF documents, academic literature, and data collected from interviews with senior US DoD and USAF officials. In terms of the US Government documents, I found many useful insights about US DoD approach in the Patriot System Performance report, published by the DSB, and the commentaries on it by academics and policy theorists. In the part related to the evolution of doctrine within USAF, I studied Air Force Doctrine Documents, Manuals, Unmanned Aircraft Systems Roadmaps, and commentaries, published primarily by the Air University. In my examination of the origins of delegating lethal authority to autonomous machines, I have relied on several excellent academic publications, including Katharine Chandler's work on the genealogical evolution of drone warfare and Madeleine Elish's work on the history of US drone operations.

The chapter is divided into three sections. The first section explores the most immediate events that led to the introduction of US DoD Directive 3000.09: the tragic experience with autonomous supervised systems called Patriots during the Operation Iraqi Freedom in 2003, when missiles shot down friendly planes.

In the second section, I trace how supervised control, despite technological deficiencies, has become employed in the application of a lethal force, as well as how the references to 'human control' have been erased from US DoD strategies. I argue that the 'connection' between supervised control and lethality opened the door towards the use of supervised AWS and semi-autonomous weapon systems for lethal purposes. Such weapons have become perceived as particularly effective means of achieving specific military and political goals.

In the third and final section, I have argued that the tipping point of this 'evolution' was the process called 'remote split' - the operation of lethal UAVs from a great distance by a remote controller. The widespread adoption of remote split operations introduced distance between the supervisor and the UAV; it removed human operators from the theatre of war; and exacerbated a difficult challenge regarding how to best allocate roles between automation and human in the wider decision-making process of complex socio-technical systems.

1. The Origin of the Problem of the Lethal Use of Autonomous Weapons

This section argues that the lethal use of AWS has initially been framed as problematic by US DoD due to the challenges associated with the operational control of humans in the use of lethal autonomous supervised weapons such as Patriots missiles. The autonomous use of force has been considered as inferior to human-operated weapons, particularly due to

technical deficiencies. Further, human operators did not trust some of the advanced robotic weapons operating in an automatic or autonomous mode. The DSB, after Patriot missiles accidents in 2003, clearly recommended the greater involvement of a human operator in future Patriot operations due to the significant risks associated with the use of supervised autonomous weapons.[520]

Despite these early warnings, the progress of autonomy within US DoD has accelerated over the years and various rules regarding the use of autonomy in weapon systems have been co-produced accordingly. The autonomous use of lethal force has become a major development goal within US DoD, and specifically USAF, while subsequent reports have gradually moved away from the requirement of human control over AWS. Directive 3000.09 was the result of a bottom-up initiative from the USDP due to a number of legal and ethical requests from various military branches that were uncertain whether they could incorporate some autonomous features into specific weapons.[521] Despite the DSB recommendation to strengthen human control measures over the use of advance robotic weapon systems,[522] the USDP came up with a policy on AWS which legitimised all existing use of autonomy in weapon systems and did not prohibit the use of AWS. The genealogy of the US DoD problematisation of AWS echoes the words of Foucault: 'What is found at the historical beginning of things is not the inviolable identity of their origin; it is the dissension of other things. It is disparity.'[523] In the early 2000s, US DoD officials recognised that a lack of tighter measures might result in the development and use of robotic weapon systems that are unsafe. The safety of military crews during supervised autonomous weapons operations was their main concern. Later, particularly

[520] Defense Science Board, 'Patriot System Performance' (n 302).
[521] Interview with Paul Scharre (n 256).
[522] Defense Science Board, 'Patriot System Performance' (n 302).
[523] Michel Foucault, 'Nietzsche, Genealogy, History', *Language, Counter-Memory, Practice: Selected Essays and Interviews* (Cornell University Press 1980) 142.

during the process of drafting Directive 3000.09 in 2011-2012, officials not only wanted to address safety concerns, but they were also worried that a lack of policy on AWS might constrain the research and development of new features of autonomous features of weapon systems.[524] US DoD started developing fully autonomous weapons, let alone supervised autonomous weapons; the problem of the autonomous use of lethal force became increasingly framed as the problem of the researchers and developers, who did not have a clear guidance regarding how to develop autonomy in weapon systems, rather than as a safety problem regarding control over the use of such weapons.

This genealogical evolution led to the creation of US DoD governmentality of AWS. The dominant mode of thinking among US DoD, and particularly USAF leaders, has been to foster the development and use of lethal AWS without any rigid limitations, particularly in the context of a direct human control. Yet different perspectives regarding the importance of direct human control,[525] or criticism of specific applications of autonomy in weapon systems,[526] or resistance among air pilots to hand over their job to AWS[527] are not entirely absent in US DoD. Rather, they form the bulk of 'subjugated knowledges' that exists as the marginalised perspective of some US military officials. In this section, I explore one such subjugated knowledge, which focuses on the research on HSC and behavioural psychology regarding human decision-making. I argue that US DoD problem construction of AWS has marginalised concerns regarding differences in 'thinking' between machines and humans, specifically that they have not appreciated enough the value

[524] Interview with Paul Scharre (n 256).
[525] C. Todd Lopez (n 3).
[526] For example, the NASCAI argues that only human beings can authorize employment of nuclear weapons. See Eric Schmidt, Robert Work, and others (n 405).
[527] Scharre (n 34) 61.

of human operators' deliberative thinking, which is different from the largely automatic 'thinking' of autonomous systems.

1.1.US DoD Lessons from Fatal Experiences with Patriots

US DoD officials claim that the policy on AWS has been a result of the recognition that the technology behind semi- and autonomous weapons, while increasingly useful in the modern theatre of war, is not yet fully operational and could create fatal accidents. The drafters of Directive 3000.09 have been particularly influenced by Patriot missiles fratricide during 'Operation Iraqi Freedom' in 2003.[528] As John Hawley and Anna Mares said 'in some respects Patriot provides a glimpse into the future of military systems and operations.'[529]

Patriot is a land-based air and missile defence system. It can operate in two different modes: semi-automatic and automatic. In the semi-automatic mode, Patriot is a human-in-the-loop system as human operator must authorize all target engagements. In the automatic mode, on the other hand, the Patriot system will fire unless a human operator stops engagement. Patriot in the automatic mode was one of the first tactical systems in US DoD's inventory to employ what has been termed within US DoD as 'lethal autonomy' in combat. 'Lethal autonomy' refers to a system capable of applying lethal force with little or minimal *direct human oversight*. According to the US DoD Directive 3000.09, Patriots in the automatic mode are considered as AWS which are *supervised*, that is, a human is on the loop.

[528] Defense Science Board, 'Patriot System Performance' (n 302). See also Dan Saxon (n 7) 90.
[529] John K. Hawley and Anna L. Mares, 'Human Performance Challenges for the Future Force: Lessons from Patriot after the Second Gulf War', *Designing Soldier Systems Current Issues in Human Factors* (1st Edition, Ashgate 2012).

The first automatic use of Patriot occurred during Operation Desert Storm in the early 1990s. Despite some 'anecdotal reports' of classification and identification problems which might have affected the attribution of a target, the use of Patriots during Operation Desert Storm was considered a success within the US Army.[530] Hawley, an engineering psychologist with the US Army Research Laboratory's Human Research and Engineering Directorate and who has extensive experience in Patriot operations, commented:

> The Army left Desert Storm very full of itself regarding Patriot and its capabilities. Self-congratulation led to complacency, which led to unwarranted trust in, and reliance on, the system's automatic operating mode.[531]

The role of Patriot in automatic mode in the Second Gulf War in 2003 during Operation Iraqi Freedom was also significant. In total, the US deployed up to 40 fire units, while allied countries deployed 22 fire units.[532] Patriot was involved in three fratricide incidents.[533] Hawley, who has been studying Patriots since the late 1970s, was not surprised by the fatal incidents:

> Those outcomes had been in the card deck, so to speak, ever since they first flipped the engagement mode switch to automatic and assumed that was all there was to the conduct of near-autonomous operations. The fratricides were incidents waiting to happen.[534]

After the first automatic use of Patriot during Operation Desert Storm, Hawley wrote in a 1992 article about potential problems with Patriot's automatic mode. He noted that Patriot's automatic mode had been adapted from the engagement control logic of the Safeguard

[530] Defense Science Board, 'Patriot System Performance' (n 302).
[531] John K. Hawley (n 301).
[532] Defense Science Board, 'Patriot System Performance' (n 302).
[533] ibid.
[534] John K. Hawley (n 301).

system.[535] Safeguard was the first operational US anti-ballistic missile system and was deployed briefly in the 1970s. Hawley argued that 'the Safeguard level of automation' was not an appropriate mode for Patriot's operating environment, particularly due to the greater potential for track classification and identification mistakes.[536] The main problem with the Patriot's automatic mode is that the system work in 'an all-or-nothing fashion.' There are only a few 'decision leverage points' that allow the operators to influence the system's engagement logic and exercise supervisory control over a mostly automated engagement process.[537] Hawley argued that this automatic mode feature was not suitable for conventional air threats, as the machine was unable to 'handle unusual or ambiguous tactical situations reliably.'[538]

Furthermore, Patriot itself is a complex system. One of the approaches to measure the size of a software's complexity is by counting the number of lines included in a program and Patriot employs more than 3.5million such lines. In addition, the Patriot system is not a standalone weapon: it is a 'system of systems' as it requires coordination with other systems for air battle management, composing what is termed an 'integrated air and missile defense system.' [539] These associated systems include warning and control systems, sea-based missile defence systems, and various space-based systems.[540]

Despite these challenges, the US Army was influenced by the success of Operation Desert Storm and started to use Patriot in the automatic mode as 'a preferred operating concept.'[541] For example, they decided to reduce the experience level of their operating

[535] ibid.
[536] ibid.
[537] ibid.
[538] ibid.
[539] John K. Hawley and Anna L. Mares (n 529).
[540] ibid.
[541] John K. Hawley (n 301).

crews and the amount of training provided to individual operators and crews. That is why the fratricide events were 'incidents waiting to happen.'[542]

The US Army Research Laboratory commissioned the Patriot System Performance report investigating the causes of these accidents that occurred in 2003 at the request of Maj Gen Michael Vane.[543] The report contained three conclusions: First, the Patriot system's ability to identify and distinguish object was 'very poor.' This deficiency had been observed during many training exercises and had never been fixed before fielding the weapon. The report stated: 'The Task Force remains puzzled as to why this deficiency never garners enough resolve and support to result in a robust fix.'[544]

Second, there was insufficient communication and coordination between the missile batteries and other systems deployed in the field. Again, the report stated:

> We tend to assume that data are routinely communicated from one system to the other, that targets are correlated, and target information is shared and assimilated by all. The Task Force believes that we are a long way from that vision.[545]

It turned out that the communication links were absent, and Patriot batteries were only able to communicate with their headquarters unit, but even that connection was sometimes weak.

Third, the procedures for human-machine interfaces were poorly designed and the operating philosophy for Patriots relied too much on automation, despite the software deficiencies.[546] The operating philosophy was primarily based on the passive supervision

[542] ibid.
[543] John K. Hawley and Anna L. Mares (n 529).
[544] Defense Science Board, 'Patriot System Performance' (n 302).
[545] ibid 2.
[546] Defense Science Board, 'Patriot System Performance' (n 302).

of human operators who were trained to trust the Patriot software with little room left for active human control.

These findings stimulated discussion within US DoD to address the problem of the growing autonomy incorporated into lethal weapon systems and the role of humans and machine in operating such systems. The major DSB recommendation was to introduce *more human operator involvement and control* in the functioning of a Patriot battery, which should follow changes in software, computer displays, and training of human operators.[547] However, even though the DSB findings recommended the greater involvement of a human operator in the future Patriot operations, US DoD was still so convinced of the Patriot system's successes that they did not want to challenge the use of the system in an automatic mode. Instead, US DoD, framed any difficulties as a purely technical, software problem: 'The claim was repeatedly made that a "technical fix" … was just around the corner,'[548] said Hawley and Mares.

Furthermore, and despite recurring warnings, after Operation Desert Storm, the US Army reduced the experience level of their operating crews and the amount of training provided to individual operators and crews, and even after the findings of the DSB report, 'they still have not fully corrected many of these deficiencies.'[549]

1.2.A Move Away from Direct Human Control towards the Notion of Human Judgment

My examination has revealed that US DoD, particularly USAF, did not adopt the DSB findings regarding the greater involvement of human operators. They did the opposite by

[547] Ibid.
[548] John K. Hawley and Anna L. Mares (n 529) 7.
[549] John K. Hawley and Anna L. Mares (n 529).

gradually shifting from direct human control to human-machine integration and human judgment.

Since 1987, automated air and missile defence systems such as Patriot have operated under Title 10 of the US Code, the section regarding weapons development and procurement. Section 226 of Pub. L. 100-108 provided the following rule:

> No agency of the Federal Government may plan for, fund, or otherwise support the development of command-and-control systems for strategic defense in the boost or post-boost phase against ballistic missile threats that would permit such strategic defenses to initiate the directing of damaging or lethal fire except by *affirmative human decision* [MF emphasised] at an appropriate level of authority.[550]

This rule, which has not changed until very recently (2021),[551] assumed that highly automated or AWS must operate under the previously discussed concept of 'direct human control.'[552]

Yet, according to Hawley's first-hand experience, the requirement of direct human control has had little impact on air and missile defence system development or operations. According to him, part of the problem is that the concept has not been well defined, and US DoD has often confidently asserted that the requirement has been met, even if that means only a passive supervision by humans (human-on-the-loop) over the process of

[550] National Defense Authorization Act for Fiscal Years 1988 and 1989 1987.
[551] The rule has been initially in the United States Code, Title 10, Subtitle A, Chapter 144, Sec. 2431 - Weapons development and procurement schedules. In 2021 Pub. L. 116–283, § 1846(h)(2), amended section generally. Pub. L. 116–283, § 1846(h)(1), renumbered section §2431 for §4205 and removed the reference to 'affirmative human decision.'
[552] It is also called 'positive human control.' See John K. Hawley (n 301).

engagement. '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.'[553]

In fact, the use of Patriot and some other semi-autonomous weapon systems has allowed US DoD to transition from an operating system of control based on a human in the loop (active supervision) to a system of human on the loop (passive) in certain operations such as air missile defence. One would expect that a military crew and specifically weapons' operators, would have been thoroughly trained and prepared for such a revolutionary change. Yet this was not the case.

> […] the Army had committed all the classic "sins" associated with the development and use of automated systems. They had trusted the system in a naïve manner; they had not adequately prepared their operators and crews for proper oversight of automated operations; and they had been unwilling or unable to confront the fact that near-autonomous operations are qualitatively different from old-style manual control (i.e., "on-the-loop" versus "in-the-loop" control).[554]

The training before the Patriot fratricides focused on 'individual tasks and components' rather than a holistic understanding of how the weapons operate in complex environments.[555] Yet US DoD has continued to push the agenda for greater autonomy of weapon systems to the extent that they have started moving away from any reference to 'affirmative,' 'direct,' or 'positive' human control. If there have been occasionally any references to human control, they have almost exclusively been in a negative sense. For example, human control has been portrayed as a kind of barrier that stalls the development of autonomy in weapon systems and should be 'reduced' or 'moved away.' A series of US

[553] ibid 9.
[554] John K. Hawley (n 301).
[555] Dan Saxon (n 7) 191.

policy roadmaps and strategies between 2005 and 2011 concentrated more on the importance of 'effective integration' of MUM systems rather than on direct human control.[556] The Unmanned Aircraft Systems Roadmap of 2005 was the last report that explicitly assumed the superiority of human control. In the 2005 report, one could still read that 'pattern recognition by software today is generally inferior to that of a human.'[557]

In the following reports, published by US DoD or USAF, the sentiment regarding the state of technology is much more optimistic. The 2005 report concentrated on the 'human computer interface' (HCI), a theme which has been replaced in the following reports by the term 'human systems integration'. The assumption behind HCI was as follows: the design of weapon systems' design must be reliable and effective. In order to achieve this objective, there must be a human able to effectively interact with the system. The following passage provides a good illustration:

> Operators, administrators, and maintainers interact with software-based information systems using the system's HCI. The HCI includes the appearance and behavior of the interface, physical interaction devices, graphical interaction objects, and other human-computer interaction methods.[558]

The concept of 'human systems integration' differs from HCI. The emphasis is less on a human's direct interface with a system, but rather on a weapon's system design that should accommodate a human user. Since 2007, USAF reports have focused on 'human systems integration' whereby human considerations, including human capabilities and limitations,

[556] See US DoD, 'Unmanned Aircraft Systems Roadmap 2005-2030' (2005); US DoD, 'Unmanned Systems Integrated Roadmap FY2007–2032' (n 43); USAF, 'Unmanned Aircraft Systems Integrated Roadmap FY2009-2034' (2009); USAF, 'Unmanned Aircraft Systems Flight Plan FY2009-2047' (2009); US DoD, 'Unmanned Systems Integrated Roadmap FY2011-2036' (n 91); US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (n 91); USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479).
[557] US DoD, 'Unmanned Aircraft Systems Roadmap 2005-2030' (n 556) 52.
[558] ibid APPENDIX E, E-14.

should be integrated into the engineering system development, design, and management. A good illustration of this concept is an excerpt from The Unmanned Aircraft Systems Flight Plan 2009, which presents how human systems integration can support a human-on-the-loop approach rather than direct human control that stems from the HCI:

> The systems' programming will be based on human intent, with humans monitoring the execution of operations and retaining the ability to override the system or change the level of autonomy instantaneously during the mission.[559]

Notably, US DoD Directive 5000.01, the directive that describes the principles governing the US DoD acquisition processes has also been updated to include human systems integration. In 2015, the Directive 5000.01 was updated to 'focus on the integration of human considerations into the system acquisition process to enhance *soldier-system design* […].'[560] The term 'human systems integration' replaced the concept formerly known as Manpower and Personnel Integration (MAPRINT) which, similarly to HCI, emphasised more the active role of personnel in the weapon's acquisition and operation. As the wording suggests, the focus of MAPRINT was primarily on the integration of manpower and personnel rather than on building a soldier-system design, in other words, a deeper integration of human factors at the level of the weapon system's design.

The transition from HCI to human systems integration coincided with the gradual removal of the concept of direct human control over the development and use of weapon systems. In the Unmanned Aircraft Systems Flight Plan 2009, there was even an explicit reasoning against the rule of 'direct human control' from Title 10 of the US Code:

[559] USAF, 'Unmanned Aircraft Systems Flight Plan FY2009-2047' (n 556) 41.
[560] Army Regulation 602–2 Human Systems Integration in the System Acquisition Process 2015.

> […] advances in [artificial intelligence] will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input.[561]

Another USAF document from the same year also stated a goal of moving away from a direct human control:

> First and foremost, the level of autonomy should continue to progress from today's fairly high level of human control/intervention to a high level of autonomous tactical behavior that enables more timely and informed human oversight.[562]

Two years later, in 2011 roadmap, the move away from human control had been leveraged as the US DoD's key objective. The 'Vision' section at the very beginning of the document stated:

> DOD envisions unmanned systems seamlessly operating with manned systems while gradually reducing the degree of human control and decision making required for the unmanned portion of the force structure.[563]

While the 2011 Roadmap suggested that in the near future, decisions regarding the use of force and the choice of which individual targets to engage with lethal force would be retained under human control, the 2013 Roadmap asserted that 'development in automation are advancing toward […] a state of autonomous systems able to make decisions and react without human interaction.'[564]

All these declarations, as well as rules codified Directive 3000.09 and Directive 5000.01, support the concept of human judgment rather than human control over the use of AWS.

[561] USAF, 'Unmanned Aircraft Systems Flight Plan FY2009-2047' (n 556) 41.
[562] USAF, 'Unmanned Aircraft Systems Integrated Roadmap FY2009-2034' (n 556) 27.
[563] US DoD, 'Unmanned Systems Integrated Roadmap FY2011-2036' (n 91) 3.
[564] US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (n 91) 15.

Yet the explicit reference to 'affirmative human control' from Title 10 of the US Code has remained unchanged until 2021.

This legal ambiguity, particularly between the reference to human judgment in Directive 3000.09 and the reference to affirmative human control from Title 10 of the US Code, links back to the topics discussed earlier about the status of Directive 3000.09 in the US legal system. It is unclear whether specific rules from the Directive 3000.09 should be treated as having the 'force and effect of law' or whether they merely 'advise the public of the manner in which the agency proposes to exercise a discretionary power'. From a governmentality perspective, however, what is at stake is not the positivist interpretation of whether rules conflict with each other, and if so, which one should be treated as superior. A governmentality approach asks what effects these various rules have on regulated subjects, in other words, whether Title 10 of the US Code or Directive 3000.09 – and indeed the concept of human control or judgment – represent a guide of conduct for regulated subjects, such as human operators. This question will be investigated in Chapter 7.

Leaving aside Title 10 of the US Code, the internal production of administrative rules by US DoD and USAF between 2005 and 2011 regarding the use of autonomy in weapon systems has evolved to move considerably away from the requirement of direct human control over AWS. The publication of these rules led to 'a tipping point within the DoD.'[565] The USDP has received questions about the legal and ethical issues associated with the use of increasingly autonomous weapon systems.[566] Different US DoD military branches have held conflicting positions regarding the adoption of autonomy in weapon systems. The US Army initially opposed it by claiming that 'they will never delegate use-of-force decisions to a robot,' while USAF was the biggest proponent of greater

[565] Dan Saxon (n 7) 195.
[566] ibid.

autonomy.[567] Researchers and developers within US DoD were increasingly hesitant to develop autonomous functions of weapons without having a guideline of what was permissible. Military practitioners were also confused as they did not trust the safety features of some of the advanced weapon systems which reached the field.[568] By 2011, it was evident within US DoD that the clearer guidance for the development and use of AWS was needed.[569]

Two main reasons motivated USDP to produce Directive 3000.09. First, US DoD recognised that the lack of policy could lead to the development and use of weapons which were unsafe, and this was particularly critical with respect to the lethal use of such weapons.[570] Hence, Directive 3000.09 established a more complex procedure for the use of such weapons. The second reason was to provide more clarity for researchers and developers so they could incorporate various autonomous functions in weapons within defined legal and ethical boundaries. Thus, the Directive's 3000.09 authors aimed at providing 'neutral language and parameters' for the design intended to encourage the development of new autonomous weapons.[571] As discussed earlier, Directive 3000.09 did not include the requirement of human control over the use of weapon systems. Rather, the drafters of the directive employed the term 'appropriate levels of human judgment' and left this concept undefined. During the UN GGE meetings on LAWS, US DoD delegation explained the difference between these two terms in the following words:

> "Human judgment over the use of force" is distinct from human control over the weapon. For example, an operator might be able to exercise meaningful control over every aspect of

[567] ibid.
[568] Stew Magnuson, 'Armed Robots Sidelined in Iraqi Fight' *National Defense* (1 May 2008) <https://www.nationaldefensemagazine.org/articles/2008/5/1/2008may-armed-robots-sidelined-in-iraqi-fight>.
[569] Dan Saxon (n 7) 196.
[570] ibid 196–197.
[571] ibid 198–199.

> a weapon system, but if the operator is only reflexively pressing a button to approve strikes recommended by the weapon system, the operator would be exercising little, if any, judgment over the use of force. On the other hand, judgment can be implemented through the use of automation. For example, […] the use of algorithms or even autonomous functions that take control away from human operators can better effect human intentions and avoid accidents.[572]

The US delegation distanced itself from the notion of human control because they thought that framing the debate regarding the use of LAWS on 'control' was too restrictive and might imply so-called *direct* human control. As I have argued in Chapter 5, one of the assumptions behind Directive 3000.09 was there should be no limits in principle regarding the technical development of AWS. The directive phrase of 'appropriate levels of human judgment over the use of force' leaves the possibility open to exercise no human involvement at all during the target engagement process. In this respect, Directive 3000.09 has codified US DoD's gradual move away from direct human control over the use of AWS.

The increasing automation of weapon systems and the gradual removal of human control is often portrayed as an inevitable 'advancement.' As already cited earlier, the US DoD Roadmap states:

> In general, research and development in automation are advancing from a state of automatic systems requiring human control toward a state of autonomous systems able to make decisions and react without human interaction.[573]

[572] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).
[573] US DoD, 'Unmanned Systems Integrated Roadmap FY2013–2038' (n 91) 15.

As pointed out by Lucy Suchman and Jutta Weber, the term 'in general' implies not only a tendency, but also a kind of inevitability.[574] While the roadmap acknowledges that at present AWS have limited operational use, the implication is that this is just a temporary state, not the end of 'progress.' This 'dominant knowledge' within US DoD, specifically USAF, has subjugated some alternative perspectives which challenge the inevitability of growing autonomy in weapon systems and the diminishing role of direct human control.

1.3.The Subjugated Knowledge of Behavioural Psychology of Decision-Making

What is striking in the genealogical account of the problem of using lethal force in an autonomous way is US DoD's overoptimism regarding the capabilities of increasingly autonomous weapon systems post-Operation Desert Storm as well as its disregard of the DSB, Hawley, and other military practitioners' recommendations to increase the involvement of a human operator in future operations of weapons in automatic or autonomous mode. In fact, one could argue that US DoD leadership has not only disregarded the individual opinions of selected military practitioners, but a whole bunch of 'knowledge' – that is, in Foucauldian language, a network of relationships that allows certain statements to be accepted as 'true' in a specific context.[575] These 'subjugated knowledges' consist of alternative perspectives that challenge US DoD consensus regarding the diminishing prominence of direct human control over the use of weapon systems. For example, Hawley's argument that the Patriot's automatic mode is not appropriate for all operating environments, such as conventional air threats, and in such situations human operators should make engagement decisions. Or more fundamental observations, for example, that certain weapon systems require *more human operator*

[574] Lucy Suchman and Jutta Weber, 'Human-Machine Autonomies', *Autonomous Weapons Systems* (Cambridge University Press 2016).
[575] Michel Foucault, *Discipline & Punish: The Birth of the Prison* (n 107).

involvement and control rather than more autonomy in order to avoid fatal consequences; and that automation should not lessen operator training requirements. In short, these various statements elevate the role of human operators and legitimises their continuous significance. This is contrary to the dominant US DoD, specifically USAF discourse, according to which human pilots are becoming less and less important on the modern battlefield.

Perspectives of USAF pilots illustrate this point. Marry Cummings, one of the first female US Navy fighter pilot and currently a professor in the Duke University Pratt School of Engineering, argues that the current generation of pilots go through a process like 'every group of people who ever had their job automated went through.'[576] She stresses the perspective of some of the pilots which is absent in the governmental roadmaps and strategies. 'The pilot has that image as sort of the last bastion of derring-do and perceives [piloting] skills as irreplaceable. We have an emotional attachment to the idea of being a pilot that is very hard to lose.'[577] Similarly, Lt Gen David Deptula, a retired F-15 pilot, was once a supporter of greater use of UAVs, but later grew dissatisfied. He argues that drones often require more human involvement given the fact that someone must analyse the data and constantly stay in the loop. This does not necessarily mean that, because an airplane can operate remotely, US DoD should use it this way. 'The essence of conflict warfare is a very human activity,' he says. 'If it was a science, we could turn it over to machine-to-machine, and whoever had the best algorithms would win.'[578] Further, a retired USAF Gen Norton Schwartz, who served as the USAF Chief of Staff, argues that 'many pilots feel the USAF stumbled by pushing them into a field they never signed up for.'[579] These comments

[576] Michael Milstein, 'Pilot Not Included' *Smithsonian Magazine* (July 2011).
[577] ibid.
[578] ibid.
[579] ibid.

are not just isolated statements; they form a whole series of subjugated knowledges that has been disqualified as inferior and has been insufficiently elaborated to undermine the dominant position in favour of growing autonomy in weapon systems and removal of direct human control.

These statements not only come from the direct experiences of USAF military practitioners, but they are also informed by a bulk of literature that has largely been disregarded by US DoD in their problem construction of AWS: research on HSC and dual processing.[580] In dual processing research, authors such as Daniel Kahneman, argues that there are two types of reasoning: System 1 and System 2. System 1 thinking happens fast and automatically, while System 2 refers to controlled, more deliberative processes. System 2, the deliberative process, often comes after System 1 thinking and thus is slower.[581] Sharkey points out that both types of reasoning have relevance to weapon systems, and both have advantages and disadvantages.[582] The advantage of automatic decision-making lies particularly in retinue tasks which require fast reactions to predictable events. However, automatic reasoning brings many disadvantages such as the neglect of ambiguity, confirmation bias, and misattribution of causes and intentions. Therefore, automatic decision making should coexist with deliberative processes. Sharky points out that, in many current human supervisory weapon systems in the US, the condition of deliberative processing is not met.[583] Specifically, he refers to passive supervised AWS (semi-autonomous weapon systems) where humans are on-the-loop and do not directly control the trajectory of the engagement. According to Sharkey, such weapon systems would only

[580] Noel Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2006) 30.
[581] Daniel Kahneman, *Thinking, Fast and Slow* (Penguin 2011).
[582] Noel Sharkey (n 549) 32.
[583] ibid 28.

reinforce automation bias and leave little room for deliberation, if any.[584] The Patriot fratricides in Iraq in 2004 illustrates this point. The operators were given 10 seconds to veto a computer solution and they did not react accordingly, in part because displays were confusing and often incorrect, but also because the operators lacked training in dynamic complex operations.[585]

Further, according to Cummings, some of the established supervised systems with a human-in-the-loop also do not meet the criteria of deliberative reasoning.[586] It is interesting to note however that Cummings is a member of DIB, a federal advisory committee within US DoD, tasked with providing 'independent recommendations on emerging technologies to the senior DoD leaders' and working with 'sponsors' inside US DoD to take action to implement these recommendations.[587] Cummings' perspective is not just an example of marginalised knowledge; it is also an example of what Foucault called as 'counter-conduct' – that is, a visible act of dissent, a struggle against the dominant form of governmentality. As pointed out by Foucault, 'points of resistance are present everywhere in the power network' and operates at the microlevel, for example in the form of disruption of the dominant narrative.[588] Cummings, by being integrated within internal US DoD structures, simultaneously challenges, and reinforces dominant power relations. She challenges US DoD dominant narrative on AWS as she argues for limits regarding the use of such weapons and supports the requirement of direct human control. Yet she

[584] ibid 32-34.
[585] Defense Science Board, 'Patriot System Performance' (n 302).
[586] Mary Cummings, 'Automation and Accountability in Decision Support System Interface Design' [2006] Journal of Technology Studies.
[587] *Carter Provides Remarks at Defense Innovation Board Meeting* (2016) <https://www.dvidshub.net/video/486245/carter-provides-remarks-defense-innovation-board-meeting>.
[588] Michel Foucault, *The History of Sexuality. Volume I: An Introduction* (Vintage Books 1990) 95.

reinforces the dominant view as she is nonetheless part of US DoD, and despite her criticism, she legitimises their work by being affiliated with the department.

2. A Genealogy of Delegating the Exercise of Lethal Force to an Autonomous System

It is peculiar how US DoD and the Campaign's discourses about the problem of applying lethal force in an autonomous way have focused on potential future weapons, while there are already existing AWS that are in use, and some of them, such as Patriots and other supervised AWS, long predates Directive 3000.09. In the wider academic literature such weapons have long been credited as introducing a 'different type of lethality,' undermining traditional human controls.[589] More strikingly, the introduction of the first mass-produced US weapon system with autonomous functionalities in the engagement area took place during the Second World War in the form of an air-dropped, passive acoustic homing torpedo called Mark-24 'Fido.'[590] This means that the important transformations regarding human-machine controls have occurred relatively unnoticed by wider public opinion and have only recently started to be exposed and investigated in the wake of increasing public pressure from the Campaign and later Future of Life Institute. This observation is also consistent with the rationale behind the drafting process of Directive 3000.09, as discussed in an earlier section.

This stage of the analysis asks the question of how this specific representation of the problem of delegating lethality to a supervised autonomous weapon has come about.

[589] Thom Shanker and James Risen, '"Raid's Aftermath: U.S. Troops Search for Clues to Victims of Missile Strike' *The New York Times* (11 February 2002) <https://www.nytimes.com/2002/02/11/world/nation-challenged-raid-s-aftermath-us-troops-search-for-clues-victims-missile.html>.
[590] E. W. Jolie, 'A Brief History of U.S. Navy Torpedo Development' (1978) NUSC Technical Document 5436.

The purpose of this analysis is to examine the origins of the US DoD problem representations of AWS and associated changes in human-machine controls.

In the academic literature, the evolution of human–machine controls over weapons usually begins with a human's exercise of direct action with limbs, via tools and powered control.[591] The notion of direct human control over a weapon has gradually evolved through various transformations. Over time, new developments of weapons have led to an increasing indirectness of the relationship between the human and the controlled item, with a decrease in required muscular strength and an increase in the role of sensing and thought. A first significant transformation was the *introduction of supervisory control*. Active supervisory control is when a human actively executes a task, for instance flying a vehicle, and the endeavour is mediated by a computer somewhere in the process. With manual control, the operator interacts directly with the task.[592] Supervisory control has introduced a distance between the operator and the weapon's effects and laid the foundations for the possibility of drones, which are today commonly described as *unmanned* aerial vehicles. In the academic literature, armed drones are sometimes credited as predecessors of AWS due to their ability to process real-time information and use for targeted killings. Autonomous control is in turn considered to be the next phase of development after the remote control of drone operators. Hawk Carlisle, a retired USAF general, echoed this sentiment when asked whether the ability to extend an aircraft's reach with AI-augmented wingmen was the next step for air combat:

[591] Thompson Chengeta (n 41) 839–846.
[592] Norm Haller, *Human-Automation Interaction Considerations for Unmanned Aerial System Integration into the National Airspace System* (The National Academies Press 2018) 27.

> This is a natural evolution, especially when you look at the capability today with respect to AI, with respect to systems, with respect to the computing power and capability you can put in a particular size.[593]

This section concentrates specifically on the emergence of supervisory control and the connection of this control with a lethal use of weapons. I argue that this 'connection' effectively allowed lethal operations to be conducted at a distance with remote control and which have gradually shifted the role of human operators to more passive supervisors. The initial control architecture based on the human-in-the-loop has been gradually replaced in many weapon systems to human-on-the loop, where humans are merely monitoring the weapon's engagement. This development has brought about a breakthrough in military affairs, which has allowed further applications of the use of force in an autonomous way and has paved the way towards LAWS. This rationale was driven by the perceived view of US DoD that such weapons are successful means to achieve military and political objectives, as they reduce the risk to human operators and are effective in carrying out specific tasks, such as targeted killings.

In contrast to the popular narrative, I argue that the connection between supervisory control and the lethal use of force has not been a linear evolution, but rather a series of experiments, often with mixed results, and that the process has been ultimately legitimised during the War on Terror with the use of targeted killings against terrorists, much before Directive 3000.09. This 'connection' would also not have been possible without several other technical breakthroughs that have occurred since the early 1960s, such as advancements in microprocessors, precision, robotics, to name a few, all of which allowed

[593] Stephen Losey, 'How Autonomous Wingmen Will Help Fighter Pilots in the next War' [2022] *DefenseNews* <https://www.defensenews.com/air/2022/02/13/how-autonomous-wingmen-will-help-fighter-pilots-in-the-next-war/>.

real-time intelligence collection and analysis, satellite data-link connection, or precision munitions.[594] While appreciating that these innovations have played a critical role in the development of unmanned lethal systems, I will trace back only the 'connection' between supervisory control and the unmanned lethal use of force.

2.1. The Emergence of Remote Control: 'Urgent Need for Radio-Controlled Aircraft for Use as an Aerial Target'

Remote control allows the operation of devices that are out of convenient reach for the direct operation of controls. In the air domain, Archibald Low, 'a father of radio guidance systems,' demonstrated already in 1916 a remote-controlled aircraft, called 'aerial torpedo'.[595] The first remote-controlled model airplane flew later in 1932, so what the public today calls 'drones' were far from the first unmanned systems built and used by militaries, although they were the first to use the term.[596] The label 'Drones' was the name of a US Navy project to build a radio-controlled aircraft in 1936 called NT Drone.[597] The first drones were not intended to serve as offensive weapons, but they were conceived as training targets. They were used for training personnel in anti-aircraft activities under conditions closely simulating action to counter aerial bombardment.[598] The drone-as-a-target is distinct from the contemporary use of drones, such as Predators, which are equipped with a camera and missile onboard.[599] For the US personnel trained in the 2000s to operate unmanned aircraft, the dissimilarity was still fundamental. Predator operators

[594] Michael Kreuzer, *Drones and the Future of Air Warfare* (Routledge 2017) 2–11.
[595] Jonathan Sale, 'The Secret History of Drones' *The Guardian* (10 February 2013) <https://www.theguardian.com/world/shortcuts/2013/feb/10/secret-history-of-drones-1916>.
[596] Katherine Chandler (n 104) 12..
[597] ibid 16.
[598] ibid 17.
[599] General Atomics Aeronautical, 'Predator XP' <https://www.ga-asi.com/remotely-piloted-aircraft/predator-xp>.

wore patches with the motto 'We're not drones... We shoot back'[600] to differentiate the systems they used from the drones dedicated to air defence training.

This disjuncture suggests that the mere introduction of supervisory control did not necessarily change the nature of lethality, at least not initially. After all, it was still the pilot in the aircraft who made a final targeting decision. In other words, the first drone flights were an extension of manned flight, rather than an ontological break with it. As argued by Chandler 'the violent consequences of unmanning cannot be attributed to mere mechanization and technological advance.'[601]

2.2.A Shift from Aerial Target and Reconnaissance to Lethal Weapons

After initial testing of drones as aerial targets, attitudes shifted, and the Chief of the US Navy argued in 1939 that it was 'reasonable technology development' to expand the role of the radio-controlled airplane from a passive one as a target to an active one as an offensive weapon.[602] The Navy officially began its Assault Drone Program in March 1940, but this did not last long. It concluded in 1944 and was deemed a failure.[603] The use of television-guided drones named *American Kamikaze* was meant to promote the US superior technology to eliminate human risk.[604] However, the program was cancelled due to technology limitations, military budget restrictions, and the internal power struggle within US DoD to define 'reasonable' technological advancement. While the Drone Assault

[600] Lt Col Timothy Cullen, 'The MQ-9 Reaper Remotely Piloted Aircraft: Humans and Machines in Action' (Massachusetts Institute of Technology 2011) <https://dspace.mit.edu/bitstream/handle/1721.1/80249/836824271-MIT.pdf?sequence=2>.
[601] Katherine Chandler (n 104) 29.
[602] Katherine Chandler, 'Drone Flight and Failure: The United States' Secret Trials, Experiments and Operations in Unmanning, 1936-1973' (The University of California 2014) <https://escholarship.org/content/qt0fg216f7/qt0fg216f7_noSplash_8dd4be278e0eefd9a44b97ef83782b98.pdf>.
[603] ibid.
[604] Katherine Chandler (n 104) 37–59.

Program ended, the technology persisted as a designation for target aircraft and remotely piloted air vehicles.[605]

Over following years, drones come into prominence again as a platform for surveillance, integrating cameras as an alternative to piloted reconnaissance flights and paving the way towards 'unmanned' characteristics. The attempts to use drones before the 1980s as both weapon systems and unmanned reconnaissance were dismissed by the US military as unsuccessful technological innovations, primarily due to challenges involving human machine interaction controls.[606] It was at that time when US DoD has started to explore the management of UAVs with the human-on-the-loop and they even tried to build a stealthy, intercontinental, fully autonomous UAV known as the Advanced Airborne Reconnaissance System (AARS).[607] While the efforts to build AARS were not successful, in part due to technology deficiencies,[608] the supervisory mode based on human-on-the-loop has been adopted by US DoD for various weapon systems.

In the academic literature, the difference between human-in-the-loop and on-the-loop can be also explained by two methods of supervisory control of UAVs which are management by consent (MBC) and management by exception (MBE).[609] MBC exhibits LOA 5 and it requires the system to ask for explicit consent from a human operator before taking any actions. MBE, on the other hand, typically exhibits LOA 6 and it allows the automation to perform actions unless overruled by a human operator.[610] For some authors, such as the Campaign, Cummings, or Sharkey, the use of UAVs with supervised control,

[605] ibid.
[606] ibid 95.
[607] Thomas Ehrhard, 'Air Force UAVs The Secret History' (Mitchell Institute for Airpower Studies 2010).
[608] ibid.
[609] Jessie Chen, Michael Barnes, and Michelle Harper-Sciarini, 'Supervisory Control of Unmanned Vehicles' (Army Research Laboratory 2010) ARL-TR-5136 10.
[610] ibid 10–11.

particularly in MBE operations, crosses the line of what should be allowed in the military, while for US DoD some weapon systems have been controlled by MBE since at least early 1990s.

The re-emergence of lethal drones in the 1980s was associated with the increasing importance of cruise missiles and improved technology, particularly regarding the amount of freedom of manoeuvre delegated to the weapon system.[611] The most relevant factor, however, was the victory of the Israeli Air Force over the Syrian Air Force in 1982, often referred to as the first successful use of drones for real-time reconnaissance.[612] The success of the Israeli Air Force was attributed to the miniaturisation of drones, but also to their ability to produce real-time information through 'a television camera.'[613] The US military administration described the battle as paradigm-changing for modern warfare and initiated a report to learn lessons from the Israeli victory. The report, prepared by the CIA, in 1986 outlined a strategic argument for the re-emergence of lethal drones motivated by the warning that terrorists could use a drone with a missile against a US embassy or other sensitive target:

> Although we have no indication that a Third World nation or terrorist group is planning such a modification, operators of RPVs [drones] can replace the surveillance equipment with high-explosive payload, effectively converting the RPV into a guided bomb capable of surprise attacks at short and medium ranges.[614]

It was the first articulation of a remotely controlled supervised weapon system as a potential countermeasure for terrorism. Indeed, in the few years after the report, the first ever targeted

[611] Dan Saxon (n 7) 190.
[612] CIA, 'Remotely Piloted Vehicles in the Third World: A New Military Capability' (1986).
[613] ibid.
[614] ibid.

killing orchestrated by an Israeli drone was framed as 'a technology solution' to counter terrorists' organisation.[615] The reasoning fits well with the CIA report and the US forces also began using lethal drones for targeted killings. The programme was thus driven by three major developments that come together: the earlier use of drones for reconnaissance, enhanced supervisory technology with real-time information for aerial targeting, and the plausible narrative supporting the development of armed drones as a-political, 'technology' solution to threats.

3. The Introduction of Remote Split Compound the Problem of Autonomous Weapons

In the previous section, I have argued that the connection between supervised control and unmanned lethal weapons opened the door to a lethal use of supervised autonomous weapon systems. Such weapons have become perceived as effective means of achieving specific military and political goals. The malfunction of these weapons led to the introduction of Directive 3000.09.

In this section, I argue that the tipping point of this process was the process called 'remote split' – the operation of lethal UAV from a great distance by a remote controller. The widespread adoption of remote split operations by the US DoD only exacerbated and accelerated the scale of challenges associated with the use of supervised AWS. I have argued earlier that the US military had a problem establishing trust in machine's autonomous capabilities, as the Patriot's example showed, that such weapons might not produce predictable outcomes. In addition to that problem, the US military had to respond

[615] Lisa Hajjar, 'Lawfare and Armed Conflicts: A Comparative Analysis of Israeli and U.S. Targeted Killing Policies and Legal Challenges against Them', *Life in the Age of Drone Warfare* (Duke University Press 2017).

to the more general problem of what the right allocation should be between automation and human in the decision-making process of complex socio-technical systems.

3.1. The Impact of Remote Split Operations on the Air Operations Decision Making

The 'Global War on Terror' has been credited as the first conflict to use lethal drones. Nearly one month after the 11 September attacks on the World Trade Center, a USAF pilot conducted a first lethal strike with a modern drone, the Predator. In the months and years following that first strike, lethal drones became the 'weapon of choice' for many states involved in armed conflicts.[616] Bode and Huelss argue that the 'War on Terror' was the most important push factor for drone technology, which had been technically available for some time.[617] Much has been written about the use of lethal drones and their contested compliance with LOAC, but what has attracted considerably less attention in the wider debate is the fact that the process of managing remote warfare led to another major invention - *remote split operations*. By 'remote split operation' I refer to the socio-technical system that allows drone operators to watch and attack a target from a great distance. UAVs do not require remote split operations and they have not always been operated as such. For example, the Predator drone used by the CIA in Afghanistan, was first utilised in 1995 in the Balkans, but was only adopted for remote split operations in 2001.[618]

In other countries, such as the UK, drones have only recently begun to operate as remote split operations.[619] Before that, UAVs were operated locally in the theatre of war,

[616] Dan Saxon, 'A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of "Appropriate Leveles of Human Judgment over the Use of Force"', *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 190.
[617] Ingvild Bode and Hendrik Huelss (n 104) 128.
[618] Madeleine Clare Elish (n 250).
[619] ibid 1101–1102. In the UK first remote split operations started in 2012.

that is inside or near the mission area. The aircrafts were controlled from the Ground Control Station (GCS) responsible for launching the aircraft and there was also a Mission Control Element (MCE) for conducting the operation. Further, most of UAVs missions were conducted as CAS, which meant that drone operations occurred in conjunction with troops on the ground.[620] These CAS operations required that the action of drones' had to be coordinated, not only between the team flying the aircraft, but also with the controller of the ground troops in the area and the local operations command team, which would at least include a battle commander, a signal intelligence officer, and a Judge Advocate General.

Remote split operations *distribute* the management of UAVs between two primary locations and two different kinds of teams.[621] The team responsible for launching the aircraft is called the Launch and Recovery Element (LRE) and consists of an USAF or defence contractor crew. Their work concentrates on launching the drone, getting it to a specified altitude, accomplishing a systems check, and handing the aircraft off via multi-user Internet relay chat or a phone call to a GCS located in the US air base.[622] Once 'hand-off' has occurred, the aircraft is controlled through a satellite data link, which causes a three-second delay (1.5 seconds each way) in transmitting information between the aircraft and the MCE, part of the so-called distributed ground control system within the US.[623] This delay, in addition to rather frequent satellite data link drops, is why an LRE element is necessary and why the aircraft must be controlled via a line-of-sight data-link during take-off and landing.[624] The MCE team in the US is often composed of multiple locations and

[620] Madeleine Clare Elish, '24/7: Drone Operations and the Distributed Work of War' (Columbia University 2018) 10 <file:///Users/mikolaj/Downloads/Elish_columbia_0054D_14452%20(2).pdf>.
[621] Madeleine Clare Elish (n 250); Madeleine Clare Elish (n 620).
[622] Madeleine Clare Elish (n 250) 1104–1105.
[623] ibid 1105.
[624] ibid.

consists of a pilot, a sensor operator, a mission commander, information analysts and support personnel involved in analysing and coordinating signal and human intelligence. The split between LRE and MCE distributes the tasks of analysis, verification, and planning across among multiple teams. It is worth stressing however that even today various types of US air weapon system are used differently. For instance, US weapon systems such as Air Force's MQ-1 Predators and MQ-9 Reapers are controlled from a remote location within the US via satellite link. However, the Army's MQ-1C Gray Eagles, even though they are manufactured by the same company as the Predator and Reaper, are operated locally in the theatre of war.

3.2.Challenges with USAF Centralised Control

The distribution of the management of UAVs between two primary locations and two different kinds of teams has complicated the established decision-making process within USAF, which was until very recently (2021) highly centralised.[625] According to longstanding rules and practices in USAF the 'centralised control' over the use of all aerospace assets lies with a senior airman in-theatre, also known as JFACC, who is appointed by and works directly for the Joint Forces Commander (JFC).[626] The forces are thus generally centrally controlled and tasked from the Air Operations Center (AOC).[627] The concentration of tasks makes the AOC a critical yet vulnerable component in US air operations.[628] Its destruction would paralyse USAF operations, as all the AOC's functions are singular and there are hardly any backups for them. AOC also has a sizeable

[625] Sandeep Mulgund, 'Evolving the Command and Control of Airpower' (21 April 2021) <https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2575321/evolving-the-command-and-control-of-airpower/>.
[626] The Joint Chiefs of Staff, 'Joint Publication 3-30, Joint Air Operations' (n 100).
[627] ibid.
[628] William A. Woodcock, 'The Joint Forces Air Command Problem' (2003) 56 Naval War College Review 126.

infrastructure, which is both cumbersome and difficult to deploy closer to the theatre of war.[629] Further, the C2 of aerospace power resides essentially in a single person. While the air operations staff supports the commanders, the JFACC remains the final authority and a single point of failure.[630]

Recognising this challenge, some USAF members, including senior members such as Gen James Holmes, started to contest the established decision-making rules.[631] He argues that the senior officer who exercises centralized control likely cannot be fully adaptive and responsive enough to the ever-changing nature of tactical operations supported by a high degree of automation. Yet in some cases ground commanders are required to ask how to act in many tactical decisions, and are usually just waiting for strike approval from the AOC, which is often located hundreds of miles away. An illustration of this problem is the example of the prosecution of 'time-sensitive targets' (TST), that is targets which require an immediate response because they pose (or will soon pose) a clear and present danger to friendly forces.[632] TSTs have been prevalent in military operations for decades, evolving from SCUD missile hunting in the Gulf War into current operations with the use of UAVs. The challenge is not new, but it has become more widespread due to the proliferation of highly mobile weapons. The USAF decision-making process for strikes against TSTs has been criticised internally because it centralises too many decisions, which, in turn, reduces the flexibility of the air strike package and sometimes leads to

[629] William A. Woodcock (n 628).
[630] ibid.
[631] Lt Col Clint Hinote, 'Centralized Control and Decentralized Execution: A Catchphrase in Crisis?' (Air University, Air Force Research Institute 2009); Sydney J. Freedberg Jr, 'Decentralize The Air Force For High-End War: Holmes' *Breaking Defense* (13 October 2017) <https://breakingdefense.com/2017/10/decentralize-the-air-force-for-high-end-war-holmes/>; William A. Woodcock (n 628); Lt Commander Matthew Quintero, USAF, 'Master and Commander in Joint Air Operations' (2019) 92 Joint Force Quarterly. Commander Gilmary Hostage III, USAF and Lt Col Larry Broadwell, Jr, USAF, 'Resilient Command and Control The Need for Distributed Control' (2014) 74 Joint Force Quarterly.
[632] Office of the Chairman of the Joint Chiefs of Staff, 'DOD Dictionary of Military and Associated Terms' (2021) 218.

mission failure. The weakness stems from USAF's inability to employ force quickly and to engage an emerging target before it disappears. This is in part because of the way the JFACC has historically been organised as a centralised source of control over the use of airpower assets. The forces are tasked from the AOC through publication of the Air Tasking Order (ATO). The ATO is the single-source plan for all air operations in an area of operations over a 24-hour period.[633] The ATO assigns to individual units their targets, weapons, and arrival times over those targets. Further, it instructs all assets about mission specifics. This process takes time: a given day's ATO takes anywhere from 36 to 48- hours to produce.[634] By the time the order is issued, most of its assumptions, analyses, and targeting decisions are out of date when dealing with TSTs.[635]

USAF leadership has been aware of this criticism, yet for the long time the mainline approach was to emphasise that commanders should delegate decisions to subordinates wherever possible and empower subordinates to take the initiative based on their commander's guidance rather than engaging in constant communications. This concept is referred to within USAF as 'decentralized execution.'[636] However, this idea has not quite been translated into military practice, according to some of USAF members.[637] Even the name 'decentralised execution' signals that it is less about taking control over military decisions and more about implementing higher-level orders. In fact, since the development of satellite communications and the Internet, joint air operations have become increasingly centralised.[638] Today's communications systems allow operational commanders to make tactical-level decisions from thousands of miles away.

[633] William A. Woodcock (n 628) 127.
[634] ibid.
[635] ibid.
[636] Lt Col Alan Docauer (n 167).
[637] William A. Woodcock (n 628); Lt Commander Matthew Quintero, USAF (n 631).
[638] Lt Commander Matthew Quintero, USAF (n 631).

Some USAF members point out, however, that this approach has significant shortcomings due to the limitations of modern communications technology. Authors such as Lt Commander Matthew Quintero argue that the AOC is vulnerable against an enemy actively working to degrade the Internet and satellite capabilities that enable it.[639] The updates on aircrafts on the battlefield heavily rely on satellite communications, which have historically proven unstable during major combat operations. Further, adversaries may force air force assets to block off communications to remain undetected, and thus the AOC will be unable to closely control tactical units. In the future, Quintero argues, conflicts may require decisions to be made within hours, minutes, or potentially seconds compared with the current multiday process, while persistent access to safeguarded networks cannot be assumed.[640] Thus, it is too risky to rely on a centralised decision-making process. AWS are perceived within US DoD as the remedy to these problems. Some of US adversaries, including China, are already developing advanced AI systems and autonomous capabilities. Future conflict may therefore occur at such a high speed that human operators will be unable to keep up. Even if communications between AOC and human operators have not been jammed or hacked, the back-and-forth process between humans will naturally be slower than the reaction of autonomous machine, for example, in the context of defending against massive drone swarms. Some military professionals argue that AWS may well 'create an environment too complex for humans to direct.'[641] Thus, any force that does not employ AWS will inevitably operate outside its enemy's 'OODA loop,' thereby losing the initiative on the battlefield.[642]

[639] ibid 92.
[640] ibid.
[641] Peter Singer, *Wired for War* (Penguin 2009) 128. (quoting Thomas Adams, Colonel (Ret.), U.S. Army).
[642] Michael N. Schmitt and Jeffrey S. Thurnher (n 176) 238–239.

USAF members argue then that to achieve the capability to respond to TSTs in the networked military conflicts requires not just decentralisation of execution, but also *distribution of control* over military assets, including the distribution of control between humans and machines on the battlefield.[643] It is argued that the JFACC must delegate responsibilities and processes even to tactical levels depending on the mission, thereby achieving not only lessened vulnerability of the AOC, but also faster decision cycles. Machine intelligence and human-machine teaming should play a critical role in this process. This is well summarized by The Air Force Next Generation ISR Dominance Flight Plan 2018-2028, which advocates a departure from the current 'industrial-age, single-domain approach' to pursue 'architecture and infrastructure to enable machine intelligence, including automation, human-machine teaming, and ultimately, artificial intelligence'.[644]

The internal debate about the architecture of the decision-making process regarding the use of air forces within USAF has intensified since the introduction of remote split operations. The distribution of the management of UAVs between two primary locations and two different kinds of teams – one within the US and another closer to the war theatre – has often led to confusion how the control of air assets should be executed.

4. A Summary of the Chapter

In this chapter, I have explored how the application of a lethal force in an autonomous way has come about as a policy problem for the US Government. I have examined what were the origins of the problem of increased risk of unintended engagements related to the use of AWS.

[643] Elsa B. Kania, '"AI Weapons" In China's Military Innovation' (The Brookings Institution 2020) 3.
[644] 'The Air Force Next Generation ISR Dominance Flight Plan 2018-2028' (USAF 2018).

I have argued that the increased risks of unintended engagements occurred since at least early 2000s due to the transformations in the way US DoD, specifically USAF, have conducted their operations relaying on increasingly sophisticated yet not regulated *lethal autonomous supervised systems* such as Patriot missiles or UAVs. As illustrated by the Patriot's fratricides during the Operation Iraqi Freedom in 2003, the US military had a problem to establish a trust in machine's autonomous capabilities. The findings from these events revealed that there were technological problems with Patriot software itself, but also the control procedures for human-machine interfaces were poorly designed for specific operations.

I have argued further that the use of lethal autonomous supervised systems can be traced back to the use of remotely controlled UAVs, which introduced novel challenges to the control of military operations. Specifically, the distance between teams on the ground in the theatre of war and remote teams in the US distributed control over the military assets and distributed control between humans and machines on the battlefield. I have argued also that these changes undermine the longstanding USAF doctrine of centralised control, according to which only a commander is responsible for a direction, coordination, and specific use of forces on the battlefield.

PART III – A CRITICAL EXAMINATION OF THE US DOD GOVERNMENTALITY OF LAWS

Chapter 7: Distributed Control. The Case Study of USAF

Chapter 7 tackles the fourth thesis's sub-question. It explores the specific effects of problematisation of AWS have on US DoD and USAF regimes of practices. As discussed in the earlier chapters, the application of lethal force in an autonomous way has become a policy problem for the US Government due to the increased risk of unintended engagements. This increased risk relates to the growing sophistication of weapons, particularly the introduction of the automatic mode of supervised systems and remote split operations. Specifically, the widespread use of remotely controlled UAVs introduced novel challenges to the control of military operations. Thus, the use of autonomous weapon systems, contrary to what the name suggests, has become a complex socio-technical system that requires trust and deep integration of human and automation factors.

As argued in Chapter 4, the notion of trust in US DoD has a double significance. First, in the socio-technical system of using AWS, there must be a trusted relationship and deep integration between humans and machines. Second, there must also be trust in a machine's autonomous capabilities to produce predictable outcomes. In other words, that the system has trustworthy autonomous capabilities. Thus, I have argued that the potential higher bar of risk of unintended consequences associated with the use of AWS is in fact

rooted in a deeper, underlying problem of how trust can be established in the decision-making process of using AWS.

In this chapter, I focus on the effects that are produced by this representation of the 'problem.' I have decided to limit the scope of investigation in two ways. First, as discussed earlier, I have decided to focus exclusively on USAF to present an in-depth study of at least one of the six US DoD military branches. USAF is a fitting example, as the air branch of the military was an early adopter of autonomous supervised systems and arguably the most vocal supporter of autonomy in weapon systems within US DoD. Moreover, USAF is the only military branch that has codified its longstanding doctrine of 'centralized control and decentralized execution' which has recently been updated and a significant part of my analysis connects the recent doctrinal changes with the effects of the problem representation. Finally, the concept of centralized control and decentralized execution, while not universally codified, have also been recognized in the C2 of other military branches in US DoD.[645]

The second limitation of this case study is focused on narrowing down the analysis to the specific set of effects that the problem representation has on *the emergence of norms* associated with the use of AWS. Thus, in the first section I study the 'first problem of trust', i.e. the allocation of functions and the authority of control between humans and machines in the decision-making process involving autonomous supervised weapon systems and I consider the effects of that problem on the established doctrine in USAF. I call these effects 'doctrinal change' as they refer to the changes within the formal organization of the USAF decision-making process in missions involving the use of AWS. Moreover, in the second section of this chapter, I explore the 'second problem of trust,' that is what effects are produced with respect to the specific problem of whether a human can trust an autonomous

[645] Specifically, the concept of decentralized execution is also recognised by the US Army and Marine Corps. See Lt Col Clint Hinote (n 631) 18.

machine to produce predictable outcomes. I call these effects 'standards change' as they refer to the emerging set of 'standards of appropriateness' within US DoD and USAF regarding specifically the use of AI algorithms in weapon systems. The sole focus on norm emergence is consistent with the objective of the thesis. My interest lies in a deeper understanding of the governmentality of AWS in US DoD, and particularly how the requirement of a human element in the use of force is problematised in the area of increasingly autonomous weapons.

One important point will be made regarding the concept of 'norms.' The concept of norms and a suite of related terms, such as 'normative statements' or 'normativity', have a very rich tradition in socio-legal studies, but the purpose of this thesis is to discuss the emergence of norms according to the Foucauldian-Bacchian approach consistent with the methodological design. While Foucault deliberately avoided defining 'norm,' I agree with Mark Kelly's analysis that Foucault understood the norm 'as a model of perfection that operates as a guide to action in any particular sphere of human activity, and normalisation correlatively as the movement by which people are brought under these norms.'[646] A norm thus establishes a relation to an ideal that others are required to conform to, while normalisation is a process of conforming to a norm.

The question of whether norms are effectuated as specific legal rules is of lesser interest. Foucault himself was less clear about this topic. He stated that norms should be differentiated from 'laws.' He argued that there was a difference between negative commands forbidding specific practices or limited positive commands requiring specific actions (which shall be called 'laws'), and positive ideals that guide actions (which shall be called 'norms').[647] I found this distinction helpful as it fits neatly with the organization of

[646] Mark Kelly, 'What's In a Norm? Foucault's Conceptualisation and Genealogy of the Norm' (2019) 27 Foucault Studies 1.
[647] Michel Foucault, *Abnormal* (Verso 2003) 48–50.

the US legal system. Foucauldian 'norms' would fall into the broad category of 'non-legislative rules,' which are rules that do not have the force of law but are issued to 'advise the public of the manner in which the agency proposes to exercise a discretionary power.'[648] These are, for example, policy statements or internal standards which detail how a specific governmental agency aims to conduct a particular action within its domain. The conceptualisation of 'non-legislative rules' thus resembles Foucauldian 'norms' which he defined as 'a prescriptive, optimal model, that is, a positive idea of how a thing should be, at its best.'[649] Norms function so as 'not to exclude and reject. Rather, [norms are] always linked to a positive technique of intervention and transformation.'[650]

In the study of norms, I draw from Bacchi's apparatus and focus on both 'discursive' and 'non-discursive effects.'[651] Discursive effects set boundaries around what can be recognised as relevant. By studying discursive effects, one can explore how the formulation of a problem establishes new concepts and limits what can be said about them. For example, in earlier chapters I discussed that Directive 3000.09 deliberately avoided the term 'human control,' instead opting for the concept of 'human judgment' as US DoD argued that the notion of 'human control' was too restrictive and might imply a so-called *direct* human control, which was said to be a less 'effective' approach to manage autonomous weapons in specific military operations. I presented, also, how after the introduction of the Directive 3000.09, the reference to the notion of 'human control' disappeared from other US DoD and USAF documents, for example, from the Unmanned Aerial Roadmaps. This example of a discursive effect clearly shows how the specific representation of the problem sets limits on what can be said. The term 'discursive practices' not only refers to the *'things said'* or *'not said'* about the problem representation, but also to the rules that *explain* how

[648] Administrative Procedure Act 5 U.S.C. § 553(b)(A).
[649] Michel Foucault, *Security, Territory, Population* (n 218) 58; Mark Kelly (n 646) 10.
[650] Michel Foucault, *Abnormal* (n 647) 50.
[651] Carol Bacchi and Susan Goodwin (n 136) 15–18.

it becomes possible to say (or know) or not say (or not know) certain things,[652] i.e. 'the rules governing a knowledge.'[653] In other words, discursive practices describe *practices* of knowledge formation by focusing on how specific 'discourses' operate and work in daily practices, thus creating a certain 'knowledge' and 'truth' (i.e. a dominant knowledge). Staying with the example of 'human control,' the choice of US DoD to use the reference to 'human judgment' as opposed to 'control' stems from the past US military experiences with operating supervisory autonomous systems, even before the introduction of Directive 3000.09.

As illustrated earlier, lethal supervisory autonomous systems, such as Patriot missiles, have been engaging targets in automatic mode since the Second Gulf War. The Patriot supervisory system allowed the human only a restricted time for veto before automatic execution. If the US Government had introduced a policy on AWS stating that autonomous and semi-autonomous weapon systems should be used in such a manner that commanders and operators exercise appropriate levels of human control over the use of force, then they would have prohibited the use of Patriots in automatic mode. The fact that, before the Directive 3000.09, various US documents, including the US Code, had stated that weapon systems should not fire without an '*affirmative human decision'* or *'human control'* only reaffirms that within US DoD, and particularly USAF, a set of military practices has diverged from the rules and guidelines of formal documents, even those considered as 'law.'

US DoD officials formed their own 'knowledge' and 'expertise' about the use of such systems. In fact, one can argue that their 'knowledge' or 'perspective' proved dominant over other subjugated discourses, as the reaction after the findings of the DSB Patriot report illustrated. US DoD officials did not follow the major recommendation from

[652] Michel Foucault, *The Archaeology of Knowledge* (Pantheon Books 1972) 182-183.
[653] Mark Cousins and Athar Hussain, *Michel Foucault* (Macmillan 1984) 94.

the DSB to introduce *more human operator involvement and control* in the operations of autonomous human supervised systems. Thus, the US military practices explain how it has become possible not to talk about the requirement of 'human control' within US DoD since the introduction of Directive 3000.09.

Non-discursive effects focus on the study of subjectification and lived effects.[654] Subjectification effects focus on the exploration of how the specific problem representation of AWS affects the transformations and creations of new subjects in the US military. Subjectification effects focus on how 'subjects' are implicated in problem representations, and how they are produced as specific kinds of subjects that denote the positions made available within particular discourses and knowledge.[655] An example is the introduction of new military staff members or novel responsibilities designed to address the formulated problem. For instance, the newly formed Office of the CDAO[656] is responsible for enabling, assessing, and tracking the implementation of the US DoD's policy on AI. Finally, lived effects, capture the material 'impact of problem representations on people's embodied existence,'[657] that is, how discursive and subjectification effects translate into people's lives. An example is the changing role of airmen in USAF when considering the use of autonomous supervisory weapon systems in remote split operations. Taking these various effects into account allows us to shed light on the complex array of implications that the US DoD problem representation of AWS entails.

My analysis necessarily excludes many other effects produced by US DoD problem representation of AWS. For example, I do not consider the effects of the problem representation outside the boundaries of US DoD and USAF, for example, on weapons' manufacturing companies and on US DoD contractors. They clearly play an important role

[654] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (n 148) 15.
[655] Carol Bacchi and Susan Goodwin (n 136) 23.
[656] CDAO is the successor organization to the Joint Artificial Intelligence Center (JAIC).
[657] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (n 148) 70.

as they usually develop weapon systems for the US Government. The framing of the problem representation of AWS may thus directly affect their work. I have also not considered the effects of the problem representation on the development of technology as such. I refer to certain technological innovations in the field of autonomous weapons only to the extent that they have influenced the norm emergence process, for example in AI models deployment. I have not, however, focused on the broad set of technical implications that can be linked to the effects of the US DoD problem representation of AWS. Finally, I have generally not considered the effects of the problem representation on the wider US Armed Forces bureaucratical organisation structure, although I refer to this topic, for example when discussing the changing role of air pilots. The topic of course is much broader than the position of air pilots and goes beyond the purpose of this thesis.

The chapter is divided into two major sections. In the first section, I study the first problem of trust, that is, the allocation of functions between humans and machines based on a specific example of a CAS mission. I explain how it was possible for USAF to update a longstanding doctrine of 'centralized control' by studying the distributed allocation of roles within USAF war-fighting missions. I argue that recent discursive changes in USAF doctrine, in other words, the change of wording from a 'centralized' to a 'distributed' concept of control, stems directly from the daily practices of USAF, particularly during missions involving TSTs. In the second section, I study the second problem of trust – that is, how one can trust a machine to produce predictable outcomes. Specifically, I focus on the problem of how US DoD can trust AI models in weapon systems. I explore US DoD and USAF practices that addresses this problem, and outline what norms are emerging and what the normalisation process looks like.

1. The Emergence of a Doctrine of Distributed Control

As discussed in the Chapter 6, US DoD desire to address the problem of AWS stems most directly from US DoD's fatal experiences with the malfunction of lethal supervisory autonomous systems, specifically Patriot missiles. The identified problem with the allocation between automation and human factors in the decision-making process of autonomous supervised weapon systems led ultimately to 'unintended engagements,' a phrase used 13 times in Directive 3000.09 signalling the major worry of US DoD policymakers associated with the use of such weapons. As further discussed, the emergence of remote split operations only exacerbated the problem of function allocation, as it introduced a greater distance between the human operator and machine. Thus, in this section, I explore the effects the problem representation on established doctrine in USAF.

1.1. USAF Notion of Control and Execution

USAF doctrine consists of fundamental principles by which military forces guide their actions in support of national objectives. It constitutes official advice but requires judgment in application.[658] Doctrine falls under what Foucault perceived as 'norms,' rather than 'law,' as doctrine constitutes official advice, but it does not have the force of law, i.e. it is not a legislative rule. While both policies and doctrine are example of 'norms,' they differ in terms of content. Policy is guidance that is directive or instructive, stating what is to be accomplished. It reflects a conscious choice to pursue a certain course of action rather than another. Within military operations, a policy may not only be expressed in terms of objectives, but also in terms of rules of engagement detailing what USAF wants to engage, with what capabilities, and under what circumstances. Policies are, however, different from doctrine as they are more mutable and may change due to various political, fiscal, or other considerations. Doctrine, on the other hand, is said to be more enduring, as it draws on the

[658] USAF, 'A Primer on Doctrine' 1.

long-term history of past military experiences and codifies best practices on how to accomplish military goals and objectives. As the USAF website states: 'It is a storehouse of analyzed experience and wisdom.'[659]

Doctrine focuses on the organisation of C2, that is the process that outlines the efforts of various entities (individuals and organisations) and resources, including information, toward the achievement of mission-related activities.[660] At the operational level, command refers to the authority exercised by a commander in the area of direction and coordination of armed forces (it answers the questions of 'who' and 'what'), while 'control' can be associated with implementing orders by communicating decisions, organising to carry them out, and evaluating the outcome to feed back into command (it answers the question of '*how*'). Thus, control can be perceived as being part of a chain of C2, rather than something that is only applied in relation to 'select' and 'engage' functions of specific weapons. This is an important detail, as it helps better understand the US DoD governmentality of AWS. The Campaign and various academics often understand military control in a very narrow sense, that is who makes a final engagement decision. US DoD and USAF, however, consider the notion of 'control' in a broader sense, i.e. as the answer to the question of how a specific mission should be exercised and how the decision-making chain should be organised.

The core concept of USAF doctrine is the so-called *mission command* which is an approach to C2 that empowers subordinate decision-making for flexibility, initiative, and responsiveness in accomplishing a commander's intent.[661] The execution of mission command was, until recently, administrated by the key tenet of the doctrine, that is the concept of 'centralized control and decentralized execution.' However, despite the

[659] ibid 4.
[660] The Joint Chiefs of Staff, 'JP 1, Doctrine for the Armed Forces of the United States' (2013).
[661] USAF, 'Air Force Doctrine Document (AFDP)' (n 197).

dominant USAF discourse that the key tenet of centralised control and decentralised execution has been 'proven over decades' and is 'the fundamental organizing principle(s) for air […],'[662] the principle has received a fair share of criticisms from USAF practitioners, who argue that the meaning of the words control and execution is unclear and question whether the two terms are distinct or one is inherent in the other[663]. Further, officers from other services have been also critical, such as US Army Maj Mark Davis who argued:

> […] from a joint perspective, centralized control and decentralized execution is illogical and cannot exist together because control is about execution and is inherent in command, as explained in Joint Publication (JP) 3-0, Doctrine for Joint Operations.

Indeed, from a historical perspective of the US air operations the concept of centralised control cannot be taken for granted. USAF has long struggle to capture the key tenet of their decision-making and was torn between either more centralisation or decentralisation.[664]

Despite these criticisms and growing popularity of remote split operations, the concept of centralised control and decentralised execution has shaped the USAF warfighting for over two decades.[665] Only very recently, in 2021, did USAF update its doctrine due to wider changes in the entire US DoD. JAIC has led to the development of the Joint All Domain Command and Control (JADC2), the postulate of decision-making that translates decisions rapidly into action and leverages capabilities across all domains with support of relevant parties.[666] The concept is based on the military's 'network of networks,' whereby separate communication networks of each branch of the US Armed

[662] USAF, 'Air Force Doctrine Document 1 (AFDD 1, 2003)' (2003) 28.
[663] Lt Col Clint Hinote (n 631).
[664] ibid 7–12.
[665] USAF, 'Air Force Doctrine Document 1 (AFDD 1, 1997)' (1997) 23. USAF, 'Air Force Doctrine Document 1 (AFDD 1, 2003)' (n 662).
[666] Congressional Research Service, 'Joint All-Domain Command and Control (JADC2)' (2022).

Forces will be integrate into a larger network to facilitate more effective use of AI and autonomous systems.[667] As a result of these changes, USAF also revised their main strategy tenet from 'centralized control' to become 'distributed control.' This new doctrine aims to empower subordinate decision-making in the accomplishment of commander's intent. It also allows sub-commanders operating in environments of increasing uncertainty and complexity to react with the freedom of action to exploit emergent opportunities. The new approach is to be implemented through centralized command, *distributed control* and decentralized execution.

1.2.The Doctrine of Distributed Control

The concept of distributed control enables commanders to delegate planning, guiding, and conducting operations to sub-commanders so that they can respond to changes in the operational environment and seize the initiative, particularly in physically or electronically contested environments or in the operations requiring dynamic targeting.[668] The term 'contested environments' refers to environments where there may be significant disruptions or degraded communications between AOC, where the Air Force conducts centralised planning, and forward operating locations.[669] As Gen Charles Brown, the USAF chief of staff, said 'distributed control will allow the force to […] address rapidly changing and increasingly challenging operating environments.'[670] This is because the doctrine of distributed control allows dispersed units in contested or degraded environments to conduct 'situationally-driven operational and tactical planning refinements.'[671] The control is said

[667] Todd Harrison, 'Battle Networks and the Future Force' (CSIS 2021).
[668] Sandeep Mulgund (n 625).
[669] Miranda Priebe at all, 'Distributed Operations in a Contested Environment' (RAND 2019) 978-1-9774-0232–5.
[670] NGAUS, 'Air Force Reveals New Airpower Doctrine' (27 April 2021) <https://www.ngaus.org/about-ngaus/newsroom/air-force-reveals-new-airpower-doctrine>.
[671] Sandeep Mulgund (n 625).

to be 'distributed,' but not 'decentralized' as sub-commanders still need to take the overall intent and context of a mission into account.

The term 'dynamic targeting,' according to US military doctrine, should be distinguished from deliberate targeting.[672] While both targeting processes consist of largely the same stages, the main difference is time. Deliberate targeting is the process when targets are identified and developed with sufficient time and detail to allow for capability analysis and force assignment. Dynamic targeting is more responsive than deliberate targeting as the process is used to prosecute targets that are either unexpected or not yet precisely detected or selected for action in sufficient time to be included in the deliberate process.[673] TSTs usually require dynamic targeting, but not always. For example, the Army may want a bridge destroyed at a specific time to create a trap. This is a pre-planned target, but also time sensitive.[674]

USAF has decided to replace the principle of centralised control with distributed control primarily to provide 'more suitable' guidance regarding the dynamic targeting operations. The air service has realised that dynamic targeting is more prevalent in contemporary conflicts where most of the attacks are conducted in dynamic, and often contested environments where centralised control has significant limitations and operators rely heavily on autonomous targeting and engagement capabilities of weapon systems.[675] A lack of clear guidance of sub-commanders and human operators regarding the planning, guiding, and conducting of operations with the use of autonomous capabilities of weapon systems constitutes what I described as the first problem of trust. After all, sub-commanders and human operators located in the theatre of war are forced to make decisions regarding the use of autonomous capabilities of weapon systems, particularly in missions requiring

[672] USAF, 'Air Doctrine Publication 3-60 Targeting' (n 97) 3.
[673] ibid 3–4.
[674] ibid 4.
[675] Commander Gilmary Hostage III, USAF and Lt Col Larry Broadwell, Jr, USAF (n 631).

dynamic targeting. In the next section, I explore a case study of a selected type of USAF mission that often requires dynamic targeting to illustrate how the principle of distributed control is exemplified in action.

1.3.Distributed Control in Action – a Case Study of CAS Missions

This section presents a case study of CAS mission to attack TSTs.[676] CAS missions aim to provide airborne firepower support for troops on the ground who may be operating near the enemy.[677] Air support may be either unanticipated (immediate) or it may be incorporated into the deliberate targeting process in certain situations where CAS is already anticipated (pre-planned).[678] My focus is on immediate air support that requires a dynamic targeting process. Dynamic execution assumes a responsive use of air assets to exploit enemy vulnerability that is likely of limited duration. Deliberate targeting in CAS missions, on the other hand, is when assets are pre-assigned, but it is unclear what targets may have to be engaged (if any), when and how. In military practice, CAS is usually requested by the ground commander, who may then provide guidance to the air force after which the dynamic targeting process starts.

By investigating the CAS decision-making in more detail, I argue that CAS missions in dynamic targeting, even before the change of doctrine, have relied on shared control between various agents, let alone the aircrew and the ground troops. I then argue that the term 'distributed control' is an example of 'discursive practices,' a codification of language which explains how it become possible to say (and know) that control exercised over the air operations has suddenly become shared rather than, as 'a longstanding USAF

[676] See a good overview of CAS case study in the context of assessing human control in operation. Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 150–170.
[677] USAF, 'Air Force Doctrine Publication 3-03 - Counterland Operations' (USAF 2020).
[678] ibid.

doctrine' has stated, centralized. These discursive practices also illustrate what not to say about the notion of 'centralisation' in the contemporary US air operations. In that sense, the discursive practices of distributed control have started to become the rules governing our current military knowledge about US air operations. I also point out to various non-discursive effects such as the experience of commanders and operators involved in CAS missions.

1.3.1. A Case Study Justification and Limitations

The focus on CAS missions attacking TSTs is justified by three factors. First, I am looking for a specific type of mission where dynamic targeting is dominant.[679] This is why I have excluded missions requiring deliberate targeting such as Air Interdiction (also known as Deep Air Support), which is an effort to attack enemy targets before they engage in combat.[680] Second, I am interested in the most generic type of missions involving dynamic targeting, rather than a specific one.[681] Finally, I also have to consider the limitations of the available documentation and sources. For this case study, I draw from USAF targeting doctrine, publicly available commentaries (including by current and former USAF members), recently published academic books and articles, and my interviews with various military practitioners, including, but not limited to, pilots, drone operators, and military lawyers. It is worth stressing that, even though I have tried to present a comprehensive picture of the decision-making process, it has been simplified as it does not consider any specific mission requirements or circumstances.

---

[679] ibid.

[680] CAS is different from Air Interdiction. CAS is directed towards targets close to friendly ground units, as closely coordinated air-strikes, in direct support of active engagement with the enemy. Air interdiction is carried out further from the active theatre of war, based more on strategic planning and less directly coordinated with ground units.

[681] There are generally two main types of counterland missions: CAS and Air Interdiction, but there are many specific applications (sub-categories) of these missions. See USAF, 'Air Force Doctrine Publication 3-03 - Counterland Operations' (n 677).

1.3.2. Centralised Control Architecture of CAS Missions

CAS missions attacking TSTs are based on dynamic targeting. According to US doctrine, the targeting process that is developed to facilitate dynamic targeting is called the F2T2EA (or F2T2E2A) cycle. F2T2EA stands for: Find – Fix – Track – Target – Engage – Assess.[682] The decision-making process for CAS missions is complex. As a joint mission involving both ground and air assets, even what sounds like a simple matter of deciding what air assets will support which ground-combat unit is the result of a tedious back-and-forth between the ground commander and the commander of the air units in the same area.

This question of which service should 'own' CAS assets has in fact been a highly contentious issue since USAF became independent in 1947.[683] Various US Army generals argued that, while USAF should own tactical air assets, the Army should exercise decentralized, operational control of these assets.[684] One of the chief arguments was that Army objected to the USAF desire to receive pre-planned CAS requests from the ground forces 24 hours in advance of the operation.[685] The Army's generals pointed out that this requirement could not be met in time-sensitive and fluid operations. Ultimately, the architecture of CAS decision-making today is based on the close coordination and communication between the ground and air forces, but the historical arguments about the difficulty of practically realising such a coordination are still valid, particularly in the context of satisfying the doctrinal requirement of centralised control.

[682] USAF, 'Air Doctrine Publication 3-60 Targeting' (n 97).
[683] Lt Col Clint Hinote (n 631). Herman Wolk, 'The Struggle for Air Force Independence, 1943-47' (1984).
[684] Maj. Michael Lewis, *Lt Gen Ned Almond, USA A Ground Commander's Conflicting View with Airmen over CAS Doctrine and Employment* (Air University Press 1997) 57–60.
[685] ibid 58.

Turning to the decision-making process, CAS missions usually begin with the JFC, a usually ground commander responsible for joint operations of the US Army and USAF.[686] The JFC delegates planning for air operations to the JFACC, who is responsible for planning, tasking and executing operations against dynamic targets. AOC personnel determine if pop-up targets meet the JFCs established criteria. If so, the JFACC then issues mission-type orders, via ATOs, to subordinate organisations to accomplish the total mission objectives set by the JFC, leaving the details of execution to be implemented on the ground. Before updating the USAF doctrine, the JFACC, in theory, retained centralized control of the assets, irrespective of whether CAS mission was pre-planned or based on dynamic targeting. In other words, 'target engagement authority' is held at the JFACC level and describes the control that those commands have over particular functions within the F2T2EA process that contribute to a commander's decision to engage a specific dynamic target.[687] While there is no doctrinal definition of target engagement authority, the term is generally understood by air services as the authority to execute the specified functions of F2T2EA.[688] These functions include combat identification, positive identification, target validation, strike asset deconfliction and assignment, collateral damage estimation, and execution order and approval.

The centralization of control required all forces to receive 'approval' from the JFACC, even before prosecuting a dynamic target. This 'knowledge,' to use a Foucauldian term, comes from the aspiration to achieve 'full-spectrum awareness,' that is the exploitation of ISR systems, in particular big data, to gather information regarding the enemy by observing their behaviour and tracking their movements. According to this

[686] USAF, 'Air Force Doctrine Publication 3-03 - Counterland Operations' (n 677). The Joint Chiefs of Staff, 'Joint Publication 3-30, Joint Air Operations' (n 100).
[687] Maj Nicholas Hall, USAF, 'Preparing for Contested War: Improving Command and Control of Dynamic Targeting' (Air University 2017) 5.
[688] ibid.

concept, the intelligence flow is organized in an upward trajectory to the JFACC, the AOC, and ultimately the policy maker.

1.3.3. USAF Criticism of Centralised Control of CAS Missions with Dynamic Targets

The dominant, formal discourse has, however, constantly been challenged within USAF in the context of dynamic targeting, particularly in terms of prosecuting TSTs. USAF service members, specifically sub-commanders on the ground, are aware that, in modern warfare, adversaries use Anti-Access/Area Denial strategies[689] that limit the ability of ISR systems to provide actionable intelligence in a timely manner to the JFACC, thus creating activity paralysis in the lower ranks as subordinate commanders await decisions from higher headquarters. In 2020 The US Air Force Special Operations Command (AFSOC) strategic guidance stated that AFSOC was no longer ready for 'full-spectrum' warfare and readiness due to the nature of contemporary conflict.[690] The guidance postulated that AFSOC must change to meet existing challenges.

1.3.4. CAS Missions with Dynamic Targets – A Practical Approach to the Control Authority

Military criticism of the concept of centralised control in dynamic targeting was informed by USAF service members first-hand experience of how, in fact, distributed control over such operations is. It is because of these daily military practices, it become possible to say (and know) that control is distributed, and in effect the new doctrine has been codified.

[689] In the military terminology Anti-Access/Area Denial strategies describe attempts to control access to and within an operating environment.
[690] USAF, 'The US Air Force Special Operations Command 2020' (2020).

Multiple parties are responsible for various elements of decision-making. CAS, being a joint operation, requires a significant level of coordination between air and ground forces to produce desired effects. CAS operations begins with a target nomination by the ground commander.[691] Once the ground commander has nominated the target, the Joint Terminal Attack Controller (JTAC) is often tasked with developing target data to ensure targets of highest priority to the ground commander are engaged. The JTAC is the person who directs the action of combat aircraft engaged in the CAS of ground troops. The JTAC is in close proximity to the ground troops. In a situation of pre-planned CAS, the aircraft operators usually have prior communication channel with the JTAC and ground forces. In a situation of dynamic targeting this is rarely the case. Instead, the operator may be briefed by the JTAC in-flight. Usually, the operator receives a brief that includes the target description, geographical information, how to manoeuvre for attack or to avoid threats and hazards), and how to exit the target area. While ground commanders determine what kinds of targets should be engaged, control over the engagement (at least to some degree) lies with the JTAC and depends on the circumstances. Joint Forces doctrine specifies three types of control (Types 1, 2, and 3):

> (1) Type 1 Control. Type 1 control is used when the JTAC/FAC(A) requires control of individual attacks and the situation requires the JTAC/FAC(A) to visually acquire the attacking aircraft and visually acquire the target for each attack.
>
> (2) Type 2 Control. Type 2 control is used when the JTAC/FAC(A) requires control of individual attacks and is unable to visually acquire the attacking aircraft at weapons release or is unable to visually acquire the target.

[691] The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (2014). Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 152.

> (3) Type 3 Control. Type 3 control is used when the JTAC/FAC(A) requires the ability to provide clearance for multiple attacks within a single engagement subject to specific attack restrictions.[692]

In the first phase of dynamic targeting – the phase called 'find' – the JTAC may sometimes have a target before requesting CAS or providing it in real-time. Thanks to a digital targeting system, the JTAC can broadcast the desired target directly to CAS aircraft mission computers via machine-to-machine interface.[693] However, when friendly ground troops are unable to provide real-time sensors to help the pilot find the dynamic targets (Control 2 and 3), the only way to find and engage those targets may be to fly around and look for them using the sensors on the aircraft and the pilot's eyes. On this type of mission, Lt Col Michael Kometer says 'the pilot essentially performs almost the whole C2 loop: he is the sensor, the targeteer, the decisionmaker, and the pilot.'[694]

Although significant delegation of control to the operator is possible, contemporary CAS operations are often supported with technologies that allow commanders to insert themselves into these decision-making processes and retain control over important targeting decisions. A valuable source of information for the aircraft operators is the Distributed Common Ground/Surface System (DCGS).[695] DCGS serves a system which produces military intelligence for the USAF collected by the UAVs such as U-2 Dragonlady, RQ-4 Global Hawk, MQ-9 Reaper and MQ-1 Predator.[696] However, the DCGS process is characterised by a lengthy target development methodology that has

[692] The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (n 691).
[693] Maj Ridge Flick, USAF, 'Winning The Counterland Battle By Enabling Sensor-to-Shooter Automation' (1 November 2021) <https://www.alsa.mil/News/Article/2822476/winning-the-counterland-battle-by-enabling-sensor-to-shooter-automation/>.
[694] Lt Col Michael Kometer, USAF, *Command in the Air - Centralized versus Decentralized Control of Combat Airpower* (Air University Press 2007) 52.
[695] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 156.
[696] Marc Schanz, 'Spy Eyes in the Sky' [2013] AIR FORCE Magazine.

slowed the kill chain as DCGS analysts generally have limited experience supporting dynamic targeting.[697]

There are few exceptions such as the Langley Target Development Cell (TDC) within DCGS which consists of targeteers from the 363rd ISR Wing responsible for USAF analysis and targeting.[698] TDC conducted target development in support of the Joint Task Force (JTF) responsible for the US military operation against Islamic State, i.e. operation 'Inherent Resolve'.[699] The advantage of the TDC is that it resides within the DCGS, at the source of exploited intelligence which shortens the time for target identification.[700] The use of TDC in DCGS is an example of military practice that resists and challenges centralised doctrine according to which a ground commander is responsible for target development. In Foucauldian language, it is an example of a counter-conduct within USAF, that is a practice of transforming the USAF way of doing things and allows for an analysis of how it has become possible to say (and know) that control exercised over the air operations is distributed.

It is important to emphasise that discursive effects are grounded in these daily practices. The practices of TDC in DCGS do not fit with the doctrine of centralised control and thus call for 'different form of conducts' regarding at least some types of air operations. Some USAF members have even argued that a distributed ISR/C2 unit with a targeting capability has the potential to support dynamic targeting more broadly in future conflicts.[701] The TDC's access to intelligence, as well proximity between targeteers, collectors, and analysts, allows for real-time refinement of requirements and quick re-tasking of sensors to identify TSTs.

[697] Maj Nicholas Hall, USAF (n 687) 7.
[698] ibid.
[699] ibid.
[700] Maj Nicholas Hall, USAF (n 687).
[701] ibid 14–17.

Once the target is found, the next two stages of dynamic targeting are the phases 'fix' and 'track,' whereby a goal is to validate a target, that is in military discourse to obtain a PID of the target ('fix') and - particularly in the case of moving targets - to update and maintain PID ('track').[702] Target validation ensures that all vetted targets meet the objectives outlined in the commander's guidance.[703] Depending on the situation, this may be done by the JTAC, another observer, the aircraft operator, or even a unit such as DCGS. During the next phase, the 'target' phase, relevant law, rules of engagement and other rules may be considered, as well as the assessment of collateral damage. According to US doctrine, at least some of these considerations should be assessed by the JTAC and the ground commander and delivered to the operator.[704] As in the discussion about the 'find' phase, some USAF service members believe that a unit such as the DCGS TDC should be given PID, CID, target validation, and collateral damage estimation authority.[705] According to US doctrine, the JTF and ground commander retains the ultimate authority for supporting fires in the respective operational area.[706] They promulgate guidance stipulating the criteria required to achieve PID and CID, validate a target, and estimate collateral damage. However, in CAS with dynamic targeting the subordinate unit has the authority to determine when those criteria are met and thus plays an important role in approving target execution. Further, if the expected collateral damage rises significantly, this may also elevate the authority to the operational or even strategic-political levels. Again, as a result,

[702] The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (n 691); Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80).
[703] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 200.
[704] The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (n 691).
[705] Maj Nicholas Hall, USAF (n 687) 13–17.
[706] The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (n 691).

the task of finding and fixing the target is often distributed across different individuals and organisations.[707]

Finally, during the 'execution' stage, the act of 'pulling the trigger' is also distributed.[708] The aircraft operator may release the missile, but the JTAC or another observer that has visually acquired the target may be the one guiding the bomb via laser to the target. After the JTAC has declared the target is 'cleared to engage,' the operator may commence engagement. Depending on the circumstances, the operator's role may be quite extensive.[709] The operator may release and guide the weapon to the target, conduct a transient collateral damage estimation (up to a certain level), and conduct most of the decision-making cycle in the absence of ground forces or third parties to serve as sensors. On the other hand, the operator may also be labelled as someone who merely 'pulls the trigger,' in other words, the person who executes the mission, but retains little control in the process. Either way, despite the fact the concept of centralization of control assumes that the JTF/ground commanders retain control over the F2T2EA functions stipulated within the target engagement authority, control over important targeting decisions is in fact also distributed across a range of individuals and organisations.

1.3.5. Autonomous Engagement by AI-Assisted UAVs

It is worth stressing that, currently, aircrafts used in CAS missions are predominantly operated by human pilots. Pilots use aircrafts such as AC-130s, legacy fighter aircrafts (F-16s and F-15s), F-35s, and A-10s.[710] However, USAF has also used UAVs for CAS missions such as MQ-9 Reapers drones, which are capable of remote controlled or

[707] Merel Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (n 80) 156.
[708] ibid 158.
[709] ibid 158–159; The Joint Chiefs of Staff, 'Joint Publication 3-09.3 Close Air Support' (n 691).
[710] Maj Kamal Kaaoush, USAF, 'The Best Aircraft for Close Air Support in the Twenty-First Century' (2016).

autonomous flight.[711] Some USAF service members even argue that the Reaper, which was viewed primarily as an ISR asset, has grown into a key CAS tool for the US military due to the system's superior strike capabilities.[712] The Reaper has even performed urban missions[713] (i.e. missions with close proximity to urban areas) and, according to Gen Hawk Carlisle, a retired head of ACC, the drone 'performed well.'[714]

Currently, military drones such as Reaper are remotely controlled and need a remote pilot to take off and land, as well as a payload operator to identify the target, let alone launch a missile. This being said, US DoD has started to equip Reaper drones with more advanced AI software.[715] The objective is for a drone to be able to carry out autonomous flight, decide whether to direct sensors, and recognise objects on the ground. Currently, if an adversary jams the satellite communications of a Reaper, the drone will circle in place or fly back to try re-establishing the communication link.[716] An AI-assisted Reapers, on the other hand, can utilise AI to navigate using landmarks, identify threats on the ground, re-direct the flight, and find relevant targets on its own.

While the AI software for Reapers is under pilot development, the US military have already deployed an AI-powered drone called Nova for defence purposes during the winter of 2018 in the Middle East.[717] Nova is an autonomous quadcopter drone produced by Shield AI, a private technology company, which is operated by the AI system called Hivemind.[718] This software allows Nova drones to manoeuvre autonomously in GPS and communication-degraded areas. The drones have already been used for reconnaissance and

---

[711] ibid 43.
[712] Colin Clark, 'Reaper Drones: The New Close Air Support Weapon' *Breaking Defense* (10 May 2017).
[713] Joe Ritter, 'MQ-9S Over Sirte: Unmanned Airpower for Urban Combat' (21 March 2022) <https://mwi.usma.edu/mq-9s-over-sirte-unmanned-airpower-for-urban-combat/>.
[714] Colin Clark, 'Drones Do Excellent Urban Close Air Support; Mideast F-35A Deployment In Several Years' *Breaking Defense* (24 February 2017).
[715] David Hambling, 'U.S. To Equip MQ-9 Reaper Drones With Artificial Intelligence' [2020] Forbes.
[716] ibid.
[717] Elliot Ackerman, 'A Navy SEAL, a Quadcopter, and a Quest to Save Lives in Combat' *Wired* (30 October 2020).
[718] Shield AI, 'Nova 2' <https://shield.ai/nova-2/>.

to solve problems such as urban room-clearing and navigating fatal funnels, (doorways, hallways, and other narrow places).[719]

Currently, AI-equipped drones such as Nova and Reaper do not carry out autonomous strike missions, irrespective of whether the communication channel is jammed.[720] Based on my interviews with practitioners and on publicly available data, they have not yet been endowed with such capabilities in any of the US military operations.[721] However, there have been reported instances of other countries using AI-assisted drones with autonomous force engagement.[722] Both USAF and DARPA have been working for a few years on swarm collaborative autonomy that can facilitate future CAS operations.[723] This is confirmed by the recent USAF Roadmap which explicitly states the ambition to use autonomous UAVs for CAS operations:

> DoD does not currently have an autonomous weapon system that can search for, identify, track, select, and engage targets independent of a human operator's input. In the future weaponization will be a crucial capability in mission sets where the unmanned system is directly supporting forces engaging in hazardous tasks.[724]

Further, as discussed earlier, Directive 3000.09 gives permission to use lethal supervised autonomous weapons for offensive purposes and, subject to further review, it leaves the door open for the use of fully autonomous weapon systems. In fact, Nova II, a new version

[719] Elliot Ackerman (n 717).
[720] ibid.
[721] Interview with Senior Force Developer for Emerging Technologies at the Office of the Under Secretary of Defense for Policy (n 426); Interview with fomer Senior US DoD official (n 345); Interview with fomer USAF member, 'Mikolaj Firlej Interview' (2 February 2021).
[722] Zachary Kallenborn, 'Was a Flying Killer Robot Used in Libya? Quite Possibly' *Bulletin of Atomic Scientists* (20 May 2021) <https://thebulletin.org/2021/05/was-a-flying-killer-robot-used-in-libya-quite-possibly/?utm_source=Twitter&utm_medium=SocialMedia&utm_campaign=TwitterPost05202021&utm_content=DisruptiveTechnology_WasAFlyingKillerRobotUsedInLibya%3F_05202021>.
[723] DARPA, 'Collaborative Operations in Denied Environment (CODE) (Archived)' (28 November 2022) <https://www.darpa.mil/program/collaborative-operations-in-denied-environment>.
[724] USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479) 22.

of the Shield AI drone, has swarming capabilities and the company has recently received a contract from USAF to develop technologies that would allow autonomous drones to partner with humans, initially for the collection of intelligence in GPS-denied environments.[725]

1.3.6. Human-Machine Teaming and the Changing Role of Air Pilots

Even before the formal changes in USAF doctrine acknowledging the distributed nature of control, USAF was already pursuing projects aimed at deep integration between autonomous systems and humans known as 'human-machine teaming'. The Shield AI contract is an example of such a project. Human-machine teaming is another discursive effect of the problem representation of increased risk involved in the operation of AWS where trust must be established in the allocation of functions and the authority of control between humans and machines in the wider decision-making process. USAF's recent *Roadmap* states:

> Establishing trust with operators in this manner will ensure that human authority remains at the center of mission approval for autonomous systems and ensures effective human-machine teaming. Without an adequate level of trust between operators/commanders and autonomous unmanned systems […] these systems will not be used in any mission set.[726]

[725] Valerie Insinna, 'Shield AI to Work on Swarming Drones, Autonomous Rotorcraft for Air Force' *Breaking Defense* (21 February 2022) <https://breakingdefense.com/2022/02/shield-ai-to-work-on-swarming-drones-autonomous-rotorcraft-for-air-force/>.
[726] USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479) 21.

This is also illustrated in a USAF report specifically on the role of human-machine teaming in airpower: 'Airmen need to develop informed trust – an accurate assessment of when and how much autonomy should be employed, and when to intervene.'[727]

The problem of how to establish trust in the socio-technical systems used in CAS operations also highlights the subjectification effects, that is, the changing role of various air force individuals, particularly commanders and operators. The doctrine of distributed control has become a dominant knowledge within USAF and is effectively recognised as a 'norm' in part because the daily practices of selected USAF individuals are no longer guided appropriately by the concept of centralised control. USAF has provided the following justification for the change of doctrine:

> Doctrine for current airpower employment and future […] must allow for the flexibility and versatility of centralization or decentralization needed, which CCDE [MF - centralized control and decentralized execution] does not express in an understandable way to diverse audiences across the joint force.[728]

As the CAS case study illustrates, the doctrine of centralised control required too broad a scope of control for some actors, such as ground commanders, while limiting the responsibility of others (e.g. JTAC, DCGS TDC, or in some cases even operators). Moreover, centralised control misrepresented the function of various contemporary technologies and the associated units which used them. As illustrated, some advanced technologies, such as Reaper drones or communication and intelligence technologies, have transformed units such as DCGS into complex socio-technical systems which increasingly

[727] Greg Zacharias, USAF, 'Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future, Volume I: Human-Autonomy Teaming' (n 103) 8.
[728] Sandeep Mulgund (n 625).

assume more control over the military decision-making process. Yet with the advent of AI-assisted drones, their role is changing too as UAVs are increasingly able to take over some of their functions.

Another important transformation relates to the role of air fighter pilots. USAF has already started conducting air operations with both a human pilot and a system that they call 'AI pilot.'[729] In the specific training operation, AI system, called ARTUμ, is able to assume control and direction of a radar on a U-2 spy plane in California as well as responsibility for tactical navigation while a human pilot flies the aircraft and coordinates with the AI agent.[730] USAF has even referred to ARTUμ as a 'working aircrew member.'[731] The test flight has been described as an exercise in developing trust between humans and AI decision-making, with the ambition to hand off responsibility to AI in tactical situations. Will Roper, the Assistant Secretary of USAF for Acquisition, Technology and Logistics said enthusiastically:

> With no pilot override, ARTUμ made final calls on devoting the radar to missile hunting versus self-protection. Luke Skywalker certainly never took such orders from his X-Wing sidekick![732]

It shall be noted, however, that the goal of the programme was not to remove a human pilot, but to *transform* it. According to DARPA, the AI system will fly the plane in partnership with the pilot, who will remain 'in the loop,' monitoring what the AI is doing and

[729] Sue Halpern, 'The Rise of A.I. Pilots' *The New Yorker* (17 January 2022) <https://www.newyorker.com/magazine/2022/01/24/the-rise-of-ai-fighter-pilots>.
[730] Will Roper, 'Exclusive: AI Just Controlled a Military Plane for the First Time Ever' *Popular Mechanics* (16 December 2020).
[731] Eric Tegler, 'An AI Co-Pilot Called "ARTUμ" Just Took Command of A U-2's Sensor Systems On A Recon Mission' *Forbes* (16 December 2020) <https://www.forbes.com/sites/erictegler/2020/12/16/an-ai-co-pilot-called-artujust-took-command-of-a-u-2s-sensor-systems-on-a-reconnaissance-mission/?sh=568d38e461f0>.
[732] Will Roper (n 730).

intervening when necessary.[733] DARPA's expectation is that a fighter jet with autonomous features will allow pilots to become 'battle managers,' directing squads of UAVs 'like a football coach who chooses team members and then positions them on the field to run plays.'[734]

The changing role of air pilots illustrates how the problem representation creates new 'subjects,' that is, augmented air pilots who are not only expected to conduct engagements, but also to manage the battlefield. This being said, the 'lived effects' of pilots reveal a more nuanced and contested picture. Air pilots differ regarding their contemporary role and responsibilities; sometimes they are required to conduct most of the decision-making cycle and other times they are considered as mere supervisors who fly a drone from a distance and just pull the trigger. USAF strategies and commentaries from current and former air pilots illustrate this rather ambivalent dichotomy.[735] On the one hand, US DoD and USAF state that air fighter pilots should remain a critical component of future operations. According to this narrative, AI systems are portrayed as tools to augment air fighters' responsibilities and transform them into 'battle managers.' On the other hand, USAF leaders claim that AI can take charge of air fighter mission, reducing the role of a human to mere supervisor. Will Roper said that putting AI in charge 'could tip those odds in our favour.'[736] Operators of UAVs such as Reaper are called 'pilots,' but, in fact, they do not physically fly the aircraft but rather they sit in front of screens and operate the aircraft remotely from a safe base in the US. The USAF report seems to confirm this ambiguity about the role of 'air pilots' more generally:

[733] Sue Halpern (n 729).
[734] ibid.
[735] See e.g. USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479).
[736] Will Roper (n 730).

> In the future, it is desirable to have each operator control multiple unmanned systems, thus shifting the human's role from operator towards mission manager. To ensure agility, the HMIs must support a range of control options whereby the human can be either "off the loop" with no control over an autonomous system, "on the loop" supervising the unmanned systems, or "in the loop" exercising commands to control a particular vehicle's path or payload.[737]

The conflicting roles and uncertain future of air pilots generates tensions between 'battle managers' and drone supervisors known as 'sensors' or 'tele-operators.'[738] Traditional air pilots must be officers and are generally believed to belong to an elite group to which nearly all aspire.[739] The supervisors are enlisted personnel, and usually disregarded by air pilots.[740] For instance, the Air Force did not officially recognise remotely piloted aircraft pilots as a career track until 2011. This being said, for some air pilots, the transition to drone supervision 'represented the potential for a brighter future' and 'an innovative way to stay relevant' in operations increasingly dominated by unmanned and AWS.[741] This perspective is not surprising in the context outlined by Chaillan, who reported that, during trainings with AI-controlled jets, the best fighter pilots 'end up losing against AI.'[742] He referred to DARPA's 'Alpha Dogfight Trials in August 2020 where an AI agent defeated a seasoned F-16 fighter pilot in a series of simulated combat engagements.[743] These 'lived effects' of operators also have implications for the emergence of norms. 'Battle managers' flying with 'AI pilots' sustain the practices of distributed control by splitting their responsibilities

[737] USAF, 'Unmanned Systems Integrated Roadmap 2017-2042' (n 479) 30.
[738] Madeleine Clare Elish (n 620).
[739] ibid 139.
[740] ibid 136–144.
[741] ibid 146–147.
[742] *US Government Must "Wake up Now" to AI Threat from China, Says Former Air Force Software Chief* (n 404).
[743] Jon Hoper, 'Pentagon Grappling With AI's Ethical Challenges' (10 November 2020) <https://www.nationaldefensemagazine.org/articles/2020/11/10/pentagon-grappling-with-ais-ethical-challenges>.

between that require their manual input and those that are handed over to machines. Other USAF members decide to transition to the role of 'sensors' supporting the work of battle managers.

The adoption of autonomous systems has also led to another tension between so-called 'strategic corporals' and 'tactical generals.'[744] Gen Charles Krulak used the term 'strategic corporal' to describe the implications that arise from the increasing responsibilities and pressures placed on tactical leaders due to the growing adoption of digital technologies.[745] In the contemporary battlefield, there is a tendency to require a broad set of responsibilities from tactical unit leaders or even from air pilots beyond what they have usually been trained for. Strategic corporals need to judge the reliability of AI predictions, determine the ethical consequences of algorithmic work, and judge in real-time whether, why, and to what degree a specific human-machine teaming system is performing well. As James Johnson writes: 'In other words, "strategic corporals" will need to become military, political, and technological "geniuses."'[746] On the other hand, there is a tendency for commanders accustomed to centralised control to act as 'tactical generals' by exploiting AI and data analytics to micromanage and take control of every aspect of tactical decision-making.

1.4.The Place of Human Judgment in Distributed Control

In the previous chapters, I have discussed the concept of the appropriate level of human judgment over the use of AWS, which is the key policy guideline behind Directive 3000.09. For some theorists, a concept of 'human judgment' is similar or even equivalent to the

[744] James Johnson, 'Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command-and-Control Decision-Making in the Digital Age' [2022] Defense Studies 3.
[745] Gen Charles Krulak, USMC, 'The Strategic Corporal: Leadership in the Three Block War' [1999] Marines Magazine.
[746] James Johnson (n 744) 14.

concept of 'human control'. In earlier chapters, I have discussed the US DoD discourse that 'judgment' is different from, and should not be conflated with, 'control' because 'human control' is too restrictive and may imply 'direct human control,' that is reflexively pressing a button to approve strikes. In contrast, US DoD has argued that the automation of weapon systems' functions could allow operators to exercise better judgment over the use of force, but the sole engagement can, in some cases, be left to machines. The use of autonomous functions that take direct control away from human operators can, as a result, better effect human intentions and avoid accidents. This discourse shifts the question of who should make the final application of lethal force – human or machine – onto the broader decision-making process where 'human intention' is effectuated by various stages of making a targeting decision. This shift is in line with the military conceptualisation of 'control' which relates not to the final engagement per se, but more broadly to the question of how a specific mission should be exercised and how the decision-making chain should be organized.

As a result, US DoD discourse has problematised AWS as weapons which should be, in principle, allowed subject to specific trust measures. As discussed, in the socio-technical system of using AWS there must be a trust between humans and machines. Even if the weapon systems can act in an autonomous way, the use of such machine still occupies certain place in a wider chain of military command. I have argued that a doctrine of distributed control has been initiated by wider changes in US DoD (JADC2), with the objective of clarifying rules applicable for the shared control of modern military operations between various teams as well as autonomous systems and humans. This 'discursive effect' of changing the USAF doctrine from centralised to distributed control reflects already certain already existing military practices of shared control in the wider targeting process. As illustrated by the CAS case study, military practices, at least in the context of dynamic targeting where AWS are often used, did not adhere to the doctrine of centralised control.

Rather, one can argue that the CAS operations in dynamic targeting were example of counter-conduct under the previous doctrine of centralised control. One can thus consider that the US DoD problematisation of AWS as weapon systems which require increased trust between humans and machines led to the normative changes in the way USAF operations should be conducted.

A concept of distributed control places the emphasise on the flexible role of socio-technical systems, rather than prescribing that there should always be direct human input at any specific stage of a targeting process, irrespective of circumstances. It is a doctrine that requires 'more trust' and 'shared control' between commanders and their subordinates; and it requires human-machine teaming where different forms of control may be considered appropriate for different situations. A doctrine whereas 'human control' is sometimes retained only at the strategic-political level while the tactical control can be left for autonomous or AI systems, however, in other circumstances control is considerably delegated to human actors on the scene.

2. The Emergence of Trustworthy AI Principles and Standards

In this section, I also explore the 'second problem of trust,' that is, what effects are produced with respect to the specific problem of whether a human can trust an autonomous machine to produce predictable outcomes. I call these effects 'standards change', as they refer to the emerging set of principles and more specific standards within USAF regarding the use of AI algorithms in weapon systems.

2.1. US DoD AI Ethical Principles

In Chapter 6, I argued that the lethal use of AWS has become initially framed as problematic by US DoD due to the lack of operational control of humans over the use of

autonomous supervised weapons such as Patriots missiles. The autonomous use of force has been considered inferior to human-operated weapons due to technical deficiencies. Human operators simply did not trust some of the advanced robotic weapons operating in automatic or autonomous mode. The DSB, after the Patriot missiles accidents in 2003, found serious flaws in the systems' software, specifically the computer vision capabilities.[747] While there have been some improvements in Patriot' software, Hawley has argued that during operational tests of Patriot software upgrades, incidents similar to the fratricide occurred in Iraq took place. He argues that such incidents often happen when events 'go off-script,' and operators have to confront situations they have not previously seen or have not been explicitly trained to address.[748] In another DSB report, military analysts pointed out that AWS also present a challenge for developers, who have to move from a hardware-oriented, vehicle-centric development process to one that addresses the primacy of software in creating autonomy.[749] The report states: 'These challenges can be characterized as a lack of trust that the autonomous functions of a given system will operate as intended in all situations.'[750]

The effect of this problem representation has been a long debate within US DoD regarding how to guide developers in developing future autonomous and AI-assisted weapons. Since the introduction of Directive 3000.09, US DoD took eight years, before finally announcing, on 24 of February 2020, a short document proclaiming, 'ethical principles for the use of AI' (Table 4).[751] During that eight-year period various branches of the Armed Forces pursued many ambitious projects in the development and application of AI. Before the publication of the *AI Principles*, the Defense Innovation Unit (DIU) granted

[747] Defense Science Board, 'Patriot System Performance' (n 302).
[748] John K. Hawley (n 301).
[749] Defense Science Board, 'Report of the Task Force on the Role of Autonomy in DoD Systems' (n 114) 2.
[750] ibid.
[751] US DoD, 'DOD Adopts Ethical Principles for Artificial Intelligence'.

a number of fast prototype contracts, known as 'Other Transactions' to private companies offering commercial solutions to national security challenges.[752] Many of these companies were developing specific AI capabilities for the US military. For example, Shield AI received their first military contract in 2016 to develop an AI-controlled drone for clearing rooms in urban areas.[753] US DoD's AI rinciples require AI capabilities to be responsible, equitable, traceable, reliable, and governable. Lt Gen Michael Groen, Director of JAIC, at the time of the announcement of the principles, justified them as 'the core of US DoD trusted-AI ecosystem.'[754] It is uncertain whether any of these principles have been applied in the evaluation process of the Shield AI contract. In fact, some members of the US DoD community have voiced concerns since the publication of the principles that they are too general, and that they need to be operationalised to be used in daily practices. Alka Patel, head of AI ethics policy at JAIC said 'There are ethics principles for the DoD, but then the next question is: well, what does that mean? What does that mean in my role? How do I ensure that I'm actually satisfying these principles?'[755]

Table 4: Ethical AI Principles[756]

Principle	Description
Responsible	DoD will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities
Equitable	DoD will take deliberate steps to minimize unintended bias in AI capabilities

[752] Defense Innovation Unit, 'Annual Report 2020' (Defense Innovation Unit 2020) 3.
[753] Kenrick Cai, 'Shield AI Rejected A Pivot To "Selfie Drones"—Now Its Drones Are Being Used By The Military Overseas' (16 July 2020) <https://www.forbes.com/sites/kenrickcai/2020/07/16/shield-ai-ryan-brandon-tseng-ai-50-interview/?sh=24d785e439db>.
[754] CDAO, 'AI Ethical Principles – Highlighting the Progress and Future of Responsible AI in the DoD' (26 February 2021) <https://www.ai.mil/blog_02_26_21-ai_ethics_principles-highlighting_the_progress_and_future_of_responsible_ai.html>.
[755] Jon Hoper (n 743).
[756] Defense Innovation Board, 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense' (2019).

Traceable	DoD AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.
Reliable	DoD AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.
Governable	DoD will design and engineer AI capabilities to fulfil their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

2.2.Ethical AI Principles as Norms

Even though US DoD took so much time to announce these principles, the department's narrative is that the principles are not new. It is argued that 'the principles are based on existing and widely accepted' ethical, legal and policy commitments under which the DoD has operated for decades such as the 2015 DoD Law of War Manual and DoD Directive 3000.09.[757] This seems like quite a far-reaching statement. I have mentioned the DSB findings regarding Patriot fratricides in Iraq, which concluded that there were serious deficiencies in the Patriot software, specifically regarding how the missile system classified targets.[758] The Patriot system applied complex computer algorithms to assess a target's speed, altitude, and (in the case of an airplane) its radio transponder signal known as

[757] ibid 6.
[758] Defense Science Board, 'Patriot System Performance' (n 302) 2.

'identification friend or foe'.[759] If the software's data matched the profile of an enemy aircraft or missile, it displayed the target as hostile on an operator's' monitor. However, the Patriot training data was not specific enough and algorithms were too 'brittle' for the system's engagement context, i.e. unable to handle unusual situations reliably.[760] The criteria programmed into the Patriot computer were based on the many different Anti-Radiation Missiles available worldwide, rather than any specific missiles which might resemble those from Iraq. Therefore, it is far-fetched to say that the Patriot's algorithms were 'equitable' as the system did not have robust enough data to learn how to distinguish objects that did not easily match the pre-defined criteria. The DSB report stated that Patriot system's ability to identify and distinguish object was 'very poor.'[761] The report confirmed that the data should have been much more specific, based on the known threat from Iraq, and concluded that the generic Anti-Radiation Missile classification criteria programmed into the Patriot computer were a contributory factor in the accident.[762]

Further, it is doubtful whether the Patriot system capabilities were reliable as it can be contested whether Patriot's automatic mode had 'explicit, well-defined use.'[763] Patriot automatic mode had been adapted from the engagement control logic of the Safeguard system. The important difference between Safeguard and Patriot is the context of their operational missions.[764] Safeguard was only meant to fight air battles against other missiles in a nuclear war context. Patriot, on the other hand, operates in a more complex and ambiguous lower-tier region of the air defence environment where the risk tolerance is lower. It is then argued that 'the Safeguard level of automation' was not an appropriate

[759] Lester Haines, 'Patriot Missile: Friend or Foe?' *The Register* (20 May 2004) <https://www.theregister.com/2004/05/20/patriot_missile/>.
[760] John K. Hawley (n 301).
[761] Defense Science Board, 'Patriot System Performance' (n 302).
[762] ibid.
[763] Defense Innovation Board (n 756).
[764] John K. Hawley (n 301).

mode for Patriot's operating environment, particularly due to the greater potential for track classification and identification mistakes.[765] This system operating in an all-or-nothing fashion with limited availability for operators to intervene may not have been suitable, at least at that period, for conventional air threats. One can also argue that the system was not reliable because its capabilities were not subject to robust 'testing and assurance within those defined uses.'[766] The system's deficiency had been observed during many training exercises and was never fixed before fielding the weapon. As the report stated: 'The Task Force remains puzzled as to why this deficiency never garners enough resolve and support to result in a robust fix.'[767]

It can be also argued that the Patriot system was not traceable, as its capabilities were not 'deployed such that relevant personnel possess an appropriate understanding of the technology.'[768] Hawley argued that highly automated systems such as Patriot should rely on expertise confronted with 'tough cases' that challenge and expand the skill level and depth of system understanding.[769] The US Army did not have such training prior to Operation Iraqi Freedom as its training had focused too much on getting crews certified in a short period of time and too little on corresponding skills development.[770]

In short, the statement that the *AI Principles* have not fundamentally changed how the Armed Forces have implemented AI and autonomy for decades does not give much credit to the latest changes that have occurred since the introduction of those principles. Irrespective of whether such a statement has any factual grounding, such a discourse reinforces the wider governmentality of AWS. The justification that AI ethical principles are 'based on existing and widely accepted ethical, legal and policy commitments under

[765] ibid.
[766] Defense Innovation Board (n 756).
[767] Defense Science Board, 'Patriot System Performance' (n 302) 2.
[768] Defense Innovation Board (n 756).
[769] John K. Hawley (n 301).
[770] ibid.

which the DoD has operated for decades' – even if it does not represent the existing state of affairs – helps organise a certain way of thinking about already fielded (semi) autonomous weapon systems and responds to a problem of how future AWS should be governed. As Foucault wrote:

> Problematisation doesn't mean the representation of a pre-existing object, nor the creation through discourse of an object that doesn't exist. It is the set of discursive and non-discursive practices that makes something enter into the play of the true and the false and constitutes it an object for thought.[771]

Thus, AI ethical principles can be considered as a discursive effect of the problem construction of AWS which call for the trustworthy use of AI capabilities. The principles constitute and reinforce the governmentality of AWS according to which already fielded AI capabilities have had to go through the same legal, ethical, and policy considerations as contemporary AI systems. At the same time, the principles are designed in such a manner as to provide 'a prescriptive, optimal model, […] how a thing should be, at its best.' [772] Their function is 'not to exclude and reject' any prior use of AI capabilities but presents 'a positive technique of intervention and transformation.'[773] This is well reflected in the words of the US DoD analyst who claimed that the principles were not new:

[771] Michel Foucault, 'The Concern for Truth', *Michel Foucault: Politics, philosophy, culture. Interviews and other writings, 1977-1984* (Routledge 1988) 257.
[772] Michel Foucault, *Security, Territory, Population* (n 218) 58; Mark Kelly (n 646) 10.
[773] Michel Foucault, *Abnormal* (n 647) 50.

> The principles should not be perceived as a set of rules or constraints that hinder AI adoption and use. Rather, they are intended to contribute to the efficiency, effectiveness, and legitimacy of the DoD's AI capabilities.[774]

This aspirational purpose of the *AI Principles* emphasises the process of governing conduct through positive means. The objective is not to establish novel concepts or rules that might challenge existing US military law or ethics, but rather the goal is to re-arrange or transform internal processes to achieve a specific end – a trusted use of AI capabilities.

2.3.Normalisation Through Standardisation

The process of transforming internal US DoD practices associated with the adoption of AI and autonomous systems has been initiated over the last few years, at least since the publication of the *AI Strategy* in 2018. The strategy initially led to the formation of JAIC, which is considered as a key unit to accelerate the delivery of AI-enabled capabilities, scale AI, and synchronise US DoD AI activities.[775] In a May 2021 memo, Deputy Secretary of Defense Kathleen Hicks explained that it was critical for US DoD to *create* a 'trusted ecosystem' in AI, that 'not only enhances our military capabilities, but also builds confidence with end-users, war fighters, and the American public.'[776] As discussed, JAIC published the *AI Principles* in February 2020. The memorandum was broadly considered as a response to the criticism aimed at US DoD, and particularly JAIC, that there had been little real progress in implementing AI ethical principles. Lt Gen Shanahan, a former director of JAIC, confirmed that the JAIC was struggling with the scale of challenge:

[774] Merel Ekelhof, 'Responsible AI Symposium - Translating AI Ethical Principles into Practice: The U.S. DoD Approach to Responsible AI' (23 November 2022) <https://lieber.westpoint.edu/translating-ai-ethical-principles-into-practice-us-dod-approach/>.
[775] US DoD, 'Summary Of The 2018 Department of Defense Artificial Intelligence Strategy' (2019).
[776] Kathleen Hicks, 'Memorandum "Implementing Responsible Artificial Intelligence in the Department of Defense"'.

'Implementing the AI ethics principles will be hard work. The Department's efforts over the next year will shape the DOD's future with AI,'[777]

In the aftermath of Hicks's call for a trusted AI ecosystem, various divisions of US DoD, particularly the DIU, have been working to translate the principles from the memorandum into concrete guidance.[778] DIU is a US DoD unit that accelerates the adoption of emerging technologies into the US military by awarding fast prototype contracts, often to vendors which are considered 'non-traditional', i.e. a vendor has never before worked with the US DoD.[779] In November 2021, DIU released the *Responsible AI* (RAI) guidance for contractors looking to partner with the US DoD. The document provides guidelines for each phase of the AI development lifecycle – planning, development, and deployment – and is intended to act as a 'starting point for operationalising' the Defense Department's AI ethical principles.[780]

Responsible AI, the initiative which aims to operationalise the *AI Principles*, is another discursive effect of the problem representation. It illustrates how the US DoD aims to normalize AI principles through a plethora of emerging standards that provide more specific guidance regarding the adoption of AI by the US DoD. According to US DoD *AI Principles* are general rules that aim to guide action, while their operationalisation aims to develop 'tools, policies, processes, systems, and guidance' that ensure that AI technology systems comply with principles.[781] For simplicity's sake, I refer to these various operationalisation mechanisms as 'standards' purely to differentiate them from the doctrinal considerations discussed earlier. This is consistent with the National Science and

[777] Jackson Barnett, 'Department of Defense AI Ethics Principles Still Lack Implementation Guidance' *FedScoop* (28 June 2021) <https://www.fedscoop.com/ai-ethics-principles-dod-military-implementation-guidance/>.
[778] Megan Lamberth, 'Putting Principles into Practice: How the U.S. Defense Department Is Approaching AI' (2022).
[779] See 10 U.S.C. § 2302(9).
[780] Jared Dunnmon and others, 'Responsible AI Guidelines in Practice' (DIU 2021).
[781] ibid.

Technology Council terminology, which describes 'standards as requirements, specifications, guidelines, or characteristics that can be used consistently to ensure that AI technologies meet critical objectives.'[782] Similarly to the discussion above in the context of principles, for Foucault, standards can also be considered as 'norms' or, to be precise, 'norms in making,' as the normalisation process of trustworthy AI capabilities has just started within US DoD. 'Trust in AI is and will increasingly be a cultural challenge until it's simply a norm — but that takes time,'[783] said David Spirk, the Pentagon's former Chief Data Officer, who was involved in the organisational changes in JAIC.

The DIU's work has resulted in developing 'process-oriented standards,' i.e. standards that could transform internal departmental assessment processes when evaluating the adoption of AI and autonomous capabilities. Jared Dunnmon, one of the authors of *Responsible AI* and a technical director for AI and ML at the DIU, said that standards were designed as an ideal to 'hold ourselves accountable for how we are running these programs.'[784] The standards also provide many novel requirements for contractors in the form of checklists that vendors need to meet before partnering with US DoD. I do not intend to present all the different standards that are part of the *Responsible AI* process, but I would like to highlight two examples which present how US DoD think about building more trustworthy AI capabilities.

First, there is a requirement to establish a baseline for any AI project prior to system development, i.e. a measure that allows for performance comparison regarding the task of interest before, during, and after a project.[785] A baseline should ideally compare

[782] NSTC, 'U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools' (2019) 8.
[783] Brandi Vincent, 'Pentagon Reaches Important Waypoint in Long Journey toward Adopting "Responsible AI"' *FedScoop* (29 June 2022).
[784] 'Interview with Jared Dunnmon' <https://federalnewsnetwork.com/artificial-intelligence/2021/12/defense-innovation-unit-issues-guidance-for-responsible-use-of-artificial-intelligence/>.
[785] Jared Dunnmon and others (n 780) 19–20.

performance on a quantitative basis, and where it cannot, it should have a well-defined qualitative measure that allows comparison to be made. Questions for vendors interested in satisfying this standard include (1) How is the task currently performed? (2) What is an acceptable minimum performance threshold? And (3) What are the most important evaluation criteria (e.g. speed, volume of data processed, quality of output, etc.)?[786]

Second, there is also a requirement to conduct harm modelling to assess the likelihood and magnitude of harm from AI decision-making.[787] Vendors are expected to identify a list of potential harms, such as physical harm, psychological harm, opportunity loss, and so on. For each harm they should consider questions such as how severe an impact a particular harm can have, and what the scale, probability and frequency of the harm might be. They should also ask questions regarding realistic worst-case scenarios in terms of how errors might impact society, individuals, and stakeholders? If things go wrong, what will the impact be at the individual and community level?[788]

According to US DoD discourse these standards help advance the process of adopting more trustworthy AI capabilities. Yet theorists and some members of the Armed Forces have voiced concerns which contest the dominant conduct. There are three major lines of criticism. First, current standards are still too general and give the impression that this is an ethics-washing project, i.e. a surface-level effort not meat to develop any constraints on AI development and use, but rather to send a signal to the public that US DoD is 'serious' about ethical considerations.[789] An example is the assessment of whether a specific AI capability provides a unique, non-marginal benefit or whether an alternative solution should be selected. Some examples where AI is not a recommended approach

[786] ibid 21.
[787] ibid 23–24.
[788] ibid 24.
[789] Matt O'Brien, 'Pentagon Adopts New Ethical Principles for Using AI in War' *AP News* (24 February 2020) <https://apnews.com/article/technology-us-news-business-artificial-intelligence-73df704904522f5a66a92bc5c4df8846>.

include cases which require subjective judgment: e.g. where different people would reasonably disagree about the best outcome; or solving existing human problems: e.g. clarifying an existing process that is confusing and/or problematic; or fixing existing problems in sets of data (such as bias).[790] These assessment guidelines are indeed broad. Questions, for example, whether drones should be equipped with AI for targeting purposes are clearly matters of subjective judgment, while such projects have been accepted by US DoD, at least for the prototype contracts. An example of a more concrete AI-relevant standard which can be found in the US Government is P1872-2015 (Standard Ontologies for Robotics and Automation), developed by the Institute of Electrical and Electronics Engineers (IEEE).[791] It outlines a standardised way of representing knowledge and a common set of terms and definitions which is said to allow for knowledge transfer among humans, robots, and other artificial systems.

Second, some argue that process-oriented standards are not sufficient, and that they should be complemented by outcome-oriented considerations such as performance benchmarks.[792] Currently, US DoD does not have any standards or accepted methodologies for the operational deployment of AI models.[793] For example, it is not clear what should be the effectiveness criteria for a computer vision algorithm designed to detect objects for autonomous drone engagement. Is an AI system considered as effective and ready for deployment when the algorithm correctly finds targets 85% or rather 95% of the time? The outcome-based measures refer to how well the algorithm was designed for its mission deployment and there is no single standard for representation anywhere in US DoD.

[790] Jared Dunnmon and others (n 780) 19.
[791] National Science & Technology Council, 'The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update' (2019) 33.
[792] ibid 33–34.
[793] Interview with A CEO of a US DoD vendor supplying a tool for validating AI models, 'Mikolaj Firlej Interview' (5 February 2021).

Finally, there is also internal tension within US DoD, even between some of its leaders, regarding the tempo of developing standards. This argument often refers not just to work on standards, but also to the general approach of US DoD efforts regarding trustworthy AI capabilities. On the one hand, Michael Groen, a former Director of JAIC argues that JAIC has deliberately adopted a 'slow and incremental' approach to ensure trust within the Armed Forces communities and contractors.[794] On the other hand, Chaillan, a former USAF's Chief Software Officer, decided to leave the US DoD and criticised the department for 'a deep, entrenched lack of institutional urgency and a degree of comfortable complacency'[795] This criticism spotlights a complicated story of recent changes within US DoD. While certainly there are many discursive effects in the form of announced standards calling for changes in the organisation of the Armed Forces, the daily practices of at least some officials and military personnel reveal a more contested picture where the institution in which they served has failed to deliver on its promise.

2.4. Subjectification and Lived Effects of Addressing the Problem of Trustworthy AI Capabilities

The role of a Chief Software Officer identifies an important aspect of building trustworthy AI capabilities in US DoD, that is the organisational changes that have occurred within the department to better streamline the AI work. The process of developing principles and standards has been associated with a major reorganisation of the Department that has produced many subjectification effects, such as new units, tasks, responsibilities, and even,

[794] Jackson Barnett, 'JAIC Chief Wants AI Progress to Be "Slow and Incremental"' (*FedScoop*, 8 October 2021) <https://www.fedscoop.com/jaic-chief-wants-ai-progress-to-be-slow-and-incremental/>.
[795] *US Government Must "Wake up Now" to AI Threat from China, Says Former Air Force Software Chief* (n 404).

to a certain extent, a new organisational culture. The focus of this thesis is not to describe these changes in detail. Rather, I argue that, when considered both discursive and non-discursive effects, the process of developing standards reveal a contested and complicated picture. Since the creation of JAIC there has been a surge of new units and roles within US DoD to support the mission to build more trustworthy AI capabilities. JAIC itself has been through three major strategy evolutions. It was initially created as a product-focused office, building AI tools tailored for specific problem sets.[796] When the leadership changed in October 2020, it declared a 'JAIC 2.0' strategy that turned the centre into an AI 'enabling' force, working throughout US DoD to find ways to field AI by coordinating with other tech-oriented offices.[797] More recently, JAIC has merged with several other US DoD components such as Office of the Chief Data Officer and Defense Digital Service – resulting in the CDAO with much broader set of responsibilities.[798] JAIC's most important initiative, besides the RAI guidelines, has been the creation of the Joint Common Foundation, a cloud-based AI development and experimentation environment that delivers tools and capabilities to support the US DoD's pursuit of AI capabilities. As the US DoD discourse presents, the idea is 'to create its own app store of sorts, with catalogues of algorithms trained and ready to be applied to new data.'[799]

In addition to the work in JAIC, US DoD and their military branches have created a group of new subjects as the abovementioned the Chief Data Officer (responsible for data management, data governance and standards processes), the Chief Information Officer (responsible for the policy and oversight of information resources management), the Chief Software Officer (responsible for IT modernisation), the Director of AI (a position separate

[796] Jackson Barnett (n 794).
[797] ibid.
[798] Brandi Vincent, '2022 in Review: What the Pentagon's CDAO Accomplished in Its Inaugural Year' (30 December 2020) <https://defensescoop.com/2022/12/30/2022-in-review-what-the-pentagons-cdao-accomplished-in-its-inaugural-year/>.
[799] Jackson Barnett (n 794).

from JAIC/CDAO and responsible for developing the AI roadmap), among others. Military branches of US DoD, such as USAF, have also created similar roles within their departments. USAF and other departments have also created their own cloud platforms with tools to exploit data for various mission-related activities. In USAF, this infrastructure is called the VAULT Platform which 'empowers airmen to work on current-state and emerging technology tools with real data in a safe and secure environment,' according to Eileen Vidrine, a Chief Data Officer for the Department of the Air Force.[800] There is also an emphasis on staff training to help them adopt AI capabilities. For instance, USAF Academy offers their staff the opportunity to study data science and conduct relevant internships in the subject.[801] USAF is also organising Datathons, that is events aimed at solving challenges by using data for war fighters.[802]

Yet despite all these 'transformative' initiatives, US DoD seems trapped by institutional inertia and struggles to change the organisational culture. An example which illustrates how the US DoD daily practices seem to diverge from – or even challenge – aspirational norms refer to the practices of project management. US DoD already committed in 2010 to transition from waterfall project management to agile practices, modelled after software development projects.[803] Traditional waterfall project management maps out a project into distinct, sequential phases, with each new phase beginning only when the previous one has been completed. In contrast, agile methodology is an iterative approach to managing projects that focuses on continuous releases and incorporating

[800] Eileen Vidrine, 'Air Force CDO: Flying High With AI' *The Wall Street Journal* (20 August 2021) <https://deloitte.wsj.com/articles/air-force-cdo-flying-high-with-ai-01629481727>.
[801] Jennifer Spradlin, 'Academy Announces New Data Science Major' (8 June 2020) <https://www.af.mil/News/Article-Display/Article/2211059/academy-announces-new-data-science-major/>.
[802] USAF, 'Air Force Chief Data Office Announces First Datathon' (9 July 2020) <https://www.af.mil/News/Article-Display/Article/2268590/air-force-chief-data-office-announces-first-datathon/>.
[803] Stephany Bellomo, 'A Closer Look at 804: A Summary of Considerations for DoD Program Managers' (Carnegie Mellon University 2011) CMU/SEI-2011-SR-015.

customer or user's feedback. In 2010 the DSB presented a report to Congress proposing that the Under Secretary of Defense should create a new acquisition process for IT systems based on agile methodology.[804] In response to the DSB recommendation, Congress passed the Section 804 NDAA for 2010 and made the use of agile mandatory for acquisition processes.[805] Yet this formal norm, despite being codified, did not translate into daily practices as US DoD was reluctant to adopt the methodology. The wider focus on providing a framework for the adoption of AI by US DoD inspired some decision-makers to re-launch the effort to transform the daily practices of project management. In 2018, US DoD renewed its plans to adopt agile methodology and in 2020 the department published a document offering advice to managers on how to develop agile projects.[806] While the document was generally well received, US DoD was again criticised for making little practical progress in adopting these guidelines, among others by the US Government Accountability Office (GAO), a legislative branch government agency that provides auditing, evaluative, and investigative services for Congress.[807] 'Agile is 22 years old and yet it is barely used,' said Chaillan who also points out that 'to this date, there is not even one hour of required agile training for warfighters.'[808] On the contrary, the curriculum of the Defense Acquisition University, the university in charge of training the US DoD acquisition workforce, include courses that are characteristic of waterfall project management.[809]

---

[804] ibid.
[805] National Defense Authorization Act for Fiscal Year 2010 2009.
[806] US DoD, 'DIB Guide: Detecting Agile BS' (2018) <https://media.defense.gov/2018/Oct/09/2002049591/-1/-1/0/DIB_DETECTING_AGILE_BS_2018.10.05.PDF>; US DoD, 'Agile Software Acquisition Guidebook' (2020).
[807] US GAO, 'DoD Software Acquisition: Status of and Challenges Related to Reform Efforts' (2021) GAO-21-105298.
[808] Nicolas Chaillan, 'Let's Catch-up with China within 6 Months' (*LinkedIn*, 24 November 2021) <https://www.linkedin.com/pulse/lets-catch-up-china-within-6-months-nicolas-m-chaillan/>.
[809] ibid.

2.5.The Place of Directive 3000.09 in Responsible AI Guidelines

I have argued earlier that AWS have become the subject of government interest because of the increased risks associated with their development and use. The problem, in part, exists because AWS users do not have enough trust in AI capabilities. One of the effects of this problem representation in the area of norms is the development of both principles and of more specific and emerging process-oriented standards that should guide US DoD acquisition work. As discussed, Directive 3000.09 established a policy of 'appropriate levels of human judgment' to guide the development and use of AWS. It has also included a procedure to ensure that AWS will function 'as anticipated in realistic operational environments […] and are sufficiently robust to minimize failures.'[810] US DoD leaders argue that the *AI Principles* and *RAI Guidelines* are complementary to Directive 3000.09 and there is no need for a major update of rules.[811] I would like, however, to specify three issues that generate a certain confusion. The first relates to terminology. RAI efforts have not led to any operational measures of what 'appropriate levels of human judgment' should entail. While there are many heuristics to translate how to implement concepts such as responsible, equitable, traceable, reliable, and governable, there is no guidance for implementing 'levels of human judgment.' Further, some relevant new terms have been adopted by US DoD that are not defined in Directive 3000.09, e.g. 'AI-enabled autonomous weapon system.'[812] This topic will be elaborated in more detail in the Chapter 8.

Second, according to the Directive 3000.09 AWS used for the application of lethal force are subject to a senior review process. However, it is uncertain how the evaluation of AI capabilities will be assessed in the senior review process. It is not clear whether this

[810] Directive 3000.09 Autonomy in Weapon Systems 4a (1) (a) (c).
[811] Defense Innovation Board (n 756) 6.
[812] Kathleen Hicks (n 776).

process is different from RAI guidelines, and if so, what specific measures will be considered.

Finally, Directive 3000.09 refers only to 'select and engage' functions of AWS, whereas RAI guidelines apply to any AI capabilities across a broad spectrum of warfighting activity, including finding and fixing potential targets. For example, Mike White, an Assistant Director for the DoD's Hypersonic Office, stated that he wants future US DoD hypersonic weapons to use autonomy and AI to 'optimize flight characteristics.'[813] As hypersonic weapons could potentially carry nuclear warheads such an application of AI and autonomy has significant consequences for safety and strategic stability. Yet such application of AI capabilities is currently not regulated by Directive 3000.09 and seems to have been left only to RAI guidelines.

3. A Summary of the Chapter

In this chapter, I have explored what specific effects the US DoD problematisation of AWS have on US DoD and USAF regimes of practices. I have decided to focus on USAF to present in depth study of at least one out of six US DoD military branches. I have narrowed down the analysis to a specific set of effects that the problem representation has on *the* emergence on norms associated with the 'trusted' use of AWS.

According to US DoD, one of the problems of trust is how to ensure a 'right' allocation of functions and the authority of control between humans and machines in the decision-making process involving increasingly autonomous weapon systems. I have illustrated that, according to the US doctrine, the concept of 'control' is broader than merely

[813] Mike White and Gillian Bussey, 'Assistant Director Mike E. White and Director Gillian Bussey Remarks to The Technology and Training Corporation on Hypersonics and Autonomous Systems' (4 November 2020) <https://www.defense.gov/News/Transcripts/Transcript/Article/2412014/assistant-director-mike-e-white-and-director-gillian-bussey-remarks-to-the-tech/>.

'who pulls the trigger'. It refers to the question of 'how' military operations should be conducted. I have argued that the use of AWS systems occupies certain place in a wider chain of military control across various stages of targeting process. USAF doctrine has long been guided by the tenet of centralized control according to which military control belongs to a mission commander, but as illustrated by the CAS case study, military practices – at least in the context of dynamic targeting where AWS are often used – did not adhere to the doctrine of centralized control. I have therefore argued that a 'discursive effect' of changing USAF doctrine from centralized to distributed control reflected already existing military practices of shared control in the wider targeting process. The normalisation of distributed control within USAF is one of the key effects of the US DoD problematisation of AWS.

I have also explored the problem with trustworthy capabilities of AWS, that is whether humans can trust an autonomous machine that it will produce predictable outcomes. I have argued that US DoD developed both *AI Principles* and more specific and emerging process-oriented RAI standards that should guide US DoD acquisition work related to AI-assisted weapon systems, including AI-assisted AWS. While this set of emerging standards is, according to US DoD leaders, 'based on existing and widely accepted' ethical, legal and policy commitments including Directive 3000.09, I have argued that in fact AI principles and RAI guidelines are not necessarily in conjunction with Directive 3000.09. It is uncertain how a key policy of human judgment over the use of AWS, from Directive 3000.09, fits with *AI Principles* and the concept has never been operationalised by RAI guidelines. Directive 3000.09 in turn is silent on AI-capabilities of AWS whereas one could argue that AI introduce different challenges than the use of autonomy in weapon systems.

Chapter 8: What Has Been Left Out of the Problem Representation?

This chapter tackles the fifth and the final thesis's sub-question. It explores what has been left unproblematic in US DoD problem representation of AWS. In other words, the issues that are often raised in academic or public discourse about AWS, but which were not addressed in Directive 3000.09. For each issue, I reflect on how this specific omission can be questioned and ultimately how it can disrupt US DoD dominant problem representation.

The chapter is divided into four sections. In the first section, I argue that the US DoD governmentality of AWS focuses heavily on the concept of autonomy, but it disregards the evaluation of advanced AI capabilities of AWS. As discussed in Chapter 7, US DoD has recently developed AI principles and RAI guidelines to assess AI capabilities, but such considerations have not been part of the initial problem representation of AWS. Rather, they can be considered as the effects of it. I argue, however, that the new AI principles and standards, precisely because they occurred later as an *after-thought*, reveal certain gaps and disjunctions with the initial problem representation of AWS. I thus argue that the US DoD problem representation of AWS can be challenged by putting an emphasis on the inclusion of the AI-specific assessment of such weapon systems.

In the second section I argue that the US DoD problem representation of AWS disregards the potential threat posed by weaponised AGI, i.e. a hypothetical intelligent agent able to understand and perform any intellectual task that a human being can. Instead, US DoD either denies or deflects the risk of applying AGI in warfare. An alternative problem representation can be based on the recognition that the potential long-term threats of weaponised AGI are relevant and should be addressed now. I discuss various strategies regarding how militaries might mitigate the potential advent of AGI.

In the third section, I argue that US DoD problem representation of AWS has not considered deep ethical concerns regarding the use of such weapons. I then argue that an alternative problem representation of AWS might include ethical considerations as certain restrictions over the development and use of such weapons. I restate that the inclusion of a direct human control requirement at the level of engagement can not only alleviate potential threats from weaponised AGI but also, to certain extent, limit ethical concerns. I also discuss various proposals to build an ethical machine which might be able to assess whether a specific lethal action is ethically permissible or – under much narrower set of assumptions – whether the action is ethically impermissible. I also explore the possibility of creating a dedicated oversight agency with the goal of providing more accountability regarding the use of AWS, potentially alleviating some fears that such weapons are being used for impermissible ethical action according to wider public morality.

Finally, in the fourth section I argue that US DoD problem representation of AWS excludes the complexity of autonomous cyber weapons. I argue that the potential inclusion of cyber considerations has merits, but one must address the problem of what mechanisms restricting the spread and use of autonomous cyber weapons should be in place, if any. I also discuss the question of whether AI-augmented autonomous cyber weapons can be used for both defensive and offensive purposes.

1. Advanced AI and Autonomous Weapons

In this section, I build and expand on the discussion in Chapter 5 about the relationship between 'autonomy' and AI. I argue that the US DoD governmentality of AWS focuses heavily on the concept of autonomy understood as a machine's independence from a human operator, but that it disregards the evaluation of the advanced AI capabilities of AWS. Directive 3000.09 senior review process for LAWS does not apply, for instance, to AI-

augmented Reaper drones used to hunt for specific targets or for loitering munitions. Such weapon systems can be made to operate in a communication-denied or electronically contested environment without direct human supervision, yet Directive 3000.09 does not place any additional limitations on their use, besides the general laws applicable to all weapon systems. I therefore argue that the US DoD problem representation of AWS can be challenged for its lack of inclusion of AI-specific assessment of AWS.

1.1.A Lack of Consideration of How AI Capabilities Should be Assessed

The US DoD problematisation of AWS concentrates on the risks associated with the use of autonomy, characterised as a machine's independence from a human operator. One can argue that autonomy as such does not bring any new qualitatively different considerations for the assessment of weapon systems. As has been discussed earlier, mines can also be classified as AWS and US DoD leaders responsible for Directive 3000.09 seem to agree with such a classification.[814] I have argued in Chapter 5 that AWS may not necessarily imply a high level of sophistication or intelligence. US DoD representatives have repeatedly asserted that 'AI and autonomy are not interchangeable. While some AWS use AI, this is not always the case.'[815] Yet there is a substantial difference between, say, landmines and advanced AI-assisted aircrafts that are able to fly and engage targets autonomously. One could argue that the application of advanced AI, such as ML techniques, is what makes a *real autonomy* possible, not the independence from human as such.

Directive 3000.09, however, gives 'a green light' to all existing uses of autonomy, including those which depend on ML capabilities. According to Directive 3000.09 there is

[814] Scharre (n 34) 50–51.
[815] 'The Policy and Law of Lethal Autonomy with Michael Meier and Shawn Steene' (n 499); Interview with Shawn Steene (n 256).

no separate formal process that would require decision-makers to assess the risk of using AI in the autonomous functions of weapon systems. As confirmed by the DIB: 'DoD Directive 3000.09 addresses autonomy in weapons systems, but it neither addresses AI as such nor AI capabilities not pertaining to weapon systems.'[816] This is surprising as US DoD has emphasised that the increased level of risk of unintended consequences of using AWS requires building two different layers of trust measures. As discussed in Chapter 7, one layer of trust relates to human-machine interaction; the second layer of trust relates to the degree of confidence in using autonomous capabilities so that users can expect predictable outcomes. US DoD has stated that at least some autonomous capabilities are, and will be, based on advanced AI.[817] One can therefore argue that various AI methods bring equally important, but qualitatively different challenges for the evaluation of weapon systems relative to the evaluation of human-machine interaction. Directive 3000.09 does not require any new T&E procedures specifically to address the second layer of trust, i.e. the adoption of AI capabilities by AWS. It does not even mention terms such as 'AI or 'ML.'

In part, this approach has been driven by the internal terminology developed by US DoD. Instead of focusing on AI as a separate and potentially transformative technology, the department adopted the 'levels of autonomy' framework, which does not directly consider AI but rather assigns different level of autonomy to a machine, depending on the extent to which that machine executes decisions and informs the human of that decision. I have illustrated this framework in the Levels of Automation Table 3.

This is of course not to argue that US DoD is not interested in AI. Even before the introduction of Directive 3000.09, AI – alongside autonomy - played an important role in national defence strategy. Yet the US DoD problematisation of AWS as weapons that

[816] Defense Innovation Board (n 756) 5.
[817] US DoD, 'U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway' (2022); US DoD, 'AI Strategy: Harnessing AI to Advance Our Security and Prosperity' (n 395).

generate the increased risks of unintended consequences assumes there should be *the same process* of mitigating risks, irrespective of whether those risks come from human-machine interaction or from the adoption of advanced AI capabilities.

1.2. How ML-Specific Assessment is Different?

By leaving AI considerations out of the problem representation of AWS, one can challenge US DoD policy construction by putting emphasis on the inclusion of advanced AI-specific assessment. The case rests on the argument that the evaluation of advanced AI-assisted AWS is fundamentally different from the assessment of AWS which are based on a classical rule-based software.

To illustrate this difference, consider how both systems are different. Note, I refer to advanced AI as solution-based systems or ML systems.[818] Both rule-based and solution-based systems are efforts of knowledge representation, i.e. they aim to represent information about the world in a form that a computer system can use to solve various tasks.[819] Rule-based system is a knowledge representation in which knowledge is stored as logical rules in the form of if-then-else statements. The promise is to elucidate the knowledge of a human expert in a specialised domain and embody it within a computer system. A human expert knows a set of facts and is able to model this knowledge in the set of explicit rules. Weapon systems based on rule-based programme are deterministic, i.e. their inputs are clearly defined and their outputs are predictable. While rule-based systems rely on explicitly stated and static formulas, solution-based systems rely on algorithms that detect patterns in data to create their own models of association between inputs and outputs.

[818] Solution-based systems are often described in the literature as sub-symbolic systems, while rule-based systems are called symbolic systems.
[819] Sara Brown, 'Machine Learning, Explained' (MIT 2021) <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.

ML systems learn from data and adapt to new situations by themselves, whereas rule-based systems require human intervention for any changes. Therefore, while the outputs of rule-based systems are deterministic, the outputs of ML systems are only probabilistic.[820]

This means that ML systems can produce uncertain outputs rather than consistently producing the same results. Thus, software updates of solution-based systems are very different from those of rule-based systems.[821] As the outcomes are only probabilistic, much depends on providing specific and large training data sets that accurately reflect the operational environment. If the operational environment changes, failing to rapidly update a ML model can lead to reduced performance or even to malfunction.[822] Further, it is currently challenging to produce formal proofs of the behaviour of ML systems, i.e. that is to verify the correctness and accuracy of the results.[823] This poses difficulties for attaining the levels of formal verification that are required for many software code-based systems, especially for systems performing critical functions on which human lives may rely.[824]

Thus, one can argue that ML-enabled lethal use of AWS against human targets as such renders too excessive risk because of the inherently probabilistic nature of such engagements. The war as such is already characterised by deep uncertainty and missions rarely go according to the plan. The adoption of ML-enabled weapons which struggle with uncertainty can only exacerbate 'a fog of war.' As argued by Cummings:

> What does the [US DoD] military does not understand about the AI is that for the most part AI struggles with uncertainty. [...] The military, and I have been here, in every mission I have been briefed never happened as we were briefed. [...] I worry that a fog of war

[820] Dusica Marijan and Arnaud Gotlieb, 'Software Testing for Machine Learning' (2020) 34 Proceedings of the AAAI Conference on Artificial Intelligence 13576.
[821] ibid.
[822] Interview with A CEO of a US DoD vendor supplying a tool for validating AI models (n 793).
[823] Jie M. Zhang and others, 'Machine Learning Testing: Survey, Landscapes and Horizons' (2022) 48 IEEE Transactions on Software Engineering 1.
[824] Ana Pereira and Carsten Thomas, 'Challenges of Machine Learning Applied to Safety-Critical Cyber-Physical Systems' (2020) 2 Machine Learning and Knowledge Extraction 579.

> combined with AI inability to handle uncertainty is a problem. […] AI should not be used in any time you need dynamic adaptation under high uncertainty.[825]

One could argue that in the context of taking someone's life one should expect a deterministic calculation whereby the correctness and accuracy of the outcome can be reasonably predicted. One could refute this logic by arguing that during a war often weapons are used in a manner that is different from predicted use or there might be situations where it is difficult to trace back whether a particular engagement was indeed according to the pre-planned manner (the latter is called a 'traceability problem' of ML systems). In such cases, humans are not fully able to verify the correctness and accuracy of the engagement outcome. Yet I argue that there is a difference between descriptive assessment of specific engagements and normative criteria that should be in place to assess the engagement of weapon systems. One could, therefore, postulate the normative criteria of using only deterministic calculation of weapon systems for the engagement of human targets.

That said, one could point out that there are already ML techniques that are able to build a strong confidence threshold of producing formal proofs of the behaviour of ML systems at least in certain operational environments.[826] In other words, such ML-enabled AWS might still be based on probabilistic outcomes but the confidence threshold of producing accurate and traceable results will be almost indistinguishable from deterministic systems, e.g. a ML-enabled AWS will exhibit over 99% accuracy in desired operational environment. In such case one could potentially grant the use of ML-enabled AWS against human targets, but such weapon should go through certain additional review process, such

[825] 'Missy Cummings Asks: Should the US Military Use AI Weapons?' <https://www.youtube.com/watch?v=yaBY3Pfmdno&t=1s>.
[826] See e.g. Alexander Lavin and others, 'Technology Readiness Levels for Machine Learning Systems' (2022) 13 Nature Communications.

as a senior review process from Directive 3000.09. Note, however, that a senior review process has never been conducted due to US DoD's use of unbounded notion of autonomy. Recall that, in Chapter 4, I have argued that the US military has established such a high bar to qualify any lethal weapon as 'autonomous' that in fact the concept of autonomy is used in an indeterminate fashion. This problem construction allows in turn to exclude the plethora of already existing weapon systems with advanced autonomous capabilities from the increased regulatory oversight, i.e. an additional review mechanism from Directive 3000.09. Thus, one can further argue that both ML-enabled lethal use of AWS *and semi-autonomous weapons* against human targets should at least go automatically to senior review process. As semi-autonomous weapons are already in a wide use by the US military, their inclusion in a senior review process could make the review process truly relevant.

One could further argue that still in some cases, even though probabilistic systems may exhibit a high accuracy threshold, the nature of their engagement is 'indiscriminate by design.'[827] For example, in the use of nuclear weapons. Thus, the alternative problem construction of AWS could include at least a clear prohibition of using ML-enabled AWS with nuclear warheads. In this context, one could even go step further and state that only human beings can authorize employment of nuclear weapons. Such limitations of using AWS could not only mitigate a risk of inadvertently triggering massive nuclear escalation, but also it will be consistent with the US policy which states that 'the decision to employ nuclear weapons requires the explicit authorization of the President of the US.'[828]

The above alternative problem representations of AWS rest on the fundamental recognition that there should be an additional layer of review specifically for advanced AI-assisted AWS *in addition* or *jointly* with the review of autonomy. A ML-specific review

[827] Jean-Marie Henckaerts and Louise Doswald-Beck (n 156) Volume II, Chapter 20, Section B, Rule 71. Weapons That Are by Nature Indiscriminate.
[828] US DoD, 'Nuclear Matters Handbook 2020' (2020) 18.

could assess and decide upon the acceptability of the process for producing new training data; test and evaluate procedures for determining whether new software updates are robust and reliable; and to explicitly state which potential features of ML-enabled AWS make a risk of using such weapons too excessive.

2. The Threat of General AI Fully Autonomous Weapon Systems

In this section, I argue that US DoD problem representation of AWS has left out of conceptualisation potential longer-term consequences of applying AI to AWS, in particular the emergence of general-level intelligence vastly superior to that of humans. I present how US DoD uses the strategy of denial or deflection when confronted with the topic of weaponised AGI. By 'weaponised AGI' I simply refer to the application of AGI in warfare. I argue that an alternative problem representation can be based on the recognition that the potential long-term threats of weaponised AGI are relevant and should be addressed now. I discuss various strategies for how militaries can mitigate the potential advent of weaponized AGI. First, one could argue that humans should always retain direct control over the force engagement decision, that is human-in-the loop. The second strategy is a non-waivered requirement that a human being should always retain the ability to stop a weapon's engagement, that is some form of human-on-the loop. Finally, a more radical view is that weaponised AGI is inevitable, and the only way humans can defend themselves is by developing a brain-computer interface to transform soldiers into AI-enhanced species that will be able to control machines.

2.1. The Directive on AWS Assumes Narrow Weapon's Applications

Today's application of AI to AWS includes engagement and targeting functions, but the progress is likely more advanced at other stages of targeting, such as 'find' and 'track.' Yet

US DoD has excluded considerations regarding the other functions of weapon systems by focusing exclusively on the 'select and engage' functions of AWS. This is another inconsistency with the *AI Principles* and *RAI Guidelines*, which are applicable to all stages of targeting. The greatest risk in the application of AI to warfighting lies, however, in the hypothetical creation of an artificial system that will be a general-level intelligent machine surpassing any human war fighter in any task and able to gain autonomy from human operators. I have discussed in Chapter 4 that by using an unbounded notion of autonomy, US DoD signals a message that the department might be interested in the creation of a fully autonomous artificial soldier. While militaries generally do not have an interest in developing weapon systems they cannot control, a military tolerance for risk could vary depending on the situations and, when advanced AI capabilities are applied to all stages of targeting, the department may inadvertently create a weaponised AGI system. In fact, US DoD in the past was interested in developing such systems.

For example, in 1983, DARPA established the Strategic Computing Initiative (SCI), a project that aimed to develop high-performance machine intelligence for military applications.[829] Instead of focusing on one specific problem, SCI was conceived to treat intelligent machines as a single problem composed of interrelated subsystems. The initiative was terminated in 1993, partly because it appeared that it would not succeed in creating AGI as originally planned; in part also because it faced controversy over the potential military use of SCI research.[830] Already back then the Computer Professionals for Social Responsibility, a non-profit organisation, called that SCI was oriented towards producing what they characterized as 'killer robots.'[831]Yet despite these past experiences, Directive 3000.09 does not include any specific safeguards against the creation of

[829] Alex Roland with Philip Shiman, *Strategic Computing DARPA and the Quest for Machine Intelligence, 1983–1993* (MIT Press 2002) 1.
[830] ibid 86.
[831] ibid 86–87.

weaponised AGI. The next sub-sections explore in more detail the concept of weaponized AGI and how one can challenge US DoD problem representation of AWS by including specific limitations against weaponised AGI.

2.2. Weaponised Artificial General Intelligence (AGI)

It is worth stressing that existing weapon systems that can activate lethal force in an autonomous way are considered as 'narrow AI,' i.e. these are systems that can achieve a specific goal, usually more effectively than a human, but they do not have any cognisance of how they relate to a goal.[832] In the academic literature, the weaponised AGI can be considered as a system that is based on the Generating Model. Leveringhaus argued that there are two models of autonomous targeting: The Generating Model and the Execution Model.[833]

According to the Generating Model, 'making an engagement decision' means that an artificial agent can evaluate whether a particular object is a morally legitimate target under certain moral criteria, such as the criteria of LOAC. In order to do this, the agent would need to be able to apply the principles of LOAC with respect to past decisions involving those principles. According to the Execution Model, the operator would assess whether certain potential targets were indeed legitimate targets whilst an artificial agent would then be able to execute the targeting decision in a manner consistent with criteria set out in its orders. There are two ways in which artificial agents, under the Execution Model, can determine whether an object is a legitimate target: conventional and unconventional. According to the conventional approach, the system makes 'decisions' to engage a target based on a programmed set of targets; if potential objects fall into the right category of

[832] Future of Life Institute, 'AI FAQ' <https://futureoflife.org/ai/ai-faq/>.
[833] Leveringhaus (n 26) 53.

targets, the machine will engage them. The unconventional approach is that machines can choose between different targets deemed legitimate by an operator.[834] Examples of such weapons are autonomous loitering munitions and LRASM. There are however no weapons which could satisfy the criteria of Generative Model. In other words, weapons of today, despite their ability to deploy a lethal force in an autonomous way, are not self-governing agents.[835] AGI is, therefore, a potential *future technology* which may greatly exceed a human's capabilities in all dimensions as it will display intelligence that is not tied to a highly specific set of tasks. Instead, it will generalise what it has learned, including generalising to contexts that are qualitatively different from those it has experienced before, or it will generally interpret its tasks in the context of the world at large.[836]

Herein lies a major risk. A system trained to maximise a certain type of performance may lead to unexpected outcomes. A good example is when an AI system, taught to play Tetris by researcher Tom Murphy at Carnegie Mellon University, was instructed not to lose. At the point where the system faced inevitable defeat the algorithm found a creative solution: pause the game and leave it paused - thus avoiding a loss.[837] This kind of indifference to broader norms about fairness may not matter much in a game but could be catastrophic in warfare. Further, a weaponised AGI may learn to override human decisions by rewriting its own code to increase its intelligence. By updating its rules, the machine can also update its values and potentially turn even against their creators.

According to the key drafters of Directive 3000.09 the development of such weapons, if it is even possible, is very distant and US DoD is not concerned with this

[834] ibid 56.
[835] Michael Robillard, 'No Such Things as Killer Robots' (2017) 35 Journal of Applied Philosophy 211.
[836] Dustin Lewis, Gabriella Blum, and Naz Modirzadeh (n 185) 18.
[837] Zachary Fryer-Biggs, 'Are We Ready for Weapons to Have a Mind of Their Own?' (*The Centre for Public Integrity*, 17 February 2021) <https://publicintegrity.org/national-security/future-of-warfare/mind-of-their-own-artificial-intelligence-weapon/>.

process. The US DoD Law of War Manual states that the laws of war 'impose obligations on persons […] not on the weapons themselves.'[838] This means that for US DoD weapon systems are conceived of as tools in the hands of people, rather than agents. Work further argues that the development of weaponised AGI is 'very, very, very far in the future because general AI hasn't advanced to that.'[839]

However, recent commercial progress in developing AI systems capable of exhibiting intelligence across many different dimensions have accelerated. For example, DeepMind's new model, Gato can solve multiple unrelated problems: it can play many different games, label images, chat, and operate a machine, among other things.[840] Just few years ago, a dominant view within the AI community was that AI systems are 'narrow' in the sense that they are capable only to solve a single problem. This is no longer true, and some authors argue that we are actually 'on the verge of artificial general intelligence and the only problem left is scale of general-purpose AI systems.'[841]

2.3.A Discourse of Denial of AGI

Stuart Russel, a prominent supporter of the Campaign, explores various discourses regarding the threat of AGI. Based on his classification, the US DoD approach can be regarded as *denial*. The major denial strategies are claims that (1) it is too soon to worry about the self-governing agents and/or (2) that the arrival of AGI is impossible.[842] US DoD representatives in their statements often refer to so-called 'AI experts' to justify their claims. These experts are primarily internal research groups or US DoD funded external

[838] DoD (n 432).
[839] Scharre (n 34).
[840] Deepmind, 'A Generalist Agent' <https://www.deepmind.com/publications/a-generalist-agent>.
[841] Mike Loukides, 'Closer to AGI? And Is Artificial General Intelligence What We Really Need?' (*O'Reilly*, 7 June 2022) <https://www.oreilly.com/radar/closer-to-agi/>.
[842] Stuart Russel (n 124) 146–152.

groups. For example, a recent report into the use of military AI belittled the threat from AGI. The report was prepared by JASON, an advisory group of US scientists funded US DoD that informs the US government on science and technology policy. In the report, JASON claims that:

> To *most* [MF emphasised] computer scientists, the claimed 'existential threats' posed by AI seem at best uninformed." They do not align with the most rapidly advancing current research directions of AI as a field, but rather spring from dire predictions about one small area of research within AI, Artificial General Intelligence (AGI). […] Where AI is oriented around specific tasks, AGI seeks general cognitive abilities. On account of this ambitious goal, AGI has high visibility, disproportionate to its size or present level of success.[843]

An alternative problem representation is advanced by the Campaign, whose representatives point out that the question of whether the engineering system called AGI is technically available is far from clear and a highly contested issue. Based on some recent surveys the majority of AI experts selected among academics and practitioners argue that AGI will be eventually created.[844] For instance, in May 2017, 352 AI experts were surveyed and estimated that there was a 50% chance that AGI would occur until 2060.[845] In 2019, 32 AI experts participated in another survey and 45% of respondents predicted the arrival of AGI before 2060, 34% of all participants predicted a date after 2060, while 21% of participants predicted that AGI would never occur.[846]

---

[843] Richard Potember, 'Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD' (JASON (the Mitre Corporation) 2017) JSR-16-Task-003 3.
[844] Nick Bostrom (n 132) 23.
[845] Katja Grace and others, 'When Will AI Exceed Human Performance? Evidence from AI Experts' <https://arxiv.org/pdf/1705.08807.pdf>.
[846] Daniel Faggella, 'When Will We Reach the Singularity? – A Timeline Consensus from AI Researchers' (*EMERJ*, 18 March 2019) <https://emerj.com/ai-future-outlook/when-will-we-reach-the-singularity-a-timeline-consensus-from-ai-researchers/>.

Further, one could argue that there is a strong probability that AGI may be created inadvertently from the evolutionary premise and the sheer nature of technological progress. Authors argue that we already know that blind evolutionary processes can produce human-level general intelligence, since they already done so at least once. Thus, evolutionary processes coupled with human intelligence should be then able to achieve a similar outcome with greater efficiency.[847] Russel argues that one of the challenges with weaponized AGI is that war involving these weapons would be intrinsically unpredictable and likely lead to the potential large destruction of civilians and environment.[848] It will be particularly challenging to address the complex adaptive system of multiple general-AI AWS, especially if they will be influenced by different ethical principles and cultures regarding what is acceptable during war. Therefore, some supporters of the Campaign, argue that such weapons should be regarded as weapons of mass destruction given how significant damage their potential use can inflict.[849]

2.4.A Discourse of Deflection of AGI

Some US DoD representatives argue that, even if the threat of AGI can be real, the development of military weapon systems should still not be restricted. This is a discourse of deflection regarding the threat of AGI and the main argument is that US DoD cannot control the development of technology in the realm of international military competition. Work, when pressed on the topic of AGI, agreed that these developments could result in significant challenges.

[847] Stuart Russel (n 124).See also for the critical discussion: Nick Bostrom (n 132) 28–35.
[848] Stuart Russel (n 124).
[849] See Peter Asaro, 'Ban Killer Robots before They Become Weapons of Mass Destruction' *Scientific American* (7 August 2015); Liz O'Sullivan, 'Side Event "The Urgent Need for a Treaty to Retain Meaningful Human Control over the Use of Force"' (2019).

> The danger is if you get a general AI system and *it can rewrite its own code* [MF emphasised]. That's the danger. We don't see ever putting that much AI power into any given weapon.' […] We will be extremely careful in trying to put general AI into an autonomous weapon.[850]

However, despite this claim, Work acknowledged that if other countries start to use general AI in the battlefield the US military may need to rethink its approach.

> The only way that we would go down that path, I think, is if turns out our adversaries do, and it turns out that we are at an operational disadvantage […]. The nature of the competition about how people use AI and autonomy is really going to be something we *cannot control,* and we cannot totally foresee at this point.[851]

This means that at least some of the architects of Directive 3000.09 are leaving open the possibility of developing and using weaponised AGI. To be clear, this is not to say that DoD's stated goal is to develop and deploy weaponised AGI. None of the official strategies evoke this concept directly and even among the architects of the Directive 3000.09 there is a degree of reservation towards such an idea.[852] This being said, US DoD leaves the door open for further significant improvements of already existing AI applications and it may build weaponised AGI inadvertently due to the technologically progress of narrow-AI AWS and the pressure of global competition.[853]

2.5.An Alternative Problem Representation to Weaponised AGI

[850] Scharre (n 34) 98.
[851] ibid 99.
[852] ibid.
[853] SeeCongressional Research Service, 'Renewed Great Power Competition: Implications for Defense—Issues for Congress' (2020)..

Taking the above into account, US DoD problem representation of AWS in the context of AGI can be summarised as follows: Narrow-AI AWS are technically controllable, and the risk of general-AI AWS is low and/or irrelevant. An alternative problem representation can be based on the recognition that the potential long-term threats of weaponised AGI are relevant and should be addressed now. There are several alternatives for how one can address the problem of weaponised AGI.

First, one can argue that humans should always retain a direct control over the engagement decision. In other words, humans should always be in the loop and engage targets by themselves either in the war theatre or remotely. As discussed, this is the Campaign's counter-discourse to the US DoD narrative and the consequence of such a view is that many of the existing semi-autonomous systems should not be allowed.

Second, a more nuanced view is that there must be a non-waivered requirement that a human being should always retain the ability to stop a weapon's engagement. In fact, one version of such a requirement is already present in Directive 3000.09, but Section 2 of Enclosure 3 of Directive 3000.09 gives an opportunity to waive the requirement in the case of urgent military operational need.'[854] As discussed in Chapter 5, the term 'urgent military operational need' is open to broad interpretation and effectively makes the requirement less mandatory as various missions can meet the criteria for urgent military need. The US DoD requirement of human intervention currently takes the following form:

> The system is designed to complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, to terminate engagements or seek additional human operator input before continuing the engagement.[855]

[854] Directive 3000.09 Autonomy in Weapon Systems Enclosure 3, 2.
[855] ibid 4a (2).

An alternative problem representation can retain such a requirement of human intervention and remove the waiver. This can be further specified by programming the requirement of human intervention in the initial code that dictates what codes AI cannot rewrite by itself.

Finally, a more radical view is that weaponised AGI is inevitable, and the only way humans can defend themselves is by transforming soldiers into AI-enhanced species that will control machines with their thoughts. In this hypothetical scenario, a brain-computer interface will be established by a direct communication pathway between the brain's electrical activity and an external device, most commonly a computer or robotic limb. While existing brain-computer interface efforts are primarily concentrated on augmenting, or repairing human cognitive or sensory-motor functions, such an application, if successful, can also be applied in weapon systems.[856]

In fact, DARPA has worked for decades on various projects related to human enhancement. In 1997, the Agency created the Controlled Biological Systems program, which is credited for, among other things, developing brain-controlled prosthetic arms for soldiers who have lost limbs.[857] The Continuous Assisted Performance project attempted to create a '24/7 soldier' who could survive without sleep for up to a week. 'Soldiers having no physical, physiological, or cognitive limitation will be key to survival and operational dominance in the future,' said one of DARPA managers.[858] Some DARPA representatives openly admitted that they were exploring the possibilities of using brain-controlled techniques to augment healthy soldiers, for example, in the area of 'memory prothesis'

[856] Robbin Miranda and others, 'DARPA-Funded Efforts in the Development of Novel Brain–Computer Interface Technologies' (2015) 244 Journal of Neuroscience Methods 52.
[857] Al Emondi, 'Revolutionizing Prosthetics (Archived)' (*DARPA*) <https://www.darpa.mil/program/revolutionizing-prosthetics>; Michael Gross, 'The Pentagon's Plans to Program Soldiers' Brains' (15 November 2018) <https://www.theatlantic.com/magazine/archive/2018/11/the-pentagon-wants-to-weaponize-the-brain-what-could-go-wrong/570841/>.
[858] Michael Gross (n 857).

where the idea is to inject memory using the precise neural codes for certain skills.[859] Various scientists doubt whether this experiment will be feasible, as humanity still has limited knowledge about the way the human brain works, let alone how to emulate it and then transmit a device to a brain in a non-invasive manner.[860] Yet these DARPA-led efforts suggest that one should at least consider that such a possibility might exist. Thus, the current problem representation of AWS can be challenged by not addressing these matters.

3. Moral Concerns over the Use of Autonomous Weapon Systems

In this section, I argue that the US DoD problem representation of AWS has left deep ethical concerns regarding the use of such weapons. I then argue that an alternative problem representation of AWS might include ethical considerations as certain restrictions over the development and use of such weapons. I restate that the inclusion of a direct human control requirement at the level of engagement can not only alleviate potential threats from weaponised AGI but also limit ethical concerns. I also discuss various proposals to build an ethical machine which might be able to assess whether a specific lethal action is ethically permissible or at least ethically impermissible. I also explore the possibility of creating a dedicated oversight agency with the goal of providing more accountability regarding the use of AWS, potentially alleviating some fears that such weapons are used for impermissible ethical action according to wider public morality.

3.1.'We Did Not Consider Ethical Issues as Relevant'

While the ethical challenges regarding AWS are highly debated, also within US DoD, the department conceptualisation of the risks associated with the use of AWS, in fact, leaves ethical concerns unaddressed. To illustrate this argument, it is worth identifying *what*

[859] ibid.
[860] ibid.

specific risks constitute the dominant US DoD discourse in the wider debate about the various types of risks associated with the use of semi-autonomous weapons supervised systems and later fully autonomous weapon systems. In the chapter 1 I outlined five common types of risks associated with AWS based on the literature review. First, the use of AWS can lead to at least partial moral disengagement of human operators. Second, AWS can generate situations in which no one can be held responsible for what a machine does. Third, the use of such weapons can lead to unintended consequences, such as the engagement of friendly forces. Fourthly, the use of AWS can, at least in specific cases, violate the main principles of LOAC and human rights, such as the principle of proportionality and distinction. Fith and finally, the use of increasingly autonomous weapon systems may lead to an AI technology arms race and pose a 'security dilemma' resulting in decreased stability and security among nations.

According to poststructuralist approach 'the risk' as such does not exist as a 'given fact' waiting to be discovered. Rather, what constitutes 'risk' is the result of contingent outcomes of a struggle between competing discourses which transform 'what is out there' (e.g. Patriot fratricide) into a socially, policy, and politically relevant issue (e.g. a specific conceptualisation of 'risk,' e.g. in purely technical terms as a software's poor image classification capabilities). The US DoD problematisation of risks associated with the use of semi-autonomous weapons supervised systems and AWS was focused on the risk of *unintended consequences* rather than other risks. While some of the identified risks may to some observers look interconnected - e.g. the existence of a responsibility gap may render the use of certain weapons incompatible with LOAC - US DoD framing of risk is different. US DoD explicitly challenges the existence of a responsibility gap and has repeatedly argued that even lethal applications of AWS can be used in a manner consistent with LOAC. The potential ethical arguments against AWS *did not* resonate much during the

drafting process of the Directive 3000.09. One of the drafters of the Directive stated in 2021:

> Initially, during the drafting process of the Directive, I did not consider ethical issues as very relevant. However, over time, my view has shifted, and I think that ethical questions are important.[861]

The major recognised risk that led to the drafting of Directive 3000.09 was the risk of *unintended consequences*, such as those illustrated by the Patriot fratricide during Operation Iraqi Freedom. However, what constitutes 'the risk of unintended consequences' is also the result of a struggle between competing discourses. Some agents such as the DSB or Hawley, or Cummings argue that it is important to emphasise direct human control over the use of Patriots given the fact that fratricides occurred in part because the procedures for human-machine interfaces were poorly designed so that Patriots relied too much on automation. Yet the dominant discourse within US DoD has formulated the risk of unintended consequences differently. Their focus was primarily on addressing the technical shortcomings of the weapon system, such as missiles image recognition classification and communication and coordination between the missile batteries and other systems deployed in the field.[862] They did not want to challenge the use of the Patriot system in the automatic mode, instead opting for a further decrease of direct human control. The dominant risk discourse was that the lack of manual human control has actually increased the overall 'level of control' over the weapon systems.[863]

[861] Interview with Paul Scharre (n 256).
[862] John K. Hawley (n 301).
[863] US DoD, 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (n 6).

The introduction of weapons with supervisory control in the automatic mode has led to the gradual removal of direct human control. Instead, the key element of risk analysis was to ensure that human judgment will be effectuated in the wider decision-making process of using weapon systems and *the systems will perform as expected*.

3.2.Ethical Arguments Against AWS

What is left unproblematic in US DoD approach towards AWS is that such weapons, even if they meet all necessary legal and technical reviews, may still be considered as unacceptable *solely on an ethical basis*. There can be different versions of this argument. First, it can be argued that lethal use of AWS undermines human dignity when deployed to engage with human targets.[864] For example, Heyns argues that autonomous killing undermines the protection of dignified life.[865] A concept of dignified life means that each person is entitled to be treated according to his or her own full merits. The consequence is that people cannot be reduced to a statistical number in algorithmic calculations. If this happens, one person's death is indistinguishable from that of so many others who happen to find themselves within the range of autonomous targeting. An illustrative example is the ruling of the German Constitutional Court that legislation allowing the Minister of Defence to authorise the shooting down of a civilian aircraft involved in a 9/11 style terror attack was unconstitutional.[866] This was despite the lives of those that would potentially be saved and based on the argument that shooting down the plane would be a violation of the right to dignity of those on board.

[864] Christof Heyns (n 48) 3–20.
[865] Christof Heyns (n 48).
[866] *Authorisation to shoot down aircraft in the Aviation Security Act void* [2006].

Second, authors such as Leveringhaus argues that the replacement of human agency with artificial agency at the point of force delivery is not morally desirable because it leads to moral disengagement.[867] The argument does not suggest that human operators are not entirely disengaged. They must ensure that the use of AWS does not lead to excessive risks. They are not, however, *fully* morally engaged. It is because to be a fully morally engaged human is more than just to respect someone else's rights. It is to act for reasons that are not entirely rights-based, such as recognition of a common humanity, a concern for the vulnerable or pity and mercy. Thus, the replacement of human agency with artificial agency leads to at least partial moral disengagement.[868]

This being said, one may argue that what has been left unproblematic in the DoD representation of the AWS is a deep ethical concern about the future of warfare. This argument could be even stronger when supported by the public opinion. For instance, according to a survey conducted by Ipsos in 2019 and commissioned by the Campaign, 61% of respondents from 26 countries said that they opposed the use of LAWS. Interestingly, among those who were opposed, 66% said that they were so because they believe LAWS cross a moral line as machines should not be allowed to kill, while only 21% of respondents cited legal concerns.[869]

3.3.A Problem Representation that Considers Ethical Arguments

Thus, one can argue for a problem representation of AWS that includes ethical considerations as certain restrictions over the development and use of AWS. Various proposals have been discussed in the academic literature. I will discuss four approaches:

[867] Leveringhaus (n 26) 90–94.
[868] ibid.
[869] IPSOS, 'Six in Ten (61%) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems' (22 January 2019) <https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons>.

(1) Direct human control at the level of engagement; (2) the ethical governor; (3) the "MinAI" ethical robot; and (4) an Oversight Agency.

Much has been already said in this thesis about the concept of direct human control. Similarly, as in the discussion about the limitations of AGI, the argument is that humans should always be in the loop and engage targets by themselves. Again, such view would against using many of existing weapon systems.

The ethical governor is a concept developed by Arkin, who argues that LAWS should be programmed in such a manner that they will be able to assess whether a specific lethal action is ethically permissible under all possible conditions, including key principles of LOAC and considerations regarding human dignity.[870] However, such a concept is difficult to code in programming language as it would require extensive ethical engineering. Thus, the implementation of the ethical governor is contingent on having the constraint application process responsible for reasoning about the active ethical constraints of the machine's action. Yet it is unreasonable to expect that such a constraint application process would be 'neutral' as humans' ethical preferences vary greatly. For example, dignity is a complex matter, and many people may understand this concept differently. Further, it can be argued that different cultures are guided by different ethical principles, as illustrated by the MIT Media Lab experiment 'Moral Machine.'[871]

A third proposition is a concept called a "MinAI" ethical robot.[872] This is another ethical constraint-driven approach where the constraint features are at the level of weapon's

[870] Ronald Arkin, Patrick Ulam, and Brittany Duncan, 'An Ethical Governor for Constraining Lethal Action in an Autonomous System' (2009) GIT-GVU-09-02.
[871] Edmond Awad and others, 'The Moral Machine Experiment' (2018) 563 Nature 59.
[872] Jason Scholz and Jai Galliott, 'The Humanitarian Imperative for Minimally-Just AI in Weapons', *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021).

design. The concept is a more modest variation of the 'ethical governor' and postulates a narrower set of constraints. MinAI deals with what is ethically impermissible, rather than what is ethically permissible. It aims to satisfy the following basic principles of morality:

(1) *Distinction of the ethically impermissible* including the avoidance of application of force against 'protected' things such as objects and persons marked with the protected symbols of the Red Cross, as well as protected locations, recognizable protected behaviours such as desire to parlay, basic signs of surrender, and potentially those that are hors de combat, or are clearly non-combatants.

(2) *Ethical Reduction in proportionality* includes a reduction in the degree of force below the level lawfully authorized if it is determined to be sufficient to meet military necessity.[873]

The argument for MinAI is that, as opposed to an ethical governor, the evaluation of ethical behaviours is limited to negative cases of things that should not be attacked and, importantly, the approach seems to be practically achievable within the current state of AI and does not engage in difficult ethical considerations of what permissible ethical action should look like. The technical implementation of the MinAI Ethical Weapon would depend on the augmentation of the weapon seeker and advanced computer vision to recognise commonly accepted signs of surrender (e.g. lowering a flag, discarding weapons, abandoning armed vehicles or aircrafts etc.).[874] Furthermore, as there is no universally agreed international sign or symbol of surrender, the practical implementation of such proposal will likely further depend on the creation of such symbol akin to the sign of Red Cross used for humanitarian purposes.[875]

[873] ibid 58–59.
[874] ibid 59.
[875] ibid 61–67.

A final proposition is the resignation from any hard ethical limitations at the expense of transparency oregarding the lethal use of AWS. The clarity over the rationale for using AWS in a specific war theatre will at least provide more accountability and potentially alleviate some fears that such weapons are used for impermissible ethical action according to the wider public morality. This proposal can be implemented by the creation of the oversight agency that would monitor the development and use of AWS with the ability to reverse the decision about the weapons' deployment.[876] The potential use of AWS would then be contingent on providing information of (1) the operational areas in which such weapon systems are being used; (2) the purpose of engagement, i.e. whether offensive or purely defensive; (3) whether human targets will be engaged; (4) what is the legal rationale for the engagement.[877] After the mission the agency should also (i) collect data on how these weapon systems were performing, particularly regarding the target identification metrics; (ii) a complete account of the provenance of data, processes, and artifacts involved in the decision-making process.

The counterargument to this proposal is that militaries often must hide behind a veil of secrecy to gain a competitive advantage over adversaries. Yet I would argue that such a transparency framework can represent good balance between legitimate military interests and holding the military accountable. It will not disclose any specific mission-parameters such as targets or specific localisation before the mission; it will also not disclose the inner workings of the weapon systems. Yet such transparency framework can still provide relevant information for the public opinion, such as whether AWS are used in complex environments where there is a higher probability of encountering civilians, or whether such

[876] Steven Barela and Avery Plaw, 'Programming Precision? Required Robust Transparency for AWS', *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021).
[877] See a similar argument made by ibid.

weapons are used for purely defensive missions. Gradually, the public will be able to form opinions about certain types of AWS operations and may put more pressure on the agency to grant certain limitations regarding the use of AWS, e.g. only for defensive purposes.

4. The Exclusion of Cyber Weapons and Their Complexities

Directive 3000.09 explicitly excludes autonomous or semi-autonomous systems for cyberspace operations.[878] One of the drafters of Directive 3000.09 explained that the reason why cyberweapons have been exempt was because:

> […] We knew bureaucratically it would be hard enough simply to create a new policy on autonomy. Adding cyber operations would have multiplied the complexity of the problem, making it very likely we would have accomplished nothing at all.[879]

Yet cyber weapons are becoming more and more potent, while countries such as the US do not have clear rules how to respond to such threats. The problem is pressing as on the same day as Directive 3000.09 was issued, DARPA announced a programme to create a 'foundational cyberwarfare' capability that would allow US DoD to better monitor, exploit, and attack an enemy's systems.[880] Moreover, cyber weapons may in fact represent an even greater threat than the isolated uses of kinetic AWS because they actually exist in various forms, have been used for offensive purposes, proliferate rapidly and have a large scale as they could generate malicious effects across the Internet.[881]

4.1. The Development and Use of Autonomous Cyber Weapons

[878] Directive 3000.09 Autonomy in Weapon Systems 2b.
[879] Scharre (n 34) 228.
[880] Jennifer Roberts, 'Plan X (Archived)' (*DARPA*) <https://www.darpa.mil/program/plan-x>.
[881] François Delerue, *Cyber Operations and International Law* (CUP 2020) 160.

There are differences between autonomous cyber and kinetic weapons. The obvious one is that cyber-weapons act in the cyber-sphere, although with the possibility of causing serious harm to physical infrastructure such as power grids, banking network systems, government records, and other critical infrastructure.[882]

Further, both kinetic AWS and cyber-weapons offer the possibility of being 'supervised' or 'autonomous' systems. It is important to emphasise that not all cyber weapons are, in fact, autonomous.[883] Cyber weapons exhibit little autonomy if they are controlled by pre-established algorithms, and conduct targeting operations according to predetermined scenarios, or when they are supervised by humans. However, some cyber weapons are considered autonomous when there is no communication with a human agent once launched. Furthermore, such autonomous cyber weapon can also be augmented with advanced AI and thus learn from external variables and adjust its target selection and target engagement. Such weapons can become powerful force multipliers enabling cyber-attacks at faster speeds across multiple military domains. They can also be engineered to self-replicate and self-propagate. Depending on the design of the weapon or on unanticipated situations AI-augmented autonomous cyber weapons may spread in unpredictable and potentially uncontrollable ways.[884] An example is the Stuxnet worm, developed by the US and Israeli intelligence agencies, which destroyed Iran's nuclear centrifuges at the Natanz nuclear facility.[885] Stuxnet was able to identify specific models of programmable logic controllers (PLCs) which allowed the facility's computers to control the centrifuges used

[882] National Audit Office, 'Investigation: WannaCry Cyber Attack and the NHS' (2017).
[883] Tanel Tammet, 'Autonomous Cyber Defence Capabilities', *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021) 37.
[884] Daniel Trusilo and Thomas Burri, 'Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems', *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021) 58.
[885] Ann Väljataga and Rain Liivoja, 'Cyber Autonomy and International Law: An Introduction', *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021) 3.

to enrich uranium.[886] The worm modified the PLCs' programming while at the same time providing computer operators with fake sensor values, causing them to believe that the PLCs were functioning normally. The Stuxnet virus was autonomous by design as, once deployed, it could not interact with its human operators since, for security purposes, Natanz is connected to the wider Internet.

4.2.Autonomous Cyber Weapons Generates Novel Problems

The development and use of AI-augmented autonomous cyber weapons generates at least four operational challenges for existing C2 architecture. First, there is the problem of offensive unpredictability.[887] An intelligent malware agent with self-learning capabilities can learn and override defensive acts to exploit any potential vulnerabilities of a system.[888] Such a system may however spread in uncontrollable way by self-replicating. The second challenge is concerned with offence undetectability, called in cyber studies 'the attribution problem'.[889] A complex malware is difficult to detect and, even when recognised, one can only provide an account for known intrusions. A challenge is to confront with a prospect of the permanent intrusion of the defender's infrastructure when the scale and scope of infiltration can raise significant security issues. A third challenge relates to the complexity of the defence system.[890] While the attacker usually only needs to understand the procedures of entry, the defender must protect the entire network against many interrelated points of vulnerabilities. Fourth and last, traditional C2 architecture is under pressure from

[886] Nicolas Falliere, Liam Murchu and Eric Chien, 'W32.Stuxnet Dossier' (Symantec 2010); Kim Zetter, 'An Unprecedented Look at Stuxnet, the World's First Digital Weapon' *Wired* (3 November 2014) <https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/>.
[887] Lucas Kello (n 178) 68–69..
[888] Miles Brundage and others, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation' (2018) 20.
[889] Lucas Kello (n 178) 69–72, 129–132.
[890] ibid 72–73.

supply chain risks, such as manufacturers who introduce vulnerabilities into specific components of a system.[891]

4.3.Regulating Autonomous Cyber Weapons

Despite these challenges, and despite the fact that autonomous cyber weapons have been excluded from US DoD problem representation, this does not mean that US DoD is not advancing such weapons, even for offensive purposes, as the Stuxnet case illustrates. Such a situation creates a legal, ethical, and policy vacuum when it comes to the potential development and use of particularly offensive autonomous cyber weapons. Hence, an alternative way of seeing the problem representation of AWS is to include considerations regarding autonomous cyber weapons. There are at least two major issues to be addressed.

First, one should consider what mechanisms restricting the spread and use of autonomous cyber weapons should be in place, if any. For example, some autonomous cyber weapons may be deployed into private or secure networks, in which case the supervisor is unable to communicate with them. In such cases, the degree of effectiveness of a supervisor's control over such weapons can be questioned. This was the case with the Stuxnet worm. However, even Stuxnet had several safeguards in place to limit its spread and effects, for instance it could not modify itself autonomously and had a self-termination date.[892] While these mechanisms were in place to contain the proliferation of the weapon, the worm eventually spread to other systems outside Natanz and across the world.

One can argue that cyber weapons are even less discriminating than currently existing AI-augmented autonomous kinetic weapons due the ability of such weapons to

[891] ibid 73–74.
[892] Scharre (n 34) 224.

self-replicate.[893] Thus, it is likely that autonomous cyber weapons without any safeguards' mechanism will violate LOAC. One might therefore consider the regulation of cyber weapons, which will include various layers of restrictions depending on the specific use cases. For example, the baseline rule can state that such weapons should be discriminate by design. Cyber weapons can have restricting attack capability to specific targets within certain networks and for a limited period. There might also be limitations regarding a weapon's area of operation or compliance with certain defined rules of engagement. Such design choices will allow the use of autonomous cyber weapon in a more deterministic and predictable fashion. A second layer of restriction may include the requirement to use a cyber weapon only under the supervision of a human operator, for instance, through a real-time monitoring mechanism that allows the operator to adjust the algorithm to modify instructions, assign new tasks, or correct software bugs. A supervisor may also have the ability to deactivate a cyber weapon if it starts to behave unexpectedly or once it has successfully completed its mission.

A second area of consideration should assess the purposes for which AI-augmented autonomous cyber weapons should be used. The US has already used such weapons for offensive purposes, while a NATO research group postulated the use of autonomous cyber weapons to defend their computer network.[894] At first sight, the application of such weapons for defensive purposes seems logical. The more advanced capabilities to detect network vulnerabilities are naturally good. However, in some situations applying AI-augmented autonomous cyber defence may be too dangerous. Consider nuclear facilities. AI applications designed to enhance cybersecurity for nuclear facilities could simultaneously make nuclear weapon system more vulnerable to cyber-attack. Adversarial

[893] Daniel Trusilo and Thomas Burri (n 884) 62–63.
[894] Alexander Kott and others, 'Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture. Release 2.0' (18 September 2019) <https://arxiv.org/abs/1803.10664>.

AI techniques that aim to compromise the use of AI models could infiltrate a nuclear weapons system, destabilise its communications, and possibility even gain control of the nuclear arsenal. Thus, it may be reasonable to provide certain limitations of the scope of using AI-augmented autonomous cyber weapons for defensive purposes. The use of these weapons for offensive purposes is even more complicated as it may lead to inadvertent escalation. Consider the following example: State A launches an AI-augmented cyber weapon to exploit State B's infrastructure, but State B has an AI-augmented defence weapon which autonomously responds to the attack by attacking State A's infrastructure. State B responds in the same manner and this escalation of operations leads to unproportionate chain reactions known as 'flash wars.'[895] In order to avoid such inadvertent escalation, one could argue that any use of AI-augmented cyber weapons for offensive purposes should go through additional review in a similar manner to the use of lethal autonomous kinetic weapon systems.

5. A Summary of the Chapter

In this chapter, I have discussed what has been left unproblematic in the US DoD problem representation about AWS. I have argued that US DoD governmentality of AWS focuses heavily on the concept of autonomy but disregards the evaluation of advanced AI capabilities of AWS. I have explored that the assessment of advanced AI capabilities, i.e. ML differs from traditional software evaluation as the outcomes of ML models are only probabilistic, not deterministic and it is challenging to verify the correctness and accuracy of the ML model outcomes. Thus, I have argued that there should be an additional layer of review specifically for ML-augmented AWS in addition to the review of autonomy.

[895] Scharre (n 34) 199–210.

The lack of consideration of AI capabilities of AWS has led me to reflect further on the potential threat posed by weaponised AGI, which has been explicitly denied or deflected by US DoD. I have recommended an alternative problem representation calling for the inclusion of measures to mitigate the potential advent of AGI. I have argued that one measure could be the inclusion of the requirement for direct human control over the weapon's engagement decision, but this requirement will be difficult to implement in practice as it would render many existing weapon systems illegal. Thus, I have provided an argument for a non-waivered requirement that a human being should always retain the ability to stop a weapon's engagement. I have also reflected on the possibility of a brain-machine interface to create autonomous soldiers who might be able to control weaponized AGI.

US DoD's disregard of the risks posed by weaponised AGI stems from their believe that AWS augmented with narrow AI capabilities are controllable in technical terms. I have exposed that the conceptualisation of risk associated with the development and use of AWS in purely technical terms excludes deep ethical concerns regarding the use of such weapons. I have thus discussed an alternative problem representation whereby ethical considerations serve as certain restrictions over the development and use of such weapons. Similarly, I have restated that the inclusion of a direct human control requirement at the level of engagement can not only alleviate potential threats of weaponised AGI but also, to certain extent, accommodate ethical concerns. I have also discussed a proposal to build an ethical machine which might be able to assess whether a specific lethal action is ethically permissible. I have concluded that the practical feasibility of such a proposal is limited, but it is more likely that a weapon system can be designed to assess at least whether the action is ethically impermissible. I have also postulated the creation of a dedicated oversight agency with the goal of providing more accountability regarding the use of AWS. Such an

agency could alleviate some fears that such weapons are being used for impermissible ethical action according to the wider public morality.

Finally, I have concentrated on another significant omission from the US DoD problem representation of AWS, that is the lack of consideration of autonomous cyber weapons. I have argued that the inclusion of cyber considerations must address the problem of what limitation mechanisms restricting the spread and use of autonomous cyber weapons should be in place and whether AI-augmented AWS can be used for both defensive and offensive purposes. I have concluded that it may be justified to provide certain limitations over the scope of use of AI-augmented autonomous cyber weapons even for defensive purposes, particularly in the context of protecting nuclear weapons.

CONCLUSION

Robotics systems play an increasingly important role in armed conflicts and there are already weapons in service that replace a human being at the point of engagement. The US is the first country that has adopted a policy on AWS in Directive 3000.09. It is a significant document not only because it is first, but also it could serve as a template of how to regulate autonomous systems more generally and how one could think about the role of human involvement over the use of such systems. Yet, despite the significance of Directive 3000.09, the US policy on AWS is poorly understood in the academic and policy circles. This DPhil thesis addressed the question of how US DoD constructs and problematises the concept of AWS. It asks the questions of how the 'problem' of AWS is constructed, what assumptions underlie this representation of the 'problem', how has this representation of the 'problem' come about, what effects are produced by this representation of the 'problem' on the emergence of new norms within US DoD, and finally what is left out of the problem representation and how could this distinct representation of the problem of AWS be questioned and disrupted.

The US DoD problematisation of AWS does not only provide more details about US DoD approach to such weapon systems, but – most centrally – it explores the role of human involvement in the use of AWS. In Chapter 1, I have argued that the focus on the role of human involvement over the use of AWS stems directly from the US definition of AWS which shifts the conceptual problem of defining AWS to the relationship between human and machine. I have further argued that, while the notion of human involvement regarding the use of AWS has been formulated in several different ways by both policymakers and academics, the various terms that are being used by different actors – such as meaningful human control or human judgment – are often conflated and lack in-depth consideration.

Authors who recognise the importance of defining and analysing the role of human-machine interaction over the use of AWS focus predominantly on defining key terms e.g. by describing whether human involvement should be called 'human control' or 'judgment' and content of these terms, e.g. describing elements of human involvement, but there is little focus on the context within which this human involvement ought to and is exercised, more specifically, who exercise control, how, and over what. By exploring the US DoD problematisation of AWS this thesis focuses not only on content of the concept of the US policy of human judgment, but primarily on how this concept relates to the wider US military understanding of 'control.'

In that, it unpacks the concept of human judgment and distinguishes it from the concept of human control. I have argued that both concepts are important in the debate on AWS as they represent alternative policy approaches to the use of such weapons. I have argued that by making these concepts more explicit, my thesis contributes to the specific and emerging academic debate about the operationalisation of the role of human involvement over the use of AWS.

My DPhil thesis presents that a poststructuralist analysis of how a specific problem has been constructed by a policy allows to better understand how this policy could work in practice in the wider institutional setting. While academic authors often criticise that the concept of 'human judgment over the use of AWS' is too general and does not specify whether US DoD would like to retain humans in the loop at the point of engagement, the study of problematisation – according to my modified Bacchi's approach – allows to situate the requirement of human judgment in the wider US DoD context and the USAF targeting practices. Further, the exploration of the US DoD problematisation of AWS does not only result in a more detailed understanding of a specific policy, but it opens a perspective that makes politics, understood as the complex strategic relations that contributed to the

introduction of this specific policy, visible. Thus, this exploration allows to reflect on how one could challenge a dominant US DoD problem representation of AWS by appreciating the political and institutional drivers of such a representation. Such a critique has meaningful practical merits. If one is indeed interested in a *policy change*, one must consider a wider political context of why a specific policy has been established, to what problem it purportedly responds, how that problem was socially constructed, and what assumptions underpin such problem construction. Based on the US DoD problematisation of AWS one could *reasonably* postulate within the US political context a policy change to include a review applicable for AI-augmented AWS or postulate the prohibition of using ML-enabled AWS with nuclear warheads. The analysis presented here, however, makes it less likely for the US policymakers to accept the introduction of the concept of meaningful human control over the use of AWS at the point of engagement or international law treaty to regulate the use of AWS. In fact, the US DoD problematisation of AWS reveals how much more influential are established domestic military practices rather than LOAC considerations for the introduction of certain limitations related to the use of such weapons.

I have been able to generate these findings thanks to the poststructuralist methodology and a concentrated case study focused on US DoD. In the academic literature, many authors focus on the analysis of legal limitations over the use of AWS, but they overwhelmingly apply a positivist approach to policy analysis according to which policies are considered as more or less self-evident responses to 'objective social problems.' They neither try to unpack and contextualize the approach of a specific country to the use of AWS, nor do they study military targeting practices in more depth. I have argued that the use of poststructuralist policy analysis with an in-depth case study of US DoD and a nested case study of USAF coupled with newly generated data from interviews with military

practitioners represents a novel contribution to the academic literature on AWS. The next section presents my findings in more detail.

1. Main Research Finding

In this DPhil thesis I have argued that US DoD constructs the problem of AWS based on their potential 'lethal use', that is the use of such weapons to kill other humans. According to US DoD, such a potential application of a lethal force in an autonomous way increases the risks of 'unintended engagements.' This perspective has been informed by mixed experiences with the use of autonomous supervised weapons, which have become complex socio-technical systems requiring heightened trust measures.

I have argued that the notion of trust has a particular significance within US DoD to address the risks of unintended engagements of AWS. US DoD identifies two layers of trust-building measurers in AWS. First, in the socio-technical system of using AWS there must be a trust and deep integration between humans and machines. Second, there must also be trust in machine's autonomous capabilities that it will produce predictable outcomes. In other words, that the system has trustworthy autonomous capabilities. Thus, I have argued that the potential higher bar of risk of unintended engagements associated with the use of AWS is in fact rooted in a deeper, underlying problem of how trust can be established in the decision-making process involving complex socio-technical systems.

US DoD addresses potential increased risks associated with a lethal use of AWS, in the form of an additional senior review mechanism. Directive 3000.09, however, does not in principle restrict a development and use of such weapons, contrary to the Campaign's discourse that any use of AWS should be prohibited. I have argued that the US military has established such a high bar to qualify any lethal weapon as 'autonomous' that in fact a

concept of autonomy is used in an indeterminate fashion. This problem construction allows, in turn, to exclude the plethora of already existing weapon systems with advanced autonomous capabilities from the increased regulatory oversight, that is an additional senior review mechanism. I have then explored various types of weapons which have not been qualified as 'lethal autonomous weapon systems' according to Directive 3000.09: Phalanx CIWS, LRASM, and loitering munitions such as Switchblade. I have argued that Directive's 3000.09 'legalisation' of their use is not without controversy in the wider academic literature. Semi-autonomous weapons are based on supervisory systems that are prone to automation bias, may cause military skills degradation, and generate moral dilemmas associated with the practice of killing from a distance. Weapon systems such as LRASM are based on advanced AI and image recognition techniques which are prone to a trade-off between performance and interpretability. Loitering munitions can operate without any human supervision even at the level of targeting and engagement which only exacerbates earlier discussed challenges.

I have explored two alternative policy responses towards the problem of lethal use of AWS, represented by US DoD and the Campaign. US DoD policy does not prohibit LAWS, but Directive 3000.09 introduces the requirement of 'appropriate levels of human judgment' over the use of such weapons. The Campaign, on the contrary, proposes to ban LAWS because they are beyond 'human control'. I have explicated that the concept of 'human judgment' is different from the requirement of 'human control' as the latter relates to so-called 'direct control' in the form of a human manually exercising control by terminating a weapon's engagement. US DoD argues for the broader understanding of 'control' that includes both manual control and control-by-design that is the ex-ante determination of weapon's capabilities.

Thus, US DoD problematisation of the role of human factors over the use of AWS is instrumental, i.e. it is about delineating of what is possible in specific operations. I have argued that there might be even situations in which a machine may perform specific targeting and engagement functions more effectively than a human, hence leaving a door open to exercise no direct, manual human control at all. This is the case, for example, in a situation of an urgent military operation need, where I have demonstrated that US DoD policy allows to supersede a direct control with a control effectuated at the level of design.

In contrast, the Campaign's problem construction is conceptual, i.e. they construct the problem of AWS as one where human control needs to be exercised in order to render a use of such weapon systems acceptable according to specific normative criteria, that is LOAC. Therefore, the discussions about the Campaign's concept of meaningful human control usually focus on content, i.e. specific elements of human control, whilst the discussions about US DoD concept of appropriate human judgment focus on the context, i.e. how human judgment ought to be exercised, and specifically who should exercise human judgment and over what so the weapon system can complete the mission effectively.

The US DoD problematisation of AWS is based on three key assumptions. First, the policy aims to strike a balance between safety considerations associated with the use of autonomy in weapon systems and the need to maintain the asymmetric combat advantage against main adversaries, such as Russia and China. Second, while US DoD recognises the increased risks associated with the use of AWS the department does not consider that the delegation of lethal authority to autonomous weapons is illegal according to LOAC and the US domestic law. In that, US DoD discourse takes a qualitatively different take on the legal problems of discrimination and responsibility relative to the Campaign's narrative. The US administration shifts the legal ramification of the problems into a technical issue by arguing

that existing AWS can be controlled primarily through technical measures, such as testing and evaluation and software validation and verification. Third, Directive 3000.09 concentrates only on the risks associated with autonomy conceived of independence from a human operator, leaving considerations regarding AI-enhanced weapons, and their self-learning capabilities, unaddressed.

US DoD problem construction of AWS as weapons which render increased risks of unintended engagements is primarily informed by the transformations in the way US military conducted their operations relaying on increasingly sophisticated yet not regulated lethal autonomous supervised systems such as Patriot missiles or UAVs. As illustrated by the Patriot's fratricides during the Operation Iraqi Freedom in 2003, the US military had a problem to establish a trust in machine's autonomous capabilities. The findings from these events revealed that there were technological problems with Patriot software itself, but also the control procedures for human-machine interfaces were poorly designed for specific operations.

I have argued that the use of lethal autonomous supervised systems can be traced back to the use of remotely controlled UAVs which introduced novel challenges to the control of military operations. Specifically, the distance between teams on the ground in the theatre of war and remote teams in the US distributed control over the military assets and distributed control between humans and machines on the battlefield. I have argued also that these changes undermine the longstanding USAF doctrine of centralized control, according to which only a commander is responsible for a direction, coordination, and specific use of forces on the battlefield.

I have then further investigated what specific effects the US DoD problematisation of AWS has on US DoD and USAF regimes of practices. I have decided to focus on a

nested case study of USAF to present in depth study of at least one of US DoD military branches. I have narrowed down the analysis to the specific set of effects that the problem representation has on *the* emergence of norms associated with the 'trusted' use of AWS.

According to US DoD, one of the problems of trust is how to ensure a 'right' allocation of functions and the authority of control between humans and machines in the decision-making process involving increasingly autonomous weapon systems. I have illustrated that, according to the US doctrine, the concept of 'control' is broader than merely 'who pulls the trigger'. It refers to the question of 'how' military operations should be conducted. I have argued that the use of AWS occupies a certain place in a wider chain of military control across various stages of targeting process. USAF doctrine has long been guided by the tenet of centralized control according to which military control belongs to a mission commander, but as illustrated by the CAS case study, military practices – at least in the context of dynamic targeting where AWS are often used – did not adhere to the doctrine of centralized control. I have thus argued that a 'discursive effect' of changing USAF doctrine from centralized to distributed control reflected already existing military practices of shared control in the wider targeting process. The normalization of distributed control within the USAF is one of the key effects of the US DoD problematisation of AWS.

I have also explored the problem with trustworthy capabilities of AWS, that is whether humans can trust an autonomous machine that it will produce predictable outcomes. I have argued that US DoD developed both AI principles and more specific process-oriented standards that should guide US DoD acquisition work related to AI-assisted weapon systems, including AI-assisted AWS. While this set of emerging standards is, according to US DoD leaders, 'based on existing and widely accepted' ethical, legal and policy commitments including Directive 3000.09, I have argued that in fact AI principles

and RAI guidelines are not necessarily in conjunction with Directive 3000.09. It is uncertain how a policy of human judgment over the use of AWS from the Directive 3000.09 fits with AI principles and the concept has not been translated by Responsible AI guidelines. Directive 3000.09 in turn is silent on AI-capabilities of AWS whereas one could argue that advanced AI introduce different challenges than the use of autonomy in weapon systems.

The exploration of US DoD approach to AWS reveals that there is a room for contestation of the current US DoD problem representation of AWS, let alone because it passed ten years since the adoption of Directive 3000.09 and some things have changed. As an obvious example, the position of Under-secretary of Defense for Acquisition, Technology, and Logistics – required to conduct a senior review process – no longer exists, and its former responsibilities are now divided between two new under-secretary positions.[896] In terms of potential substantial changes, I have argued that the US DoD governmentality of AWS focuses heavily on the concept of autonomy conceived as an independence from a human operator, but it disregards the evaluation of advanced AI capabilities of AWS. I have explored that the assessment of advanced AI capabilities, i.e. ML differs from traditional software evaluation as the outcomes of ML models are only probabilistic, not deterministic and it is challenging to verify the correctness and accuracy of the ML model outcomes. Thus, I have argued that there should be an additional layer of review specifically for advanced AI-augmented AWS in addition to the review of autonomy.

The lack of consideration of AI capabilities of AWS led me further to reflect on the potential threat of weaponised AGI which has been explicitly denied or deflected by US

[896] Gregory Allen, 'DOD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread' (Centre for Strategic and International Studies 2022).

DoD. I have recommended an alternative problem representation calling for the inclusion of measures to mitigate the potential advent of AGI. I have argued that one of the measurers can be the inclusion of the requirement for a direct human control over the weapon's engagement decision, but this requirement will be difficult to implement in practice as it would render many of the existing weapon systems illegal. Thus, I have provided an argument for a non-waivered requirement that a human being should always retain the ability to stop a weapon's engagement. I have also reflected on the possibility of brain-machine interface to create autonomous soldiers who might be able to control weaponised AGI.

US DoD's disregard of the risks of weaponised AGI stems from their believe that AWS augmented with narrow AI capabilities are controllable in technical terms. I have exposed that the conceptualisation of risk associated with the development and use of AWS in purely technical terms excludes ethical concerns over the use of such weapons. Thus, I have discussed an alternative problem representation whereby ethical considerations serve as certain restrictions over the development and use of such weapons. Similarly, I have restated that the inclusion of a direct human control requirement at the level of engagement can alleviate not only potential threats of weaponised AGI but, to certain extent, accommodate ethical concerns. I have also discussed a proposal to build an ethical machine which might be able to conduct assessment whether a specific lethal action is ethically permissible. I have concluded that a practical feasibility of such a proposal is limited, but it is more likely that a weapon system can be designed to assess at least whether the action is ethically impermissible. I have also postulated the creation of a dedicated oversight agency with the goal of providing more accountability regarding the use of AWS. Such an agency could alleviate some fears that such weapons are used for impermissible ethical action according to the wider public morality.

Finally, I have concentrated on another significant omission from US DoD problem representation of AWS, that is the lack of consideration of autonomous cyber weapons. I have argued that the inclusion of cyber considerations must address the problem of what limitation mechanisms restricting the spread and use of autonomous cyber weapons should be in place and whether AI-augmented AWS can be used both for defensive and offensive purposes. I have concluded that it may be justified to provide certain limitations over the scope of use of AI-augmented autonomous cyber weapons even for defensive purposes, particularly in the context of protecting nuclear weapons. The use of these weapons for offensive purposes should go through additional review in a similar manner as the use of lethal autonomous kinetic weapon systems as it may lead to inadvertent escalation.

2. Limitations of Research Finding

The thesis is based on the current stage of technological development of AWS and is based primarily on publicly available information. Such information may only partially reveal the real development in the domain of AWS. Even though I have tried to present a comprehensive picture of US DoD and USAF decision-making process, it is simplified as it does not consider any specific mission requirements and circumstances. Further, a study of specific targeting practices is limited to a generic type of CAS mission in dynamic targeting, and it does not consider in-depth rules of other branches of US DoD and other types of USAF missions.

BIBLIOGRAPHY

'2022 United States Military Strength' (*Global Firepower 2022*, September 2022) <https://www.globalfirepower.com/country-military-strength-detail.php?country_id=united-states-of-america>

AeroVironment, 'Switchblade 600 Loitering Missile' <https://www.avinc.com/tms/switchblade-600>

Al Emondi, 'Revolutionizing Prosthetics (Archived)' (*DARPA*) <https://www.darpa.mil/program/revolutionizing-prosthetics>

Alex Roland with Philip Shiman, *Strategic Computing DARPA and the Quest for Machine Intelligence, 1983–1993* (MIT Press 2002)

Alexander Kott and others, 'Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture. Release 2.0' (18 September 2019) <https://arxiv.org/abs/1803.10664>

Alexander Lavin and others, 'Technology Readiness Levels for Machine Learning Systems' (2022) 13 Nature Communications

Ana Pereira and Carsten Thomas, 'Challenges of Machine Learning Applied to Safety-Critical Cyber-Physical Systems' (2020) 2 Machine Learning and Knowledge Extraction 579

Andreas Haslbeck and Hans-Juergen Hoermann, 'Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills Among Airline Pilots' (2016) 58 Human Factors 533

Andreas Matthias, 'The Responsibility Gap - Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175

Andrew Eversden, 'Meet Anduril's New Loitering Munitions, the Firm's First (but Not Last) Weapons Program' *Breaking Defense* (6 October 2022)

Anduril Industries, 'Anduril Announces Best In Class Loitering Munition' (6 October 2022) <https://blog.anduril.com/anduril-announces-best-in-class-loitering-munition-8b00a72aba2a>

——, 'Altius' <https://www.anduril.com/hardware/altius/>

Ann Väljataga and Rain Liivoja, 'Cyber Autonomy and International Law: An Introduction', *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021)

Anne Dienelt, 'The Shadowy Existence of the Weapons Review and Its Impact on Disarmament' *S+F Sicherheit und Frieden / Security and Peace*

Armin Krishnan, *Legality and Ethicality of Autonomous Weapons* (Routledge 2009)

Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (Article 36 2013) <http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf.>

——, 'Autonomous Weapons, Meaningful Human Control and the CCW' (Article 36 2014) <http://www.article36.org/weapons-review/autonomous-weapons-meaningful-human-control-and-the-ccw/>

——, 'Killing by Machine: Key Issues for Understanding Meaningful Human Control' (Article 36 2015) <http://www.article36.org/autonomous-weapons/killing-by-machine-key-issues-for-understanding-meaningful-human-control/>

——, 'Key Elements of Meaningful Human Control' (Article 36 2016) <https://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf>
Associated Press, 'Putin: Leader in Artificial Intelligence Will Rule the World' (1 September 2017)

Austin Wyatt, *The Disruptive Impact of Lethal Autonomous Weapons Systems Diffusion* (Routledge 2022)

Ben Emmerson, 'Report of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism' (2014) A/HRC/25/59

Bernard Harcourt, 'An Answer to the Question: 'What Is Poststructuralism?' (2007) 156 University of Chicago Public Law & Legal Theory Working Paper

Bonnie Docherty, 'Losing Humanity. The Case Against Killer Robots' (Human Rights Watch 2012) 978-1-6231–32408

——, 'Making the Case. The Danger of Killer Robots and the Need for a Preemptive Ban' (2016)

Brandi Vincent, '2022 in Review: What the Pentagon's CDAO Accomplished in Its Inaugural Year' (30 December 2020) <https://defensescoop.com/2022/12/30/2022-in-review-what-the-pentagons-cdao-accomplished-in-its-inaugural-year/>

——, 'Pentagon Reaches Important Waypoint in Long Journey toward Adopting "Responsible AI"' *FedScoop* (29 June 2022)

Brent Reininger and others, 'Assessing the Obama Administration's Pivot to Asia' (2016)
Bryan Derballa, 'How Rogue Techies Armed the Predator, Almost Stopped 9/11, and Accidentally Invented Remote War' *Wired* (17 December 2015)

C. Todd Lopez, 'Defense Official Discusses Unmanned Aircraft Systems, Human Decision-Making, AI' (*Department of Defense News*, 3 February 2021) <https://www.defense.gov/News/News-Stories/Article/Article/2491512/defense-official-

discusses-unmanned-aircraft-systems-human-decision-making-ai/>

Campaign to Stop Killer Robots, 'Key Elements of a Treaty on Fully Autonomous Weapons' (2019)

——, 'Country Views on Killer Robots' (Campaign to Stop Killer Robots 2020)
Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (1st edition, Pearson 2009)

——, 'The Issue of Intentionality in Frame Theory: The Need for Reflexive Framing', *The Discursive Politics of Gender Equality: Stretching, bending and policymaking* (Routledge 2009)

——, 'Why Study Problematizations? Making Politics Visible' (2012) 2 Open Journal of Political Science 1

——, 'Problematizations in Health Policy: Questioning How "Problems" Are Constituted in Policies' (2016) 6 SAGE Open

——, 'Meanings of Problematization' (18 February 2018) <https://carolbacchi.com/2018/02/18/meanings-of-problematization/>
Carol Bacchi and Susan Goodwin, *Poststructural Policy Analysis – A Guide to Practice* (Palgrave Macmillan 2016)

Carter Provides Remarks at Defense Innovation Board Meeting (2016) <https://www.dvidshub.net/video/486245/carter-provides-remarks-defense-innovation-board-meeting>

CDAO, 'AI Ethical Principles – Highlighting the Progress and Future of Responsible AI in the DoD' (26 February 2021) <https://www.ai.mil/blog_02_26_21-ai_ethics_principles-highlighting_the_progress_and_future_of_responsible_ai.html>
Center for a New American Security, 'People' <https://www.cnas.org/people?group=full-time-staff>

Center for the Study of the Drone, 'Loitering Munitions' (2017)

Charles Glaser, 'The Security Dilemma Revisited' (1997) 50 World Politics

Charles Trumbull IV, 'Autonomous Weapons: How Existing Law Can Regulate Future Weapons' (2020) 34 Emory International Law Review 533

Charlie Gao, 'Loitering Munitions (A.K.A. Suicide Drones) Are Getting Deadlier Every Year' (26 November 2020) <https://nationalinterest.org/blog/reboot/loitering-munitions-aka-suicide-drones-are-getting-deadlier-every-year-173454>

Cheryl Pellerin, 'Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence' (*DoD News*, 31 October 2016) <https://www.defense.gov/News/Article/Article/ 991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence>

Christof Heyns, ‘Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions’ (2013) GE.13-12776

——, ‘Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: An African Perspective’, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016)

CIA, ‘Remotely Piloted Vehicles in the Third World: A New Military Capability’ (1986)
Claude Lévi-Strauss, *Structural Anthropology* (Allen Lane 1968)

Colin Clark, ‘Drones Do Excellent Urban Close Air Support; Mideast F-35A Deployment In Several Years’ *Breaking Defense* (24 February 2017)

——, ‘Reaper Drones: The New Close Air Support Weapon’ *Breaking Defense* (10 May 2017)
Colin Gordon, ‘Afterword’, *Power/Knowledge: Selected Interviews & Other Writings 1972-1977 by Michel Foucault* (Pantheon Books 1980)

Colin Koopman and Tomas Matza, ‘Putting Foucault to Work: Analytic and Concept in Foucaultian Inquiry’ 39 Critical Inquiry 817

Commander Gilmary Hostage III, USAF and Lt Col Larry Broadwell, Jr, USAF, ‘Resilient Command and Control The Need for Distributed Control’ (2014) 74 Joint Force Quarterly

Congressional Research Service, ‘Lethal Autonomous Weapon Systems: Issues for Congress’ (2016)

——, ‘General Policy Statements: Legal Overview’ (2016)

——, ‘Artificial Intelligence and National Security’ (2019)

——, ‘Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems’ (2020)

——, ‘Renewed Great Power Competition: Implications for Defense—Issues for Congress’ (2020)

——, ‘Joint All-Domain Command and Control (JADC2)’ (2022)

Cris Shore and Susan Wright, ‘Technologies of Governance and the Politics of Visibility’, *Policy worlds: Anthropology and the analysis of contemporary power* (Berghahn Books 2011)

Dan Saxon, ‘A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of “Appropriate Leveles of Human Judgment over the Use of Force”’, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016)

Dan Steinbock, ‘The Challenges for America’s Defense Innovation’ (2014)

Daniel Dennett, ‘When HAL Kills, Who’s to Blame? Computer Ethics’, *Hal’s Legacy:*

2001's Computer as Dream and Reality (MIT Press 1997)

Daniel Faggella, 'When Will We Reach the Singularity? – A Timeline Consensus from AI Researchers' (*EMERJ*, 18 March 2019) <https://emerj.com/ai-future-outlook/when-will-we-reach-the-singularity-a-timeline-consensus-from-ai-researchers/>

Daniel Kahneman, *Thinking, Fast and Slow* (Penguin 2011)

Daniel Trusilo and Thomas Burri, 'Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems', *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021)

DARPA, 'Collaborative Operations in Denied Environment (CODE) (Archived)' (28 November 2022) <https://www.darpa.mil/program/collaborative-operations-in-denied-environment>

David Akerson, 'The Illegality of Offensive Lethal Autonomy', *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff 2013)
David Gunning and others, 'DARPA's Explainable AI (XAI) Program: A Retrospective' (2021) 2 Applied AI Letters 1

David Hambling, 'U.S. To Equip MQ-9 Reaper Drones With Artificial Intelligence' [2020] Forbes

David Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy* (Anchor; 1st edition 2010)

David Mindell, *Our Robots, Ourselves* (Penguin 2015)

David Schiff, 'Socio-Legal Theory: Social Structure and Law' (1976) 39 The Modern Law Review

Deepmind, 'A Generalist Agent' <https://www.deepmind.com/publications/a-generalist-agent>

Defense Innovation Board, 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense' (2019)

Defense Innovation Unit, 'Annual Report 2020' (Defense Innovation Unit 2020)
Defense Science Board, 'Patriot System Performance' (2005) D.C. 20301-3140

——, 'Report of the Task Force on the Role of Autonomy in DoD Systems' (2012)

——, 'The Role of Autonomy in DoD Systems' (2012)

——, 'Summer Study on Autonomy' (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics 2016) 20301–3140

Defense Science Board Task Force, 'Fulfillment of Urgent Operational Needs' (2009) 20301–3140

Department of the Army, ‘ADP 1-01 Doctrine Primer’ (2019)

Department of the Navy, ‘Secretary of the Navy Instruction 5000.2E, Department of the Navy Implementation and Operation of the Defense Acquisition System and the Joint Capabilities Integration and Development System’

Deutsche Welle, ‘Regulate Killer Robots, Says Heiko Maas’ *Deutsche Welle* (15 March 2019)

DoD, ‘Law of War Manual’ (Office of General Counsel DoD 2015)

Don Harris and Wen-Chin Li, *Decision Making in Aviation* (Routledge 2017)

Douglas Birkey, Lt Gen David Deptula, USAF (Ret.) and Maj Gen Lawrence Stutzriem, USAF (Ret.), ‘Manned-Unmanned Aircraft Teaming: Taking Combat Airpower to the Next Level’ (2018) 15 Mitchell Institute Policy Papers

Dusica Marijan and Arnaud Gotlieb, ‘Software Testing for Machine Learning’ (2020) 34 Proceedings of the AAAI Conference on Artificial Intelligence 13576

Dustin Lewis, Gabriella Blum, and Naz Modirzadeh, ‘War-Algorithm Accountability’ (2016)

E. W. Jolie, ‘A Brief History of U.S. Navy Torpedo Development’ (1978) NUSC Technical Document 5436

Edmond Awad and others, ‘The Moral Machine Experiment’ (2018) 563 Nature 59

Eileen Vidrine, ‘Air Force CDO: Flying High With AI’ *The Wall Street Journal* (20 August 2021) <https://deloitte.wsj.com/articles/air-force-cdo-flying-high-with-ai-01629481727>

Elliot Ackerman, ‘A Navy SEAL, a Quadcopter, and a Quest to Save Lives in Combat’ *Wired* (30 October 2020)

Emily Jones, ‘A Posthuman-Xenofeminist Analysis of the Discourse on Autonomous Weapons Systems’ (2018) 44 Australian Feminist Law Journal 93

Eric Schmidt, Robert Work, and others, ‘Final Report of the National Security Commission on Artificial Intelligence’ (2021)

Eric Tegler, ‘An AI Co-Pilot Called “ARTUμ” Just Took Command of A U-2’s Sensor Systems On A Recon Mission’ *Forbes* (16 December 2020) <https://www.forbes.com/sites/erictegler/2020/12/16/an-ai-co-pilot-called-artujust-took-command-of-a-u-2s-sensor-systems-on-a-reconnaissance-mission/?sh=568d38e461f0>

Evan Ackerman, ‘Lethal Microdrones, Dystopian Futures, and the Autonomous Weapons Debate’ *IEEE Spectrum* (2017)

Filippo Santoni de Sio and Jeroen van den Hoven, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account' [2018] Frontiers in Robotics and AI

François Delerue, *Cyber Operations and International Law* (CUP 2020)

Fred Martin Jr., 'Technologies of Sovereign Power? Private Military Corporations, Drones, and Lethal Autonomous Robots - A Critical Security Studies Perspective' (Ohio University 2015) <http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1428937559>

Freedberg Jr. S, 'MOSA: The Invisible, Digital Backbone Of FVL' *Breaking Defense* (13 March 2020) <https://breakingdefense.com/2020/03/mosa-fvls-invisible-digital-backbone/>

——, 'Kill Chain In The Sky With Data: Army's Project Convergence' *Breaking Defense* (14 September 2020) <https://breakingdefense.com/2020/09/kill-chain-in-the-sky-with-data-armys-project-convergence/>

——, '"Improvised Mode": The Army Network Evolves In Project Convergence' *Breaking Defense* (21 September 2020) <https://breakingdefense.com/2020/09/improvised-mode-the-army-network-evolves-in-project-convergence/>

Future of Life Institute, 'Open Letter on Autonomous Weapon' <https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/>
——, 'AI FAQ' <https://futureoflife.org/ai/ai-faq/>

Garrett Reim, 'Anduril Introduces Loitering Munition Warheads For Altius Drones' *Aviation Week Network* (7 October 2022) <https://aviationweek.com/shows-events/ausa/anduril-introduces-loitering-munition-warheads-altius-drones>

Gen Charles Brown Jr., USAF, 'Accelerated Change or Lose' (USAF 2020)

Gen Charles Krulak, USMC, 'The Strategic Corporal: Leadership in the Three Block War' [1999] Marines Magazine

Gen Mark Welsh III, USAF, 'The World's Greatest Air Force—Powered by Airmen, Fueled by Innovation. A Vision for the United States Air Force' (2013)

——, 'Global Vigilance, Global Reach, Global Power for America' (USAF 2013)

——, 'America's Air Force: A Call to the Future' (USAF 2014)

——, 'Air Force Future Operating Concept' (USAF 2015)

Gen Mark Welsh III, USAF and Deborah Lee James, USAF, 'USAF Strategic Master Plan' (USAF 2015)

General Atomics Aeronautical, 'MQ-9A "Reaper"' <https://www.ga-asi.com/remotely-piloted-aircraft/mq-9a>

——, ‘Predator XP’ <https://www.ga-asi.com/remotely-piloted-aircraft/predator-xp>

George E. Katsos, ‘U.S. Joint Doctrine Development and Influence on NATO’ 101 Joint Force Quarterly

Government of Austria, Brazil, and Chile, ‘Proposal for a Mandate to Negotiate a Legally- Binding Instrument That Addresses the Legal, Humanitarian and Ethical Concerns Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (2018) CCW/GGE.2/2018/WP.7

Government of the Russian Federation, ‘National Strategy for Artificial Intelligence Development’ (2019)

Graham Webster and others, ‘Full Translation: China’s “New Generation Artificial Intelligence Development Plan”’ (Stanford Cyber Policy Center 2017)

Greg Zacharias, USAF, ‘Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future, Volume I: Human-Autonomy Teaming’ (USAF 2015) AF/ST TR 15-01

——, ‘Autonomous Horizons The Way Forward’ (Air University 2019)

Gregory Allen, ‘DOD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread’ (Centre for Strategic and International Studies 2022)

Harry Surden, ‘Artificial Intelligence and Law: An Overview’ (2019) 35 Georgia State University Law Review 1305

Heather Roff, ‘Meaningful or Meaningless Control’ (*The Duck of Minerva*, 25 November 2014)

——, ‘Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons’ (2016)

——, ‘Survey of Autonomous Weapon Systems’ <ttps://globalsecurity.asu.edu/robotics-autonomy>

Herbert Gottweis, ‘Theoretical Strategies of Poststructuralist Policy Analysis: Towards an Analytics of Government’, *Deliberative Policy Analysis: Understanding Governance in the Network Society* (Cambridge University Press 2003)

Herbert Hart, *The Concept of Law* (2nd edition, Clarendon Press 1994)

Herman Wolk, ‘The Struggle for Air Force Independence, 1943-47’ (1984)

HRW, ‘Shaking the Foundations. The Human Rights Implications of Killer Robots’ (2014)

——, ‘Mind the Gap: The Lack of Accountability for Killer Robots’ (2015)

——, ‘Stopping Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control’ (2020)

HRW and others, ‘Killer Robots and the Concept of Meaningful Human Control’ (2016)

IAI, ‘Harpy Autonomous Weapon for All Weather’ <https://www.iai.co.il/p/harpy>

ICRAC, ‘Open Letter in Support of Google Employees and Tech Workers’

ICRC, ‘What Is International Humanitarian Law?’ (2004)

——, ‘A Guide to the Legal Review of New Weapons, Means and Methods of Warfare’ (2006)

——, ‘Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects’ (2014)

——, ‘Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons’ (ICRC 2016)

——, ‘What You Need to Know about Autonomous Weapons’ (26 July 2022) <https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons#:~:text=Mines%20can%20be%20considered%20rudimentary,anti%2Dpersonnel%20mines%20in%201997.>

Indian Ambassador Amandeep Singh Gill, ‘Chart 2 Consideration of the Human Element in the Use of Lethal Force; Aspects of Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (UN GGE 2018)

Ingvild Bode and Hendrik Huelss, *Autonomous Weapons Systems and International Law* (McGill-Queen’s University Press 2022)

Interview with A CEO of a US DoD vendor supplying a tool for validating AI models, ‘Mikolaj Firlej Interview’ (5 February 2021)

Interview with fomer Senior US DoD official, ‘Mikolaj Firlej Interview’ (7 February 2021)

Interview with fomer USAF member, ‘Mikolaj Firlej Interview’ (2 February 2021)

Interview with Paul Scharre, ‘Mikolaj Firlej Interview’ (5 February 2021)

Interview with Senior Force Developer for Emerging Technologies at the Office of the Under Secretary of Defense for Policy, ‘Mikolaj Firlej Interview’ (5 February 2021)
Interview with Shawn Steene, ‘Mikolaj Firlej Interview’ (12 January 2021)

IPSOS, ‘Six in Ten (61%) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems’ (22 January 2019) <https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons>

Jackson Barnett, 'Department of Defense AI Ethics Principles Still Lack Implementation Guidance' *FedScoop* (28 June 2021) <https://www.fedscoop.com/ai-ethics-principles-dod-military-implementation-guidance/>

——, 'JAIC Chief Wants AI Progress to Be "Slow and Incremental"' (*FedScoop*, 8 October 2021) <https://www.fedscoop.com/jaic-chief-wants-ai-progress-to-be-slow-and-incremental/>

Jai Galliott, Duncan MacIntosh, and Jens Ohlin, *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021)

Jakob Kellenberger, 'The Relevance of International Humanitarian Law in Contemporary Armed Conflicts' (Committee of legal advisers on public international law, 28th meeting Lausanne, 13-14 September 2004, 14 September 2004)

James Johnson, '"Catalytic Nuclear War" in the Age of Artificial Intelligence & Autonomy: Emerging Military Technology and Escalation Risk between Nuclear-Armed States' [2021] Journal of Strategic Studies

——, 'Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command-and-Control Decision-Making in the Digital Age' [2022] Defense Studies

James Walsh, 'Political Accountability and Autonomous Weapons' (2015) 2 Research and Politics

Jared Dunnmon and others, 'Responsible AI Guidelines in Practice' (DIU 2021)

Jason DeSon, 'Automating the Right Stuff? The Hidden Ramifications of Ensuring Autonomous Aerial Weapon Systems Comply with International Humanitarian Law' (2015) 72 Air Force Law Review 85

Jason Scholz and Jai Galliott, 'The Humanitarian Imperative for Minimally-Just AI in Weapons', *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021)

Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law* (Cambridge University Press 2012)

Jean-Paul Metzger, *Discourse: A Concept for Information and Communication Sciences* (Wiley 2019)

Jeffrey S. Thurnher, 'The Law That Applies to Autonomous Weapon Systems' (2013) 17 ASIL Insights

Jennifer Roberts, 'Plan X (Archived)' (*DARPA*) <https://www.darpa.mil/program/plan-x>

Jennifer Spradlin, 'Academy Announces New Data Science Major' (8 June 2020) <https://www.af.mil/News/Article-Display/Article/2211059/academy-announces-new-

data-science-major/>

Jeremy Scahill, ‘The Assassination Complex’ (*The Intercept*, 2015)

Jessie Chen, Michael Barnes, and Michelle Harper-Sciarini, ‘Supervisory Control of Unmanned Vehicles’ (Army Research Laboratory 2010) ARL-TR-5136

Jie M. Zhang and others, ‘Machine Learning Testing: Survey, Landscapes and Horizons’ (2022) 48 IEEE Transactions on Software Engineering 1

Joe Ritter, ‘MQ-9S Over Sirte: Unmanned Airpower for Urban Combat’ (21 March 2022) <https://mwi.usma.edu/mq-9s-over-sirte-unmanned-airpower-for-urban-combat/>

John K. Hawley, ‘Patriot Wars: Automation and the Patriot Air and Missile Defense System’ (CNAS 2017)

John K. Hawley and Anna L. Mares, ‘Human Performance Challenges for the Future Force: Lessons from Patriot after the Second Gulf War’, *Designing Soldier Systems Current Issues in Human Factors* (1st Edition, Ashgate 2012)

John Markoff, ‘Fearing Bombs That Can Pick Whom to Kill’ *The New York Times* (11 November 2014) <https://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html>

John Mearsheimer, *Conventional Deterrence* (University of Chicago Press 1987)

——, *Liddell Hart and the Weight of History* (Cornell University Press 1988)

John Simon, ‘A Conversation with Michael Foucault’ (1971) 38 Partisan Review

Jon Hoper, ‘Pentagon Grappling With AI’s Ethical Challenges’ (10 November 2020) <https://www.nationaldefensemagazine.org/articles/2020/11/10/pentagon-grappling-with-ais-ethical-challenges>

Jonathan Sale, ‘The Secret History of Drones’ *The Guardian* (10 February 2013) <https://www.theguardian.com/world/shortcuts/2013/feb/10/secret-history-of-drones-1916>

Jürgen Altmann and Frank Sauer, ‘Autonomous Weapon Systems and Strategic Stability’ (2017) 59 Survival

Katherine Chandler, ‘Drone Flight and Failure: The United States’ Secret Trials, Experiments and Operations in Unmanning, 1936-1973’ (The University of California 2014) <https://escholarship.org/content/qt0fg216f7/qt0fg216f7_noSplash_8dd4be278e0eefd9a44b97ef83782b98.pdf>

——, *Unmanning: How Humans, Machines and Media Perform Drone Warfare* (Rutgers University Press 2020)

Kathleen Hicks, 'Memorandum "Implementing Responsible Artificial Intelligence in the Department of Defense"'

Katja Grace and others, 'When Will AI Exceed Human Performance? Evidence from AI Experts' <https://arxiv.org/pdf/1705.08807.pdf>

Katrina Manson, 'US Has Already Lost AI Fight to China, Says Ex-Pentagon Software Chief' *Financial Times* (10 October 2021) <https://www.ft.com/content/f939db9a-40af-4bd1-b67d-10492535f8e0>

Ken Dilanian, Dan De Luce, and Courtney Kube, 'Biden Admin Will Provide Ukraine with Killer Drones Called Switchblades' (*NBC News*, 16 March 2022) <https://www.nbcnews.com/politics/national-security/ukraine-asks-biden-admin-armed-drones-jamming-gear-surface-air-missile-rcna20197>

Kenneth Anderson and Matthew Waxman, 'Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can' (2013) 11 American University Washington College of Law Research Paper

Kenrick Cai, 'Shield AI Rejected A Pivot To "Selfie Drones"—Now Its Drones Are Being Used By The Military Overseas' (16 July 2020) <https://www.forbes.com/sites/kenrickcai/2020/07/16/shield-ai-ryan-brandon-tseng-ai-50-interview/?sh=24d785e439db>

Kim Zetter, 'An Unprecedented Look at Stuxnet, the World's First Digital Weapon' *Wired* (3 November 2014) <https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/>

Larry Lewis, 'Killer Robots Reconsidered: Could AI Weapons Actually Cut Collateral Damage?' [2020] Bulletin of Atomic Scientists

Lester Haines, 'Patriot Missile: Friend or Foe?' *The Register* (20 May 2004) <https://www.theregister.com/2004/05/20/patriot_missile/>

Leveringhaus A, *Ethics and Autonomous Weapons* (Palgrave Macmillan 2016)

Lisa Hajjar, 'Lawfare and Armed Conflicts: A Comparative Analysis of Israeli and U.S. Targeted Killing Policies and Legal Challenges against Them', *Life in the Age of Drone Warfare* (Duke University Press 2017)

Liz O'Sullivan, 'Side Event "The Urgent Need for a Treaty to Retain Meaningful Human Control over the Use of Force"' (2019)

Lockheed Martin, 'Long Range Anti-Ship Missile (LRASM)' <https://www.lockheedmartin.com/en-us/products/long-range-anti-ship-missile.html>
Long Range Anti-Ship Missile (LRASM) (Directed by Lockheed Martin, 2016) <https://www.youtube.com/watch?v=h449oIjg2kY>

Louis Del Monte, *Genius Weapons: Artificial Intelligence, Autonomous Weaponary, and the Future of Warfare* (Prometheus Books 2018)

‘Low Cost Autonomous Attack System (LOCAAS) Miniature Munition Capability’ <https://man.fas.org/dod-101/sys/smart/locaas.htm>

Lt Col Alan Docauer, ‘Peeling the Onion Why Centralized Control / Decentralized Execution Works’ [2014] Air & Space Power Journal

Lt Col Clint Hinote, ‘Centralized Control and Decentralized Execution: A Catchphrase in Crisis?’ (Air University, Air Force Research Institute 2009)

Lt Col Daniel Lindley, USAF, ‘Assessing China’s Motives: How the Belt and Road Initiative Threatens US Interests’ [2022] Journal of Indo-Pacific Affairs

Lt Col Michael Kometer, USAF, *Command in the Air - Centralized versus Decentralized Control of Combat Airpower* (Air University Press 2007)

Lt Col Timothy Cullen, ‘The MQ-9 Reaper Remotely Piloted Aircraft: Humans and Machines in Action’ (Massachusetts Institute of Technology 2011) <https://dspace.mit.edu/bitstream/handle/1721.1/80249/836824271-MIT.pdf?sequence=2>

Lt Commander Matthew Quintero, USAF, ‘Master and Commander in Joint Air Operations’ (2019) 92 Joint Force Quarterly

Lucas Kello, *The Virtual Weapon and International Order* (Yale University Press 2017)

Lucy Suchman and Jutta Weber, ‘Human-Machine Autonomies’, *Autonomous Weapons Systems* (Cambridge University Press 2016)

Madeleine Clare Elish, ‘Remote Split: A History of US Drone Operations and the Distributed Labor of War’ (2017) Vol. 42(6) Science, Technology, & Human Values 1100

——, ‘24/7: Drone Operations and the Distributed Work of War’ (Columbia University 2018) <file:///Users/mikolaj/Downloads/Elish_columbia_0054D_14452%20(2).pdf>

Maj John Merriam, US Army, ‘Affirmative Target Identification – Operationalising the Principle of Distinction for U.S. Warfighters’ (2016) 56 Virginia Journal of International Law

Maj Kamal Kaaoush, USAF, ‘The Best Aircraft for Close Air Support in the Twenty-First Century’ (2016)

Maj. Michael Lewis, *Lt Gen Ned Almond, USA A Ground Commander’s Conflicting View with Airmen over CAS Doctrine and Employment* (Air University Press 1997)

Maj Nicholas Hall, USAF, ‘Preparing for Contested War: Improving Command and Control of Dynamic Targeting’ (Air University 2017)

Maj Ridge Flick, USAF, ‘Winning The Counterland Battle By Enabling Sensor-to-

Shooter Automation' (1 November 2021) <https://www.alsa.mil/News/Article/2822476/winning-the-counterland-battle-by-enabling-sensor-to-shooter-automation/>

Marc Schanz, 'Spy Eyes in the Sky' [2013] AIR FORCE Magazine

Marcello Guarini and Paul Bello, 'Robotic. Warfare: Some Challenges in Moving from. Noncivilian to Civilian Theaters', *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2012)

Mark Cousins and Athar Hussain, *Michel Foucault* (Macmillan 1984)

Mark Imperial, 'Implementation Structures: The Use of Top-Down and Bottom-Up Approaches to Policy Implementation', *The Oxford Encyclopedia of Public Administration : 2-Volume Set* (Oxford University Press 2022)

Mark Kelly, 'What's In a Norm? Foucault's Conceptualisation and Genealogy of the Norm' (2019) 27 Foucault Studies 1

Mark Roorda, 'NATO's Targeting Process: Ensuring Human Control Over and Lawful Use of "Autonomous" Weapons' (2015) 13 Amsterdam Law School Research Paper

Marra W and McNeil S, 'Understanding "the Loop" Regulating the Next Generation of War Machines' (2013) 36 Harvard Journal of Law and Public Policy

Marta Kosmyna on behalf of the Campaign to Stop Killer Robots, 'Statement to the UN General Assembly First Committee on Disarmament and International Security' (2019)

Mary Cummings, 'Automation and Accountability in Decision Support System Interface Design' [2006] Journal of Technology Studies

——, 'The Human Role in Autonomous Weapon Design and Deployment' [2014] Duke University

——, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (Routledge 2015)

——, 'The Human Role in Autonomous Weapon Design and Deployment', *Lethal Autonomous Weapons* (Oxford University Press 2021)

Mary Manjikian, 'Becoming Unmanned: The Gendering Of Lethal Autonomous Warfare Technology' 16 International Feminist Journal of Politics 48

Matt O'Brien, 'Pentagon Adopts New Ethical Principles for Using AI in War' *AP News* (24 February 2020) <https://apnews.com/article/technology-us-news-business-artificial-intelligence-73df704904522f5a66a92bc5c4df8846>

Matthias Leese, 'Configuring Warfare. Automation, Control, Agency', *Technology and Agency in International Relations* (Routledge 2019)

Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Knopf 2017)

Maziar Homayounnejad, 'Autonomous Weapon Systems, Drone Swarming and the Explosive Remnants of War' [2018] TLI Think!

Megan Lamberth, 'Putting Principles into Practice: How the U.S. Defense Department Is Approaching AI' (2022)

Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control' (*ICRC Blog*, 15 August 2018)

——, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (PhD Thesis, Vrije Universiteit Amseterdam 2019) <https://research.vu.nl/en/publications/the-distributed-conduct-of-war-reframing-debates-on-autonomous-we>

——, 'Responsible AI Symposium - Translating AI Ethical Principles into Practice: The U.S. DoD Approach to Responsible AI' (23 November 2022) <https://lieber.westpoint.edu/translating-ai-ethical-principles-into-practice-us-dod-approach/>

Michael Gross, 'The Pentagon's Plans to Program Soldiers' Brains' (15 November 2018) <https://www.theatlantic.com/magazine/archive/2018/11/the-pentagon-wants-to-weaponize-the-brain-what-could-go-wrong/570841/>

Michael Kreuzer, *Drones and the Future of Air Warfare* (Routledge 2017)

Michael Milstein, 'Pilot Not Included' *Smithsonian Magazine* (July 2011)

Michael N. Schmitt and Jeffrey S. Thurnher, '"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict.' (2013) 4 Harvard National Security Journal 231

Michael Robillard, 'No Such Things as Killer Robots' (2017) 35 Journal of Applied Philosophy 211

Michael Schmitt, 'The Principle of Discrimination in 21st Century Warfare' (1999) 2 Yale Human Rights and Development Law

——, 'Precision Attack and International Humanitarian Law' (2005) 87 IRRC

——, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' [2013] Harvard National Security Journal

Michel Foucault, *The Archaeology of Knowledge* (Pantheon Books 1972)

——, *Language, Counter-Memory, Practice: Selected Essays and Interviews* (Cornell University Press 1977)

——, 'Nietzsche, Genealogy, History', *Language, Counter-Memory, Practice: Selected*

Essays and Interviews (Cornell University Press 1980)

——, 'Subject and Power' (1982) 8 Critical Inquiry 777

——, *The Use of Pleasure. The History of Sexuality (Vol 2)* (Viking 1986)

——, *Madness and Civilization: A History of Insanity in the Age of Reason* (Vintage Books 1988)

——, 'The Concern for Truth', *Michel Foucault: Politics, philosophy, culture. Interviews and other writings, 1977-1984* (Routledge 1988)

——, *The History of Sexuality. Volume I: An Introduction* (Vintage Books 1990)

——, 'Governmentality', *The Foucault effect: Studies in Governmentality* (University of Chicago Press 1991)

——, 'Questions of Method', *The Foucault effect: Studies in Governmentality* (University of Chicago Press 1991)

——, *Discipline & Punish: The Birth of the Prison* (Vintage Books 1995)

——, *Abnormal* (Verso 2003)

——, *Security, Territory, Population* (Palgrave Macmillan 2007)

Mike Loukides, 'Closer to AGI? And Is Artificial General Intelligence What We Really Need?' (*O'Reilly*, 7 June 2022) <https://www.oreilly.com/radar/closer-to-agi/>

Mike White and Gillian Bussey, 'Assistant Director Mike E. White and Director Gillian Bussey Remarks to The Technology and Training Corporation on Hypersonics and Autonomous Systems' (4 November 2020)

<https://www.defense.gov/News/Transcripts/Transcript/Article/2412014/assistant-director-mike-e-white-and-director-gillian-bussey-remarks-to-the-tech/>

Mikolaj Firlej, 'Interview with a Former Senior DoD Official'

Miles Brundage and others, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation' (2018)

Miranda Priebe at all, 'Distributed Operations in a Contested Environment' (RAND 2019) 978-1-9774-0232–5

Mitchell Dean, *Governmentality: Power and Rule in Modern Society* (2nd edition, SAGE Publications 2010)

Monica Hakimi and Jacob Cogan, 'The Two Codes on the Use of Force' (2016) 27 European Journal of International Law 257

Nathan Leys, ‘Autonomous Weapon Systems and International Crises’ (2018) 12 Strategic Studies Quarterly

National Audit Office, ‘Investigation: WannaCry Cyber Attack and the NHS’ (2017)

National Science & Technology Council, ‘The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update’ (2019)

Nehal Bhuta and others, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016)

Nehal Bhuta and Stavros-Evdokimos Pantazopoulos, ‘Autonomy and Uncertainty: Increasingly Autonomous Weapons Systems and the International Regulation of Risk’, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016)

NGAUS, ‘Air Force Reveals New Airpower Doctrine’ (27 April 2021) <https://www.ngaus.org/about-ngaus/newsroom/air-force-reveals-new-airpower-doctrine>

Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014)

Nick Statt, ‘Google Reportedly Leaving Project Maven Military AI Program after 2019’ *The Verge* (1 June 2018)

Nicolas Chaillan, ‘Let’s Catch-up with China within 6 Months’ (*LinkedIn*, 24 November 2021) <https://www.linkedin.com/pulse/lets-catch-up-china-within-6-months-nicolas-m-chaillan/>

Nicolas Falliere, Liam Murchu and Eric Chien, ‘W32.Stuxnet Dossier’ (Symantec 2010)

Nicole Beltran, ‘Artificial Intelligence in Lethal Automated Weapon Systems - What’s the Problem?: Analysing the Framing of LAWS in the EU Ethics Guidelines for Trustworthy AI, the European Parliament Resolution on Autonomous Weapon Systems and the CCW GGE Guiding Principles’ (Uppsala Universitet 2020) <http://uu.diva-portal.org/smash/get/diva2:1436188/FULLTEXT01.pdf>

Nikolas Rose, Pat O’Malley and Mariana Valverde, ‘Governmentality’ (2006) 2 Annual Review of Law and Social Science 83

Noel Sharkey, ‘Grounds for Discrimination: Autonomous Robot Weapons’ (2008) 11 RUSI Defence Systems

——, ‘Weapons of Indiscriminate Lethality’ [2009] FIfF-Kommunikation

——, ‘Automating Warfare: Lessons Learned from the Drones’ (2011) 21 Journal of Law, Information and Science

——, ‘Automating Warfare: Lessons Learned from the Drones’ (2012) 21 Journal of Law, Information and Science 140

——, ‘Killing Made Easy’, *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2012)

——, ‘The Evitability of Autonomous Robot Warfare’ (2012) 94 International Review of the Red Cross 787

——, ‘Staying in the Loop: Human Supervisory Control of Weapons’, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016)

Nolan L, ‘AI Enabled Kill Webs and the Slippery Slope towards Autonomous Weapons Systems’ *Campaign to Stop Killer Robots* (12 October 2020) <https://stopkillerrobots.medium.com/ai-enabled-kill-webs-and-the-slippery-slope-towards-autonomous-weapons-systems-68440dbf8423>

Norm Haller, *Human-Automation Interaction Considerations for Unmanned Aerial System Integration into the National Airspace System* (The National Academies Press 2018)

Norman Fairclough, ‘Intertextuality in Critical Discourse Analysis’ (1992) 4 Linguistics and Education 269

NSTC, ‘U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools’ (2019)

Office of the Chairman of the Joint Chiefs of Staff, ‘DOD Dictionary of Military and Associated Terms’ (2021)

Paolo Carozza, ‘Human Dignity’, *The Oxford Handbook of International Human Rights Law* (Oxford University Press 2015)

Paul Faulkner and Thomas Simpson, *The Philosophy of Trust* (Oxford University Press 2017)

Paul Sabatier, ‘Top-down and Bottom-up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis’ (1986) 6 Journal of Public Policy 21

Paul Scharre, *Army of None* (W W Norton & Company 2018)

——, ‘Autonomous Weapons and Operational Risk’ (CNAS 2016)

——, ‘Debunking the AI Arms Race Theory’ (2021) 4 Texas National Security Review

Paul Scharre and Michael Horowitz, ‘An Introduction to Autonomy in Weapon Systems’ (CNAS 2015)

——, ‘Meaningful Human Control in Weapon Systems: A Primer’ (CNAS 2015)

Peter Asaro, ‘On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision Making’ (2012) 94 International Review of

the Red Cross 687

——, ‘Ban Killer Robots before They Become Weapons of Mass Destruction’ *Scientific American* (7 August 2015)

Peter Miller and Nikolas Rose, ‘Governing Economic Life’ (1990) 19 Economy and Society

Peter Singer, *Wired for War* (Penguin 2009)

P.J. Mitchell, Mary Cummings, and Thomas Sheridan, ‘Human Supervisory Control Issues in Network Centric Warfare’ (Massachusetts Institute of Technology 2004) HAL2004-01

Potember R, ‘Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD’ (JASON (the Mitre Corporation) 2017) JSR-16-Task-003

Raytheon, ‘Phalanx Weapon System’ <https://www.raytheonmissilesanddefense.com/what-we-do/naval-warfare/ship-self-defense-weapons/phalanx-close-in-weapon-system>

Reaching Critical Will, ‘CCW Group of Governmental Experts on Lethal Autonomous Weapon Systems’ <https://reachingcriticalwill.org/disarmament-fora/ccw>

Rebecca Crootof, ‘A Meaningful Floor for “Meaningful Human Control"’ (2016) 30 Temple International and Comparative Law Journal 53

Rich Smith, ‘AeroVironment Will Upgrade the Switchblade’ (*The Motley Fool*, 11 May 2016) <https://www.fool.com/investing/general/2016/05/11/aerovironment-will-upgrade-the-switchblade.aspx>

Richard Matland, ‘Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation’ (1995) 5 Journal of Public Administration Research and Theory 145

Rob Watts, ‘Government and Modernity: An Essay in Thinking Governmentality’ (1994) 2 Arena Journal 103

Robbin Miranda and others, ‘DARPA-Funded Efforts in the Development of Novel Brain–Computer Interface Technologies’ (2015) 244 Journal of Neuroscience Methods 52

Robert Haddick, ‘Stopping Mobile Missiles: Top Picks For Offset Strategy’: *Breaking Defense* (23 January 2015)

Robert Jervis, ‘Cooperation Under the Security Dilemma’ (1978) 30 World Politics

Robert Nakamura and Frank Smallwood, *The Politics of Policy Implementation* (St Martin’s Press 1980)

Robert Sparrow, ‘Killer Robots’ (2007) 24 Journal of Applied Philosophy

——, ‘Robots and Respect: Assessing the Case Against Autonomous Weapon Systems’ (2016) 30 Ethics and International Affairs 93

Robert Trager and Laura Luca, ‘Killer Robots Are Here—and We Need to Regulate Them’ *Foreign Policy* (11 May 2022)

Robert Work, ‘Remarks by Deputy Secretary Work on Third Offset Strategy’ (28 April 2016) <https://www.defense.gov/News/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/>

——, ‘Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities’ (CNAS 2021)

Roberta Arnold, ‘Legal Challenges Posed by LAWS: Criminal Liability for Breaches of IHL by (the Use of) LAWS’, *Lethal Autonomous Weapon Systems* (German Federal Foreign Office 2016)

Roger Deacon, ‘Theory as Practice: Foucault’s Concept of Problematization’ (2000) 118 Telos 127

Roland Bathes, *Critical Essays* (Northwestern University Press 1972)

Ronald Arkin, ‘Lethal Autonomous Systems and the Plight of the Non-Combatant’ (2013) 137 AISB Quarterly

——, ‘Warfighting Robots Could Reduce Civilian Casualties, So Calling for a Ban Now Is Premature’ [2015] IEEE Spectrum

Ronald Arkin and others, ‘A Path Towards Reasonable Autonomous Weapons Regulation’ (IEEE Spectrum 2019)

Ronald Arkin, Patrick Ulam, and Brittany Duncan, ‘An Ethical Governor for Constraining Lethal Action in an Autonomous System’ (2009) GIT-GVU-09-02

Sandeep Mulgund, ‘Evolving the Command and Control of Airpower’ (21 April 2021) <https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2575321/evolving-the-command-and-control-of-airpower/>

Sara Brown, ‘Machine Learning, Explained’ (MIT 2021) <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

Sean Lynn-Jones, ‘Offense-Defense Theory and Its Critics’ [1995] Security Studies
Shane Darcy, *Judges, Law and War: The Judicial Development of International Humanitarian Law* (Cambridge University Press 2014)

Shannon Vallor, ‘Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character’ (2015) 28 Philosophy and Technology 107
Shield AI, ‘Nova 2’ <https://shield.ai/nova-2/>

Solveg Badillo and others, ‘An Introduction to Machine Learning’ (2020) 107 Clinical Pharmacology & Therapeutics 871

Stanford Encyclopedia of Philosophy, ‘Michel Foucault’ (*Stanford Encyclopedia of Philosophy*, April 2003) <https://plato.stanford.edu/entries/foucault/>

Stavros Atlamazoglou, ‘This New Technology May Be the Future of Close Air Support’ (*Sofrep*, 2 February 2019) <https://sofrep.com/news/this-new-technology-may-be-the-future-of-close-air-support/>

Stephany Bellomo, ‘A Closer Look at 804: A Summary of Considerations for DoD Program Managers’ (Carnegie Mellon University 2011) CMU/SEI-2011-SR-015

Stephen Ball, *Politics and Policy Making in Education* (Routledge 1990)

——, ‘What Is Policy? 21 Years Later: Reflections on the Possibilities of Policy Research’ (2015) 36 Discourse: Studies in the Cultural Politics of Education 306

Stephen Losey, ‘How Autonomous Wingmen Will Help Fighter Pilots in the next War’ [2022] *DefenseNews* <https://www.defensenews.com/air/2022/02/13/how-autonomous-wingmen-will-help-fighter-pilots-in-the-next-war/>

Stephen Walt, ‘The Relationship between Theory and Policy in International Relations’ [2005] Annual Review of Political Science

Steven Barela and Avery Plaw, ‘Programming Precision? Required Robust Transparency for AWS’, *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021)

Steven Levy, ‘Palmer Luckey Says Working With Weapons Isn’t as Fun as VR’ *Wired* (14 March 2022) <https://www.wired.com/story/palmer-luckey-drones-autonomous-weapons-ukraine/>

Stew Magnuson, ‘Armed Robots Sidelined in Iraqi Fight’ *National Defense* (1 May 2008) <https://www.nationaldefensemagazine.org/articles/2008/5/1/2008may-armed-robots-sidelined-in-iraqi-fight>

Stuart Russel, *Human Compatible* (Viking 2019)

Sue Halpern, ‘The Rise of A.I. Pilots’ *The New Yorker* (17 January 2022) <https://www.newyorker.com/magazine/2022/01/24/the-rise-of-ai-fighter-pilots>

Susan Goodwin, ‘Women, Policy and Politics: Recasting Policy Studies’, *Engaging with Carol Bacchi Strategic Interventions and Exchanges* (University of Adelaide Press 2012)

Sydney J. Freedberg Jr, ‘Decentralize The Air Force For High-End War: Holmes’ *Breaking Defense* (13 October 2017) <https://breakingdefense.com/2017/10/decentralize-the-air-force-for-high-end-war-holmes/>

——, ‘Artificial Intelligence, Lawyers And Laws Of War’ *Breaking Defense* (23 April 2021)

Tanel Tammet, ‘Autonomous Cyber Defence Capabilities’, *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021)

‘The Air Force Next Generation ISR Dominance Flight Plan 2018-2028’ (USAF 2018)
The Joint Chiefs of Staff, ‘JP 1, Doctrine for the Armed Forces of the United States’ (2013)

——, ‘Joint Publication 3-09.3 Close Air Support’ (2014)

——, ‘Joint Publication 3-30, Joint Air Operations’ (2019)

——, ‘JP 1-02, DoD Dictionary of Military and Associated Terms’ (2020)

The White House, ‘National Security Strategy’ (2010)

——, ‘National Security Strategy’ (2015)

——, ‘National Security Strategy’ (2017)

——, ‘National Security Strategy’ (2022)

Thom Shanker and James Risen, ‘“Raid’s Aftermath: U.S. Troops Search for Clues to Victims of Missile Strike’ *The New York Times* (11 February 2002) <https://www.nytimes.com/2002/02/11/world/nation-challenged-raid-s-aftermath-us-troops-search-for-clues-victims-missile.html>

Thomas Bächle and Jascha Bareis, ‘“Autonomous Weapons” as a Geopolitical Signifer in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies’ (2022) 10 European Journal of Futures Research 1

Thomas Ehrhard, ‘Air Force UAVs The Secret History’ (Mitchell Institute for Airpower Studies 2010)

Thomas Sheridan, *Telerobotics, Automation and Human Supervisory Control* (MIT Press 1992)

Thomas Sheridan and William Verplank, *Human and Computer Control of Undersea Teleoperators* (MIT Press 1978)

Thompson Chengeta, ‘Defining the Emerging Notion of “meaningful Human Control"’ (2016) 49 NYU Journal of International Law and Politics

——, ‘Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law’ (2017) 45 Denver Journal of International Law and Policy

Timothy Schultz, *The Problem with Pilots: How Physicians, Engineers, and Airpower Enthusiasts Redefined Flight* (Johns Hopkins University Press 2018)

Todd Harrison, ‘Battle Networks and the Future Force’ (CSIS 2021)

Tom Clark, ‘Attorney General’s Manual on the Administrative Procedure Act’ (1947)

Tom Mitchell, *Machine Learning* (McGraw-Hill Education 1997)

‘Towards a “Compliance-Based” Approach to LAWS’ (Government of Switzerland 2016)

U C Jha, *Killer Robots Lethal Autonomous Weapon Systems Legal, Ethical and Moral Challenges* (Vij Books India 2016)

UK Ministry of Defence, ‘Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems’ (2011)

——, ‘Joint Doctrine Publication 0-30.2. The UK Approach to Unmanned Aircraft Systems’ (UK Ministry of Defence 2017)

UN GGE, ‘Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems’ (2014) CCW/MSP/2014/3

——, ‘Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems’ (2016)

——, ‘Report of the 2017 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (UN GGE 2017) CCW/GGE.1/2017/CRP.1

——, ‘Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (UN GGE 2018) CCW/GGE.1/2018/3

UNIDIR, ‘The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward’ (UNIDIR 2014)

United Nations Office for Disarmament Affairs, ‘Background on LAWS in the CCW’ <https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/>

United States Senate Committee on Armed Forces, ‘Summary of the Fiscal Year 2023 National Defense Authorization Act’ (2022)

US DoD, ‘Unmanned Aircraft Systems Roadmap 2005-2030’ (2005)

——, ‘Unmanned Systems Integrated Roadmap FY2007–2032’ (2007)

——, ‘The Quadrennial Defense Review’ (2010)

——, ‘National Military Strategy’ (2011)

——, ‘Unmanned Systems Integrated Roadmap FY2011-2036’ (2011)

——, ‘Defense Strategic Guidelines’ (2012)

——, ‘The Defense Innovation Initiative (Memorandum)’ <https://defenseinnovationmarketplace.dtic.mil/wp-content/uploads/2018/04/DefenseInnovationInitiative.pdf>

——, ‘The Quadrennial Defense Review’ (2014)

——, ‘Unmanned Systems Integrated Roadmap FY2013–2038’ (2014)

——, ‘National Military Strategy’ (2015)

——, ‘Autonomy in Weapon Systems’ (UN GGE 2017) CCW/GGE.1/2017/WP.6

——, ‘AI Strategy: Harnessing AI to Advance Our Security and Prosperity’ (2018)

——, ‘Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (US Government 2018) CCW/GGE.2/2018/WP.4

——, ‘National Defense Strategy’ (2018)

——, ‘DIB Guide: Detecting Agile BS’ (2018) <https://media.defense.gov/2018/Oct/09/2002049591/-1/-1/0/DIB_DETECTING_AGILE_BS_2018.10.05.PDF>

——, ‘Summary Of The 2018 Department of Defense Artificial Intelligence Strategy’ (2019)

——, ‘Nuclear Matters Handbook 2020’ (2020)

——, ‘DOD Adopts Ethical Principles for Artificial Intelligence’

——, ‘Agile Software Acquisition Guidebook’ (2020)

——, ‘National Defense Strategy’ (2022)

——, ‘U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway’ (2022)

US GAO, ‘DoD Software Acquisition: Status of and Challenges Related to Reform Efforts’ (2021) GAO-21-105298

US Government Must “Wake up Now” to AI Threat from China, Says Former Air Force Software Chief (Directed by Government Matters on 7 News, 2021) <https://www.youtube.com/watch?v=7tqS9_AN9y0>

USAF, ‘Air Force Doctrine Document 1 (AFDD 1, 1997)’ (1997)

——, ‘Air Force Doctrine Document 1 (AFDD 1, 2003)’ (2003)

——, ‘Unmanned Aircraft Systems Flight Plan FY2009-2047’ (2009)

——, ‘Unmanned Aircraft Systems Integrated Roadmap FY2009-2034’ (2009)

——, ‘Department of the Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities’

——, ‘Air Force Strategic Environment Assessment: 2014–2034’ (2015)

——, ‘Unmanned Systems Integrated Roadmap 2017-2042’ (2017)

——, ‘The US Air Force Special Operations Command 2020’ (2020)

——, ‘Air Force Chief Data Office Announces First Datathon’ (9 July 2020) <https://www.af.mil/News/Article-Display/Article/2268590/air-force-chief-data-office-announces-first-datathon/>

——, ‘Air Force Doctrine Publication 3-03 - Counterland Operations’ (USAF 2020)

——, ‘Air Force Doctrine Document (AFDP)’ (2021)

——, ‘Air Force Doctrine Document 1 (AFDD)’ (2021)

——, ‘Air Doctrine Publication 3-60 Targeting’ (USAF 2021)

——, ‘A Primer on Doctrine’

Valerie Insinna, ‘Shield AI to Work on Swarming Drones, Autonomous Rotorcraft for Air Force’ *Breaking Defense* (21 February 2022) <https://breakingdefense.com/2022/02/shield-ai-to-work-on-swarming-drones-autonomous-rotorcraft-for-air-force/>

Vincent Boulanin and Maaike Verbruggen, ‘Mapping the Development of Autonomy in Weapon System’ (2017)

——, ‘SIPRI Compendium on Article 36 Reviews’ (2017)

Vincent Müller and Nick Bostrom, ‘Future Progress in Artificial Intelligence: A Survey of Expert Opinion’, *Fundamental Issues of Artificial Intelligence* (Springer 2016)

Vincent Müller and Thomas Simpson, ‘Autonomous Killer Robots Are Probably Good News’ (2014) 273 Frontiers in Artificial Intelligence and Applications 297
Weapon Systems, ‘Mark 60 CAPTOR’ <https://weaponsystems.net/system/449-Mark+60+CAPTOR>

Will Roper, ‘Exclusive: AI Just Controlled a Military Plane for the First Time Ever’

Popular Mechanics (16 December 2020)

William A. Woodcock, 'The Joint Forces Air Command Problem' (2003) 56 Naval War College Review

William Barker, 'Guideline for Identifying an Information System as a National Security System' (National Institute of Standards and Technology 2003) NIST Special Publication 800–59

William Wagner, *Lightning Bugs and Other Reconnaissance Drones: The Can-Do Story of Ryan's Unmanned Spy Plane* (Aero Publishers 1982)

WJ Hennigan, 'Islamic State's Deadly Drone Operation Is Faltering, but U.S. Commanders See Broader Danger Ahead' (*LA Times*, 2017)

Zachary Fryer-Biggs, 'Are We Ready for Weapons to Have a Mind of Their Own?' (*The Centre for Public Integrity*, 17 February 2021) <https://publicintegrity.org/national-security/future-of-warfare/mind-of-their-own-artificial-intelligence-weapon/>

Zachary Kallenborn, 'Was a Flying Killer Robot Used in Libya? Quite Possibly' *Bulletin of Atomic Scientists* (20 May 2021) <https://thebulletin.org/2021/05/was-a-flying-killer-robot-used-in-libya-quite-possibly/?utm_source=Twitter&utm_medium=SocialMedia&utm_campaign=TwitterPost05202021&utm_content=DisruptiveTechnology_WasAFlyingKillerRobotUsedInLibya%3F_05202021>

'Missy Cummings Asks: Should the US Military Use AI Weapons?' <https://www.youtube.com/watch?v=yaBY3Pfmdno&t=1s>
'Interview with Jared Dunnmon' <https://federalnewsnetwork.com/artificial-intelligence/2021/12/defense-innovation-unit-issues-guidance-for-responsible-use-of-artificial-intelligence/>

'Starting Project Maven with Lt. Gen. Jack Shanahan' <https://aneyeonai.libsyn.com/episode-45-jack-shanahan>

'The Policy and Law of Lethal Autonomy with Michael Meier and Shawn Steene' <https://madsciblog.tradoc.army.mil/305-the-convergence-the-policy-and-law-of-lethal-autonomy-with-michael-meier-and-shawn-steene/>

Administrative Procedure Act 1946

Allied Joint Doctrine for Joint Targeting, AJP-3.9 2021

Army Regulation 602–2 Human Systems Integration in the System Acquisition Process 2015

Department of the Army Regulation 27-53, Review of Legality of Weapons Under International Law 1979

Directive 3000.09 Autonomy in Weapon Systems 2012

Directive 5000.01 The Defense Acquisition System 2003

Directive 5500.15 Review of Legality of Weapons under International Law, US Department of Defense 1974

Geneva Convention Relative to the Treatment of Prisoners of War (Third Geneva Convention) 1949

National Defense Authorization Act for Fiscal Year 2010 2009

National Defense Authorization Act for Fiscal Year 2023 2022

National Defense Authorization Act for Fiscal Years 1988 and 1989 1987

The Statute of the International Court of Justice 1946